



FAIRVIEW  
UNIVERSITY COLLEGE

# Journal of Educational Leadership and Innovation

Volume 3

Issue 1

March/April 2026

MY IB World School  
FAIRVIEW  
INTERNATIONAL SCHOOL



**e-ISSN: 2805-4504**

**Issue: Volume 3, Issue 1 of 1, 2026**

The Journal of Educational Leadership and Innovation (JELI) is an open-access online journal published biannually in January to February and August to September.

Employing a double-blind, peer-review system, JELI ensures content quality. We invite article submissions in English, focusing on educational leadership, pedagogical best practices, innovations in teaching and learning, and related topics. Submission deadlines should be at least 30 days before the nearest publication date; otherwise, they will be considered for the next round.

**Editor:**

Professor Datuk Dr. Sufean Bin Hussin

**Editorial Board:**

Professor Dr. Tee Siok Hoon  
Professor Gopinathan A/L Raman Kutty  
Dr. Janice Chan Mei Shen  
Dr. Vincent Chian Kwok Liang  
Dr. Kenny Cheah Soon Lee  
Mr. Mark Adrian Wright  
Ms. Chang Yann Lyn  
Ms. Elaine Wong Hui Chin

**Copyright Published By:**

Faculty of Education, University College Fairview  
Blok B,4178, Jalan 1/27D, Section 6,  
Wangsa Maju 53300 Kuala Lumpur  
E-mail: enquiries @fairview.edu.my  
Tel:03-4142088

## CONTENTS

<b>Role-Playing as a Teaching Strategy in IB Chemistry: A Mixed-Method Study.....</b>	<b>1</b>
<b>Technological Implementation: AI Adoption in Kerala CBSE Schools Amid Systemic Barriers.....</b>	<b>24</b>
<b>A Conceptual Framework for Reflexive Autoethnography in CFL Teacher Research... </b>	<b>40</b>
<b>Culturally Inclusive Practices in International Schools: An Ecological Mixed-Methods Study.....</b>	<b>61</b>
<b>AI-Assisted Grading and Feedback in International Baccalaureate Middle Years Programme Physics: A Mixed-Methods Investigation of Learning Outcomes, Feedback Quality, and the Moderating Role of English Language Proficiency.....</b>	<b>83</b>
<b>Assessment Practices and Challenges in Biology Education at a China-Based International School: A Mixed-Methods Case Study.....</b>	<b>111</b>



## Role-Playing as a Teaching Strategy in IB Chemistry: A Mixed-Method Study

Samuel Lee

University College Fairview

### **Abstract**

*Drama pedagogy, particularly role-playing, offers promise for making abstract chemical concepts tangible through embodied learning. However, evidence for its effectiveness across different age groups remains limited, especially in multilingual International Baccalaureate (IB) contexts. This mixed-methods study investigated when and how role-playing effectively supports learning of abstract chemical concepts, examining developmental readiness as a critical factor. Seventy-eight students (66 Grade 9, 12 Grade 12) from an international school in Shenzhen, China participated. A between-subjects design with Grade 9 students compared role-playing ( $n=44$ ) against conventional instruction ( $n=22$ ) for chemical reaction types. A within-subjects design with Grade 12 students examined role-playing for nucleophilic substitution mechanisms. Data included achievement tests, motivation/engagement surveys, 30 student interviews, and 10 systematic lesson observations. For Grade 9, the experimental groups who received role-playing showed negligible achievement gains compared to the control group who received conventional instruction ( $d = 0.022$ ), despite high behavioural engagement and enjoyment. For Grade 12, where all students served as their own controls in a pre-post design, role-playing produced very large improvements in understanding reaction mechanisms (Cohen's  $d_z = 1.209$ ,  $p = .002$ ). English proficiency significantly predicted achievement ( $\eta^2 = .220$ ) but did not moderate intervention effects. The two cohorts are not directly comparable due to differences in topic, design, sample size, and assessment. The Zone of Pedagogical Receptivity (ZPR) is presented only as a speculative hypothesis for future investigation; it is not a finding of this study. Findings challenge assumptions that engaging pedagogies necessarily improve outcomes, offering evidence-based guidance for implementing drama in chemistry education.*

**Keywords:** *drama pedagogy, chemistry education, developmental readiness, mixed-methods, SDG 4 (Quality Education)*

## 1. Introduction

Chemistry learning requires students to navigate cognitively between macroscopic observations, submicroscopic processes, and symbolic representations. This is what Johnstone (1991) called the chemistry triplet. Students consistently struggle with visualising molecular-level processes, particularly reaction mechanisms where electron movement, spatial arrangements, and dynamic interactions occur at the same time (Talanquer, 2011). Traditional instruction that relies on two-dimensional diagrams often fails to convey the three-dimensional nature of molecular interactions, leading to rote memorisation rather than conceptual understanding (Eichler & Peebles, 2016). This challenge is made worse in linguistically diverse classrooms where students must process unfamiliar scientific vocabulary and abstract chemical concepts through an additional language.

The gap between the cognitive demands of chemistry and the representational tools available to students has significant consequences for science education. Research has consistently shown that students who fail to develop strong mental models of molecular interactions in secondary school still have those misconceptions when they reach tertiary education (Taber, 2013). Despite the fact that reforms over the past decades have pushed for inquiry-based and student-centred approaches, most chemistry classrooms worldwide still rely on teacher-centred instruction and static visual representations.

### **Drama Pedagogy as a Potential Solution**

Drama pedagogy, particularly role-playing, offers a promising alternative based on constructivist principles (Piaget, 1972; Vygotsky, 1978). When students physically enact chemical reactions, abstract concepts become easier to understand through physical experience (Braund, 2014; Dorion, 2009). Embodied cognition theory suggests that physical engagement enhances cognitive processing by creating multiple memory pathways (Wilson, 2002). This is relevant because traditional chemistry instruction typically relies on

two-dimensional diagrams to teach concepts that are three-dimensional and dynamic in nature.

Drama-related activities have generally achieved high participation rates and positive student attitudes, but evidence that they actually improve conceptual understanding has been mixed. Otter (2020) conducted a focused role-playing study in A-level chemistry and found statistically significant gains in student understanding of organic mechanisms. Abed (2016) showed that dramatisation improved science achievement among primary students, and Najami et al. (2019) reported positive effects on student attitudes toward science. However, none of these studies examined whether the same approach works equally well across different age groups. There is also limited research on role-playing in linguistically diverse IB contexts, where students learn content through an additional language.

### **The Research Gap**

While existing studies show that drama pedagogy has potential, most do not specify whether they involve redesigning the entire curriculum or simply adding drama activities into existing lessons. This distinction matters for teachers who want to use drama but need to work within their current curriculum structure. The IB context, with its emphasis on inquiry-based learning and international-mindedness, is therefore a suitable context for investigating these questions, since linguistic diversity is built into these settings.

Much of the available literature focuses on single age groups. The lack of comparative studies across different developmental stages is a significant gap in science education research. Piaget's cognitive development theory (Piaget, 1972) and research on embodied cognition (Barsalou, 2008) suggest that age-related differences in cognitive processing should affect how well embodied teaching methods work, but this has not been empirically tested in chemistry education. This study therefore addresses these gaps.

### **Research Questions**

This study addressed four research questions:

How does role-playing compare to conventional teaching in improving student understanding of chemical reaction types (Grade 9) and reaction mechanisms (Grade 12) in IB Chemistry?

What impact does role-playing have on student motivation and engagement?

How does English language proficiency influence role-playing effectiveness?

What implementation challenges emerge, and what strategies address them?

## **2. Literature Review**

### **Drama Pedagogy in Science Education**

Drama-based instructional strategies have gained increasing attention in science education as researchers seek pedagogies that engage students more deeply with abstract content (Ødegaard, 2003). The application of drama in science ranges from brief kinaesthetic activities to full theatrical productions. Role-playing, which places students as active participants representing scientific entities or processes, sits between these extremes as a practical approach with enough depth to be useful (Dorion, 2009). In chemistry, role-playing has been used to teach topics ranging from states of matter at the primary level to organic reaction mechanisms at advanced secondary and university levels. Braund (2014) conducted a comprehensive review of drama in science education and found that there is still a gap between what embodied learning promises in theory and what the evidence actually shows in practice. While drama activities consistently produced high student engagement and positive attitudes, evidence of improved conceptual understanding was uneven.

Otter (2020) addressed this gap through a focused study of role-playing in A-level chemistry, finding statistically significant improvements in students' understanding of organic mechanisms. However, Otter's study examined only one age group, leaving questions about developmental readiness unanswered.

## **Embodied Cognition and Chemistry Learning**

Embodied cognition theory suggests that cognitive processes are shaped by the body's interactions with its environment (Wilson, 2002; Barsalou, 2008). In chemistry education, this suggests that physical enactment of molecular processes may help students form better mental models than visual or verbal instruction alone. Kontra et al. (2015) provided neuroscientific evidence for this, showing through functional magnetic resonance imaging that students who physically manipulated angular momentum apparatus showed greater activation in sensorimotor brain regions during subsequent problem solving, compared to students who merely observed.

This is directly relevant to chemistry education. Molecular processes such as nucleophilic substitution involve three-dimensional spatial transformations that are difficult to represent through static diagrams (Eichler & Peeples, 2016). When students physically enact these processes, they create mental models that show the dynamic and spatial nature of molecular interactions in ways that arrow-pushing notation on paper cannot show.

## **Developmental Readiness and Pedagogical Approaches**

Piaget's (1972) theory of cognitive development distinguishes between concrete operational thinking (ages 7 to 11) and formal operational thinking (ages 11 and beyond). Formal operational thinkers can reason abstractly, consider hypothetical scenarios, and coordinate multiple variables at the same time. While research has modified Piaget's original age ranges, recognising that the transition to formal operations is gradual and context-dependent (Kuhn, 2008), the fundamental distinction between concrete and abstract reasoning capacity remains empirically supported. For drama pedagogy, this raises an important question: does effective embodied learning of abstract concepts require consolidated formal operational thinking, or can it support the development of such thinking in transitional learners?

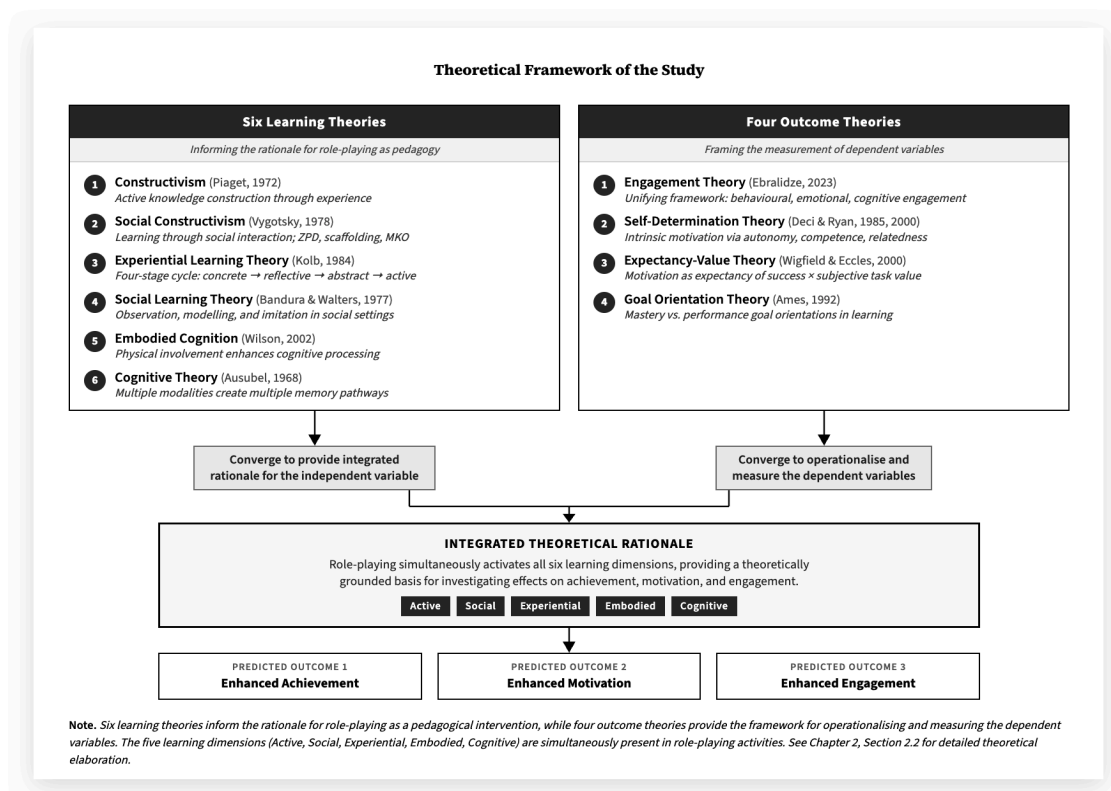
The existing literature does not directly address this question. Studies of drama in science education typically examine either primary-age students with concrete concepts (Abed, 2016) or senior secondary and university students with abstract concepts (Otter, 2020), without examining multiple age groups within the same study. This gap motivated this study's dual-cohort design, which examines role-playing with two different age groups studying different chemistry topics. While this cross-sectional design does not track the same students

over time, it allows for observation of how role-playing functions with students at different points in their schooling, with different levels of prior knowledge and different curricular demands.

---

### **3. Theoretical Framework**

This study integrates multiple theoretical perspectives that together support drama pedagogy as a suitable instructional approach for chemistry education. The framework draws on six foundational learning theories and four outcome theories to provide a basis for investigating role-playing's effectiveness. Constructivist foundations (Piaget, 1972) describe learners as active participants who build knowledge through experience. In role-playing, students must actively manipulate concepts rather than receiving information passively. Social constructivist extensions (Vygotsky, 1978) explain how peer collaboration and teacher guidance during role-playing enable students to accomplish tasks beyond individual capacity through the Zone of Proximal Development (ZPD). Experiential Learning Theory (Kolb, 1984) maps onto well-designed role-playing activities where students enact reactions, reflect on their representations, connect to chemical principles, and apply understanding to new contexts through a cyclical process of concrete experience, reflective observation, abstract conceptualisation, and active experimentation.



**Figure 1. Convergence of Theoretical Perspectives Underpinning the Study**

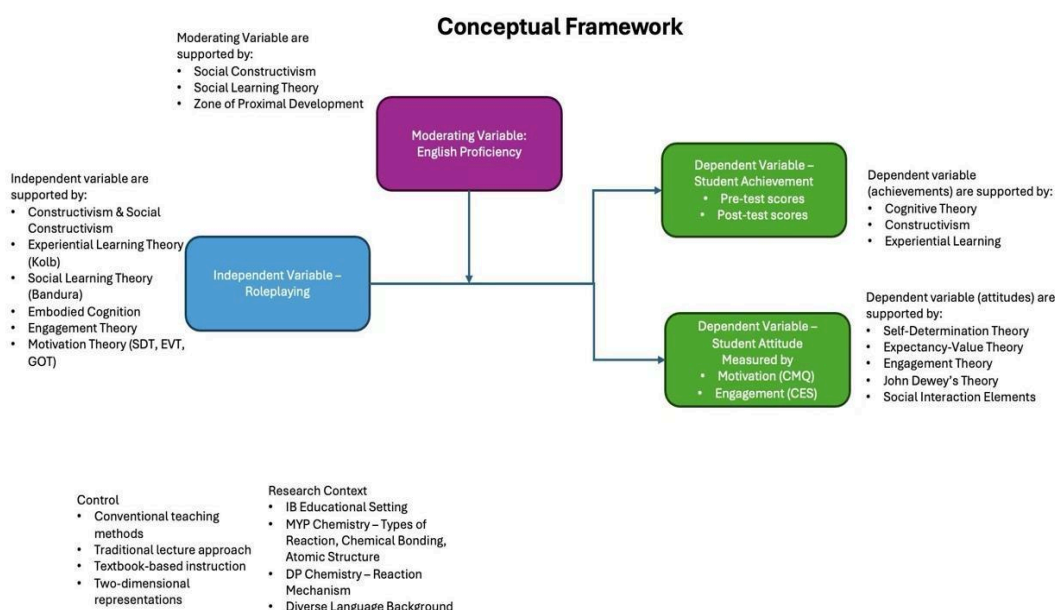
Embodied Cognition (Wilson, 2002; Barsalou, 2008) argues that how people think is connected to how their bodies physically interact with the environment. Physical representation of abstract concepts creates richer, more memorable learning experiences, as demonstrated by Kontra et al. (2015) who provided direct evidence that physical experience enhances science learning. Social Learning Theory (Bandura and Walters, 1977) further supports the approach through observational learning and modelling, while Cognitive Theory explains how multisensory input creates multiple memory pathways, which helps students to remember and recall information better (Atkinson & Shiffrin, 1968).

The study draws on three complementary motivation theories. Self-Determination Theory (Deci & Ryan, 2000) proposes that when students feel a sense of autonomy, competence, and relatedness, their intrinsic motivation increases. Expectancy-Value Theory (Eccles & Wigfield, 2002) explains motivation as a product of how likely students think they are to succeed and how much they value the task.

Goal Orientation Theory (Dweck, 1986) distinguishes mastery goals from performance goals. Engagement encompasses behavioural, emotional, and cognitive dimensions (Fredricks et al., 2004), and these are interconnected. For example, when a student is emotionally engaged, they are more likely to put in cognitive effort, which can then lead to better achievement.

#### 4. Conceptual Framework

The conceptual framework describes the relationships among the study's variables. The independent variable, drama pedagogy through role-playing, is implemented through structured activities for chemical reaction types (Grade 9) and reaction mechanisms (Grade 12).



**Figure 2. Conceptual Framework: Variables and Relationships in the Study**

The dependent variables are student achievement, measured through pre- and post-tests, and student attitudes toward the subject, measured through standardised motivation and engagement surveys. English language proficiency serves as a moderating variable that may

influence how effective role-playing is in this multilingual setting. The control group received conventional instruction, i.e. lectures, diagrams and textbook exercises.

The framework also accounts for implementation challenges through strategic topic selection and professional development. Rather than requiring a complete overhaul of the course, role-playing activities were fitted into the existing IB structure. The framework hypothesises that embodied pedagogy may interact with language proficiency in complex ways: physical enactment could potentially compensate for linguistic barriers by giving students physical ways to show meaning without relying on words, or conversely, the additional verbal demands of collaborative role-playing could add to challenges for lower-proficiency students. Testing this moderation effect is important in order to understand whether drama pedagogy works in internationally diverse classrooms.

## **5. Methodology**

### **Research Design and Philosophy**

Grounded in critical realism (Bhaskar, 1975), this study employed a convergent mixed-methods design (Creswell & Plano Clark, 2017). A between-subjects experimental design with Grade 9 students compared role-playing against conventional instruction for chemical reaction types, with Class A serving as the control group and Classes B and C as experimental groups. A within-subjects pre-post design with Grade 12 students examined role-playing for nucleophilic substitution mechanisms, with each student serving as their own control. The two cohorts studied different chemistry topics, received different role-playing activities, and were assessed with different achievement tests appropriate to their respective curricula. The Grade 9 topic (chemical reaction types) involved concrete, observable phenomena, while the Grade 12 topic (nucleophilic substitution mechanisms) involved abstract spatial and temporal reasoning. These differences reflect authentic curricular distinctions between the grade levels and mean that direct comparison of effect sizes across cohorts should be interpreted with caution. Qualitative data collection (interviews, observations) ran concurrently to explain quantitative patterns and document implementation

processes. The convergent design enabled triangulation of findings across data sources, strengthening the validity of the conclusions.

### Participants and Context

This study was conducted at an IB continuum international school in Shenzhen, China. Participants comprised 78 students: 66 Grade 9 students (ages 14–15) and 12 Grade 12 students (ages 17–18). Grade 9 students were organised into three intact classes of 22 students each, pre-assigned by school administration to ensure balanced demographics.

Demographic Variable	Grade 9 (n=66)	Grade 12 (n=12)
Gender		
Female	32 (49%)	7 (58.3%)
Male	34 (51%)	5 (41.7%)
Student Background		
Chinese (Hong Kong)	43 (65.2%)	5 (41.7%)
Chinese (Taiwan/Macau)	10 (15.2%)	0
American/Canadian	6 (9.1%)	3 (25%)
Other	7 (10.6%)	4 (33.3%)
English Proficiency		
Level 1–2	9 (13.6%)	0
Level 3	23 (34.8%)	0
Level 4–5	34 (51.5%)	12 (100%)

**Table 1. Participant Demographics**

**Grade 9 Intervention (Weeks 3–8):** Experimental groups (Classes B and C) received role-playing instruction for Atomic Structure, Chemical Bonding, and Chemical Reactions. Activities included "Electron Hotel" for electron configuration, "Electron Transfer Ceremonies" for ionic bonding, and "Atomic Speed Dating" for synthesis reactions. Each activity was designed to turn abstract concepts into physical, embodied experiences. The control group (Class A) received conventional instruction covering identical content through lectures, diagrams, and textbook exercises, ensuring content equivalence across conditions.

**Grade 12 Intervention (Weeks 1–4):** All students received conventional instruction for nucleophilic substitution (SN1 and SN2) mechanisms during Weeks 1–2, followed by role-playing instruction during Weeks 3–4. Activities included "Synchronised Molecular Dance" for the SN2 concerted mechanism and "Two-Step Drama" for the SN1 stepwise mechanism with carbocation intermediate. Students physically enacted nucleophilic attack, bond breaking, and stereochemical inversion, translating arrow-pushing notation into physical movement.

Achievement tests for Grade 9 (50 marks) assessed Periodic Table, Atomic Structure, Chemical Bonding, and Chemical Reactions. Grade 12 tests (50 marks) assessed nucleophilic substitution mechanisms. All tests underwent expert validation by an experienced IB educator and were moderated systematically through independent marking, with disagreements resolved by consensus. The modified Chemistry Motivation Questionnaire II (CMQ-II) measured intrinsic motivation, career motivation, self-determination, self-efficacy, grade motivation, and English language integration using a 5-point Likert scale ( $\alpha > .80$  for all subscales).

### **Data Collection Instruments**

Chemistry Engagement Survey (CES), with Cronbach's alphas ranging from .76 to .84, measured cognitive, emotional, and behavioural engagement. Qualitative data were collected through semi-structured interviews with 30 students selected using maximum variation sampling, along with 2 teacher-observers and 10 lessons documented using structured observation protocols.

### **Data Analysis**

Quantitative analysis employed independent samples t-tests, ANCOVA, paired samples t-tests, and linear mixed models. Effect sizes (Cohen's d, partial eta-squared) were calculated with 95% confidence intervals to provide practical significance measures alongside statistical significance. Qualitative analysis followed Braun and Clarke's (2006) thematic analysis procedure, through collaborative coding by the researcher and an independent Teacher-Observer using Atlas.ti software. Integration occurred through joint displays and meta-inference development, consistent with best practices for convergent mixed-methods designs.

## Ethical Considerations

Informed consent was received from all participants in this study; for students under 18 years, consent was also obtained from a parent or guardian. Participation was voluntary, and participants could withdraw at any time without penalty. All data were anonymised using alphanumeric codes (e.g., G9B17, G12A04) and stored securely with restricted access. The researcher's dual role as teacher-investigator was acknowledged and addressed through reflexive journaling, independent coding verification, and transparent reporting of potential bias.

## 6. Findings

### Achievement Outcomes

Independent samples t-test revealed no significant difference between control and experimental groups in Grade 9. The negligible effect size ( $d = 0.022$ ) indicated that performance was virtually identical across the conditions, suggesting that role-playing did not enhance chemistry achievement for this age group.

**Table 2. Grade 9 Achievement Test Results by Group (Scores range 0–50)**

Group	n	Pre-Test M (SD)	Post-Test M (SD)	Gain	d
Control (A)	22	33.23 (9.60)	34.36 (9.63)	+1.1 4	—
Experimental B	22	33.18 (9.25)	34.32 (9.36)	+1.1 4	—
Experimental C	22	33.27 (9.40)	34.86 (9.42)	+1.5 9	—
Combined Exp	44	33.39 (9.17)	34.57 (9.16)	+1.1 8	0.022

ANCOVA with pre-test scores and English proficiency as covariates confirmed no treatment effect ( $F(2,61) = 0.047$ ,  $p = .954$ , partial  $\eta^2 = .002$ ). Pre-test scores dominated the model

(partial  $\eta^2 = .982$ ), indicating that prior knowledge was the main factor determining achievement regardless of instructional approach. In contrast, for Grade 12 students, a paired samples t-test demonstrated significant improvement with a very large effect size.

Eleven of 12 students (91.7%) improved from pre-test to post-test, indicating that gains were consistent at the individual level. Qualitative findings helped to explain the quantitative contrast. Grade 9 students demonstrated what one teacher-observer termed "busy hands, empty minds", suggesting that students were engaged but not necessarily learning.

**Table 3. Grade 12 Pre-Post Achievement Comparison**

Measure	Value	95% CI
Pre-Test Mean (SD)	36.25 (7.11)	[31.73, 40.77]
Post-Test Mean (SD)	38.75 (7.44)	[34.02, 43.48]
Mean Difference	2.50 (2.07)	[1.19, 3.81]
Paired t-test	t(11) = 4.190, p = .002	—
Cohen's dz	1.209	[0.440, 1.949]

Student G9B17 shared: "When we did that precipitation reaction role-play, I was focused on remembering my lines and when to move and not bumping into people. The actual chemistry concept got kind of lost." In contrast, Grade 12 students showed a qualitatively different type of engagement. Student G12A04 explained: "The pause between steps, that's when it clicked. For the first time, I understood WHY it's two steps, not just that it IS two steps."

### Motivation and Engagement

Linear mixed models revealed significant cognitive engagement increases for Grade 12 students (+14.2%, p = .018) but minimal changes for Grade 9. Emotional engagement showed no significant changes across groups, which challenges the assumption that drama automatically increases enjoyment.

**Table 4. Chemistry Engagement Mean Scores by Factor and Time (1–5 Likert scale)**

Class	Cog Pre	Cog Post	Emo Pre	Emo Post	Beh Pre	Beh Post
-------	---------	----------	---------	----------	---------	----------

A (Control)	2.94	3.05	2.92	3.01	2.89	2.98
B (Exp)	2.76	2.87	2.77	2.89	2.72	2.83
C (Exp)	2.56	2.68	2.59	2.72	2.53	2.65
D (Gr 12)	2.95	3.37	2.96	3.25	2.92	3.21

CMQ-II results showed significant self-efficacy gains ( $F(1,74) = 7.89, p = .006, \eta^2 = .096$ ) and intrinsic motivation increases ( $F(1,74) = 5.23, p = .025, \eta^2 = .066$ ), primarily driven by Grade 12 students. Change in motivation correlated with change in achievement for Grade 12 ( $r = .31, p = .006$ ), suggesting that the motivational benefits of embodied learning led to actual improvements in academic results for the older students.

### Language Proficiency Moderation

English proficiency significantly predicted achievement ( $F(1,61) = 17.19, p < .001$ , partial  $\eta^2 = .220$ ), accounting for 22% of variance in post-test scores after controlling for pre-test performance. However, the Group  $\times$  English proficiency interaction was non-significant ( $F(2,59) = 0.683, p = .509$ , partial  $\eta^2 = .023$ ), indicating role-playing neither amplified advantages for high-proficiency students nor compensated for challenges faced by lower-proficiency learners. Qualitative data also showed that students used translanguaging successfully. Student G12A08 explained: "We'd quickly check our understanding in Chinese, '这是反转吗?' (Is this inversion?). It helped having someone to verify my thinking in my native language before trying to act it out in English."

### Implementation Challenges

Time was the main constraint, with setup, execution, and cleanup requiring 30–45 minutes per session. Space limitations in chemistry laboratories, designed for bench work rather than movement, created safety concerns. Teacher-Observer 1 noted: "A 10-minute explanation became 30 minutes with role-play. Space constraints, not designed for movement activities! Nearly knocked over the Bunsen burners."

Behavioural management challenges included Grade 9 students treating activities as "free play." Student G9B20 observed: "Some people didn't take it seriously, others were too shy to

participate properly." Successful adaptations included pre-arranged furniture, student leadership roles, differentiated participation options, and explicit connections between embodied activities and assessment requirements. Teachers reported that establishing clear routines over the first two sessions significantly reduced setup time and behavioural disruptions in subsequent sessions, suggesting that spending time on setting up clear procedures early on leads to better efficiency later. These findings show that successful implementation of drama pedagogy requires careful attention to logistics and classroom management, not just pedagogical design.

## 7. Discussion

### Interpreting the Contrast Between Grade 9 and Grade 12 Outcomes

The contrast between Grade 9 ( $d = 0.022$ ) and Grade 12 ( $d_z = 1.209$ ) outcomes requires careful interpretation given that the two cohorts studied different topics, received different treatments, and were assessed with different tests. Several explanations are possible: the difference may reflect topic difficulty, the nature of the role-playing activities, sample size differences, or differences in the students' cognitive readiness for embodied instruction. **These two cohorts are not directly comparable** due to the following methodological differences:

Feature	Grade 9	Grade 12
Design	Between-subjects (control vs. experimental)	Within-subjects (pre-post, no control)
Topic	Chemical reaction types (concrete, observable)	Nucleophilic substitution (abstract, spatial, temporal)
Treatment duration	6 weeks	2 weeks
Sample size	66	12
Statistical power	Adequate for medium effects	Underpowered for small effects
Prior knowledge	Low (novices)	High (advanced)

Given these differences, the contrast **cannot be attributed solely—or even primarily—to developmental readiness**. Alternative explanations include topic difficulty  $\times$  pedagogy

interaction, statistical artefacts from the small Grade 12 sample (95% CI for dz: 0.44–1.95), regression to the mean in within-subjects design, practice effects, and different outcome measures.

### **The Zone of Pedagogical Receptivity (ZPR): A Speculative Hypothesis, Not a Finding**

Drawing on these observations alongside the qualitative data, the Zone of Pedagogical Receptivity (ZPR) framework (Figure 3) is proposed only as a tentative hypothesis that identifies possible conditions under which embodied learning may be more or less effective. This framework is not a confirmed finding of the study but rather a conceptual tool generated from the pattern of results that requires further empirical testing.

#### **Figure 3. The Zone of Pedagogical Receptivity Model** (original figure retained)

ZPR comprises three interrelated components. First, formal operational thinking (Piaget, 1972) requires students to coordinate concrete actions with abstract concepts simultaneously. Grade 9 students, typically in the concrete-to-formal operational transition, struggled with this dual processing, managing the demands of performing while trying to understand the chemistry. Grade 12 students, with consolidated formal operations, were better able to connect the physical actions to the chemical concepts. Second, metacognitive awareness (Flavell, 1979) enables students to think about how the physical representations relate to the chemical principles. Grade 12 students were able to explain on their own how the physical acting helped them understand the mechanisms, whereas Grade 9 students focused on performance logistics rather than conceptual connections. Third, epistemic maturity (Hofer & Pintrich, 2012) requires students to recognise embodied exploration as legitimate scientific practice rather than play. Grade 9 students' perception of role-playing as "fun but not real learning" meant that there was a gap between what the students were doing and what they were supposed to be learning.

However, it is important to acknowledge that this study did not track the same students over time, and the different topics and treatments used for each cohort mean that the observed contrast cannot be attributed solely to cognitive readiness. The framework is offered as a starting point for future longitudinal research that could test these ideas more rigorously by following the same students across age groups or by comparing different topics within the same age group. It identifies cognitive abilities that may need to be in place for embodied

pedagogy to work, but further research is needed to confirm whether these factors are indeed the ones responsible for the pattern observed here. Readers should not treat ZPR as an empirically supported model.

### **Theoretical Implications**

The findings challenge the assumption that engaging, active pedagogies necessarily improve learning outcomes (Kirschner et al., 2006). The Grade 9 results suggest that behavioural activity does not automatically equate to cognitive processing. This could be due to what cognitive load theory (Sweller, 1988) describes as extraneous load: students who do not yet have strong background knowledge may find it too mentally demanding to handle the performance aspects of role-playing (remembering lines, coordinating movements, managing social interactions), leaving not enough mental capacity for actually learning the chemistry. Conversely, Grade 12 students' existing knowledge schemas may have had enough mental capacity left over to think more deeply through the physical activities. This is consistent with the expertise reversal effect in instructional design research, where methods that benefit advanced learners can be ineffective or even counterproductive for novices.

The language proficiency findings offer specific implications for multilingual science education. The significant main effect of English proficiency ( $\eta^2 = .220$ ) confirms that language facility has a strong effect on chemistry achievement. However, the absence of moderation effects reveals that role-playing pedagogy neither amplifies advantages for high-proficiency students nor compensates for challenges faced by lower-proficiency learners. This suggests that embodied pedagogy works together with language demands, not as a replacement for them. The physical aspects of role-playing (gesture, movement, spatial positioning) add to but do not replace the language-heavy parts of chemistry learning (Paivio, 1990). Language scaffolding remains necessary regardless of pedagogical approach.

## **Practical Implications**

For Grade 9–10 educators, role-playing should be used cautiously with simple, concrete concepts. Activities should be brief (10–15 minutes), heavily scaffolded, and immediately connected to chemical principles through explicit reflection. For Grade 11–12 educators, role-playing can greatly improve understanding of complex mechanisms. Extended activities (30–45 minutes) allow deep exploration, and student-directed elements and peer teaching can be incorporated successfully. Across all levels, implementation recommendations include differentiated roles to accommodate diverse comfort levels, clear behavioural expectations to maintain academic focus, pre-arranged furniture to reduce time lost to logistics, and explicit links between embodied activities and assessment criteria to ensure learning transfer.

## **Limitations**

Several limitations must be acknowledged. The single-school context limits generalisability to other educational settings and curricula. The Grade 12 sample size ( $N = 12$ ), while yielding large effect sizes, restricts statistical power and the strength of conclusions for that cohort. Teacher-rated English proficiency, while it reflects real classroom conditions, is not as precise as standardised language tests. The absence of a delayed post-test means retention durability cannot be assessed. Finally, the researcher's dual role as teacher-investigator introduces potential bias, though this was mitigated through reflexive practices, independent coding verification, and transparent reporting.

**Critical limitation specific to cross-cohort comparison:** This study was not designed to test developmental differences. The two cohorts differ in topic, design, treatment length, and assessment—confounding any developmental interpretation. The ZPR framework is withdrawn as a finding of this study and is presented only as a speculative hypothesis requiring independent validation.

## 8. Conclusion

This study examined role-playing as a teaching strategy with two different cohorts studying different chemistry topics. For Grade 9, role-playing for chemical reaction types produced negligible achievement gains compared to conventional instruction, despite high engagement. For Grade 12, role-playing for nucleophilic substitution mechanisms produced very large pre-post gains, with students reporting that physical enactment helped them understand spatial and temporal aspects that had been difficult to grasp through conventional diagrams alone. The contrasting results across the two cohorts led to the proposal of the Zone of Pedagogical Receptivity (ZPR) framework only as a speculative hypothesis suggesting that cognitive factors such as formal operational thinking, metacognitive awareness, and epistemic maturity may influence how well students benefit from embodied pedagogy. This hypothesis requires further testing through longitudinal or more tightly controlled comparative designs. ZPR is not a finding of this study.

For educators, the findings suggest that engagement alone should not be taken as evidence of learning. The null result for Grade 9 is just as informative as the positive result for Grade 12, though the different topics and treatments used mean that these results should not be directly compared as if all other variables were equal. What can be said is that role-playing worked well for Grade 12 students studying abstract reaction mechanisms but did not produce measurable learning gains for Grade 9 students studying chemical reaction types. Educators considering drama pedagogy should therefore attend carefully to how well the specific activity matches the cognitive demands of the topic and the readiness of the students.

Future research should look at whether the ZPR framework applies to scientific disciplines beyond chemistry, and how embodied learning may work differently at different developmental stages from a neurocognitive perspective. There is also a need for assessment approaches that can capture what students learn through multiple modes, not just written tests. Longitudinal studies that follow students through developmental transitions would help to clarify exactly when and how pedagogical receptivity develops. Cross-cultural replication studies are also needed to test whether the ZPR framework holds in educational systems with different curricular structures and language contexts.

## References

- Abed, O. (2016). *Drama-based science teaching and its effect on students' understanding of scientific concepts and their attitudes towards science learning*. *International Education Studies*, 9(10), 163–173. <https://doi.org/10.5539/ies.v9n10p163>
- Atkinson, R. C., & Shiffrin, R. M. (1968). *Human memory: A proposed system and its control processes*. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). Academic Press.
- Bandura, A., & Walters, R. H. (1977). *Social learning theory* (Vol. 1). Prentice Hall.
- Barsalou, L. W. (2008). *Grounded cognition*. *Annual Review of Psychology*, 59(1), 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>
- Bhaskar, R. (1975). *A realist theory of science*. Leeds Books.
- Braun, V., & Clarke, V. (2006). *Using thematic analysis in psychology*. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braund, M. (2014). *Drama and learning science: An empty space?* *British Educational Research Journal*, 41(1), 102–121. <https://doi.org/10.1002/berj.3130>
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Deci, E. L., & Ryan, R. M. (2000). *The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior*. *Psychological Inquiry*, 11(4), 227–268. [https://doi.org/10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01)
- Dorion, K. (2009). *Science through drama: A multiple case exploration*. *International Journal of Science Education*, 31(8), 1077–1099. <https://doi.org/10.1080/09500690802056522>

Dweck, C. S. (1986). *Motivational processes affecting learning*. *American Psychologist*, 41(10), 1040–1048. <https://doi.org/10.1037/0003-066X.41.10.1040>

Eccles, J. S., & Wigfield, A. (2002). *Motivational beliefs, values, and goals*. *Annual Review of Psychology*, 53, 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>

Eichler, J. F., & Peeples, J. (2016). *Flipped classroom modules for large enrollment general chemistry courses*. *Journal of Chemical Education*, 93(9), 1566–1573. <https://doi.org/10.1021/acs.jchemed.5b01063>

Flavell, J. H. (1979). *Metacognition and cognitive monitoring*. *American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). *School engagement: Potential of the concept, state of the evidence*. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>

Hofer, B. K., & Pintrich, P. R. (2012). *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Routledge.

Johnstone, A. H. (1991). *Why is science difficult to learn?* *Journal of Computer Assisted Learning*, 7(2), 75–83. <https://doi.org/10.1111/j.1365-2729.1991.tb00230.x>

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). *Why minimal guidance during instruction does not work*. *Educational Psychologist*, 41(2), 75–86. [https://doi.org/10.1207/s15326985ep4102\\_1](https://doi.org/10.1207/s15326985ep4102_1)

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.

Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). *Physical experience enhances science learning*. *Psychological Science*, 26(6), 737–749.

<https://doi.org/10.1177/0956797615569355>

Kuhn, D. (2008). *Formal operations from a twenty-first century perspective*. *Human Development*, 51(1), 48–55. <https://doi.org/10.1159/000113155>

Najami, N., Hugerat, M., Khalil, K., & Hofstein, A. (2019). *Effectiveness of teaching science by drama*. *Creative Education*, 10(1), 97–110. <https://doi.org/10.4236/ce.2019.101007>

Ødegaard, M. (2003). *Dramatic science: A critical review of drama in science education*. *Studies in Science Education*, 39(1), 75–101. <https://doi.org/10.1080/03057260308560196>

Otter, C. (2020). *The use of drama in A Level Chemistry [Doctoral dissertation]*. University of York.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Piaget, J. (1972). *Intellectual evolution from adolescence to adulthood*. *Human Development*, 15(1), 1–12. <https://doi.org/10.1159/000271225>

Sweller, J. (1988). *Cognitive load during problem solving: Effects on learning*. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)

Taber, K. S. (2013). *Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education*. *Chemistry Education Research and Practice*, 14(2), 156–168. <https://doi.org/10.1039/C3RP00012E>

Talanquer, V. (2011). *Macro, submicro, and symbolic: The many faces of the chemistry triplet*. *International Journal of Science Education*, 33(2), 179–195.

<https://doi.org/10.1080/09500693.2010.530551>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wilson, M. (2002). *Six views of embodied cognition*. *Psychonomic Bulletin & Review*, 9(4), 625–636. <https://doi.org/10.3758/BF03196322>

## **Technological Implementation: AI Adoption in Kerala CBSE Schools Amid Systemic Barriers**

Mariya Kutty

University College Fairview

### ***Abstract***

*This sequential explanatory mixed-methods study investigated Artificial Intelligence (AI) adoption among Central Board of Secondary Education (CBSE) teachers in Kerala, India (N = 106). Drawing on the Technology Acceptance Model and UNESCO's human-centred AI framework, the study examines teacher motivation moderated by ecosystem constraints. A four-stage integration pathway (Awareness-Efficiency-Augmentation-Transformation) is proposed to guide equitable implementation.*

*Survey results revealed a pronounced awareness–practice gap: while 75.1% of teachers (80/106) reported understanding AI's educational potential, only 28.3% (30/106) said they used it regularly in classrooms—a 46.8 percentage-point difference. Adoption patterns highlighted systemic inequities, with a disparity between urban private schools (64.3%, 9/14) and rural government schools (7.7%, 1/13). Teachers identified three interconnected barriers: infrastructural precarity (77% in rural government schools, 10/13), insufficient practice-oriented training (62% overall, 66/106), and the absence of school-level AI governance frameworks (77%, 82/106). Qualitative interviews (n = 20), analysed through reflexive thematic analysis, generated four explanatory themes: infrastructural precarity, training-to-practice disconnect, ethical uncertainty, and cautious optimism. Findings challenge deficit-oriented narratives of teacher resistance and argue that AI integration in developing contexts requires coordinated investment in infrastructure, professional development, and governance reform aligned with National Education Policy 2020 priorities.*

**Keywords:** *artificial intelligence; teacher adoption; CBSE; Kerala; digital equity; mixed methods*

## 1. Introduction

Artificial Intelligence has moved rapidly from speculative innovation to everyday reality in education. In recent years, generative AI systems, adaptive learning platforms, and automated assessment tools have begun to reshape classrooms worldwide (Holmes & Tuomi, 2022; Chiu et al., 2023). Advocates highlight AI's potential to personalise learning, enhance formative feedback, and reduce administrative workload (Tapalova & Zhiyenbayeva, 2022). Yet adoption remains uneven. High-income contexts report increasing integration, while low- and middle-income regions face persistent barriers such as weak infrastructure, limited professional development, and unclear governance (Scherer et al., 2023; Selwyn, 2022). This disparity raises urgent equity concerns: will AI help close learning gaps, or will it deepen them?

In India, the National Education Policy (NEP) 2020 positions technology-enabled learning as a pathway to equity and systemic efficiency (Ministry of Education, 2020). AI is framed as a transformative tool for personalised learning within the CBSE system—one of the world's largest, serving over 27,000 schools and 21 million students (CBSE, 2024). Yet policy aspiration does not automatically translate into classroom practice. The translation of national vision into local action depends on multiple factors: infrastructure readiness, teacher preparation, and institutional governance.

Kerala provides a compelling case study. Known for its high literacy rate (96.2%) and strong government-led technology initiatives such as the Kerala Infrastructure and Technology for Education programme (KITE, 2024), the state nonetheless reveals a visible digital divide. Its diverse CBSE school types—urban private, urban government, rural private, and rural government—reflect socio-economic variation within a high-performing system, offering a microcosm for examining structural mediators of AI adoption. Understanding adoption patterns in this context can illuminate both barriers and pathways relevant to other developing regions.

Despite policy enthusiasm for AI integration in Indian education, empirical evidence on actual adoption patterns remains limited. Most discussions occur at the level of policy discourse

rather than classroom practice. Without an empirical baseline, efforts to promote AI integration risk being misdirected or reinforcing existing inequities.

This research provides the first empirical baseline of AI adoption in Kerala's CBSE schools, offering data that can inform policy and practice. By examining disparities across school contexts, it highlights equity dimensions often overlooked in technology integration discourse. By centering teacher perspectives, it challenges deficit narratives that attribute low adoption to teacher resistance, revealing instead the structural constraints within which teachers work. The study also proposes a four-stage integration pathway offering practical guidance for schools and policymakers seeking equitable AI implementation aligned with NEP 2020 priorities.

This study focuses on CBSE school teachers in Kerala, India. The sample includes 106 survey respondents and 20 interview participants, drawn from urban, semi-urban, and rural contexts across government, aided, and private schools. Limitations include that the sample may not be representative of all CBSE schools in Kerala or India; self-report data may be subject to social desirability bias; the cross-sectional design captures adoption at a single point; and the study focuses on teacher perspectives without including student voices or direct classroom observations.

## **2. Literature Review**

Artificial Intelligence technologies in education include generative content tools, intelligent tutoring systems, adaptive learning platforms, and automated assessment engines. Studies consistently show that when these tools are used thoughtfully, they can improve personalisation, student engagement, and formative feedback (Holmes & Tuomi, 2022; Chiu et al., 2023). At the same time, scholars caution against assuming technology alone will transform learning. Outcomes are shaped by structural inequality, infrastructural readiness, digital literacy, and governance frameworks (Selwyn, 2022). In settings with limited digital access, AI may widen divides rather than close them. Systematic reviews confirm that adoption patterns are highly context-dependent, and that Global South contexts remain underrepresented in empirical research (Zawacki-Richter et al., 2019; Bond et al., 2024).

The Technology Acceptance Model (Davis, 1989) argues that perceived usefulness and ease of use predict whether people adopt new technologies. Meta-analyses confirm its relevance for AI adoption in schools (Scherer et al., 2019, 2023). Yet TAM is often applied as if these perceptions exist in isolation from context. In resource-variable settings, "ease of use" is not just a belief—it is shaped by infrastructure stability, device availability, and policy clarity. Recent studies suggest that ecosystem constraints moderate the relationship between intention and behaviour, requiring contextual extensions of TAM (Chiu et al., 2023; Karakose et al., 2023). This study adopts TAM as a foundational framework while extending it to account for structural mediators.

The TPACK framework (Mishra & Koehler, 2006) identifies the knowledge teachers need to integrate technology effectively: technological knowledge, pedagogical knowledge, and content knowledge, and their intersections. In AI integration, teachers require not only technical skills but also pedagogical understanding of how AI tools can support learning goals, and content knowledge to evaluate AI-generated materials critically. Recent research suggests that many professional development programmes focus on technological knowledge while neglecting pedagogical and content dimensions, limiting teachers' ability to apply AI meaningfully (Ayanwale et al., 2022).

UNESCO (2021, 2023) has emphasised human-centred AI principles such as transparency, accountability, privacy, and oversight. Ethical governance frameworks are essential for building trust. In schools without clear policies, teachers hesitate to experiment, worried about data misuse or liability. Selwyn (2022) notes that governance clarity is central to teacher confidence. Recent studies in developing contexts confirm that policy ambiguity is a major barrier to AI integration (Ayanwale et al., 2022; Celik et al., 2024). Teachers need institutional guidance to navigate complex ethical terrain: data privacy, algorithmic fairness, academic integrity, and student safety.

Research on the digital divide shows that technology adoption often mirrors socio-economic stratification (Selwyn, 2022; van Dijk, 2020). Without deliberate intervention, AI risks reinforcing inequities. Comparative studies in Asia and Africa reveal that rural schools face compounded barriers: unreliable electricity, limited internet, and scarce devices (Scherer et al., 2023; Ayanwale et al., 2022). Kerala's CBSE system, with its mix of urban and rural schools, provides a microcosm for examining these inequities. The concept of digital divide

has evolved from simple access distinctions to encompass skills, usage, and outcomes—what van Dijk (2020) terms the deepening divide.

Despite growing literature, three gaps remain clear. First, few empirical studies quantify rural–urban disparities in CBSE AI adoption. Second, TAM is rarely integrated with ethics frameworks in K–12 contexts. Third, no staged implementation pathway has been proposed for equitable AI integration in Indian schools. This study addresses these gaps by providing empirical baseline data, integrating TAM with UNESCO's ethical framework, and proposing a four-stage pathway for adoption.

### **3. Methodology**

#### **Research Design**

This study adopted a sequential explanatory mixed-methods design (Creswell & Plano Clark, 2018). Phase 1 involved a quantitative survey of CBSE teachers (N = 106), followed by Phase 2 qualitative interviews (n = 20). The design allowed broad statistical mapping of adoption patterns and then deeper exploration of teacher experiences. Using mixed methods ensured both breadth and nuance, with triangulation across data sources strengthening validity.

#### **Context and Setting**

The research was conducted in Kerala, India. Kerala's CBSE schools span urban private, urban government, rural private, and rural government institutions, reflecting diverse socio-economic contexts. This heterogeneity provided a natural setting to examine how structural conditions shape AI adoption.

#### **Participants**

The survey sample included 106 teachers: 78 from urban schools (73.6%) and 28 from rural schools (26.4%); 38 primary teachers (35.8%) and 68 secondary teachers (64.2%); 65 with prior AI training (61.3%) and 41 without (38.7%). Teaching experience ranged from early-career (0–5 years: 17.0%) to veteran (16+ years: 38.7%).

For qualitative interviews, 20 teachers were purposively selected to capture variation across location, grade level, experience, and AI use. The sample included 8 urban, 6 semi-urban, and 6 rural teachers; 9 primary and 11 secondary teachers; 5 early-career, 8 mid-career, and 7 veteran teachers; and 6 non-users, 8 occasional users, and 6 regular users.

## **Instruments**

The survey measured AI awareness, frequency of use, perceived usefulness, infrastructure access, training exposure, and policy awareness. Questions were adapted from established scales (Scherer et al., 2019) and contextualised for CBSE schools. Content validity was established through expert review by five independent specialists. A pilot with 15 teachers not included in the main sample yielded strong reliability (Cronbach's  $\alpha = 0.89$ ).

Semi-structured interview protocols explored lived experiences of AI experimentation, perceived barriers, ethical concerns, and institutional support, following established qualitative research guidelines (Braun & Clarke, 2021).

## **Research Questions**

This study addressed the following research questions:

What is the current level of AI awareness and adoption among CBSE teachers in Kerala?

How do adoption patterns vary across school contexts (urban/rural, government/private)?

What barriers do teachers perceive as limiting AI integration in their classrooms?

How do teachers experience and make sense of AI experimentation in their professional practice?

What systemic factors enable or constrain the translation of AI awareness into classroom practice?

## **Data Collection**

Surveys were distributed electronically via Google Forms and in paper format to reduce digital divide bias. Data collection took place between October 2024 and February 2025. Twenty individual interviews (45–60 minutes each) were conducted in English or Malayalam. All sessions were audio-recorded with consent and transcribed verbatim.

### **Data Analysis**

Quantitative survey data were analysed using SPSS Version 28. Descriptive statistics summarised adoption levels, awareness rates, and barrier distributions.

Qualitative data were analysed using reflexive thematic analysis (Braun & Clarke, 2021). Initial coding produced 248 codes, consolidated into 18 categories and finally four core themes. Thematic saturation was reached after the fifteenth interview.

Quantitative patterns were explained through qualitative insights at the interpretation stage, ensuring coherence across strands.

### **Ethical Considerations**

Ethical approval was granted by the University College Fairview Research Ethics Committee. Informed consent was obtained from all participants. Confidentiality was maintained through anonymisation of transcripts and secure data storage.

### **Trustworthiness**

Quantitative validity was ensured through expert review, pilot testing, and reliability analysis ( $\alpha = 0.89$ ). Qualitative trustworthiness was strengthened through member checking ( $n = 5$ ), peer debriefing, prolonged engagement, and a detailed audit trail (Lincoln & Guba, 1985). Triangulation across survey and interview data enhanced confirmability.

## **4. Results**

Survey responses were obtained from 106 CBSE teachers across Kerala. Due to small cell sizes when disaggregating by both location and management type, raw counts are reported alongside percentages for categories with fewer than 10 respondents.

**The readiness–implementation gap.** Survey findings revealed a striking paradox. Three out of four teachers (75.1%, 80/106) reported that they understood AI's role in education, and nearly one in five (19.2%, 20/106) described themselves as strongly confident. Yet when asked how often they actually used AI in their teaching, fewer than one in three (28.3%, 30/106) reported regular use, and only 14.2% (15/106) said they used AI daily or weekly. Another quarter (27.9%, 30/106) described themselves as experimenting—trying AI occasionally without integrating it into routine practice. This left a gap of 46.8 percentage points between those who felt ready and those who actually used AI. One secondary teacher captured this tension: "We know AI is coming. We even know how it could help. But when I return to my classroom, I don't have the tools or the time to make it real."

**Adoption patterns across school contexts.** Analysis revealed stark disparities. In urban private schools, 64.3% (9/14) of teachers reported regular AI use. In rural government schools, the figure was 7.7% (1/13)—a substantial difference between the most advantaged and the most disadvantaged contexts. Table 1 presents the full breakdown.

**Table 1. AI Adoption by School Context (N = 106)**

School Context	N	Regular Use (%)	Raw Count	Awareness (%)
Urban Private	14	64.3	9/14	89.2
Urban Aided	12	33.3	4/12	78.5
Urban Government	18	22.2	4/18	71.3
Semi-urban Private	10	30.0	3/10	74.2
Semi-urban Aided	9	22.2	2/9	70.1
Semi-urban Government	15	20.0	3/15	65.8
Rural Private	8	12.5	1/8	58.4
Rural Aided	7	14.3	1/7	62.7
Rural Government	13	7.7	1/13	72.4
<b>Overall</b>	<b>106</b>	<b>28.3</b>	<b>30/106</b>	<b>75.1</b>

*Note: Percentages for categories with N < 10 are reported alongside raw counts to avoid misleading precision.*

A teacher in a rural government school explained: "We hear about AI, but here we are still struggling with chalk and board. It feels like another world." In contrast, an urban private school teacher described using AI to gamify lessons: "My students were disappointed when the bell rang. They wanted to keep playing the AI quiz."

**Barriers in daily practice.** When teachers explained what held them back, three themes emerged: infrastructure, training, and trust.

**Infrastructure:** Half of all teachers (50.5%, 54/106) pointed to inadequate resources—unreliable internet, insufficient devices, classrooms where technology was more aspiration than reality. Among rural government teachers, 76.9% (10/13) reported infrastructure barriers. One teacher put it simply: "Electricity is available only four hours a day."

**Training:** Slightly more than half of teachers (52.4%, 56/106) felt that professional development had not prepared them for real classroom use. Rural teachers reported training inadequacy at 78.6% (22/28). Teachers distinguished between knowing what AI is and knowing how to use it. As one explained: "Training showed us PowerPoint slides about AI, not how to actually use it in the classroom."

**Trust:** About a third of teachers (33.3%, 35/106) expressed concerns about whether AI could be relied on. They worried about accuracy: "Sometimes AI produces wrong or irrelevant content." They worried about fairness: "Will it treat all my students fairly?"

**Grade-level differences.** Grade level also shaped adoption. Secondary teachers (33.8%, 23/68) used AI at nearly twice the rate of primary teachers (18.4%, 7/38). More than a third of secondary teachers (36.8%, 25/68) had never tried AI; among primary teachers, the figure was three-fifths (60.5%, 23/38). Interviews suggested why. Secondary teachers often saw AI as useful for preparing assessments, analysing student performance, and gamifying lessons. Primary teachers, by contrast, worried about age-appropriateness and the risk of reducing creativity. One primary teacher reflected: "My children need to draw, sing, and play. AI cannot replace that."

**Training and adoption.** Training emerged as a critical factor. Among teachers who had received prior AI training (61.3%, 65/106), 41.5% (27/65) reported regular AI use. Among those without training (38.7%, 41/106), only 12.2% (5/41) reported regular use. Yet

interviews revealed frustration with the quality of training. Teachers described sessions that remained abstract, focusing on definitions rather than applications. One teacher explained: "Training should show us real classroom examples. Otherwise, it feels like a lecture, not preparation."

**Experience and adoption.** Teaching experience showed variation. Early-career teachers (0–5 years) reported the highest regular use (38.9%, 7/18), followed by mid-career teachers (29.8%, 14/47), and veteran teachers (19.5%, 8/41). Interviews suggested that younger teachers often viewed AI as a lifeline for managing workload. One early-career teacher explained: "I am new, I have so much to prepare. AI helps me survive." A veteran teacher countered: "I have taught for 20 years. I know my students. A machine cannot understand a child's personal struggle."

**Ethical concerns and policy gaps.** More than half of teachers (57.6%, 61/106) felt their schools had adequate physical resources for AI, but fewer than half (43.6%, 46/106) reported having formal ethics policies. More than three-quarters (77.4%, 82/106) said their schools had no guidelines at all for AI use. Teachers in schools without policies expressed anxiety: "Where is the data from my students going? Who has access to it? The school has no policy. We are flying blind." Teachers in schools with policies felt more secure: "At least we know the rules. It gives us confidence."

**Which tools teachers use.** Among those who did use AI, ChatGPT dominated, mentioned by 42.9% (12/28) of users, used for lesson planning, creating materials, designing assessments, and drafting parent communications. Canva AI followed (28.6%, 8/28), popular for creating presentations and visual resources. A smaller number had discovered specialised educational AI tools such as Magic [school.ai](#) (14.3%, 4/28) and Gemini (11.9%, 3/28). Teachers described these tools as helpful but limited. One said: "ChatGPT saves me time, but sometimes it gives wrong answers. I must check everything."

**Perceived benefits of AI.** Despite barriers, teachers saw real potential in AI. More than four-fifths believed AI could reduce administrative workload (88.7%, 94/106), enable personalised instruction (84.9%, 90/106), and enhance feedback (82.1%, 87/106). A secondary teacher shared: "I used a gamified AI platform for a history quiz, and for the first time, my students were disappointed when the bell rang." A rural science teacher described

virtual labs as a lifeline: "We can now do experiments with no equipment. They are not just learning; they are doing science."

**Future outlook.** When asked about the future, nearly all teachers (95.3%, 101/106) said AI would be essential in CBSE schools within five years. As one teacher reflected: "AI cannot replace me. But it can help me be a better teacher."

## 5. Discussion

**Teacher motivation amid systemic barriers.** Teachers in this study expressed genuine enthusiasm for AI's potential—its ability to save time, personalise learning, and spark engagement. Nearly all believed it would soon become essential in their classrooms. Yet between this vision and daily practice stood persistent barriers: unreliable infrastructure, training that remained abstract, and the absence of clear policies. Teachers are not resisting AI, nor are they sceptical without reason. Their motivation is real, but it is hemmed in by circumstances beyond their control (Karakose et al., 2023). Similar dynamics have been observed in other resource-variable contexts, where enthusiasm is tempered by systemic misalignment (Fu, Weng, & Wang, 2024; Celik et al., 2024).

**The readiness–implementation gap.** The gap between conceptual understanding and regular use—75% ready, 28% using—is the central finding of this study. Teachers know what AI is. They have attended workshops, heard lectures, and seen demonstrations. They can describe its potential. But this declarative knowledge has not translated into procedural knowledge—the ability to apply AI meaningfully in their own classrooms (Ayanwale et al., 2022). This distinction matters for professional development. Teachers are not asking for more awareness campaigns. They are asking for training that moves beyond theory to practice, beyond PowerPoint slides to classroom application (Chiu et al., 2023).

**The structural nature of inequality.** The disparity between urban private schools (64.3%) and rural government schools (7.7%) reveals something fundamental about AI integration in India. This is not about enthusiasm or resistance. It is about access. Teachers in well-resourced schools have reliable electricity, working internet, available devices, and supportive administrators. Teachers in under-resourced schools do not (Scherer et al., 2023).

As one rural teacher explained: "Electricity is available only four hours a day." These are not excuses—they are descriptions of daily reality.

**AI as partner, not replacement.** Teachers consistently positioned AI as an augmentative tool rather than a replacement for human judgment. They welcomed its help with routine tasks because it freed them to focus on students. They valued its ability to personalise learning and appreciated its creativity. Yet they also recognised what AI cannot do. A veteran teacher reflected: "I have taught for 20 years. I know my students. A machine cannot understand a child's personal struggle." These voices challenge narratives that frame AI as threatening teachers' roles (Tapalova & Zhiyenbayeva, 2022; Selwyn, 2022).

**The ethical vacuum.** The absence of clear policies is not a minor issue. More than three-quarters of teachers (77%) reported that their schools had no guidelines at all for AI use. Teachers are navigating complex ethical terrain—data privacy, algorithmic fairness, academic integrity—alone (UNESCO, 2021). One teacher's words capture the anxiety this creates: "Where is the data from my students going? Who has access to it? The school has no policy. We are flying blind." Teachers in schools with policies felt more secure. Policy clarity does more than provide guidance—it builds trust (UNESCO, 2023; Celik et al., 2024).

**A four-stage pathway for equitable AI integration.** Based on findings, a four-stage pathway for equitable AI integration in schools is proposed. Stage 1 (Awareness and Access) requires ensuring basic infrastructure and building foundational awareness. Without reliable access, later stages cannot proceed. Stage 2 (Efficiency and Experimentation) requires hands-on opportunities for teachers to experiment with AI tools in low-stakes contexts, focusing on reducing workload. Stage 3 (Augmentation and Integration) involves teachers integrating AI deliberately into pedagogical practice with pedagogical support. Stage 4 (Transformation and Innovation) involves teachers moving beyond using AI for existing tasks to reimagining teaching and learning, requiring ethical literacy and institutional support. This pathway acknowledges different starting points. For well-resourced urban schools, Stages 1 and 2 may already be accomplished. For under-resourced rural schools, Stage 1 remains the immediate priority.

**Theoretical contributions.** This study extends existing frameworks. First, it highlights the distinction between declarative knowledge (knowing what AI is) and procedural knowledge (knowing how to use it). Second, it shows that perceived ease of use—a central construct in

TAM—is shaped by structural conditions. Third, it provides empirical evidence for an ethical policy vacuum in Indian educational AI implementation. Fourth, it captures the current state of AI integration in low-resource contexts: motivated teachers constrained by systemic barriers.

**Practical implications.** For school leaders, infrastructure must be treated as a foundation, professional development must move from awareness to application, schools must develop clear ethical policies, and leaders should respond to different teacher starting points. For policymakers, the findings underscore the urgency of addressing structural inequities. The disparity between urban private and rural government schools is not sustainable. Policy attention must focus on closing the infrastructure divide.

**Limitations and future research.** This study has several limitations. The cross-sectional design captures adoption at a single point. The study focuses on teacher perspectives without including student voices or classroom observations. The sample, while diverse, is limited to Kerala. The proposed four-stage pathway requires empirical validation through implementation research. Future research should trace teachers' AI adoption longitudinally, examine student perspectives and learning outcomes, compare adoption across Indian states, and validate the proposed pathway.

## 6. Conclusion

**Summary of the study.** AI adoption among CBSE teachers in Kerala reveals motivated teachers constrained by systemic barriers. Survey data from 106 teachers showed an awareness–practice gap of 46.8 percentage points. Adoption disparities were substantial. Teachers identified infrastructure inadequacy, insufficient practice-oriented training, and absence of governance frameworks as key barriers. Qualitative interviews generated four explanatory themes: infrastructural precarity, training-to-practice disconnect, ethical uncertainty, and cautious optimism.

**Implications.** Infrastructure must be treated as a prerequisite. Professional development must shift to practice-oriented, subject-specific application. Schools must develop clear ethical

guidelines for AI use. Equity must be central to AI integration efforts; without deliberate intervention, AI will widen existing achievement gaps.

**Recommendations for future research.** Future research should trace teachers' AI adoption longitudinally, examine student perspectives and learning outcomes, compare adoption across Indian states, and validate the proposed four-stage pathway through implementation research.

**Concluding thoughts.** AI adoption in Kerala's CBSE schools reflects motivated teachers facing systemic barriers that prevent consistent practice. This challenges deficit-oriented narratives of teacher resistance. Sustainable AI integration requires coordinated reform across infrastructure, professional development, and governance ecosystems. As one teacher reflected: "AI cannot replace me. But it can help me be a better teacher." Ensuring that all teachers have that opportunity is the work ahead.

## References

Ayanwale, M. A., Adeyemi, O. J., & Olatunji, O. A. (2022). *Teachers' readiness and barriers to AI adoption in African schools. Education and Information Technologies, 27(8), 11245–11262.* <https://doi.org/10.1007/s10639-022-11145-7>

Bond, M., Zawacki-Richter, O., & Marín, V. I. (2024). *Equity and artificial intelligence in education: A systematic review. International Journal of Educational Technology in Higher Education, 21(1), 12.* <https://doi.org/10.1186/s41239-024-00421-3>

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide.* Sage.

Celik, I., Yildiz, E., & Aydin, S. (2024). *Barriers and opportunities in AI adoption in K–12 education: A comparative study. Education and Information Technologies, 29(2), 145–162.* <https://doi.org/10.1007/s10639-023-12345>

Chiu, T. K. F., et al. (2023). *Opportunities and challenges of AI in education. Computers & Education: Artificial Intelligence, 4, 100118.* <https://doi.org/10.1016/j.caeai.2023.100118>

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research (3rd ed.)*. Sage.

Davis, F. D. (1989). *Perceived usefulness, perceived ease of use, and user acceptance of information technology*. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>

Fu, Q., Weng, X., & Wang, Y. (2024). *AI adoption in Asian schools: A systematic review of equity and infrastructure*. *British Journal of Educational Technology*, 55(1), 45–63. <https://doi.org/10.1111/bjet.13321>

Holmes, W., & Tuomi, I. (2022). *State of the art and practice in AI in education*. *European Journal of Education*, 57(4), 542–570. <https://doi.org/10.1111/ejed.12533>

Karakose, T., et al. (2023). *A systematic review of artificial intelligence in education in the 21st century*. *Education Sciences*, 13(2), 178. <https://doi.org/10.3390/educsci13020178>

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Ministry of Education. (2020). *National Education Policy 2020*. Government of India.

Mishra, P., & Koehler, M. J. (2006). *Technological pedagogical content knowledge: A framework for teacher knowledge*. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>

Scherer, R., Siddiq, F., & Tondeur, J. (2019). *The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education*. *Computers & Education*, 128, 13–35. <https://doi.org/10.1016/j.compedu.2018.09.009>

Scherer, R., Siddiq, F., & Tondeur, J. (2023). *Meta-analysis of AI adoption in schools: Psychological and contextual predictors*. *Computers & Education*, 195, 104623. <https://doi.org/10.1016/j.compedu.2023.104623>

Selwyn, N. (2022). *Education and technology: Key issues and debates (3rd ed.)*. Bloomsbury.

Tapalova, A., & Zhiyenbayeva, L. (2022). *AI in education: Teachers' perspectives on opportunities and risks*. *Education and Information Technologies*, 27(6), 7893–7912.  
<https://doi.org/10.1007/s10639-022-10987-2>

UNESCO. (2021). *AI and education: Guidance for policymakers*. UNESCO Publishing.

UNESCO. (2023). *Human-centred AI in education: Policy recommendations*. UNESCO Publishing.

van Dijk, J. (2020). *The digital divide*. Polity Press.

Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). *Systematic review of research on artificial intelligence applications in higher education*. *International Journal of Educational Technology in Higher Education*, 16(1), 39.  
<https://doi.org/10.1186/s41239-019-0171-0>

## **A Conceptual Framework for Reflexive Autoethnography in CFL Teacher Research**

Lee Boon Keat

University College Fairview

### ***Abstract***

*The global expansion of Chinese as a Foreign Language (CFL) education has led to increasingly diverse and multicultural classrooms, particularly in regions like Southeast Asia and the Middle East. Within these contexts, CFL teachers navigate complex linguistic and cultural landscapes, yet the internal decision-making processes that shape their pedagogical approaches remain critically underexplored. This conceptual paper addresses this gap by proposing a robust theoretical and methodological framework for investigating CFL teacher cognition through the lens of critical autoethnography. Synthesizing literature from teacher cognition, intercultural communication, and transformative learning theory, the paper argues that autoethnography offers a uniquely powerful tool for accessing the nuanced, context-dependent knowledge of CFL practitioners. The proposed framework integrates three core pillars: transformative learning theory, the principles of reflective practice, and the model of intercultural communicative competence. It positions the teacher as a reflexive instrument of inquiry, capable of generating rich, situated knowledge about the challenges and affordances of teaching CFL in multicultural settings. The paper proposes a novel integration of ICC as a teacher competency (not merely student goal) and operationalizes transformative learning through classroom micro-events, concluding with methodological implications.*

**Keywords:** *Autoethnography, CFL Education, Teacher Cognition, Reflective Practice, Intercultural Competence, Conceptual Framework*

## **Introduction**

The teaching of Chinese as a Foreign Language (CFL) has undergone a profound transformation in the 21st century, propelled by China's growing economic and geopolitical influence (Luo & Huang, 2020; Wang, F., 2023). Initiatives such as the Belt and Road Initiative have catalyzed a surge in Mandarin learning worldwide, creating classrooms that are more culturally and linguistically diverse than ever before.

In regions like Southeast Asia, particularly Malaysia with its multicultural fabric of Malay, Chinese, and Indian communities, CFL teachers encounter a complex tapestry of learner backgrounds, prior linguistic knowledge, and cultural expectations (Goh & Yunus, 2021; Tan & Abdullah, 2022). Similarly, in contexts such as Saudi Arabia, CFL educators must navigate the intersection of Chinese language instruction with deeply rooted Islamic cultural norms and educational traditions (Ahmad & Mustaffa, 2021; Al-Sudairi, 2020).

The paper is structured as follows: Section 1 introduction provides the context, Section 2 reviews the relevant literature on CFL in multicultural contexts, teacher cognition, and autoethnography. Section 3 presents the proposed conceptual framework, integrating transformative learning theory, reflective practice, and intercultural communicative competence, complete with a visual representation and illustrations from practice. Section 4 discusses the methodological implications of the framework, outlining how it can guide data collection, analysis, and the establishment of rigor in autoethnographic inquiry. Section 5 concludes with a summary and a discussion of the framework's implications for research and practice.

## **Role of the Teacher**

Within these dynamic environments, the CFL teacher is not merely a transmitter of linguistic knowledge but a cultural mediator, a curriculum negotiator, and a reflective practitioner constantly adapting to the needs of their learners (Li & Chen, 2021; Kaur & Chu, 2022). Despite the critical role of the teacher in shaping the learning experience, the internal cognitive and affective processes that underpin their pedagogical decisions in multicultural contexts remain significantly underexplored.

## **Research Gap**

Research on teacher cognition in language education has established that what teachers think, know, and believe profoundly influences their classroom practice (Borg, 2003; Borg, 2015; Kubanyiova & Feryok, 2015). However, much of this research has relied on third-person methodologies such as interviews and observations. While valuable, these methods may not fully capture the deeply personal, context-dependent, and often tacit nature of teacher knowledge (Mann & Walsh, 2017). What remains missing is an insider's perspective—a first-person account of how teachers actually experience, process, and learn from the moment-by-moment challenges of teaching in culturally complex settings.

## **Need for the Present Study**

This conceptual paper addresses this methodological and theoretical gap by proposing a novel framework for studying CFL teacher cognition. It argues that autoethnography, a qualitative research methodology that connects the personal to the cultural (Ellis, 2004; Ellis et al., 2011), offers a uniquely powerful lens through which to examine the lived experiences of CFL teachers in multicultural settings. By placing the researcher's own subjectivity at the center of inquiry, autoethnography enables a deep, nuanced exploration of how teachers make sense of critical incidents, negotiate cultural differences, and develop their professional identity over time (Chang, 2016; Denzin, 2014).

## **Purpose and Significance**

The purpose of this paper is to synthesize key theoretical perspectives from the literature to build a coherent conceptual framework for conducting reflexive autoethnographic research in CFL education. By integrating three core theoretical pillars, the framework provides a structured yet flexible guide for inquiry. It is significant because it offers CFL teacher-researchers a way to systematically investigate their own practice, generating knowledge that is both personally transformative and professionally relevant.

## Literature Review

The global spread of Mandarin has created diverse learning environments that challenge traditional, monolithic approaches to language teaching. In Malaysia, for example, CFL classrooms are inherently multicultural, bringing together learners from Malay, Chinese, and Indian backgrounds, each with distinct linguistic repertoires and cultural orientations towards learning (Goh & Yunus, 2021; Tan & Abdullah, 2022). This diversity presents both opportunities and challenges. Research by Rahim and Ooi (2020) highlights how Malaysian CFL teachers actively integrate local cultural content into their materials to make the language more relevant and accessible to their students. Similarly, studies on differentiated instruction in this context demonstrate the need for teachers to adapt their pedagogical strategies to accommodate the varying proficiency levels and learning styles of heritage and non-heritage learners (Lim, 2022; Tan & Abdullah, 2022).

In other parts of the world, CFL teachers face equally complex dynamics. In Muslim-majority contexts such as Saudi Arabia and Malaysia, educators must be attuned to cultural and religious sensitivities, ensuring that their teaching materials and classroom interactions are respectful and culturally responsive (Ahmad & Mustaffa, 2021; Al-Sudairi, 2020). The challenge is not merely about avoiding offense but about proactively building bridges between Chinese culture and local cultural frameworks. This requires teachers to possess a high degree of intercultural awareness and the ability to act as cultural mediators (Byram, 1997; Yuan & Xie, 2020).

Furthermore, the linguistic diversity of learners presents specific pedagogical challenges. Research by Wang and Nor (2019) on multilingual transfer in Malaysian Mandarin learning reveals how learners' prior knowledge of Malay, English, and other Chinese dialects can positively and negatively influence their acquisition of Mandarin. This finding underscores the need for CFL teachers to have a sophisticated understanding of comparative linguistics and to develop strategies that leverage learners' existing linguistic resources. The teaching of tones (Li & Lee, 2020) and Chinese characters (Zhang & Hu, 2022) remain perennial challenges, particularly for learners from alphabetic language backgrounds. This complex landscape demands that CFL teachers are not only linguistically proficient but also highly adaptive, reflective, and culturally attuned.

## **Teacher Cognition in Language Education**

The field of language teacher cognition has made significant strides in understanding the complex mental lives of teachers. Borg's (2003) seminal review defined teacher cognition as "the unobservable cognitive dimension of teaching – what teachers know, believe, and think" (p. 81). Subsequent research has demonstrated that these cognitions are shaped by a variety of factors, including teachers' own experiences as language learners, their teacher education programs, their classroom practice, and the broader institutional and societal context (Borg, 2015; Kubanyiova & Feryok, 2015).

In the context of CFL education, emerging research has begun to explore teacher cognition. Kaur and Chu (2022) conducted a qualitative study on reflective practice and teacher cognition in multicultural CFL classrooms in Malaysia, finding that teachers' beliefs about language learning and cultural difference were deeply intertwined with their instructional choices. Similarly, Zhao and Lin (2022) examined teacher cognition in international education, highlighting the dynamic interplay between teachers' personal and professional identities.

However, much of this research has been conducted using traditional qualitative methods, such as semi-structured interviews and classroom observations. These methods capture teachers' stated beliefs and observed behaviors but may not fully access the nuanced, moment-by-moment reflective processes that underpin their practice (Mann & Walsh, 2017). There is a growing recognition of the need for methodologies that can capture the "inner dialogue" of teachers as they navigate the complexities of the classroom (Farrell, 2022).

## **Autoethnography as a Methodology for Teacher Research**

Autoethnography is an approach to research and writing that seeks to describe and systematically analyze personal experience in order to understand cultural experience (Ellis et al., 2011). It treats research as a political, socially just, and socially conscious act (Adams & Jones, 2022). A researcher uses tenets of autobiography and ethnography to do and write autoethnography. Thus, as a method, autoethnography is both process and product.

The methodology has gained increasing traction in applied linguistics and teacher education as a means of exploring the lived experiences of practitioners (Keles, 2022; Hamilton et al., 2021). For language teachers, autoethnography offers a powerful tool for critically reflecting on their own practice, examining their identity formation, and making sense of the cultural and political dimensions of their work (Barkhuizen, 2021). By documenting their personal experiences, teachers can generate rich, contextualized knowledge that is often absent from more traditional research paradigms.

Several forms of autoethnography exist. Evocative autoethnography, championed by Ellis and Bochner (Bochner & Ellis, 2016), emphasizes the emotional and aesthetic power of storytelling to connect with readers and evoke empathy. Analytic autoethnography, proposed by Anderson (2006), aims to use personal experience as a basis for developing theoretical insights and contributing to existing knowledge. Critical autoethnography, as articulated by Boylorn and Orbe (2021), explicitly foregrounds issues of power, identity, and social justice, examining how broader cultural structures shape personal experience. For a CFL teacher researcher navigating multicultural classrooms, a critical and reflexive approach to autoethnography is particularly apt, as it allows for an examination of how one's own cultural positionality intersects with that of their students within a broader socio-political context.

However, autoethnography is not without its critics. Concerns have been raised about its potential for self-indulgence, lack of rigor, and ethical challenges related to representing others (Tolich, 2010). These critiques have led to the development of robust criteria for ensuring quality and trustworthiness in autoethnographic work, including the need for explicit reflexivity, transparency about the research process, and a commitment to relational ethics (Ellis, 2007; Lapadat, 2017). This paper's proposed framework directly addresses these concerns by embedding strategies for rigor within its very structure.

### **Conceptual Framework: Three Pillars of Reflexive Autoethnographic Inquiry**

Building on the literature reviewed above, this section presents an integrated conceptual framework for conducting reflexive autoethnographic research on CFL teacher cognition in multicultural contexts. The framework is not intended as a rigid prescription but as a flexible guide—a set of interconnected theoretical lenses and methodological commitments that can

inform and enrich the inquiry process. It rests on three pillars: Transformative Learning Theory, Reflective Practice, and Intercultural Communicative Competence.

### **Pillar 1: Transformative Learning Theory**

Mezirow's (1991) transformative learning theory provides a powerful lens for understanding how adults revise their meaning perspectives through critical reflection on experience. A "disorienting dilemma"—an experience that does not fit one's existing framework of understanding—can trigger a process of self-examination, critical assessment of assumptions, exploration of new roles, and ultimately, a transformation in perspective that leads to more inclusive and discriminating action.

For a CFL teacher in a multicultural classroom, disorienting dilemmas abound. Consider the experience of a teacher moving from Malaysia to Saudi Arabia:

*"When I played a video of families gathering for the New Year's Eve dinner, and the screen showed wine glasses, the entire classroom fell into silence. Not the kind of amazed silence—but the heavy, uncomfortable silence of something gone wrong. A student asked: 'Teacher, is wine haram? Why do Chinese people drink wine?'"* (Journal Entry, Week 1)

This moment represents a classic disorienting dilemma. The teacher's assumption—that materials effective in multicultural Malaysia would transfer seamlessly to conservative Saudi Arabia—was fundamentally challenged. The student's question forced a confrontation with unexamined assumptions about cultural neutrality in teaching materials.

By using autoethnography to systematically document and analyze such dilemmas, the teacher-researcher can trace their own process of perspective transformation. The reflective journal becomes a space for not only recording events but for excavating the assumptions underlying their reactions and for charting the evolution of their understanding over time. This pillar provides a theoretical rationale for focusing on critical incidents as key sites of learning and growth (Teh & Ching, 2023).

### **Pillar 2: Reflective Practice**

The second pillar, reflective practice, provides the methodological mechanism for the transformative process described above. Drawing on the work of scholars such as Farrell

(2022) and Mann and Walsh (2017), this framework conceptualizes reflection not as a casual afterthought but as a systematic, disciplined practice. Reflective practice involves deliberately and critically examining one's actions, beliefs, and the contexts in which they occur, with the goal of improving practice and deepening understanding.

In this framework, reflection operates at multiple levels. It encompasses reflection-in-action, the spontaneous, intuitive adjustments a teacher makes in the midst of teaching (Schön, 1983). It also includes reflection-on-action, the retrospective analysis of past events. Crucially, for autoethnography, it involves reflection-on-reflection, or reflexivity: a critical awareness of how the researcher's own subjectivity, values, and positionality shape the entire research process, from the questions asked to the interpretations made (Pillow, 2019). This meta-cognitive layer is what distinguishes a rigorous autoethnography from a simple memoir.

The journal entries demonstrate this reflective process in action. After the wine glass incident, the teacher reflects:

*"Later, in reflection, I realized I had been self-righteous. I had transplanted what worked in multicultural Malaysia—where many students are familiar with Chinese festivals—without sufficient cultural mediation. I hadn't previewed the materials through a Saudi lens. Those wine glasses weren't just props; they were symbols conflicting with deeply held values."* (Journal Entry, Week 1)

This is not mere description but critical analysis—an interrogation of assumptions and a recognition of the teacher's own role in the pedagogical failure. It is the "critical" in "critical reflexivity," ensuring that the researcher is constantly interrogating their own role in constructing the knowledge they present.

### **Pillar 3: Intercultural Communicative Competence**

The third pillar, Byram's (1997) model of intercultural communicative competence (ICC), provides the substantive focus for the inquiry. Byram's model moves beyond linguistic proficiency to encompass the attitudes, knowledge, and skills necessary for effective and appropriate communication across cultural boundaries. It includes five key components:

Component	Description
<b>Attitudes</b>	Curiosity and openness, readiness to suspend disbelief about other cultures and belief about one's own
<b>Knowledge</b>	Of social groups and their products and practices in one's own and in one's interlocutor's country, and of the general processes of societal and individual interaction
<b>Skills of interpreting and relating</b>	The ability to interpret a document or event from another culture, to explain it and relate it to documents or events from one's own
<b>Skills of discovery and interaction</b>	The ability to acquire new knowledge of a culture and cultural practices and the ability to operate knowledge, attitudes and skills under the constraints of real-time communication and interaction
<b>Critical cultural awareness</b>	An ability to evaluate, critically and on the basis of explicit criteria, perspectives, practices and products in one's own and other cultures and countries

For the CFL teacher, ICC is not just a goal for students but a professional imperative. The teacher's own ICC directly shapes their ability to create an inclusive classroom environment, select culturally appropriate materials, mediate cultural misunderstandings, and model intercultural sensitivity for their learners (Ahmad & Mustaffa, 2021; Yuan & Xie, 2020). By framing the autoethnographic inquiry within Byram's model, the teacher-researcher can systematically examine their own intercultural development.

The journal entry about teaching the character "十" (ten) illustrates this development:

*"A student raised his hand: 'Teacher, this looks like a cross. The cross is haram.' I paused and decided not to be defensive. 'Thank you for telling me, Y. You're right, this shape does resemble a cross. But this is the Chinese character for the number 'ten', as ordinary as our 1, 2, 3. It has no religious meaning whatsoever in Chinese culture; it simply represents the number 10.' Then I turned to the whiteboard and wrote the Arabic numeral '10' and the Roman numeral 'X' next to it. 'See, the Roman numeral for 10 is also X, like crossed lines. Different writing systems sometimes have similar shapes, but completely different meanings.'" (Journal Entry, Week 3)*

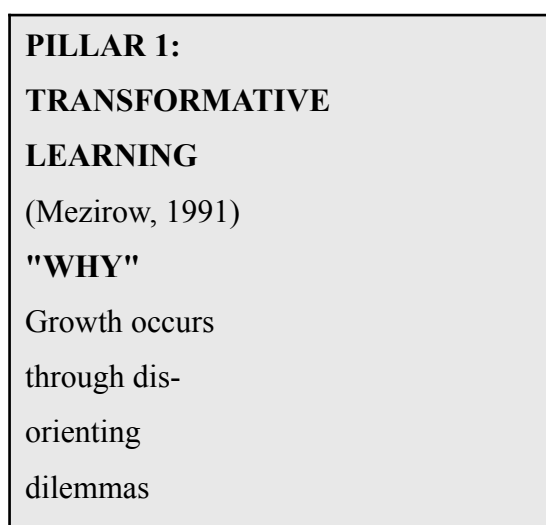
Yet I must confess: three weeks later, I made a similar mistake with the character '田', not recognizing its visual parallel to a religious symbol until a student pointed it out. Reflection does not guarantee immediate transformation; some lessons require repeated disorientation before new perspectives stabilize.

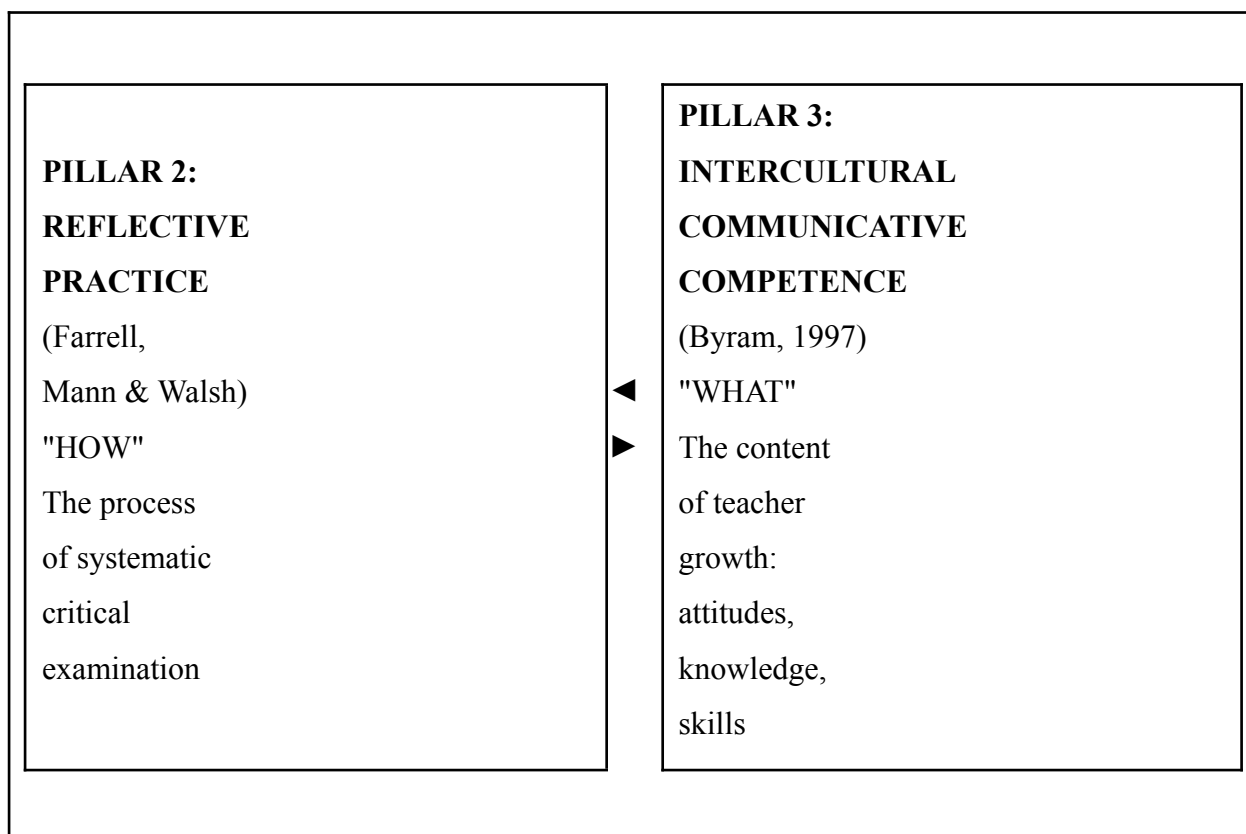
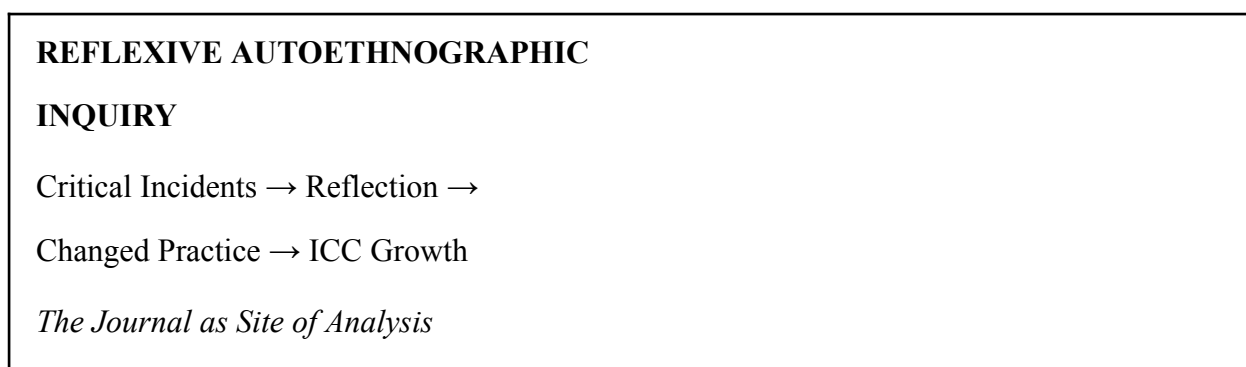
This response demonstrates several ICC components: attitudes of openness rather than defensiveness, skills of interpreting and relating (connecting the character to familiar Roman numerals), and critical cultural awareness (understanding why the shape might be sensitive). The reflective journal becomes a record of the ongoing development of one's own intercultural competence as it is enacted and refined in the crucible of classroom practice.

### **Integrating the Pillars: A Visual Framework**

The relationship between the three pillars is not linear but cyclical and mutually reinforcing. Figure 1 illustrates this dynamic interplay. Critically, these three pillars operate not as static categories but as a dynamic system. An ICC challenge (e.g., cultural misreading) surfaces as a disorienting dilemma when reflected upon, triggering perspective transformation; this transformed perspective then reshapes subsequent reflective capacity and ICC. Autoethnography's strength lies in capturing this non-linear, diachronic development through sustained journaling.

**Figure 1: Conceptual Framework for Reflexive Autoethnographic Inquiry in CFL Teacher Research**





As shown in Figure 1, the framework is anchored by **Transformative Learning Theory** (Pillar 1), which provides the theoretical rationale for *why* critical incidents matter and *why* growth occurs. A "disorienting dilemma" triggers a process of critical reflection. This reflection is not casual but disciplined and systematic, drawing on the principles of **Reflective Practice** (Pillar 2), which provides the *how*—the methodological mechanism for transformation.

The *content* of this reflection—*what* the teacher is actually learning to do better—is captured by **Intercultural Communicative Competence** (Pillar 3). Each critical incident becomes an opportunity to develop Byram's five components. The three pillars are bidirectional: developing ICC deepens the quality of reflective practice; reflective practice surfaces new

disorienting dilemmas; transformative learning reshapes the teacher's capacity for both reflection and intercultural mediation.

The autoethnographic method—structured journaling focused on critical incidents—makes this entire process visible and analyzable. The journal becomes both the site of reflection and the data for analysis, allowing the teacher-researcher to trace their own developmental trajectory across time.

### **Illustrations from Practice**

The calligraphy breakthrough entry provides a powerful example of the framework in action, demonstrating how the three pillars work together:

*"K raised his hand. 'Teacher,' he said, 'this is like when we learn to write beautifully. My grandfather does this with Arabic. He says the pen must be held with respect, just like you said about the brush.' Connection. He made the connection himself. I asked him to explain to the class. He stood up—this usually quiet boy—and demonstrated how his grandfather taught him to prepare the reed pen, how the ink must flow, how patience is everything. Other students leaned forward. For the next twenty minutes, we had a conversation—not me teaching them about Chinese culture, but us sharing what it means to treat writing as an art form, a spiritual practice, across our different traditions." (Journal Entry, Week 12)*

This moment can be analyzed through the framework as follows:

- **Transformative Learning:** The teacher's perspective shifted from "teaching Chinese culture" to "facilitating dialogue about a universal human practice." This represents a fundamental reorientation of professional identity.
- **Reflective Practice:** The teacher's conscious choice to frame calligraphy as "the art of writing" rather than "Chinese art" was a deliberate pedagogical decision informed by prior reflection on earlier incidents.
- **Intercultural Communicative Competence:** The moment demonstrates all five ICC components in action—attitudes of openness, knowledge of both cultures, skills of interpreting and relating, skills of discovery and interaction, and critical cultural awareness.

The pillars are not separate; they are mutually reinforcing. The framework makes visible what might otherwise remain invisible: the complex interplay of cognition, reflection, and cultural learning that constitutes teacher development in multicultural contexts.

### **Methodological Implications of the Framework**

This conceptual framework carries clear implications for the design and execution of an autoethnographic study. It suggests a methodology that is systematically reflective, focused on critical incidents, explicitly reflexive, rigorously analyzed, and ethically grounded.

#### **Systematically Reflective**

Data collection should be structured around regular, disciplined reflective journaling that goes beyond simple description to include critical analysis of assumptions, feelings, and evolving understandings. Prompts based on the three pillars can guide this process.

#### **Example Journal Prompts Aligned with the Framework:**

Pillar	Sample Prompt
Transformative Learning	"What assumptions about teaching and learning did I bring to this lesson? Was any assumption challenged?"
Reflective Practice	"How did I adapt my teaching in the moment when I sensed student confusion? What will I do differently?"
Intercultural Competence	"What does this interaction reveal about my own intercultural skills of discovery and interaction?"

#### **Focused on Critical Incidents**

The framework directs attention to specific, impactful moments in practice. These critical incidents become the primary units of analysis. The researcher documents the incident in rich detail, then systematically analyzes it through the lens of the three pillars.

A critical incident is defined by its impact, not its drama. A moment of silence, a student's question, a failed lesson, a breakthrough connection—each can serve as a focal point for deeper exploration of teacher cognition and development. A template for documenting critical incidents is provided in Appendix B.

### **Explicitly Reflexive**

The framework demands that the researcher continuously interrogate their own positionality. This involves keeping a separate reflexive journal to document how their personal history, cultural background, and evolving understanding of the research are shaping the inquiry. This meta-layer of reflection is essential for establishing trustworthiness (Brinkmann, 2022).

Reflexivity involves asking questions such as:

- How does my cultural background shape what I notice and what I miss?
- How do my emotional responses to incidents influence my interpretation?
- How am I changing as a teacher through this process of inquiry?

The teacher's acknowledgment of being "self-righteous" and having "transplanted what worked in multicultural Malaysia without sufficient cultural mediation" exemplifies this reflexivity.

### **Analysis**

Data analysis moves beyond simple thematic coding to a more theoretically informed interpretation. Drawing on reflexive thematic analysis (Braun & Clarke, 2022; Byrne, 2022), the researcher actively constructs themes that represent patterns of meaning related to transformative learning, reflective practice, and ICC.

The analysis should:

- Trace developmental trajectories across time
- Identify moments of disorientation and resolution

- Map evidence of ICC development
- Search for negative cases that complicate emerging themes
- Connect findings back to the theoretical framework

The goal is not just to describe what happened but to generate theoretical insights into the process of teacher development in multicultural contexts.

### **Ethically Grounded**

The framework is inherently committed to relational ethics (Ellis, 2007). The autoethnographer must be constantly mindful of their responsibilities to the students, colleagues, and others who appear in their narrative. Key ethical considerations include:

Consideration	Application
<b>Informed Consent</b>	Obtain consent from institution and, where possible, from students/parents
<b>Anonymity</b>	Use pseudonyms for all individuals and institutions
<b>Relational Ethics</b>	Consider the potential impact of writing on those represented; strive to portray them with fairness and respect
<b>Transparency</b>	Be clear with participants about the research purpose and how data will be used
<b>Ongoing Vigilance</b>	Ethics is not a one-time approval but a continuous process of reflection

The journal entries in this paper demonstrate this ethical awareness, with students identified only by initials and portrayed with respect and fairness, even when describing challenging moments.

### **Conclusion**

#### **Summary of the Framework**

This conceptual paper has proposed an integrated framework for conducting reflexive autoethnographic research on CFL teacher cognition in multicultural settings. By weaving together three core pillars—transformative learning theory, reflective practice, and intercultural communicative competence—the framework provides a coherent theoretical and methodological guide for exploring the rich, complex inner world of the CFL practitioner.

The framework positions autoethnography not as a soft or self-indulgent alternative to traditional research but as a rigorous and deeply illuminating methodology. It is capable of generating knowledge that is both personally meaningful and professionally significant. The three pillars work together cyclically: disorienting dilemmas trigger critical reflection; systematic reflective practice enables examination of assumptions and development of new perspectives; intercultural competence provides the content and goal of that development.

### **Implications for Practice and Research**

The implications of this framework are far-reaching across multiple domains.

**For CFL Teacher Education:** The framework suggests the value of incorporating autoethnographic approaches into training programs to foster critical reflection and intercultural awareness from the outset. Novice teachers could be guided to document and analyze their own critical incidents, learning to see disorienting dilemmas not as failures but as opportunities for growth. Teacher educators might use the three pillars as organizing principles for reflective assignments.

**For Curriculum Development:** The framework highlights the importance of creating materials and pedagogies that are responsive to the specific cultural contexts in which they are used. The teacher's experience in Saudi Arabia demonstrates that materials effective in one multicultural context cannot be assumed to transfer to another. Curriculum developers must build in space for local adaptation and teacher mediation, recognizing that the teacher's intercultural competence is as important as the materials themselves.

**For Future Research:** This framework offers a foundation for empirical autoethnographic studies that can provide rich, detailed accounts of teacher development in a variety of CFL settings. Future research directions include:

- Longitudinal autoethnographies tracking teacher development over multiple years

- Comparative autoethnographies examining teachers transitioning between different contexts
- Collaborative autoethnographies involving multiple teachers navigating similar transitions
- Studies incorporating student perspectives alongside teacher autoethnography
- Intervention studies examining whether structured autoethnographic reflection enhances teacher adaptation

### **Concluding Thoughts**

Ultimately, by illuminating the lived experience of the teacher, this framework contributes to a more nuanced, human-centered understanding of what it means to teach Chinese in an increasingly interconnected and culturally diverse world. As one journal entry concludes:

*"This is the reflective practice I wrote about—the cycle of experience, dissonance, critical reflection, and changed action. I'm living my own theoretical framework. It's hard. And it's necessary."* (Journal Entry, Week 16)

The framework proposed here offers a way to make that living, that difficulty, and that necessity visible. It provides a structured approach to transforming the inevitable struggles of cross-cultural teaching into opportunities for growth—for both teacher and students. In doing so, it affirms that the most valuable resource in multicultural CFL education is not a textbook or a curriculum, but a reflective, culturally-aware teacher willing to learn alongside their students.

## References

- Adams, T. E., & Jones, S. L. (2022). *Handbook of autoethnography* (2nd ed.). Routledge.
- Ahmad, J., & Mustaffa, M. S. (2021). Cultural responsiveness in teaching Mandarin as a foreign language: Perspectives from Muslim-majority classrooms. *International Journal of Language Education*, 5(2), 112-128.
- Al-Sudairi, M. (2020). *The Saudi-Chinese partnership: A new chapter in educational cooperation*. King Faisal Center for Research and Islamic Studies.
- Anderson, L. (2006). Analytic autoethnography. *Journal of Contemporary Ethnography*, 35(4), 373–395.
- Barkhuizen, G. (2021). Identity dilemmas of a teacher of Chinese as a foreign language. *System*, 103, 102645.
- Bochner, A. P., & Ellis, C. (2016). *Evocative autoethnography: Writing lives and telling stories*. Routledge.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), 81-109.
- Borg, S. (2015). *Teacher cognition and language education: Research and practice*. Bloomsbury.
- Boylorn, R. M., & Orbe, M. P. (Eds.). (2021). *Critical autoethnography: Intersecting cultural identities in everyday life* (2nd ed.). Routledge.
- Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.
- Brinkmann, S. (2022). Doing autoethnography. In *Qualitative inquiry and research design* (5th ed.). SAGE.
- British Educational Research Association (BERA). (2018). *Ethical guidelines for educational research* (4th ed.).

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity*, 56(3), 1391-1412.

Chang, H. (2016). *Autoethnography as method*. Routledge.

Denzin, N. K. (2014). *Interpretive autoethnography* (2nd ed.). SAGE.

Ellis, C. (2004). *The ethnographic I: A methodological novel about autoethnography*. AltaMira Press.

Ellis, C. (2007). Telling secrets, revealing lives: Relational ethics in research with intimate others. *Qualitative Inquiry*, 13(1), 3-29.

Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: An overview. \*Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 12\*(1).

Farrell, T. S. C. (2022). *Reflective practice in language teaching*. Cambridge University Press.

Goh, Y. S., & Yunus, M. M. (2021). The use of CLIL in Malaysian Chinese as a foreign language classrooms: A systematic review. *Journal of Language and Education*, 7(3), 45-59.

Hamilton, M. L., Smith, L., & Worthington, K. (2021). Fitting the methodology with the research: An exploration of narrative, self-study and autoethnography. *Studying Teacher Education*, 17(3), 251-265.

Kaur, D., & Chu, M. Y. (2022). Reflective practice and teacher cognition in CFL: A qualitative study in multicultural contexts. *Language Teaching Research*, 26(1), 88-105.

Keles, U. (2022). Autoethnography as a recent methodology in applied linguistics. *The Qualitative Report*, 27(2), 417-433.

Kubanyiova, M., & Feryok, A. (2015). Language teacher cognition in applied linguistics research: Revisiting the territory, redrawing the boundaries, reclaiming the relevance. *The Modern Language Journal*, 99(3), 435-449.

- Lapadat, J. C. (2017). Ethics in autoethnography and collaborative autoethnography. *Qualitative Inquiry*, 23(8), 589-603.
- Li, X., & Chen, W. (2021). Global Mandarin and the role of teacher mediation in non-immersive contexts. *Journal of Chinese Language Education*, 19(2), 34-50.
- Li, Y., & Lee, H. (2020). Tonal acquisition and pedagogical strategies for alphabetic language speakers. *Foreign Language Annals*, 53(4), 789-805.
- Lim, J. H. (2022). *Differentiated instruction in multilingual CFL classrooms: A Malaysian case study*. *RELC Journal*, 53(1), 145-162.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Luo, J., & Huang, X. (2020). *The Belt and Road Initiative and the global spread of Mandarin*. *Journal of Asian Pacific Communication*, 30(1-2), 123-145.
- Mann, S., & Walsh, S. (2017). *Reflective practice in English language teaching: Research-based principles and practices*. Routledge.
- Mezirow, J. (1991). *Transformative dimensions of adult learning*. Jossey-Bass.
- Pillow, W. (2019). Reflexivity. In P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, & R.A. Williams (Eds.), *SAGE research methods foundations*. SAGE.
- Rahim, A., & Ooi, C. L. (2020). *Integrating local culture in CFL materials: A study of teacher practices in Malaysia*. *Journal of Multilingual and Multicultural Development*, 41(8), 675-690.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Tan, K. E., & Abdullah, A. (2022). *Differentiated instruction and intercultural learning in Malaysian CFL classrooms*. *International Journal of Inclusive Education*, 26(5), 512-528.
- Teh, W. L., & Ching, L. S. (2023). *Transformative learning through reflective practice: A study of CFL teachers in Southeast Asia*. *Teaching and Teacher Education*, 121, 103912.

Tolich, M. (2010). A critique of current practice: Ten foundational guidelines for autoethnographers. *Qualitative Health Research, 20*(12), 1599-1610.

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837-851.

Wang, F. (2023). *Global CFL education: Policy, pedagogy, and challenges*. Springer.

Wang, L., & Nor, N. M. (2019). *Multilingual transfer in Mandarin language learning: Evidence from Malaysia*. *Journal of Psycholinguistic Research, 48*(4), 901-917.

Yuan, H., & Xie, M. (2020). *Exploring intercultural awareness in Mandarin teaching: A study of international classrooms*. *Journal of Intercultural Communication Research, 49*(4), 386–402.

Zhang, X., & Hu, W. (2022). *Learner challenges in Chinese character acquisition: Implications for CFL teaching*. *Language Teaching Research Quarterly, 14*(2), 22–35.

Zhao, Y., & Lin, X. (2022). *Re-examining teacher cognition in international education: A comparative study*. *Teaching and Teacher Education, 108*, 103555.

## **Culturally Inclusive Practices in International Schools: An Ecological Mixed-Methods Study**

Hemant Goyal

University College Fairview

### ***Abstract***

*This study examines how teachers in international schools understand and enact culturally inclusive practices (CIP), and the institutional conditions shaping their work. Framed by Bronfenbrenner's Ecological Systems Theory, CIP is conceptualised as an ecological phenomenon emerging from interactions between teacher beliefs, classroom practice, and institutional structures.*

*An explanatory sequential mixed-methods design was employed. Survey data from 53 teachers in Singapore and Malaysia identified patterns in understanding, enactment, perceived barriers, and institutional support. Follow-up interviews with six teachers explored how these patterns were experienced in practice.*

*Findings indicate strong conceptual commitment to CIP, but uneven enactment. A recurring commitment–capacity gap emerged, with implementation strongest in flexible classroom practices and weaker in structurally constrained domains such as assessment and instructional adaptation. Key constraints included time, access to culturally diverse resources, and variability in professional learning and institutional coherence.*

*The study reframes culturally inclusive practice as a systemically enabled endeavour and highlights the importance of alignment between teacher commitment and institutional conditions in sustaining inclusive pedagogy in international school contexts.*

**Keywords:** *culturally inclusive practices; international schools; ecological systems theory; teacher perspectives; mixed-methods research*

## **Introduction**

The 21st century has witnessed unprecedented human mobility, reshaping global demographics. Over 281 million people now live outside their countries of birth, with a further 725 million migrating within national borders (International Organization for Migration, 2024). These shifts continue to impact education systems, especially international schools which are characterised by high levels of ethnic, linguistic, and cultural diversity among both students and teachers (Bunnell, 2016; Hayden & Thompson, 2013). In response, international curricula such as the International Baccalaureate (IB) emphasise global citizenship, intercultural understanding, and international-mindedness as core educational aims (IBO, 2017).

Realising these aims, however, requires more than symbolic recognition of diversity. It calls for pedagogical approaches that meaningfully integrate students' cultural identities into teaching and learning. Such approaches are conceptualised as culturally inclusive practices (CIP), which extend beyond classroom strategies to encompass broader curricular, institutional, and relational dimensions of inclusion (Bennett, 2010; Holliday, 2011). In increasingly diverse international school contexts, CIP has become central to ensuring that all students experience belonging, engagement, and equitable access to learning.

## **Research Problem**

Despite widespread endorsement of inclusive values, the enactment of culturally inclusive practices remains uneven. While teachers often express strong commitment to inclusion, translating these principles into consistent classroom practice presents ongoing challenges (Civitillo et al., 2018; Romijn et al., 2021). This disconnect is particularly visible in areas requiring deeper pedagogical or structural adaptation, such as instructional design and assessment.

Several factors contribute to this gap. First, cultural inclusion remains a concept that is interpreted variably across contexts, leading to differences in how teachers understand and operationalise it (Glock et al., 2019; Knoblauch, 2023). Second, school-level approaches to diversity are frequently limited to symbolic or surface-level practices, with limited integration

into core pedagogy (Sleeter, 2001). Third, teachers operate within institutional environments shaped by time constraints, resource availability, collaboration structures, and policy expectations, all of which influence what is feasible in practice (Heng & Lim, 2021).

These challenges are amplified in Southeast Asian international school contexts, where high student mobility, multilingualism, and curricular hybridity create complex conditions for inclusive teaching. Yet empirical research examining how teachers in these contexts understand and enact culturally inclusive practices remains limited.

### **Purpose and Research Questions**

This study addresses this gap by examining how teachers in international schools conceptualise and enact culturally inclusive practices, the barriers they encounter, and the institutional conditions shaping their work. Guided by Bronfenbrenner's Ecological Systems Theory (1979), the study positions CIP as an ecological phenomenon shaped by interactions across classroom, relational, and institutional systems.

The study is guided by four research questions: what are international school teachers' understandings of culturally inclusive practices; how do teachers translate their understanding of culturally inclusive practices into classroom pedagogy; what barriers do teachers encounter when implementing culturally inclusive practices; and what forms of institutional support and professional learning do teachers perceive as effective in overcoming these barriers.

### **Significance of the Study**

This study contributes to the literature in three ways. First, it reframes culturally inclusive practice as a systemically shaped phenomenon rather than an individual pedagogical choice, foregrounding the interaction between teacher beliefs and institutional conditions. Second, it introduces the concept of a commitment–capacity gap to explain the discrepancy between teachers' inclusive intentions and their enacted practice. Third, it provides empirical insight into international school contexts in Southeast Asia, which remain underrepresented in research on culturally inclusive education.

## **Theoretical and Conceptual Framework**

This study is situated within literature on culturally inclusive and culturally responsive pedagogy, and is analytically framed using Bronfenbrenner's Ecological Systems Theory (EST). Together, these perspectives support an understanding of culturally inclusive practices (CIP) as both pedagogical and systemic, shaped through interactions between teacher beliefs, classroom practices, and institutional conditions.

### **Culturally Inclusive Practices in International School Contexts**

Culturally inclusive practices are broadly understood as pedagogical approaches that recognise, value, and integrate students' cultural and linguistic identities into teaching and learning processes (Banks, 2015; Gay, 2013). While often discussed alongside culturally responsive pedagogy (CRP), CIP in this study is conceptualised as extending beyond representation or celebration of diversity towards the sustained integration of students' identities within curriculum, instruction, and assessment.

This distinction is particularly significant in international school contexts, which are characterised by linguistic diversity, high student mobility, and heterogeneous cultural identities. In such environments, inclusion cannot be reduced to discrete cultural events or symbolic recognition, but must be embedded within everyday pedagogical decision-making. However, existing literature suggests that while teachers often express strong commitment to inclusive values, the translation of these values into consistent classroom practice remains uneven (Civitillo et al., 2018; Romijn et al., 2021).

### **An Ecological Perspective on Inclusive Practice**

To move beyond individualised accounts of teacher practice, this study adopts an ecological lens. Bronfenbrenner's Ecological Systems Theory conceptualises human development as shaped by interactions across nested systems, including the microsystem (immediate environments), mesosystem (relationships between settings), and exosystem (broader institutional structures).

Applied to culturally inclusive pedagogy, this framework enables analysis of how inclusion is shaped across multiple levels. At the microsystem level, teachers' beliefs, knowledge, and

reflective practices influence how inclusion is conceptualised and enacted. At the mesosystem level, collaboration, professional relationships, and shared routines mediate how practices are developed and sustained. At the exosystem level, leadership priorities, policy frameworks, resourcing, and professional learning structures shape the conditions under which inclusive practices are supported or constrained.

This ecological perspective shifts the analytical focus from individual teacher disposition to the broader conditions that enable or limit inclusive practice. It provides a conceptual foundation for examining why strong commitment to inclusion does not necessarily translate into consistent enactment.

**Table 1**

System	Definition	Application to CIP
Microsystem	Immediate settings such as classrooms and peer relationships	Teacher-student interactions, culturally responsive pedagogy, differentiation
Mesosystem	Interconnections between microsystems	Collaboration between teachers, engagement with families, team teaching
Exosystem	Indirect influences such as school policies or leadership decisions	Availability of inclusive curriculum, professional development opportunities, school-wide support
Macrosystem	Broader cultural and societal ideologies	Dominant cultural values, societal attitudes toward inclusion and diversity
Chronosystem	Temporal dimension, life transitions or historical events	Changes in teachers' experience, evolving diversity discourses, policy reforms over time

*Application of Ecological Systems Theory to Culturally Inclusive Practices***Analytical Framework**

Building on this ecological perspective, the study introduces two complementary analytical distinctions to examine culturally inclusive practices in greater depth.

First, teachers' understanding of CIP is conceptualised across three dimensions: declarative understanding (valuing inclusion and distinguishing it from surface-level cultural

celebration); applied understanding (the ability to articulate what inclusion looks like within one's classroom context); and reflective understanding (the ongoing interrogation of personal cultural assumptions and their influence on practice).

Second, the enactment of CIP is interpreted across three levels: surface enactment (adaptations to materials and classroom interactions); instructional enactment (adjustments to teaching approaches and communication patterns); and structural enactment (modifications to assessment practices and formal expectations of learning).

These distinctions provide analytical clarity in examining variation within and across teachers' practices. They also enable a more precise interpretation of the relationship between understanding and enactment, particularly in identifying where alignment or disjunction occurs.

Together, the ecological framing and these analytical dimensions underpin the study's central proposition: that culturally inclusive practice is not solely a function of teacher intention, but is shaped by the interaction between individual commitment and the systemic conditions in which teaching is embedded.

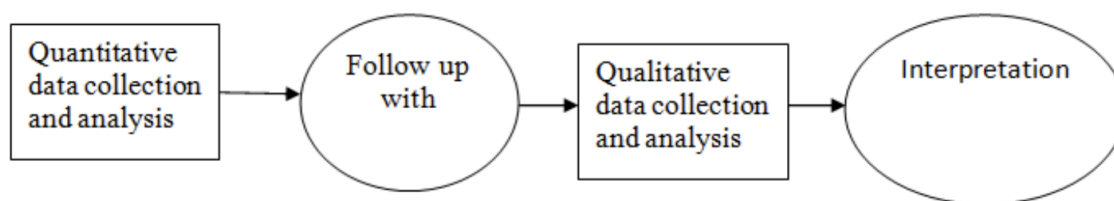
## **Methodology**

### **Research Paradigm and Design**

This study adopts a pragmatic paradigm and employs an explanatory sequential mixed-methods design, in which quantitative findings inform subsequent qualitative inquiry (Creswell & Plano Clark, 2007; Ivankova et al., 2006). This design is illustrated in Figure 1.

#### **Figure 1**

*Explanatory Sequential Design*



The quantitative phase identified patterns in teachers' understandings, classroom practices, perceived barriers, and institutional support related to culturally inclusive practices (CIP). The qualitative phase then explored how these patterns were experienced and interpreted in practice, enabling both breadth and depth of analysis.

### **Participants and Context**

**Quantitative phase.** Fifty-three teachers from international schools in Singapore (68%) and Malaysia (28%) participated in the survey. Participants represented a range of teaching experience (34% with 7–9 years; 32% with 10+ years) and curriculum frameworks (57% IB; 21% British; 9% American). Most participants (79%) were multilingual, and 64% had undertaken diversity or inclusion-related professional learning within the past year.

**Qualitative phase.** Six teachers were purposively selected from survey volunteers to ensure variation in teaching experience, curriculum context, and prior exposure to inclusion-focused professional learning.

### **Data Collection**

**Survey instrument.** A 24-item survey using a five-point Likert scale was developed to assess four constructs: understanding of CIP, classroom implementation, perceived barriers, and institutional support. Items were adapted from established instruments (Glock et al., 2019; Civitillo et al., 2018; Romijn et al., 2021) and aligned with the study's research questions. The instrument was piloted with six educators and refined for clarity and usability. Internal consistency was acceptable to good across constructs (Cronbach's  $\alpha = 0.787$ – $0.868$ ).

**Interviews.** Semi-structured interviews (45–60 minutes) explored teachers' conceptualisations of CIP, classroom practices, encountered barriers, and perceptions of institutional support. Interview questions were informed by survey findings and the ecological framework.

### **Data Analysis**

Quantitative data were analysed using descriptive statistics (means, standard deviations, and percentages) in Microsoft Excel and SPSS to identify patterns across participants.

Qualitative data were analysed using thematic analysis following Braun and Clarke's (2006) six-phase approach. Coding combined deductive categories informed by the research questions and ecological framework with inductive codes emerging from participants' accounts.

Integration occurred at the interpretation stage, with qualitative findings used to explain and contextualise quantitative patterns.

### **Ethical Considerations**

Ethical protocols followed BERA (2018) guidelines. Participation was voluntary and based on informed consent. Survey responses were anonymous, and interview data were anonymised using pseudonyms. All data were stored securely on password-protected devices.

### **Findings**

#### **Reliability of Survey Constructs**

Internal consistency was assessed using Cronbach's alpha. All constructs exceeded the recommended threshold of .70, indicating acceptable to good reliability for exploratory research (Taber, 2018).

**Table 2**

*Cronbach's Alpha for Survey Constructs*

Construct	Number of Items	Cronbach's Alpha ( $\alpha$ )	Interpretation
Understanding of CIP	4	0.868	Good
Implementation Practices	4	0.833	Good
Barriers to Implementation	4	0.808	Good
Support and Professional Learning	4	0.787	Acceptable

**Teacher Understanding of Culturally Inclusive Practices**

**Table 3**

\*Item-level descriptive statistics for teacher understanding (n = 53; all items compulsory)\*

#	Item	Mean	Standard Deviation	% Agree / Strongly Agree
1	I can clearly distinguish between celebrating cultural diversity and embedding cultural inclusion in classroom teaching.	4.23	0.72	83%
2	I feel confident explaining what culturally inclusive practices mean in my classroom context.	4.15	0.76	81%
3	I regularly reflect on my own cultural assumptions and how they influence my lesson planning.	4.04	0.82	72%
4	I believe culturally inclusive teaching contributes to students' sense of belonging and global citizenship.	4.34	0.67	89%

Teachers demonstrated strong conceptual commitment to culturally inclusive practices. Agreement was highest for the value of inclusion (M = 4.34) and the ability to distinguish between cultural celebration and deeper pedagogical inclusion (M = 4.23). Confidence in explaining CIP was similarly high (M = 4.15).

Reflective engagement showed greater variability ( $M = 4.04$ ), indicating that while inclusive principles are widely endorsed, sustained self-reflection is less consistently enacted.

Qualitative findings reinforced this pattern. Teachers described CIP as relational and embedded in everyday practice, often framed as "making sure every child feels seen." At the same time, several participants indicated that articulating and sustaining this understanding remained a work in progress. One teacher noted, "I can explain it, but I'm still figuring out how to express it simply," while another reflected, "I catch myself planning from my own perspective first."

Together, these findings indicate strong conceptual alignment with inclusion, alongside uneven depth in reflective practice.

### Classroom Implementation of Culturally Inclusive Practices

**Table 4**

\*Item-level descriptive statistics for classroom implementation ( $n = 53$ ; all items compulsory)\*

#	Item	Mean	Standard Deviation	% Agree / Strongly Agree
1	I adapt lesson materials and examples to reflect the cultural and linguistic backgrounds of my students.	4.00	0.73	77%
2	Whenever possible, I encourage students to share their home languages and cultural knowledge in class activities.	3.89	0.74	70%
3	I adjust my teaching approaches to suit students' varied communication patterns and learning preferences influenced by culture.	3.77	0.79	58%
4	I design or adapt assessment tasks to allow students to express understanding through culturally relevant examples or perspectives.	3.81	0.85	68%

Reported implementation was consistently lower than conceptual understanding, indicating a gap between inclusive intent and enacted practice. Teachers most frequently reported adapting lesson materials (M = 4.00) and encouraging students to share cultural and linguistic knowledge (M = 3.89).

Practices requiring deeper pedagogical or structural change showed lower endorsement. Adjusting teaching approaches (M = 3.77) and adapting assessment (M = 3.81) were less consistently enacted.

Interview data clarified this pattern. Teachers described material and interactional adaptations as manageable within existing routines. One participant noted, "Even if the unit is fixed, I change the examples so students recognise themselves in the learning." In contrast, assessment and instructional adaptations were described as more constrained. As one teacher explained, "I want to adapt more, but I'm not always sure what I'm allowed to change."

These findings suggest a gradient in implementation, with inclusive practices more readily enacted in flexible classroom domains and less consistently enacted where practices intersect with curriculum expectations and assessment structures.

### Perceived Barriers to Implementing Culturally Inclusive Practices

**Table 5**

\*Item-level descriptive statistics for perceived barriers (n = 53; all items compulsory)\*

#	Item	Mean	Standard Deviation	% Agree / Strongly Agree
1	I lack access to appropriate teaching materials and classroom examples that reflect cultural diversity.	3.62	0.96	58%
2	I find it challenging to meet the learning needs of students who are not native English speakers.	3.74	0.93	58%
3	There are limited opportunities to collaborate with colleagues on culturally inclusive teaching.	3.57	0.84	57%

4	School policies or schedules limit the time and flexibility needed to plan inclusive lessons.	3.62	0.92	58%
---	---	------	------	-----

Teachers reported moderate but consistent barriers to implementation. The most strongly endorsed challenge was supporting multilingual learners (M = 3.74), followed by limited access to culturally diverse materials and time constraints (both M = 3.62). Opportunities for collaboration were also reported as limited (M = 3.57).

Qualitative findings illustrated how these barriers interact in practice. Teachers described inclusive planning as requiring additional time and effort within already demanding workloads. As one participant explained, "I want to do more, but most days I'm just trying to keep up." Limited collaboration further constrained implementation, with one teacher noting, "We talk about inclusion in general terms, but not cultural inclusion."

These findings indicate that barriers are not isolated challenges but interconnected constraints related to time, resources, and instructional support.

### **Institutional Support and Professional Learning Needs**

**Table 6**

\*Item-level descriptive statistics for institutional support (n = 53; all items compulsory)\*

#	Item	Mean	Standard Deviation	% Agree / Strongly Agree
1	My school offers regular, practical professional development sessions focused on inclusion and cultural diversity.	3.57	0.79	57%
2	School leaders model and monitor the implementation of culturally inclusive teaching practices.	3.53	0.86	49%
3	I would benefit from structured mentorship or ongoing coaching to strengthen culturally inclusive teaching.	3.66	0.73	62%

4	There is a clear, school-wide framework or policy guiding culturally inclusive practices.	3.47	0.84	47%
---	---	------	------	-----

Perceived institutional support was moderate and lower than levels of conceptual understanding and implementation. The strongest endorsement was for the need for structured mentorship or ongoing coaching ( $M = 3.66$ ), indicating a clear demand for sustained, practice-focused professional learning.

Lower agreement was observed for leadership modelling ( $M = 3.53$ ) and the presence of a clear school-wide framework ( $M = 3.47$ ), suggesting variability in how inclusion is operationalised within schools.

Interview data reinforced these patterns. Teachers expressed a preference for practice-embedded learning over one-off workshops. As one participant noted, "I don't need another lecture. I need someone to sit with me and help me think through my unit." The absence of shared frameworks created uncertainty, with inclusion described as "valued but not clearly defined in practice."

These findings highlight the importance of coherent institutional structures in supporting sustained inclusive practice.

### **Integrated Summary**

Across all four research questions, a consistent pattern emerged. Teachers demonstrated strong conceptual commitment to culturally inclusive practices, yet their capacity to enact these practices consistently was shaped by time constraints, resource availability, collaboration structures, and institutional support.

This pattern reflects a gap between commitment and capacity. Teachers expressed clear alignment with inclusive values and reported active efforts to adapt classroom practice. However, implementation was strongest in flexible domains such as materials and interaction, and less consistent where practices required instructional or structural change.

Qualitative findings clarified that this gap is not driven by lack of motivation, but by the conditions under which teaching is organised. Barriers related to time, resources, and collaboration, alongside uneven institutional support, shape what is feasible in practice.

Taken together, the findings indicate that culturally inclusive practice is influenced by interactions between teacher beliefs, classroom practice, and institutional structures, rather than individual intention alone.

## **Discussion**

### **Teachers' Understanding of Culturally Inclusive Practices**

Findings indicate that teachers' understanding of culturally inclusive practices is differentiated rather than uniform. While teachers demonstrated strong declarative understanding (valuing inclusion and distinguishing it from cultural celebration) and applied understanding (articulating CIP within their classroom context), reflective engagement was less consistently enacted.

This distinction is significant. Reflective capacity underpins teachers' ability to interrogate their own cultural assumptions and adapt practice responsively (Civitillo et al., 2018; Gay, 2013). Without sustained reflection, inclusive intentions risk remaining at the level of principle rather than becoming embedded in pedagogical decision-making. The findings therefore extend existing literature by showing that conceptual commitment alone does not ensure depth of practice; it must be accompanied by structured opportunities for reflection.

From an ecological perspective, this differentiation is located within the microsystem, where teacher beliefs, prior experiences, and reflective habits shape how inclusion is understood. However, the unevenness of reflective engagement suggests that individual cognition is insufficient in isolation. Reflective practice requires reinforcement through professional dialogue and institutional routines if it is to be sustained over time.

### **The Commitment–Capacity Gap in Classroom Enactment**

The findings reveal a consistent gap between teachers' inclusive intentions and their enacted practice. This commitment–capacity gap reflects not a lack of motivation, but the conditions under which teaching is organised.

Teachers most readily enacted CIP through adaptations to materials and classroom interactions. These forms of practice are flexible, visible, and largely within teacher control. In contrast, enactment was less consistent in areas requiring deeper pedagogical or structural change, particularly instructional adaptation and assessment design. These domains are more tightly bound to curriculum expectations, shared norms, and accountability structures.

This gradient in enactment clarifies the nature of the gap. As practice moves from surface-level adjustments to instructional and structural domains, teachers' agency becomes increasingly mediated by institutional conditions. Bronfenbrenner's framework helps explain this pattern: microsystem-level agency is progressively shaped by mesosystem expectations and exosystem constraints.

This interpretation extends prior research by reframing uneven implementation not as inconsistency in teacher practice, but as a function of how different domains of teaching are structured and regulated. The implication is that strengthening inclusive practice requires expanding capacity at the instructional and structural levels, rather than focusing solely on teacher intention.

### **Barriers as Ecological Constraints**

Perceived barriers to culturally inclusive practice are best understood as systemic constraints rather than individual shortcomings. While teachers identified challenges such as supporting multilingual learners, limited access to resources, and restricted collaboration, these were experienced as interconnected rather than isolated.

Time emerged as a central organising constraint, shaping teachers' ability to plan and enact inclusive practices deliberately. Resource limitations and limited collaboration structures further compounded this constraint, reinforcing a pattern in which inclusion becomes an additional layer of work rather than an integrated aspect of pedagogy.

These findings align with literature highlighting the gap between institutional rhetoric and classroom reality (Heng & Lim, 2021; Holliday, 2011), but extend it by demonstrating how constraints operate cumulatively across ecological levels. Challenges associated with multilingual learners are located within the microsystem, but are intensified by mesosystem limitations in collaboration and exosystem constraints related to time, policy, and resource allocation.

This ecological interpretation shifts the analytical focus from teacher deficit to system design. Teachers' accounts consistently reflected commitment to inclusion, but also revealed the structural conditions that delimit what is possible in practice.

### **Institutional Support as Enabling Infrastructure**

Findings position institutional support as the key enabling condition for sustained culturally inclusive practice. Teachers consistently emphasised the importance of mentorship, collaborative routines, and coherent school-wide frameworks.

Professional learning emerged as most effective when it was sustained, practice-embedded, and directly connected to classroom decision-making. Teachers expressed a clear preference for mentorship and coaching over one-off workshops, reinforcing literature that emphasises iterative, contextually grounded professional development (Romijn et al., 2021; Rogahang et al., 2024).

At the same time, the absence of clear school-wide frameworks created variability in how CIP was interpreted and enacted. Where inclusive values were articulated but not operationalised, teachers were required to rely on individual judgement, leading to inconsistency across classrooms.

From an ecological perspective, institutional support operates across the mesosystem and exosystem, shaping whether inclusive intentions can be translated into sustained practice. When professional learning, collaboration, and policy alignment are coherent, teachers are better able to extend practice beyond surface-level adaptations towards deeper instructional and structural enactment.

### **Theoretical Contribution: Reframing Inclusion Ecologically**

This study contributes to the literature by reframing culturally inclusive practice as an ecological phenomenon. Rather than locating inclusion within individual teacher competence, the findings demonstrate how practice is shaped through interactions across systems.

The concept of a commitment–capacity gap provides a useful analytical lens for understanding the discrepancy between inclusive intentions and enacted practice. This gap

reflects the alignment, or misalignment, between teacher beliefs (microsystem), professional relationships and routines (mesosystem), and institutional conditions (exosystem).

In addition, the study differentiates between dimensions of teacher understanding (declarative, applied, reflective) and levels of enactment (surface, instructional, structural), offering a more granular framework for analysing inclusive practice. Together, these contributions extend the application of ecological systems theory to inclusion in international school contexts.

### **Practical Implications**

This study suggests that strengthening culturally inclusive practice requires a shift from individual capacity-building to system-level alignment. While teachers demonstrate strong conceptual commitment, the findings indicate that consistent enactment depends on the conditions under which teaching is organised.

For school leadership, this highlights the importance of translating inclusive values into clear pedagogical expectations. The absence of coherent frameworks leaves inclusion open to individual interpretation, resulting in variability across classrooms. Establishing shared language, expectations, and exemplars can reduce this variability and support more consistent practice.

Professional learning design also emerges as a critical lever. The findings suggest that one-off workshops are insufficient to support sustained inclusive practice. Instead, teachers benefit from ongoing, practice-embedded support such as mentorship, coaching, and collaborative planning. These structures enable teachers to move beyond surface-level adaptations towards deeper instructional and assessment-level enactment.

At the classroom level, the findings reinforce the importance of collaborative professional culture. Opportunities for teachers to share strategies, resources, and reflective insights appear to support both confidence and implementation. Embedding such collaboration within routine practice, rather than leaving it to individual initiative, is likely to strengthen collective capacity.

Taken together, these implications reinforce the central argument of the study: culturally inclusive practice is most effectively developed not through isolated teacher effort, but

through coherent alignment between beliefs, professional learning, and institutional structures.

### **Limitations and Future Research**

This study has several limitations that should be considered when interpreting the findings.

First, the study is geographically bounded, with participants drawn from international schools in Singapore and Malaysia. While these contexts are characterised by high levels of cultural and linguistic diversity, the findings may not be directly transferable to other international school settings with different structural or cultural conditions.

Second, the study relies on self-reported data, which may be influenced by social desirability or prevailing professional norms around inclusion. Although the mixed-methods design strengthens interpretive depth, the findings reflect teachers' perceptions and accounts rather than direct observation of classroom practice.

Third, the qualitative phase involved a small number of participants. While appropriate for an explanatory design, this limits the breadth of perspectives captured and the extent to which variation across contexts can be explored.

These limitations point to several directions for future research. Cross-regional studies could examine how culturally inclusive practices are shaped under different institutional and policy conditions. Longitudinal research could explore how teachers' understanding and enactment of inclusion evolve over time, particularly in response to sustained professional learning. Classroom-based studies, including observation and artefact analysis, would provide deeper insight into how culturally inclusive practices are enacted in situ. Finally, extending the focus beyond teachers to include students, school leaders, and families would strengthen understanding of how inclusion is experienced and supported across the wider school ecology.

### **Conclusion**

This study examined how teachers in international schools understand and enact culturally inclusive practices (CIP), the barriers they encounter, and the institutional conditions shaping

their work. Across both quantitative and qualitative findings, a consistent pattern emerged: teachers demonstrate strong conceptual commitment to culturally inclusive practices, yet their capacity to enact these practices consistently remains uneven.

This commitment–capacity gap reflects not a lack of motivation, but the ecological conditions within which teaching is organised. While teachers exercise agency in adapting materials and classroom interactions, their capacity to enact deeper instructional and structural changes is shaped by time constraints, resource availability, assessment expectations, and the coherence of institutional support.

These findings lead to a clear integrative conclusion. Culturally inclusive practice is not an individual pedagogical choice, but a system-dependent endeavour shaped by alignment across ecological levels. Teacher commitment, while necessary, is insufficient in isolation. Sustained inclusion depends on the extent to which institutional structures, professional learning systems, and collaborative routines enable and normalise inclusive practice.

Reframing inclusion through an ecological lens shifts the focus from individual effort to system coherence. Inclusive practice is shaped not only within the classroom, but through interactions across professional relationships, leadership priorities, policy frameworks, and resourcing decisions. Teachers' agency remains central, but it is exercised within, and often constrained by, these intersecting systems.

Sustainable progress therefore depends less on exhorting teachers to do more and more on building the conditions that make inclusive practice possible, visible, and consistent. When professional learning is embedded, mentorship is sustained, expectations are clear, and institutional support is coherent, teachers are better positioned to translate inclusive values into everyday practice.

In this sense, culturally inclusive practice in international schools is best understood as a collective, systemically enabled endeavour, shifting the focus from individual compliance to shared responsibility, and from aspiration to sustained enactment.

## References

Banks, J. A. (2015). *Cultural diversity and education (0 ed.)*. Routledge.  
<https://doi.org/10.4324/9781315622255>

Bennett, C. I. (2010). *Comprehensive multicultural education: Theory and practice (7th ed., international ed.)*. Pearson.

Braun, V., & Clarke, V. (2006). *Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2)*, 77–101. <https://doi.org/10.1191/1478088706qp063oa>

British Educational Research Association. (2018). *Ethical guidelines for educational research. (4th ed.)*.  
<https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018>

Bronfenbrenner, U. (1979). *The ecology of human development*. Harvard University Press.

Bunnell, T. (2016). *Teachers in international schools: A global educational 'precariat'?* *Globalisation, Societies and Education, 14(4)*, 543–559.  
<https://doi.org/10.1080/14767724.2015.1068163>

Civitillo, S., Juang, L. P., & Schachner, M. K. (2018). *Challenging beliefs about cultural diversity in education: A synthesis and critical review of trainings with pre-service teachers. Educational Research Review, 24*, 67–83. <https://doi.org/10.1016/j.edurev.2018.01.003>

Creswell, J.W. and Creswell, J.D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research*. SAGE.

Gay, G. (2013). *Teaching to and through cultural diversity. Curriculum Inquiry, 43(1)*, 48–70.  
<https://doi.org/10.1111/curi.12002>

Glock, S., Kovacs, C., & Pit-ten Cate, I. (2019). *Teachers' attitudes towards ethnic minority students: Effects of schools' cultural diversity*. *British Journal of Educational Psychology*, 89(4), 616–634. <https://doi.org/10.1111/bjep.12248>

Gonzalez, A., Christiaensen, L., & Robalino, D. (2019). *Migration and jobs: Issues for the 21st century*. World Bank, Washington, DC. <https://doi.org/10.1596/1813-9450-8867>

Hayden, M. C., & Thompson, J. J. (2013). *International schools: Antecedents, current issues and metaphors for the future*. In R. Pearce (Ed.), *International education and schools: Moving beyond the first 40 years*. London: Bloomsbury Academic.

Heng, T. T., & Lim, L. (2021). *Diversity, difference, equity: How student differences are socially constructed in Singapore*. *Cambridge Journal of Education*, 51(6), 693–712. <https://doi.org/10.1080/0305764X.2021.1910204>

Holliday, A. (2011). *Intercultural communication and ideology*. SAGE.

International Baccalaureate Organization. (n.d.). *Mission and philosophy*. Retrieved from [ibo.org](http://ibo.org)

International Organization for Migration. (2024). *World migration report 2024*. United Nations.

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). *Using mixed-methods sequential explanatory design: From theory to practice*. *Field Methods*, 18(1), 3–20. <https://doi.org/10.1177/1525822X05282260>

Knoblauch, C. (2023). *Cultural and religious diversity in early childhood education implications of socialization and education for the geographies of childhood*. *Religions*, 14(4), 555. <https://doi.org/10.3390/rel14040555>

Ladson-Billings, G. (1995). *Toward a theory of culturally relevant pedagogy*. *American Educational Research Journal*, 32(3), 465–491. <https://doi.org/10.3102/00028312032003465>

Ladson-Billings, G. (2014). *Culturally Relevant Pedagogy 2.0: A.k.a. the Remix*. *Harvard Educational Review*, 84(1), 74–84. <https://doi.org/10.17763/haer.84.1.p2rj131485484751>

Lincoln, Y. S., & Guba, E. G. (1986). *But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation*. *New Directions for Program Evaluation*, 1986(30), 73–84. <https://doi.org/10.1002/ev.1427>

OECD. (2018). *Preparing our youth for an inclusive and sustainable world: The OECD PISA global competence framework*. <https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/global-competence/Handbook-PISA-2018-Global-Competence.pdf>

Rogahang, S. S. N., Paramansyah, A., Zaelani, K., Iqbal, M., & Judijanto, L. (2024). *Inclusive education practices: Fostering diversity and equity in the classroom*. *Global International Journal of Innovative Research*, 1(3), 260–266. <https://doi.org/10.59613/global.v1i3.46>

Romijn, B. R., Slot, P. L., & Leseman, P. P. M. (2021). *Increasing teachers' intercultural competences in teacher preparation programs and through professional development: A review*. *Teaching and Teacher Education*, 98, 103236. <https://doi.org/10.1016/j.tate.2020.103236>

Sleeter, C. E. (2001). *Preparing teachers for culturally diverse schools: Research and the overwhelming presence of whiteness*. *Journal of Teacher Education*, 52(2), 94–106. <https://doi.org/10.1177/0022487101052002002>

Taber, K. S. (2018). *The use of cronbach's alpha when developing and reporting research instruments in science education*. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>

UNESCO. (2015). *Global citizenship education: Topics and learning objectives*. UNESCO. <https://doi.org/10.54675/DRHC3544>

*United Nations Convention on the Rights of the Child*, November 20, 1989, <https://www.ohchr.org/en/professionalinterest/pages/crc.aspx>

**AI-Assisted Grading and Feedback in International Baccalaureate Middle Years  
Programme Physics: A Mixed-Methods Investigation of Learning Outcomes, Feedback  
Quality, and the Moderating Role of English Language Proficiency**

Aminurrashid Bin Abu Bakar

University College Fairview

## Abstract

This sequential explanatory mixed-methods study examined the impact of two artificial intelligence (AI) assessment modules on student learning outcomes in International Baccalaureate (IB) Middle Years Programme (MYP) Physics at an IB World School in Shenzhen, China. Participants comprised 101 Grade 9 and Grade 10 students, of whom 91% were non-native English speakers, alongside three science teachers, over a 24-week implementation period. Revision Village was deployed for Criterion A assessments, while ChatGPT supported Criteria B, C, and D. Quantitative data from pre- and post-intervention MYP criterion-referenced assessments, two Feedback Quality Assessment Rubrics, and Technology Acceptance Surveys were analysed using paired-samples *t*-tests, one-way ANOVA, and Hayes' PROCESS macro (Models 1 and 4). Qualitative data from semi-structured teacher interviews, student focus groups, and classroom observations were analysed using Braun and Clarke's six-phase thematic analysis. Results revealed statistically significant improvements across all four MYP criteria, with effect sizes ranging from Cohen's  $d = 0.43$  (Criterion B) to  $d = 0.74$  (Criterion D), and overall MYP grade improvement from 4.58 to 5.27 ( $d = 0.64$ ). Feedback quality characteristics significantly mediated 48.3% of the relationship between AI implementation and learning outcomes (bootstrap 95% CI [0.802, 2.047]). English language proficiency significantly moderated implementation effectiveness ( $B = -0.782$ ,  $p = .010$ ), with Low- and Mid-proficiency students demonstrating significantly greater gains than High-proficiency students. The study contributes evidence-based guidance for integrating AI assessment modules in linguistically diverse IB Physics classrooms and proposes a revised AI-Human Feedback Partnership Model in which teacher mediation functions as the central bridging mechanism between AI-generated feedback and learner outcomes.

*Keywords:* AI-assisted grading, feedback quality, International Baccalaureate physics, English language proficiency, mixed-methods research, educational technology

## **AI-Assisted Grading and Feedback in IB MYP Physics: A Mixed-Methods Investigation**

The integration of artificial intelligence (AI) into educational assessment represents one of the most consequential developments in contemporary classroom practice (Crompton & Burke, 2023; Wongvorachan et al., 2022). Systematic analyses indicate that publications on AI in education in 2021–2022 nearly tripled the volume of preceding years (Wang et al., 2024), and AI-assisted grading platforms have demonstrated capabilities extending well beyond simple answer verification—including automated analysis of reasoning, real-time identification of misconceptions, and generation of personalised feedback (Luckin et al., 2016). Within this expanding landscape, however, the application of AI assessment to inquiry-based science education, and to the International Baccalaureate (IB) Physics curriculum in particular, remains critically underexamined.

The IB Middle Years Programme (MYP) Sciences curriculum emphasises conceptual understanding, inquiry-based methodology, and criterion-referenced assessment across four dimensions: Knowing and Understanding (Criterion A), Inquiring and Designing (Criterion B), Processing and Evaluating (Criterion C), and Reflecting on the Impacts of Science (Criterion D) (International Baccalaureate Organization, 2023). Effective formative assessment in this context requires feedback that is timely, specific, and aligned with the cognitive demands of each criterion (Brookhart, 2013; Hattie & Timperley, 2007). Yet traditional teacher-led assessment imposes substantial cognitive and time burdens on educators, with feedback turnaround in pre-implementation contexts often approaching ten days (Carless, 2006; Molin et al., 2021). This temporal gap undermines the formative purpose of assessment because students lose the cognitive context required to act productively on feedback (Wisniewski et al., 2020). Emerging research suggests that AI tools can deliver scalable, criterion-referenced feedback in hours rather than days (Kortemeyer, 2023; Latif et

al., 2024), but the conditions under which such tools translate into measurable learning gains remain poorly understood.

A critical research gap exists at the intersection of three under-examined dimensions: AI-assisted grading in inquiry-based science assessment, the IB MYP Physics curriculum, and the linguistic-cognitive realities of students learning in English-medium international school contexts where most learners are non-native English speakers. The present study addresses this gap through a sequential explanatory mixed-methods investigation conducted at an IB World School in Shenzhen, China, where approximately 91% of students are non-native English speakers. The study examines two AI assessment modules—Revision Village, an IB-calibrated platform, and ChatGPT, a generative AI tool—and evaluates their impact on student learning outcomes, the feedback quality mechanisms through which any effects operate, and the contextual factors that condition their effectiveness.

Four research questions guided the investigation:

RQ1. To what extent does AI-assisted grading implementation improve student learning outcomes in MYP Physics education?

RQ2. How do feedback quality characteristics mediate the relationship between AI-assisted grading implementation and student learning outcomes in MYP Physics?

RQ3. What challenges and opportunities do physics teachers and students encounter when implementing AI-assisted grading tools?

RQ4. How do contextual factors, including English language proficiency, shape the effectiveness of AI assessment modules on student learning outcomes?

## Literature Review

### Theoretical Framework

This study integrates four theoretical perspectives. Hattie and Timperley's (2007) feedback model provides the foundation, identifying three critical feedback functions—feed-up (clarification of learning objectives), feedback (progress monitoring), and feed-forward (guidance for subsequent learning)—and four feedback levels (task, process, self-regulation, and self), with self-regulation feedback yielding the strongest effects on durable learning (Mandouit & Hattie, 2023). Self-Determination Theory (Deci & Ryan, 2017) explains how AI feedback systems may support or undermine the basic psychological needs for autonomy, competence, and relatedness; immediate AI feedback can support competence by enabling observable mastery progression, while the absence of relational warmth may undermine relatedness for some learners (Fathali & Okada, 2018). The Technology Acceptance Model (Davis, 1989) and its extensions (Venkatesh et al., 2012) provide a framework for understanding teacher and student adoption of AI grading systems through perceived usefulness, perceived ease of use, and facilitating conditions (Dwivedi et al., 2019). Finally, Task–Technology Fit theory (Goodhue & Thompson, 1995) frames the alignment between AI-tool affordances and the cognitive demands of MYP criterion-based assessment.

### AI Grading Effectiveness in STEM Contexts

Empirical research on AI grading in physics has produced promising but context-dependent results. Kortemeyer (2023) demonstrated that AI systems could grade introductory physics problem solutions with correlation coefficients of  $R^2 = 0.84$  using a MathPix and GPT-4 workflow compared to human graders. A subsequent psychometric validation by Kortemeyer and Nöhl (2024) developed a threshold-based human–AI collaboration model achieving  $R^2 \approx 0.91$  when AI handled half the grading load. At the secondary level, Altal and Abo Ehsaiyan (2025) reported a large effect (Cohen's  $d = 1.21$ ) for

AI-assisted instruction incorporating GPT-4 in Grade 11 Advanced Physics in the United Arab Emirates. A recent meta-analysis by Alqahtani et al. (2023) reported a significant positive overall effect for AI integration (Hedges'  $g = 0.86$ ), with chatbots and generative AI showing the largest impact on student learning outcomes (effect size = 1.02).

However, studies also reveal important limitations. Kortemeyer (2023) found that AI systems demonstrated high accuracy for structured physics problems but struggled with open-ended investigations and creative problem-solving approaches valued in inquiry-based physics education. Chen and Wan (2024) showed that GPT-3.5 could achieve human-level accuracy for conceptual physics questions through careful prompt engineering, but struggled with multi-step problems requiring complex mathematical reasoning. Crucially, these studies have been conducted predominantly in contexts where students are native English speakers; the performance of AI grading tools in linguistically diverse classrooms remains largely uncharted.

### **English Language Proficiency and Science Learning Outcomes**

Students learning physics through English-medium instruction face dual cognitive–linguistic demands. Cummins' (2000) framework distinguishes Basic Interpersonal Communicative Skills (BICS), which typically develop within two years, from Cognitive Academic Language Proficiency (CALP), which requires five to seven years of sustained exposure. This timeline has direct implications for AI-mediated assessment, since AI-generated feedback is invariably produced in academic English calibrated against English-language criteria. Research on technology-mediated feedback for non-native English speakers reveals complex relationships between English proficiency, digital literacy, and engagement with AI-mediated assessment. Yaseen et al. (2025) found that the impact of adaptive learning technologies and personalised feedback on student engagement is shaped by digital literacy, suggesting that English proficiency and digital literacy interact in shaping

students' access to AI-generated feedback. Warschauer and Matuchniak (2010) further cautioned that educational technologies may exacerbate inequalities by favouring already-advantaged learners, a concern that motivates the present study's explicit attention to language as a moderating variable.

### **The Research Gap**

Despite the growing volume of AI-in-education research, three critical gaps remain. First, physics-specific AI grading research is sparse, particularly studies evaluating scientific reasoning in inquiry-based contexts (Zawacki-Richter et al., 2019). Second, considerations of non-native English speakers are notably absent from AI grading research, despite the rapid global growth of English-medium instruction. Third, implementation research examining authentic classroom contexts, rather than controlled experimental tasks, remains limited (Cavalcanti et al., 2021). The present study addresses these converging gaps within the underexamined context of IB MYP Physics.

## **Methodology**

### **Research Design**

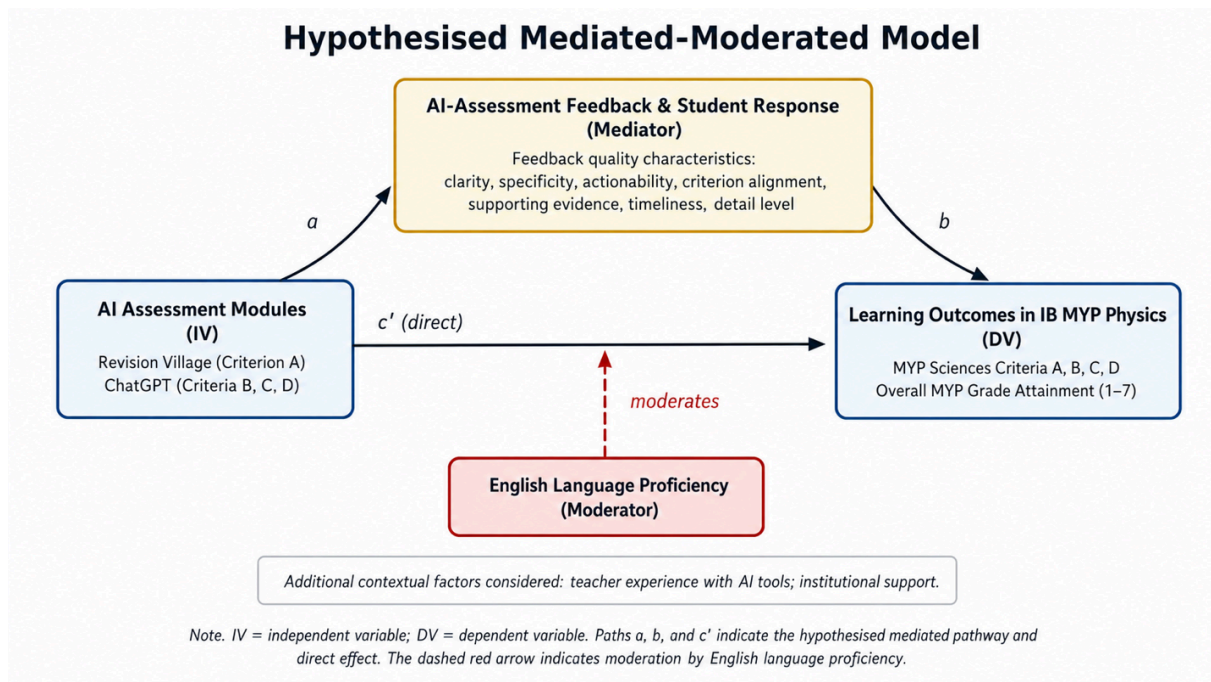
The study employed a sequential explanatory mixed-methods design embedded within a case study framework (Creswell & Plano Clark, 2018). Quantitative data were collected and analysed first, followed by qualitative data collection to explain and elaborate quantitative findings. Integration occurred through joint displays at the level of each research question. This design was appropriate for examining both the extent and the mechanisms of AI implementation effects in an authentic classroom context.

Figure 1 presents the hypothesised conceptual model that guided the inferential analyses, depicting AI-assisted grading implementation as the independent variable, feedback quality characteristics as a hypothesised mediator (RQ2), student learning outcomes as the

dependent variable (RQ1), and English language proficiency as a hypothesised moderator (RQ4).

**Figure 1**

*Hypothesised Mediated-Moderated Model*



## Research Setting and Participants

The study was conducted at an IB World School in Shenzhen, China, serving a diverse international community. The MYP Sciences and Physics pathway encompassed Grade 9 MYP Sciences (52 students across three classes) and Grade 10 MYP Physics (49 students across three classes). Of 111 enrolled students, 101 (91%) were non-native English speakers and constituted the study sample. Participants' first languages included Mandarin ( $n = 81$ , 80.2%), Cantonese ( $n = 11$ , 10.9%), Korean ( $n = 5$ , 5.0%), and other languages ( $n = 4$ , 4.0%). English language proficiency was classified as Low ( $n = 32$ , 31.7%), Mid ( $n = 43$ , 42.6%), or High ( $n = 26$ , 25.7%) based on course placement records and triangulated teacher evaluation. Three MYP Science teachers participated: Teacher A (male, 10 years IB experience, extensive prior AI experience), Teacher B (male, 4 years experience, moderate AI experience), and

Teacher C (female, 6 years experience, minimal prior AI exposure). Table 1 summarises participant characteristics.

**Table 1**

*Student Participant Characteristics by English Language Proficiency Group (N = 101)*

Characteristic	Low (n = 32)	Mid (n = 43)	High (n = 26)	Full Sample (N = 101)
Grade 9 (MYP Sciences)	19	22	11	52 (51.5%)
Grade 10 (MYP Physics)	13	21	15	49 (48.5%)
First language: Mandarin	25	35	21	81 (80.2%)
First language: Cantonese	4	5	2	11 (10.9%)
First language: Korean	2	2	1	5 (5.0%)
First language: Other	1	1	2	4 (4.0%)

*Note.* English language proficiency was classified using course placement records and triangulated teacher evaluation. Percentages are computed against the full sample (N = 101).

### **AI Tools Under Investigation**

**Revision Village** was deployed for Criterion A (Knowing and Understanding) assessments. This IB-calibrated platform provides structured, algorithm-driven feedback aligned with MYP Physics criterion descriptors, including adaptive questioning, progress tracking, and detailed solution explanations.

**ChatGPT** (OpenAI GPT-4) was deployed for Criteria B, C, and D, which require open-ended, conversational feedback suitable for evaluative and reflective tasks. Standardised prompt templates and grading instructions were developed to ensure consistent application across teachers and assessments.

### **Data Collection Instruments**

*MYP Sciences Criterion-Based Assessments.* Pre- and post-intervention assessments measured student learning outcomes across the four MYP criteria (A–D, each scored 0–8), producing a total score out of 32 and an MYP grade of 1–7.

*Feedback Quality Assessment Rubric.* Two separate rubrics—one for Revision Village (Criterion A) and one for ChatGPT (Criteria B, C, and D)—assessed AI-generated and matched human teacher feedback across four dimensions: clarity (linguistic accessibility), specificity, actionability, and alignment with learning objectives. Each dimension was scored on a 4-point scale. Inter-rater reliability for rubric scoring was established at Cohen's  $\kappa = 0.84$  across 25% of randomly selected feedback pairs.

*Technology Acceptance Surveys.* Separate surveys for teachers and students, grounded in the Technology Acceptance Model (Davis, 1989), measured perceived usefulness, perceived ease of use, learning effectiveness, and behavioural intention on 5-point Likert scales. Cronbach's alpha values ranged from .82 (Perceived Usefulness) to .88 (Behavioural Intention).

*Qualitative Instruments.* Semi-structured teacher interviews (45–60 minutes), student focus groups (60–75 minutes; 6–8 students per group), and 90-minute classroom observations were conducted using standardised protocols.

## **Procedure**

The 24-week implementation was structured into three phases: quantitative data collection (Weeks 1–12), comprising pre-intervention assessments, surveys, the active intervention introducing AI tools, and post-intervention assessments; qualitative data collection (Weeks 13–20), comprising interviews, focus groups, and observations; and integration (Weeks 21–24). Feedback turnaround was tracked across 808 assignments (eight per student).

## **Data Analysis**

Quantitative analysis used IBM SPSS Statistics (Version 28) with paired-samples *t*-tests, one-way ANOVA with Tukey HSD post hoc comparisons, Pearson correlations, and Hayes' PROCESS macro (Model 4 for mediation; Model 1 for moderation) with 5,000 bootstrap resamples. Effect sizes (Cohen's *d*) and partial eta-squared ( $\eta^2$ ) were reported. Qualitative data were analysed using Braun and Clarke's (2021) six-phase thematic analysis, with inter-rater reliability for qualitative coding established at  $\kappa = 0.81$ .

## **Ethical Considerations and Trustworthiness**

The study received ethics approval from the host institution, and parental consent and student assent were obtained for all participants. Pseudonymisation was applied throughout data handling and reporting. Methodological rigour was addressed through the criteria appropriate to each strand. For the quantitative strand, internal validity was supported through standardised assessment protocols and pre-post within-subjects design, reliability through internal-consistency checks (Cronbach's  $\alpha = .82-.88$ ), and inferential robustness through bootstrap-based confidence intervals. For the qualitative strand, trustworthiness was addressed using Lincoln and Guba's (1985) criteria of credibility (member checking with two teacher participants), transferability (thick description of context), dependability (audit trail of coding decisions), and confirmability (independent inter-rater coding,  $\kappa = 0.81$ ). Inferential quality at the mixed-methods level (Creswell & Plano Clark, 2018) was supported through joint-display integration of quantitative and qualitative strands for each research question.

## **Results**

### **RQ1: AI Implementation and Student Learning Outcomes**

Paired-samples *t*-tests revealed statistically significant improvements across all four MYP criteria following AI-assisted grading implementation (Table 2). The largest effect was

observed in Criterion D, Reflecting on the Impacts of Science ( $d = 0.74$ ), approaching a large effect size and suggesting that AI-assisted feedback particularly supported students in developing evaluative and reflective thinking. Criterion C, Processing and Evaluating ( $d = 0.58$ ), and Criterion A, Knowing and Understanding ( $d = 0.54$ ), demonstrated medium effects. Criterion B, Inquiring and Designing, showed the most modest but still statistically significant improvement ( $d = 0.43$ ). The pre-intervention mean MYP grade of 4.58 improved to 5.27 post-intervention, with a total effect size of  $d = 0.64$ —a medium effect both statistically and practically meaningful within a single academic term.

**Table 2**

*Pre- and Post-Intervention MYP Criterion Scores for Full Sample (N = 101)*

Measure	Pre <i>M</i>	Pre <i>SD</i>	Post <i>M</i>	Post <i>SD</i>	Gain	<i>t</i> (100)	Cohen's <i>d</i>
Criterion A: Knowing & Understanding	4.62	1.58	5.35	1.45	0.73	−5.11***	0.54
Criterion B: Inquiring & Designing	4.45	1.55	5.00	1.48	0.55	−4.32***	0.43
Criterion C: Processing & Evaluating	4.28	1.52	5.04	1.44	0.76	−5.83***	0.58
Criterion D: Reflecting on Impacts	4.49	1.48	5.27	1.41	0.78	−7.12***	0.74
Total Score (0–32)	17.84	4.48	20.66	4.12	2.82	−6.43***	0.64
MYP Grade (1–7)	4.58	0.84	5.27	0.79	0.69	—	—

*Note.* Criteria were scored on a 0–8 scale. Total score range = 0–32; MYP grade range = 1–7. All *t* values reflect paired-samples *t*-tests (*df* = 100). Cohen's *d* was computed as the mean difference divided by the pooled *SD* of the difference scores. All tests two-tailed. \*\*\*  $p < .001$ .

Differential analysis by English language proficiency group revealed that Mid-proficiency students recorded the largest absolute gain (3.35 total score points), followed closely by Low-proficiency students (3.16 points). High-proficiency students recorded the smallest gain (1.50 points), consistent with a ceiling effect given their high pre-intervention

mean (24.12 out of 32). These descriptive patterns are formally tested in the moderation analysis presented for RQ4 below.

## RQ2: Feedback Quality as Mediator

### *Feedback Quality Comparison*

Paired-samples *t*-tests compared AI-generated and human teacher feedback across the four rubric dimensions for Revision Village (40 matched pairs, Criterion A) and ChatGPT (120 matched pairs, Criteria B, C, and D). Table 3 presents the results.

**Table 3**

*Feedback Quality Rubric Scores: AI Tools vs. Human Teacher Feedback*

Rubric Dimension	AI <i>M</i>	Human <i>M</i>	Mean Diff.	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
<b>Revision Village (n = 40 pairs, Criterion A)</b>						
Specificity	3.42	2.61	+0.81	7.77	< .001	1.23
Alignment with Objectives	3.56	2.47	+1.09	9.38	< .001	1.48
Clarity (Linguistic Accessibility)	2.65	3.38	-0.73	-6.89	< .001	-1.09
Actionability	3.14	3.06	+0.08	1.04	.305	0.16
<b>ChatGPT (n = 120 pairs, Criteria B, C, D)</b>						
Specificity	3.18	2.76	+0.42	7.43	< .001	0.68
Alignment with Objectives	3.08	2.72	+0.36	5.71	< .001	0.52
Clarity (Linguistic Accessibility)	2.71	3.24	-0.53	-10.07	< .001	-0.92
Actionability	3.09	3.01	+0.08	1.57	.119	0.14

*Note.* Scores on a 1–4 scale: 1 = Poor, 2 = Developing, 3 = Proficient, 4 = Exemplary. Positive *d* indicates AI advantage; negative *d* indicates human teacher advantage. Cohen's *d* reflects the mean difference in matched pairs divided by the *SD* of the difference scores. The Clarity dimension specifically measures linguistic accessibility for non-native English speakers.

Both AI tools scored significantly higher than human teacher feedback on specificity and alignment with learning objectives, and significantly lower on clarity. Effect sizes for Revision Village were substantially larger than for ChatGPT (specificity *d* = 1.23 vs. 0.68;

alignment  $d = 1.48$  vs.  $0.52$ ), consistent with Revision Village's explicit calibration to MYP Physics curriculum standards. No statistically significant differences were found for actionability for either tool. It is worth noting that the Clarity dimension specifically measures linguistic accessibility for the study's predominantly non-native English speaker sample; AI-generated feedback produced grammatically well-formed English but was less linguistically accessible to the target student population than teacher-mediated alternatives.

### ***Feedback Timeliness***

Pre-implementation teacher grading across 808 assignments averaged 9.9 days turnaround, with zero assignments returned within 48 hours. Post-implementation, Revision Village reduced mean turnaround to 0.3 days (100% within 48 hours), while ChatGPT reduced turnaround to 2.4 days (16.2% within 48 hours; 94.7% within one week). This dramatic compression of the feedback cycle has direct implications for the mediation analysis presented next, as feedback timeliness contributes to the overall feedback quality composite operationalised in the rubric.

### ***Mediation Analysis***

Hayes' PROCESS Model 4 was used to test whether feedback quality characteristics mediated the relationship between AI implementation and learning outcomes. AI implementation was coded as a dichotomous variable (0 = pre-intervention, 1 = post-intervention), feedback quality characteristics were operationalised as a composite of the four rubric dimensions, and student learning outcomes were measured as total MYP score change. Bootstrap confidence intervals were computed using 5,000 resamples. Table 4 presents the results.

#### **Table 4**

*Mediation Analysis: Feedback Quality Characteristics as Mediator (Hayes PROCESS Model 4, N = 101)*

Path	<i>B</i>	<i>SE</i>	<i>t</i>	95% CI (Bootstrap)
Path <i>a</i> : AI → Feedback Characteristics	0.847	0.142	5.96***	[0.565, 1.129]
Path <i>b</i> : Feedback Characteristics → Outcomes	1.624	0.287	5.66***	[1.055, 2.193]
Direct Effect ( <i>c'</i> ): AI → Outcomes	1.474	0.458	3.22**	[0.564, 2.384]
Indirect Effect ( <i>ab</i> ): via Feedback Characteristics	1.376	—	—	[0.802, 2.047]*
Total Effect ( <i>c</i> ): AI → Outcomes	2.850	0.441	6.46***	[1.975, 3.725]
Proportion Mediated ( <i>Pm</i> )	48.3%	—	—	—

*Note.* Bootstrap 95% CIs based on 5,000 resamples.  $R^2$  for mediator model = .286;  $R^2$  for outcome model = .412. \*Indirect effect CI excludes zero, confirming significant mediation. \*\*  $p < .01$ . \*\*\*  $p < .001$ .

The mediation analysis confirmed statistically significant partial mediation. Path *a* was significant ( $B = 0.847$ ,  $p < .001$ ), confirming that AI implementation significantly improved feedback quality characteristics. Path *b* was also significant ( $B = 1.624$ ,  $p < .001$ ), confirming that improved feedback characteristics were significantly associated with improved learning outcomes, controlling for AI implementation. The bootstrap 95% confidence interval for the indirect effect ( $ab = 1.376$ ) excluded zero [0.802, 2.047], and the direct effect ( $c' = 1.474$ ) remained significant ( $p = .002$ ), indicating partial rather than full mediation. Feedback quality characteristics accounted for 48.3% of the total effect of AI implementation on learning outcomes.

### RQ3: Implementation Challenges and Opportunities

Thematic analysis of teacher interviews, student focus groups, and classroom observations revealed three superordinate themes.

**Opportunities.** Teachers reported substantial reductions in grading time, with Teacher A estimating a 60% reduction. This time gain enabled redirection of professional energy towards one-on-one student support and instructional design. AI tools also provided more consistent application of MYP rubric standards across large numbers of students, reducing

fatigue-driven criterion drift. Students described self-directed revision cycles enabled by immediate feedback: *“I can try again the same night. Before, I had to wait, and by then we were already on a new topic”* (Student G9-08, Mid proficiency).

**Challenges.** All three teachers described a verification burden, particularly during early implementation. Teacher C noted: *“At first, I checked every single comment line by line. It took almost as long as writing the feedback myself.”* Linguistic and cultural blind spots in AI feedback required teacher mediation; Teacher B reported instances where ChatGPT used culturally Western examples or idiomatic expressions that confused Mandarin-speaking students. Low-proficiency students consistently reported difficulty understanding the academic English in AI-generated feedback, requiring peer translation or dictionary use.

**Institutional gaps.** All three teachers identified the absence of formal professional development specifically focused on AI-assisted assessment as the most significant institutional constraint. Teacher C’s self-reported trajectory from cautious verification in early weeks to efficient implementation by week six illustrated that the initial productivity deficit was not inevitable but reflected the absence of pre-implementation training—a modifiable institutional condition.

#### **RQ4: English Language Proficiency as a Moderator**

One-way ANOVA revealed a statistically significant effect of English proficiency group on learning gains,  $F(2, 98) = 5.67, p = .005, \eta^2 = .104$ , representing a medium effect. Tukey HSD post hoc comparisons showed that both Low- and Mid-proficiency groups demonstrated significantly greater gains than the High-proficiency group (Low vs. High,  $p = .012$ ; Mid vs. High,  $p = .003$ ); the difference between Low and Mid groups was not significant ( $p = .891$ ). Hayes’ PROCESS Model 1 confirmed a statistically significant interaction ( $B = -0.782, SE = 0.298, t = -2.62, p = .010, 95\% \text{ CI } [-1.373, -0.191]$ ), accounting for an additional 4.8% of variance beyond main effects ( $\Delta R^2 = .048$ ). The

conditional effect of AI implementation was statistically significant at all three proficiency levels but was largest for Low-proficiency students ( $B = 3.16, p < .001$ ), followed by Mid- ( $B = 2.38, p < .001$ ), and smallest for High-proficiency students ( $B = 1.59, p < .001$ ).

Technology Acceptance Survey results showed statistically significant differences across proficiency groups for all four constructs, with High-proficiency students consistently reporting the highest acceptance and Low-proficiency students the lowest. Table 5 presents gain scores and TAM results by proficiency group.

**Table 5**

*Learning Gains and Technology Acceptance by English Language Proficiency Group (N = 101)*

Measure	Low (n = 32)	Mid (n = 43)	High (n = 26)	ANOVA <i>p</i>
<b>Learning Gains</b>				
Gain Total Score, <i>M (SD)</i>	3.16 (2.84)	3.35 (2.61)	1.50 (2.12)	.005**
<b>Technology Acceptance (1–5 scale)</b>				
Perceived Usefulness, <i>M (SD)</i>	3.52 (0.78)	3.81 (0.69)	4.08 (0.61)	.025*
Perceived Ease of Use, <i>M (SD)</i>	3.61 (0.82)	3.88 (0.71)	4.09 (0.58)	.007**
Learning Effectiveness, <i>M (SD)</i>	3.41 (0.84)	3.72 (0.74)	3.94 (0.62)	.022*
Behavioural Intention, <i>M (SD)</i>	3.68 (0.76)	3.95 (0.66)	4.17 (0.54)	.034*

*Note.* Gain Total Score = post-intervention minus pre-intervention MYP total score (range 0–32).

Technology Acceptance scores on 1–5 Likert scale (1 = *Strongly Disagree*; 5 = *Strongly Agree*).

Cronbach’s  $\alpha$ : Perceived Usefulness = .82; Perceived Ease of Use = .85; Learning Effectiveness = .83;

Behavioural Intention = .88. \*  $p < .05$ . \*\*  $p < .01$ .

Qualitative data illuminated the mechanism underlying this moderation. Teacher A captured the central asymmetry: “*My Low proficiency students benefited the most from AI feedback in absolute terms, because they had the most room to grow. But they also needed the*

*most support from me to actually use the feedback. The Mid-level students were the sweet spot: they could read the AI feedback mostly on their own, and they had enough room to improve that the gains were visible.”* This statement encapsulates the qualitatively grounded interpretation of the negative interaction term: a ceiling effect limited measurable growth for High-proficiency students, while scaffolded teacher mediation enabled Low-proficiency students to access AI feedback benefits despite linguistic barriers.

## **Discussion**

This study examined the implementation and impact of AI-assisted grading in IB MYP Physics, yielding four principal findings that extend theoretical understanding and inform educational practice. First, AI implementation produced statistically significant improvements across all four MYP criteria. Second, approximately half of this effect was attributable to enhanced feedback quality characteristics, with the remainder operating through additional pathways. Third, AI tools and human teachers demonstrated complementary rather than substitutable strengths. Fourth, contrary to one plausible prediction, lower-proficiency students gained more than higher-proficiency students.

### **AI-Assisted Grading Improves Learning Outcomes Through Feedback Quality**

The finding that AI implementation produced statistically significant improvements across all four MYP criteria, with effect sizes from  $d = 0.43$  to  $d = 0.74$ , aligns with the broader literature on automated feedback in STEM education (Alqahtani et al., 2023; Cavalcanti et al., 2021; Zawacki-Richter et al., 2019), although the effect sizes observed here are more modest than those reported in some quasi-experimental studies (e.g., Altal & Abo Ehsaiyan, 2025). The largest effect for Criterion D, Reflecting on the Impacts of Science, is theoretically coherent: this criterion requires higher-order evaluative thinking, and AI feedback operating at Hattie and Timperley’s (2007) self-regulation level may be particularly effective for developing metacognitive awareness of assessment criteria.

The partial mediation finding (48.3%) confirms that feedback quality characteristics constitute a primary mechanism of AI implementation effects, consistent with Hattie and Timperley's feedback model. The remaining direct effect, however, suggests additional pathways including the motivational benefits of iterative revision cycles described by students. This is consistent with Self-Determination Theory's (Deci & Ryan, 2017) proposition that competence experiences—such as observing mastery progression through multiple submission attempts—generate intrinsic motivation that extends beyond feedback content *per se*.

### **The AI–Human Feedback Partnership**

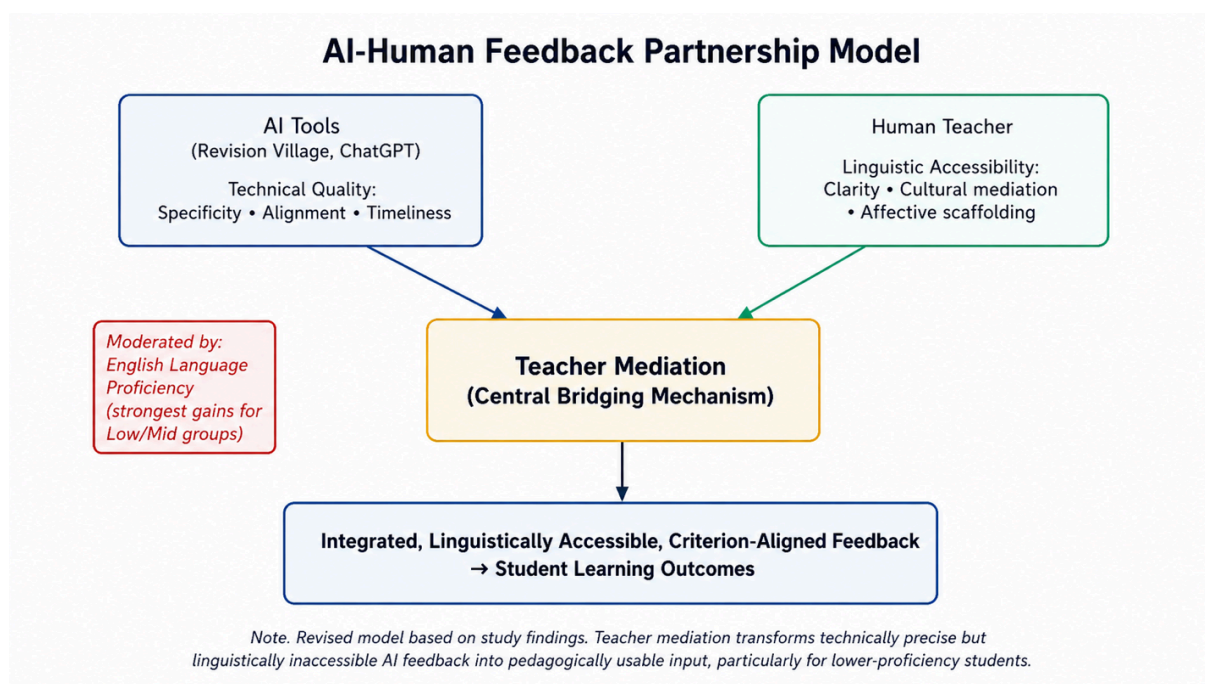
The rubric comparison findings reveal complementary strengths rather than AI superiority or deficiency. AI tools excelled at specificity and criterion alignment—dimensions where consistency and scale matter most—while human teachers excelled at linguistic clarity, the dimension most consequential for non-native English speakers. Importantly, the Clarity dimension in this study specifically measured linguistic accessibility for students whose first language was not English. AI-generated feedback was syntactically well-formed in English and would likely be perceived as clear by fluent English speakers; the clarity gap therefore reflects a context-specific limitation rather than a general deficit. This complementary pattern suggests that the effective unit of AI-assisted grading is not the AI tool alone but the AI–human partnership, in which teacher mediation transforms technically precise but linguistically inaccessible feedback into pedagogically usable input.

This partnership extends Task–Technology Fit theory (Goodhue & Thompson, 1995) by specifying that fit is not solely a property of tool–task alignment but is co-constructed through teacher mediation. Figure 2 presents the revised AI–Human Feedback Partnership Model derived from this study's findings, in which teacher mediation functions as the central bridging mechanism connecting AI-generated feedback (strong in technical quality) and the

linguistic accessibility required for non-native English speakers to translate feedback into learning gains.

## Figure 2

### *Revised AI–Human Feedback Partnership Model*



*Note.* The revised model decomposes feedback quality into technical quality (specificity, alignment, timeliness)—where AI tools consistently outperform human teachers—and linguistic accessibility (clarity, cultural mediation, affective scaffolding), where human teacher mediation provides indispensable complementary value, particularly for students with lower English language proficiency. Teacher mediation is positioned as the central bridging mechanism.

### **English Language Proficiency as a Contextual Moderator**

The significant moderation effect of English language proficiency ( $p = .010$ ) addresses a critical gap in the AI assessment literature. Contrary to concerns that AI tools may exacerbate educational inequality by favouring already-advantaged learners (Warschauer & Matuchniak, 2010), this study found that Low- and Mid-proficiency students demonstrated significantly larger gains than High-proficiency students. This finding is theoretically

generative because it implicates two simultaneous mechanisms: a ceiling effect that constrained measurable growth for High-proficiency students whose pre-intervention scores were already high, and a scaffolded-mediation mechanism through which lower-proficiency students gained substantial benefit when teacher mediation made AI feedback linguistically accessible. Crucially, these gains for lower-proficiency students were not frictionless; they required substantially more teacher scaffolding than gains for higher-proficiency students. The equity implication is that proportional investment in teacher mediation for lower-proficiency students is necessary for equitable outcomes, rather than a uniform tool deployment.

This finding also extends the Technology Acceptance Model (Davis, 1989) by demonstrating that language proficiency functions as a boundary condition for perceived usefulness and ease of use in English-medium instructional contexts. Technology acceptance research has rarely examined contexts where the technology delivers content in a language other than users' first language; the present study suggests that language proficiency moderates the core TAM relationships in such contexts.

### **Limitations**

Several limitations warrant consideration. First, the single-site case study design limits the generalisability of findings to other IB World Schools with different demographic, linguistic, or infrastructural characteristics. Second, the absence of a control group receiving only human teacher feedback constrains causal inference; although the within-subjects pre–post design provides reasonable internal validity, maturation, curriculum-alignment effects, and Hawthorne effects cannot be entirely ruled out. Third, the teacher sample of three participants precluded formal quantitative testing of teacher-level variables, although triangulated qualitative evidence provides convergent support for the role of teacher experience and institutional factors. Fourth, the 24-week implementation period does not

address the durability of gains over multiple academic terms, leaving open important questions about long-term outcomes. Finally, the proficiency classification used in this study, while triangulated, did not employ a standardised English language proficiency instrument; future replications would benefit from formal CALP measurement to support stronger inferences regarding linguistic mechanisms.

### **Implications for Practice and Policy**

For physics teachers, these findings support the adoption of differentiated feedback protocols calibrated to student proficiency levels, the integration of AI feedback into lesson cycles as structured revision windows, and the explicit cultivation of student AI literacy as part of assessment practice. For school administrators, the findings support investment in pre-implementation professional development for AI-assisted assessment—identified as the most significant institutional gap by all three teacher participants—and the establishment of living AI assessment policies subject to annual review. For IB curriculum developers and AI tool providers, IB-specific calibration of feedback tools, alongside the development of linguistically accessible feedback generation suitable for English-medium contexts with substantial multilingual populations, represents a high-leverage priority.

### **Conclusion**

This mixed-methods study provides empirical evidence that AI-assisted grading, implemented as an AI–human partnership, can improve student learning outcomes in IB MYP Physics while reducing feedback turnaround time from days to hours. Feedback quality characteristics partially mediated this effect, with AI tools providing superior specificity and criterion alignment, and human teachers providing superior linguistic accessibility. English language proficiency significantly moderated implementation effectiveness, with lower-proficiency students showing larger gains but requiring greater teacher mediation to access them. The proposed AI–Human Feedback Partnership Model positions teacher

mediation as the central bridging mechanism between AI-generated feedback and learner outcomes, particularly in linguistically diverse classrooms.

Future research should pursue multi-site replication across diverse IB World Schools, longitudinal designs examining the durability of gains, and experimental comparisons of AI-only, human-only, and integrated feedback conditions. The partnership model proposed here generates testable hypotheses for future investigation and offers a principled framework for responsible AI integration in linguistically diverse STEM classrooms. As AI assessment tools continue to evolve and their use in international school contexts continues to grow, the findings of this study suggest that the most consequential design choice is not which AI tool to deploy but how to embed teacher mediation within its deployment.

## References

- Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., Bin Saleh, K., & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy, 19*(8), 1236–1242. <https://doi.org/10.1016/j.sapharm.2023.05.016>
- Altal, A., & Abo Ehsaiyan, H. (2025). Evaluating AI-assisted instructional systems in secondary physics: A quasi-experimental study from the United Arab Emirates. *Inquisiva Open*.
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE.
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233. <https://doi.org/10.1080/03075070600572132>
- Cavalcanti, A. P., Barbosa, A., Costa, F. A., Tucci, C., & Gaona, M. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence, 2*, 100038. <https://doi.org/10.1016/j.caeai.2021.100038>
- Chen, Z., & Wan, T. (2024). Achieving human-level partial credit grading of written responses to physics conceptual questions using GPT-3.5 with only prompt engineering. *arXiv*. <https://arxiv.org/abs/2407.15251>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(1), 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Multilingual Matters.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Deci, E. L., & Ryan, R. M. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford.
- Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2019). Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model. *Information Systems Frontiers*, 21, 719–734. <https://doi.org/10.1007/s10796-017-9774-y>
- Fathali, S., & Okada, T. (2018). Technology acceptance model in technology-enhanced OCLL contexts: A self-determination theory approach. *Australasian Journal of Educational Technology*, 34(4). <https://doi.org/10.14742/ajet.3629>
- Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 19(2), 213–236. <https://doi.org/10.2307/249689>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- International Baccalaureate Organization. (2023). *MYP sciences guide*. International Baccalaureate Organization.

- Kortemeyer, G. (2023). Toward AI grading of student problem solutions in introductory physics: A feasibility study. *Physical Review Physics Education Research*, 19(2), 020163. <https://doi.org/10.1103/PhysRevPhysEducRes.19.020163>
- Kortemeyer, G., & Nöhl, J. (2024). Assessing confidence in AI-assisted grading of physics exams through psychometrics: An exploratory study. *Physical Review Physics Education Research*, 21(1), 010136.
- Latif, E., Lee, G. G., Neumann, K., Kastorff, T., & Zhai, X. (2024). G-SciEdBERT: A contextualized LLM for science assessment tasks in German. *arXiv*. <https://arxiv.org/abs/2402.06584>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
- Mandouit, L., & Hattie, J. (2023). Revisiting “The power of feedback” from the perspective of the learner. *Learning and Instruction*, 84, 101718. <https://doi.org/10.1016/j.learninstruc.2022.101718>
- Molin, F., Haelermans, C., Cabus, S., & Groot, W. (2021). Do feedback strategies improve students’ learning gain? Results of a randomized experiment using polling technology in physics classrooms. *Computers & Education*, 175, 104339. <https://doi.org/10.1016/j.compedu.2021.104339>
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178. <https://doi.org/10.2307/41410412>
- Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 252, 124167. <https://doi.org/10.1016/j.eswa.2024.124167>

- Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education, 34*, 179–225. <https://doi.org/10.3102/0091732X09349791>
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, 487662. <https://doi.org/10.3389/fpsyg.2019.03087>
- Wongvorachan, T., Lai, K. W., Bulut, O., Tsai, Y. S., & Chen, G. (2022). Artificial intelligence: Transforming the future of feedback in education. *Journal of Applied Testing Technology, 23*, 95–116.
- Yaseen, H., Mohammad, A. S., Ashal, N., Abusaimh, H., Ali, A., & Sharabati, A. A. (2025). The impact of adaptive learning technologies, personalized feedback, and interactive AI tools on student engagement: The moderating role of digital literacy. *Sustainability, 17*(3), 1133. <https://doi.org/10.3390/su17031133>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, 39. <https://doi.org/10.1186/s41239-019-0171-0>

## Assessment Practices and Challenges in Biology Education at a China-Based International School: A Mixed-Methods Case Study

Liu Yan Tin

University College Fairview

### Abstract

This study examined how various assessment types are perceived and ranked by biology students and teachers in a China-based international school, and how these assessments are associated with student engagement and conceptual understanding. Operating under the Cambridge IGCSE curriculum, the school faces a fundamental tension: its student-centred, formative assessment philosophy conflicts with China's entrenched examination culture. Using an explanatory sequential mixed-methods design grounded in pragmatism, quantitative data were collected from 55 IGCSE biology students and 6 teachers via structured questionnaires, followed by semi-structured qualitative interviews with 20 students and 5 teachers. Findings indicate that formative and summative assessments were rated most effective by both students (54.5% and 45.5% respectively) and teachers (66.7% each), while oral presentations and individual projects received strongly negative ratings. Three mediating variables — assessment-learning alignment ( $r = 0.831$ ), student autonomy ( $r = 0.787$ ), and feedback quality ( $r = 0.689$ ) — showed strong associations with student engagement. Students' primary challenges centred on linguistic barriers and memorisation load; teachers identified authentic assessment in the AI era and balancing multiple curriculum demands as core concerns. Six improvement opportunity categories were identified, including assessment format diversification, enhanced feedback systems, and guided technology integration. This study contributes evidence-based guidance for international school educators and policymakers seeking to integrate global pedagogical standards with localised educational expectations.

**Keywords:** biology assessment; formative assessment; student engagement; international schools; China; mixed-methods; IGCSE

## 1. Introduction

Assessment is a foundational element of teaching and learning, shaping instructional decisions, measuring student progress, and influencing the overall quality of education (Wulandari et al., 2020). In biology education, where conceptual understanding must be integrated with practical reasoning and scientific inquiry, the design and implementation of assessments carry particular importance. The multidimensional nature of biological knowledge — spanning cognitive, psychomotor, and attitudinal domains — demands a similarly multifaceted approach to assessment (Clark & Hsu, 2023).

Despite widespread consensus on the benefits of formative assessment for enhancing learning (Black & Wiliam, 1998; Wiliam, 2011), practice in many science classrooms continues to privilege summative, examination-centred approaches. This misalignment is especially pronounced in international schools operating in China, where global pedagogical ideals meet a deeply rooted examination culture centred on high-stakes testing (Yu, 2023). International schools following curricula such as the Cambridge IGCSE thus occupy a contested educational space, simultaneously promoting inquiry-based, formative practices while navigating the normative pressures of Chinese academic culture.

Research on assessment in international schools remains sparse, particularly in contexts that bridge Western curriculum frameworks with Asian educational expectations (Hayden, 2011). Even fewer studies have examined how these dynamics unfold specifically in the biology classroom, where authentic scientific reasoning and inquiry skills are explicit curriculum goals (National Research Council, 2014). This study addresses that gap.

The present study examined assessment practices and challenges at one China-based international school, guided by three objectives: (1) to explore how various assessment types are perceived and ranked in terms of effectiveness by IGCSE biology students and teachers; (2) to investigate the associations between assessment types and students' engagement and conceptual understanding; and (3) to identify challenges and opportunities associated with implementing assessment practices that align with desired learning outcomes.

The study is grounded epistemologically in pragmatism (Creswell & Plano Clark, 2018) and draws on an integrated theoretical framework comprising Constructivist Theory, Socio-Cultural Theory, Self-Determination Theory, and Expectancy-Value Theory. Together, these frameworks illuminate the cognitive, social, motivational, and value-related dimensions of assessment and learning.

## **2. Literature Review**

### **2.1 Assessment in Biology Education**

Assessment in science education has undergone significant transformation in recent decades, shifting from a primary focus on knowledge recall towards evaluating higher-order cognitive skills including application, synthesis, and scientific reasoning (Pellegrino, 2014). In biology specifically, this evolution reflects recognition that disciplinary competence extends beyond factual knowledge to encompass practical skills, critical thinking, and the capacity to engage with complex biological systems (Goldey et al., 2012).

Formative assessment — designed to monitor and improve learning during the instructional process — has attracted sustained scholarly attention as a mechanism for enhancing student outcomes (Black & Wiliam, 1998). By providing timely, specific feedback that students can use to adjust their learning strategies, formative practices are theorised to support metacognitive development and intrinsic motivation (Ismail et al., 2022). Summative assessments, while often positioned as primarily accountability tools, can also promote learning when designed thoughtfully (Looney, 2011). The relative effectiveness of these two broad approaches, however, remains empirically contested, with context-dependent factors playing a decisive moderating role (Drew et al., 2021).

Alternative assessment formats — including projects, oral presentations, and authentic assessments — have been advocated as means of evaluating competencies not easily captured through written examinations (Kulasegaram & Rangachari, 2018). However, the literature cautions that such formats deliver their intended benefits only when implemented with explicit expectations, adequate scaffolding, and clear alignment to learning outcomes (DeLuca et al., 2013). Without these conditions, alternative assessments may increase student anxiety without contributing meaningfully to learning.

### **2.2 Assessment in International Schools and China's Educational Context**

International schools in China operate at the intersection of two pedagogically distinct traditions. Western curricula such as the Cambridge IGCSE advocate student-centred, formative approaches to assessment, while Chinese educational culture is characterised by examination-centrism, content coverage emphasis, and preparation for high-stakes national examinations such as the Gaokao (Yu, 2023). Even students in international schools who are

not subject to the Gaokao are embedded in cultural contexts that normalise examination-oriented learning (Kirkpatrick & Zhang, 2011).

This cultural duality creates structural tensions for both teachers and students. Teachers must navigate the gap between international curriculum standards and local expectations, while students face the challenge of adapting to assessment paradigms that differ markedly from their prior schooling experiences (Jing, 2023). Research on these dynamics remains limited, with most existing literature focusing on broader questions of international school identity or aggregate student outcomes rather than the granular dynamics of classroom assessment (Hayden, 2011; Yemini, 2012).

### **2.3 Theoretical Framework**

This study is guided by four complementary theoretical frameworks. Constructivist Theory (Piaget; Bruner, 1960) positions assessment as a tool for scaffolding progressive conceptual development, suggesting that effective assessments should challenge students at the edge of their existing understanding rather than merely confirming memorised knowledge. Socio-Cultural Theory (Vygotsky) emphasises the social dimensions of learning and assessment, highlighting the role of collaborative tasks and dialogic feedback in extending students' zones of proximal development (Allal & Ducrey, 2000).

Self-Determination Theory (Ryan & Deci, 2020) provides a motivational lens, positing that assessment designs which support students' sense of autonomy, competence, and relatedness will promote intrinsic motivation and deeper engagement, while controlling or anxiety-inducing assessments will undermine these outcomes. Finally, Expectancy-Value Theory (Wigfield et al., 2016) frames students' assessment engagement as a function of their expectations of success and their perception of task value, foregrounding the importance of authentic, meaningful assessment tasks.

## **3. Methodology**

### **3.1 Research Design**

An explanatory sequential mixed-methods design was employed (Creswell & Plano Clark, 2018). In this design, quantitative data collection and analysis are conducted first and are followed by qualitative data collection that contextualises and elaborates upon the quantitative findings. This design was selected because the study's objectives required both the breadth of quantitative pattern-detection across a defined cohort and the depth of

qualitative insight into participants' lived experiences of assessment. The study is grounded in pragmatism, privileging empirical usefulness over methodological orthodoxy.

### **3.2 Research Setting and Participants**

The study was conducted at an international school in metropolitan mainland China offering Cambridge IGCSE and International Baccalaureate programmes. For the quantitative phase, 55 IGCSE biology students from Grades 9 and 10 (predominantly local Chinese, ages 14–16) completed structured questionnaires, representing approximately 27.5% of the IGCSE biology cohort. All 6 biology teachers completed parallel questionnaires. For the qualitative phase, 20 students and 5 teachers participated in semi-structured interviews. Participants were selected through a combination of convenience and purposive sampling to ensure diversity across gender, grade level, teaching experience, and nationality. Student interview data reached thematic saturation at the 16th interview, with subsequent interviews serving a confirmatory function (Braun & Clarke, 2021).

### **3.3 Instruments**

Both student and teacher questionnaires were developed from established instruments, including the Student Engagement Instrument (Appleton et al., 2008), the National Survey of Student Engagement (Kuh, 2003), and Bloom's Taxonomy-Based Assessment frameworks (Anderson & Krathwohl, 2001). The instruments incorporated Likert-type scales and rank-order items covering eight assessment types: formative assessment, summative assessment, authentic assessment, self-assessment, technology-enhanced assessment, laboratory practical assessment, individual or group projects, and oral presentations. Content validity was confirmed by an expert panel (Content Validity Index = 0.89) and internal consistency was established through Cronbach's alpha (student questionnaire  $\alpha = 0.84$ ; teacher questionnaire  $\alpha = 0.85$ ).

Semi-structured interview protocols explored participants' experiences with assessment effectiveness, engagement, understanding, and challenges, with protocols refined through pilot testing to enhance clarity and manageability. Inter-rater reliability for qualitative coding was confirmed by Cohen's kappa (teachers:  $\kappa = 0.8457$ ; students:  $\kappa = 0.8143$ ), both exceeding the 0.80 benchmark for substantial agreement.

### **3.4 Data Analysis**

Quantitative data were analysed using descriptive statistics, Kruskal-Wallis H tests, Spearman correlations, and Pearson correlations. Spearman correlations were selected for

item-level association analyses given the ordinal nature of Likert-scale data; Pearson correlations were used for composite mediating variable scores. Qualitative data were analysed using Braun and Clarke's (2021) reflexive thematic analysis, with themes verified through member checking with 6 students and 2 teachers, and triangulation with quantitative findings.

## 4. Findings

### 4.1 Research Question 1: Assessment Effectiveness Perceptions

Students and teachers converged strongly on the perceived effectiveness of conventional assessment types. Formative assessment received the highest combined positive rating from students (54.5% 'Most Effective' + 'Very Effective'), while summative assessment followed closely (45.5%). Teachers rated both formative and summative assessment equally high (66.7% combined positive). Notably, formative assessment's high student rating was driven by a 40.0% 'Very Effective' frequency — the single highest category score recorded across all assessment types and respondent groups.

At the opposite end, oral presentations received the most negative ratings in the dataset, with 89.1% of students and 83.3% of teachers rating them negatively ('Ineffective', 'Very Ineffective', or 'Most Ineffective'). Individual or group projects attracted similarly negative assessments (63.6% student negative ratings; 83.3% teacher negative ratings). Technology-enhanced assessment occupied a low-middle position (9.1% student combined positive).

A significant divergence emerged regarding self-assessment: 20.0% of students rated it positively, whereas teachers awarded it no positive ratings and 33.3% rated it 'Ineffective'. Kruskal-Wallis tests revealed no statistically significant grade-level differences across all eight assessment types (all  $p > .05$ ), justifying reporting of aggregate findings.

Qualitative findings contextualised these patterns. Students preferred formative assessments because they provided clear feedback and reduced examination anxiety. The strong dislike of oral presentations and projects was attributed to vague expectations, limited scaffolding, and high perceived stakes relative to perceived learning benefit. Students and teachers alike acknowledged a gap between their stated preference for authentic, real-world assessment and the reality that written examinations continued to dominate classroom practice.

#### **4.2 Research Question 2: Assessment Types and Student Engagement**

Students' highest reported engagement drivers were: student autonomy in approaching assessments (78.2% combined agreement), perception of assessment value for learning (72.7%), and opportunity to pursue personally interesting topics (67.3%). Teachers identified class discussion promotion (83.4% combined agreement) and authentic, meaningful assessments (66.7%) as the key engagement mechanisms.

Three mediating variables — assessed through Pearson correlations with overall engagement scores — demonstrated substantial associations: assessment-learning alignment ( $r = 0.831$ ,  $p < .001$ ), student autonomy and agency ( $r = 0.787$ ,  $p < .001$ ), and feedback quality ( $r = 0.689$ ,  $p < .001$ ). The strength of the alignment correlation was particularly striking, indicating that students who perceived assessments as directly connected to their learning goals reported markedly higher overall engagement.

A notable divergence emerged between quantitative and qualitative data: while 74.5% of students endorsed creative assessments as valuable in principle (Likert item), they simultaneously gave the most negative effectiveness ratings to oral presentations and projects — formats that are typically classified as creative. Qualitative data resolved this paradox: students endorsed creative engagement abstractly but described the current implementation of projects and presentations as poorly scaffolded, generating anxiety rather than authentic creativity.

Teacher and student perspectives diverged on engagement mechanisms: teachers emphasised structural autonomy (83.3% prevalence in teacher interviews) while students prioritised authentic content connections and real-world relevance (55.0% prevalence in student interviews). This divergence suggests a professional development gap wherein teachers design autonomy structures without anchoring them to content contexts that students find motivating.

#### **4.3 Research Question 3: Assessment Types and Conceptual Understanding**

Quantitative data revealed a systematic decline in assessment effectiveness across Bloom's taxonomy levels: knowledge-level assessment effectiveness was endorsed by 74.5% of students, falling to 49.1% at the synthesis level — a 25.4 percentage-point gap. This pattern suggests that current assessments are reasonably effective at supporting lower-order cognitive tasks but substantially less effective at developing the synthesis and evaluation capabilities required by the IGCSE biology curriculum.

Feedback quality ( $r = 0.689$  with overall engagement) and assessment-learning alignment ( $r = 0.831$ ) emerged as key mediating mechanisms through quantitative and qualitative analysis. Students who received specific, actionable feedback reported greater ability to revise their understanding and close conceptual gaps. However, the majority of students described feedback as delayed, grade-focused, and insufficiently developmental — arriving after the assessed work had been returned with no opportunity for revision.

Teachers expressed concern about the capacity of existing assessments to distinguish genuine understanding from surface recall, a challenge amplified by the emergence of AI tools that can generate scientifically plausible responses to standard biology questions.

#### **4.4 Student Challenges (RQ4) and Teacher Challenges (RQ5)**

Five categories of student challenges were identified: (1) linguistic and vocabulary difficulties, particularly for Chinese students completing assessments in English; (2) memorisation load and content volume, which incentivised surface-level retention over deep understanding; (3) conceptual understanding barriers, especially in transitioning from factual recall to examination-response formats; (4) knowledge application and examination response challenges; and (5) practical and technical implementation challenges. Vocabulary and memorisation load were the most pervasive barriers, appearing across all twenty student interview transcripts.

Three teacher challenge categories were identified: (1) ensuring authentic understanding in the AI era, as teachers were uncertain whether student submissions reflected genuine comprehension or AI-assisted generation; (2) balancing multiple assessment dimensions — content coverage, skill development, practical application, and examination preparation — within curriculum time constraints; and (3) language and terminology barriers in distinguishing linguistic difficulties from conceptual misunderstandings among multilingual learners.

#### **4.5 Opportunities for Improvement (RQ6)**

Students and teachers demonstrated high convergence on improvement priorities, identifying six opportunity categories: (1) assessment format diversification beyond written examinations, incorporating oral alternatives, projects, and laboratory components — with emphasis on purposeful variety supported by clear rubrics; (2) enhanced feedback systems prioritising timeliness, specificity, and actionability; (3) practical and experiential learning enhancement; (4) technology integration and guided AI use that preserves assessment rigour;

(5) differentiation and personalisation of assessment approaches; and (6) alignment between assessment design and higher-order curriculum goals.

## **5. Discussion**

### **5.1 The Paradox of Conventional Assessment Effectiveness**

The finding that both students and teachers rated formative and summative assessments as most effective — while simultaneously expressing a desire for more authentic, real-world assessment — points to a structural paradox that extends beyond this specific context. This convergence challenges the dominant narrative in assessment literature that positions formative assessment as unequivocally superior to summative assessment (Black & Wiliam, 1998). Instead, the data suggest that students and teachers recognise the complementary functions of these two broad assessment paradigms (Looney, 2011), while acknowledging that neither currently realises its pedagogical potential in this setting.

The strong negative ratings for oral presentations and projects are particularly theoretically significant. Progressive education literature frequently assumes that these alternative formats inherently promote deeper learning (MacMath et al., 2010). This study's data suggest, rather, that implementation quality — not format selection — is the critical variable. As DeLuca et al. (2013) argue, alternative assessment formats deliver learning benefits only when supported by transparent expectations, sufficient scaffolding, and authentic alignment to learning outcomes, conditions that were clearly absent in the present context.

### **5.2 Mediating Variables as Leverage Points**

The three Pearson correlations — assessment-learning alignment ( $r = 0.831$ ), student autonomy ( $r = 0.787$ ), and feedback quality ( $r = 0.689$ ) — are theoretically coherent with the integrated framework employed. Alignment corresponds to Constructivist theory's proposition that effective assessments scaffold progressive understanding; autonomy maps onto Self-Determination Theory's need for self-direction (Ryan & Deci, 2020); and feedback quality relates to Vygotsky's concept of the Zone of Proximal Development, wherein assessment feedback serves as the scaffolding that enables students to accomplish tasks beyond their current independent capacity (Allal & Ducrey, 2000).

Crucially, all three mediating variables are amenable to deliberate pedagogical design. This finding has practical significance: meaningful assessment reform does not require

dismantling the IGCSE examination architecture. Rather, it requires strategically strengthening alignment, autonomy, and feedback quality within existing structural constraints. Schools that invest in these three areas can improve student outcomes without awaiting curriculum-level reform.

### **5.3 AI as an Emergent Assessment Challenge**

Teachers' concern about authentic understanding in the AI era represents a frontier challenge inadequately addressed in current assessment theory. The rapid proliferation of generative AI tools capable of producing scientifically plausible responses to standard biology questions has destabilised traditional assumptions of assessment validity (Swiecki et al., 2022; Wang & Li, 2024). The finding that teachers are uncertain whether student submissions reflect genuine understanding or AI assistance signals an urgent need for assessment redesign that foregrounds skills AI cannot readily replicate: experimental reasoning, ethical argumentation, oral explanation, and real-time problem-solving.

### **5.4 Language Barriers and Assessment Validity**

The pervasive linguistic challenges identified among Chinese students completing IGCSE biology assessments in English carry important implications for assessment validity. Where students are assessed in a language in which they are not proficient, there is a risk of conflating language barriers with conceptual misunderstanding, systematically disadvantaging multilingual learners (Gram et al., 2013). International assessment systems that assume linguistic uniformity may inadvertently measure language proficiency alongside — or instead of — biological understanding, an issue that requires attention both at the level of individual classroom practice and at the level of international curriculum design.

## **6. Conclusion**

This study examined assessment practices and challenges in IGCSE biology education at a China-based international school, using an explanatory sequential mixed-methods design with 55 students and 6 teachers. Key findings reveal a structural paradox: students and teachers share aspirations for meaningful, higher-order assessment but operate within a system where examination demands compress the time, flexibility, and agency required to realise those aspirations. Three mediating variables — assessment-learning alignment, student autonomy, and feedback quality — emerged as the most influential determinants of student

engagement and understanding, and all three are addressable through deliberate pedagogical and institutional design.

The study contributes to assessment theory by demonstrating that single-paradigm advocacy (formative over summative, or vice versa) misrepresents how practitioners perceive assessment effectiveness in complex, cross-cultural settings. It extends the assessment-for-learning literature by identifying AI as an emergent validity threat and by documenting the understudied context of international schools in China. For practice, it offers three high-priority interventions: systematic enhancement of feedback quality, professional development that connects autonomy structures to authentic content, and reconfiguration of assessment tasks to explicitly develop higher-order thinking.

The study's single-site design limits its generalisability. Future research should adopt multi-site comparative designs, incorporate observational and artefact-based evidence, and develop more sophisticated cultural measures beyond nationality proxies. Longitudinal designs would illuminate whether the assessment perceptions documented here change as students progress through international school programmes. The AI challenge, in particular, warrants sustained empirical attention as both a practical and theoretical frontier in assessment research.

## References

- Allal, L., & Ducrey, G. P. (2000). Assessment of — or in — the zone of proximal development. *Learning and Instruction*, 10(2), 137–152. [https://doi.org/10.1016/s0959-4752\(99\)00025-0](https://doi.org/10.1016/s0959-4752(99)00025-0)
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's educational objectives*. Longman.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, 48(1), 12–19.
- Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in Schools*, 45(5), 369–386.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE.
- Bruner, J. S. (1960). *The process of education*. Harvard University Press.

- Clark, N., & Hsu, J. L. (2023). Insight from biology program learning outcomes: Implications for teaching, learning, and assessment. *CBE—Life Sciences Education*, 22(1). <https://doi.org/10.1187/cbe.22-09-0177>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- DeLuca, C., Chávez, T. R., Bellara, A. P., & Cao, C. (2013). Pedagogies for preservice assessment education: Supporting teacher candidates' assessment literacy development. *The Teacher Educator*, 48(2), 128. <https://doi.org/10.1080/08878730.2012.760024>
- Drew, S., Thomas, A., & Hardy, I. (2021). Student conceptions of effective assessment practice: A review and analysis. *Assessment & Evaluation in Higher Education*, 46(3), 391–408.
- Goldey, E. S., Abercrombie, C. L., Ivy, T. M., Kusher, D. I., Moeller, J. F., Rayner, D. A., & Steinberg, S. L. (2012). Transformative learning: Personal empowerment in biological education. *Journal of College Biology Teaching*, 38(1), 3–14.
- Gram, M., Jæger, K., Jun-yang, L., Li, Q., & Wu, X. (2013). Chinese students making sense of problem-based learning and Western teaching: Pitfalls and coping strategies. *Teaching in Higher Education*, 18(7), 761–772. <https://doi.org/10.1080/13562517.2013.836096>
- Hayden, M. (2011). Transnational spaces of education: The growth of the international school sector. *Globalisation, Societies and Education*, 9(2), 211–224.
- Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: Impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia*, 12(1). <https://doi.org/10.1186/s40468-022-00191-4>
- Jing, H. (2023). Understanding the IB curriculum reform in China: An exploratory study on stakeholders' perceptions. *Frontiers in Education*, 8. <https://doi.org/10.3389/educ.2023.1135787>
- Kirkpatrick, R. J., & Zhang, Y. (2011). The negative influences of exam-oriented education on Chinese high school students: Backwash from classroom to child. *Language Testing in Asia*, 1. <https://doi.org/10.1186/2229-0443-1-3-36>
- Kuh, G. D. (2003). What we're learning about student engagement from NSSE. *Change: The Magazine of Higher Learning*, 35(2), 24–32.
- Kulasegaram, K., & Rangachari, P. K. (2018). Beyond 'formative': Assessments to enrich student learning. *Advances in Physiology Education*, 42(1), 5–14. <https://doi.org/10.1152/advan.00122.2017>
- Looney, J. W. (2011). Integrating formative and summative assessment: Progress toward a seamless system? *OECD Education Working Papers*, No. 58. OECD Publishing.
- MacMath, S., Roberts, J., Wallace, J., & Chi, X. (2010). Curriculum and teaching dialogue: Perceptions of authentic assessment. *Curriculum and Teaching Dialogue*, 12(1), 49–61.

- National Research Council. (2014). *Developing assessments for the next generation science standards*. National Academies Press.
- Pellegrino, J. W. (2014). Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress. *Psicología Educativa*, 20(2), 65–77.
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Swiecki, Z., Khosravi, H., Chen, G., Martínez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075. <https://doi.org/10.1016/j.caeai.2022.100075>
- Wang, L., & Li, W. (2024). The impact of AI usage on university students' willingness for autonomous learning. *Behavioral Sciences*, 14(10), 956. <https://doi.org/10.3390/bs14100956>
- Wigfield, A., Tonks, S. M., & Klauda, S. L. (2016). Expectancy-value theory. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of Motivation at School* (2nd ed., pp. 55–74). Routledge.
- Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.
- Wulandari, T. R., Ramli, M., & Muzzazinah, M. (2020). The assessment of understanding biological concepts: A systematic review. *Journal of Physics: Conference Series*, 1563. <https://doi.org/10.1088/1742-6596/1563/1/012030>
- Yemini, M. (2012). Conceptualising the international in Israeli schools. *Comparative Education*, 48(2), 151–164.
- Yu, J. (2023). Exam culture and formative assessment in China: The Gaokao reform and its sociocultural hindrance. *Journal of Education, Humanities and Social Sciences*, 23, 291–301. <https://doi.org/10.54097/ehss.v23i.12900>