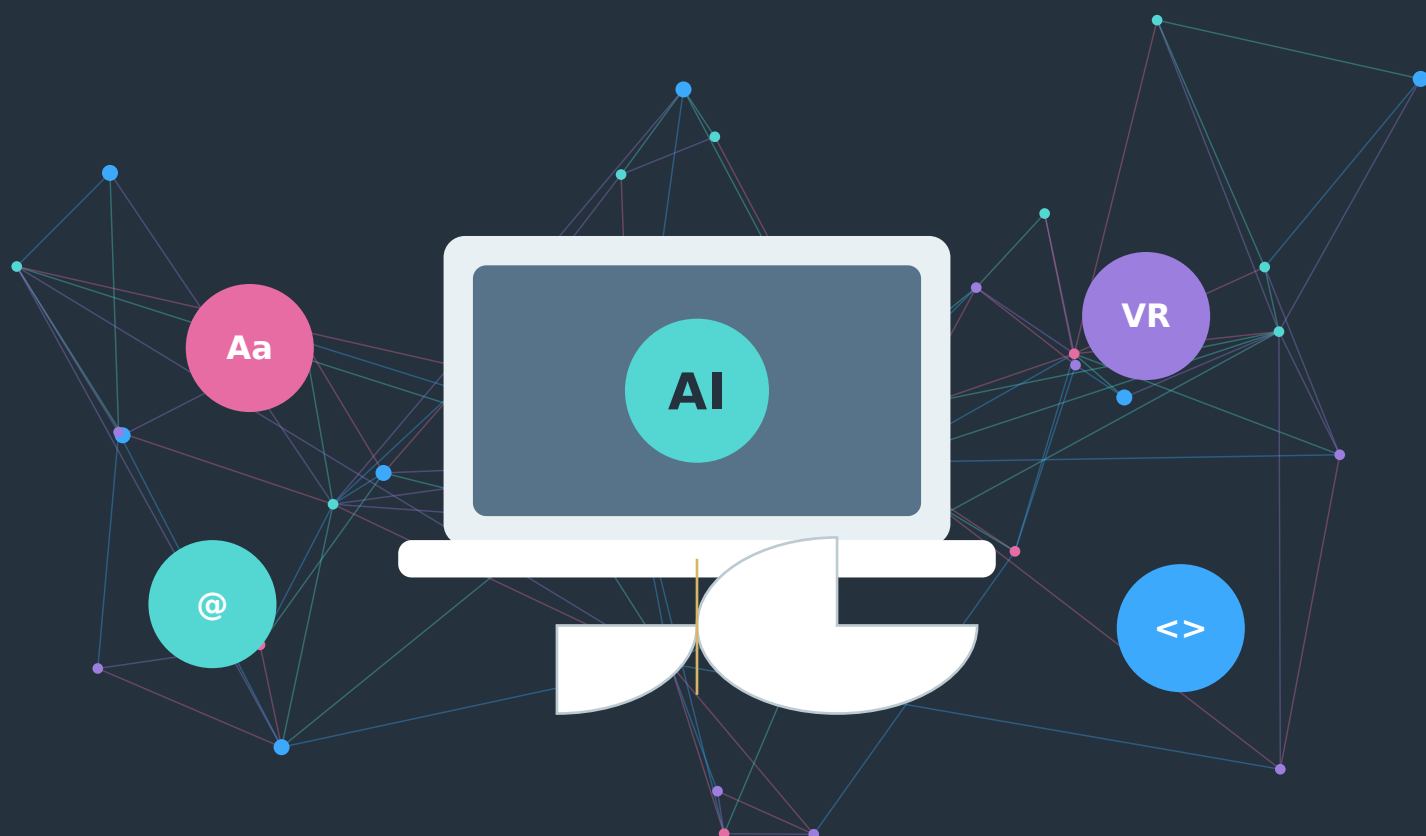


PROCEEDINGS

7th International Conference on Technology-Enhanced Language Learning **ICTELL2026**

16-17 July 2026 | Online



ICTELL2026 PROCEEDINGS

7th International Conference on Technology-Enhanced Language Learning

PROCEEDINGS

7th International Conference on Technology-Enhanced Language Learning

ICTELL2026

16-17 July 2026

Online Conference

Editor

Harwati Hashim

PUBLISHED BY

Faculty of Education
Universiti Kebangsaan Malaysia
2026

PUBLICATION AND eISSN INFORMATION

Cetakan Pertama / First Printing 2026

Hak Cipta Universiti Kebangsaan Malaysia / Copyright Universiti Kebangsaan Malaysia, 2026

Hak cipta terpelihara. Tiada bahagian daripada penerbitan ini boleh diterbitkan semula, disimpan dalam sistem capaian semula, atau dipindahkan dalam sebarang bentuk atau melalui sebarang cara, sama ada elektronik, mekanikal, fotokopi, rakaman atau sebaliknya, tanpa kebenaran bertulis daripada Fakulti Pendidikan, Universiti Kebangsaan Malaysia.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission from the Faculty of Education, Universiti Kebangsaan Malaysia.

Diterbitkan di Malaysia oleh / Published in Malaysia by

FAKULTI PENDIDIKAN

UNIVERSITI KEBANGSAAN MALAYSIA

43600 UKM Bangi, Selangor D.E., MALAYSIA

<https://www.ukm.my/fpend>

E-mail: fpending@ukm.edu.my

PROCEEDINGS OF THE 7th INTERNATIONAL CONFERENCE ON TECHNOLOGY-ENHANCED LANGUAGE LEARNING (ICTELL2026) - ONLINE

Conference Chair: Prof. Dr. Harwati Hashim

eISSN 3083-9513

KOD JALUR / BARCODE



eISSN 3083-9513

PREFACE

It is my pleasure to present the proceedings of the 7th International Conference on Technology-Enhanced Language Learning (ICTELL2026), held online on 16-17 July 2026. The conference brought together educators, researchers, practitioners and postgraduate scholars to share ideas and evidence on the evolving relationship between technology, language learning and educational practice.

The papers in this volume reflect both the breadth and the growing maturity of the field. They address artificial intelligence in language learning and assessment, digital learning design, learner engagement, speaking and reading development, digital citizenship, culturally responsive pedagogy, and the responsible integration of technology in heritage and interdisciplinary education. Across these varied contexts, a common message emerges: technology is most meaningful when it is guided by sound pedagogy, ethical awareness and a clear understanding of learners and their communities.

This collection also demonstrates the value of connecting innovation with careful inquiry. The studies and conceptual papers move beyond the adoption of tools to examine how digital approaches can support participation, feedback, communication, inclusion and authentic learning. I hope the volume will encourage readers to continue questioning not only what technology can do, but also how and why it should be used in particular educational settings.

I extend my sincere appreciation to all authors for contributing their work, to the reviewers for their thoughtful evaluations, and to the organising committee, speakers, session chairs and participants whose commitment made ICTELL2026 possible. Their collective efforts have enabled this proceedings volume to serve as a record of the conference and a resource for future teaching, research and collaboration.

May these papers inspire continued dialogue and responsible innovation in technology-enhanced language learning, within Malaysia and across the wider international community.

Prof. Dr. Harwati Hashim

Conference Chair, ICTELL2026

Faculty of Education, Universiti Kebangsaan Malaysia

CONTENTS

AI-Assisted Infographic Poster Preparation in Technical English 2: Students' Perceptions of NotebookLM	1
Suzianah Sahar; Noor Azua Mohd Bahatera @ Mohd Putera; Diana Ahmad Busra	
Negotiating Digital and Academic Registers: Influence of TikTok Slang on Academic Language Use among University Students	15
Ahmad Syahmi bin Salehuddin; Haliza Harun	
AI Tool Perception and Self-Regulated English Learning among Polytechnic Students: Examining Perception Versus Frequency of Use	26
Sharifah Nozi Syed Hamid	
Students' Acceptance of CIDOS in Education 5.0-Oriented Blended Communicative English Learning: A Conceptual Framework for Malaysian TVET	37
Nurfarizah binti Ishak; Masniza binti Mansor; Diana binti Ahmad Busra	
Empowering Responsible Digital Citizens: A Survey of Cyberbullying and Digital Citizenship Among Adolescents in an International School	54
Katherine Chong Pui Kei; Shanti Ramalingam; Syeda Rabia Tahir	
Constructing a Sustainable ICH-STEAM Curriculum Architecture in Higher Education: An Expert Consensus via Delphi Method	73
Sun Yanhu; Nuriah Abd Majid	
Navigating the Implementation Constraints of Cultural Heritage Education: A Qualitative Thematic Analysis of University Instructors	89
Sun Yanhu; Nuriah Abd Majid	
Evaluating Curriculum Priorities in Traditional Vocal Heritage: An Analytic Hierarchy Process (AHP) Approach on Mongolian Khoomei	98
Sun Yanhu; Nuriah Abd Majid	
Optimizing Student Engagement in Interdisciplinary Heritage Education: A Fuzzy-Set Qualitative Comparative Analysis (fsQCA)	107
Sun Yanhu; Nuriah Abd Majid	

CONTENTS CONTINUED

Sustaining the Intangible: The Culture-Technology-Ethics Symbiotic Model (CTESM) and Eight-Dimensional Assessment Framework (EDAF) for Heritage Education	114
Sun Yanhu; Nuriah Abd Majid	
Designing CEFR-Aligned Content-Based Speaking Materials with SDG 4 Themes for Islamic Higher Education	124
Kalsum; Rahimah Saimin; Engku Muhammad Syafiq Engku Shafruddin	
Onomatopoeic-Based Approach for Teaching Personification in Poetry Among Malaysian Lower Secondary Proficiency ESL Learners	134
Pei Yi Tan; Nur Izzah Azman; Nur Haziqah Nadhrah Mohd Fuad; Hanita Hanim Ismail	
Wild Wonders: A CALL-Based Digital Learning Module for Enhancing Reading Comprehension among Year 4 ESL Learners	153
Reshantini Rajendran; Nurul Juni Melissa Muhammad Akif Mark; Harwati Hashim; Nur Syafiqah Yaccob; Hanita Hanim Ismail	
Digital Literacy Knowledge and Reported Application of Digital Practices Among Politeknik Nilai Students: A Pilot Study	165
Arnie Chiah Liling; Fazlind binti Mohd Shuraimi; Nor Farina binti Jamaluddin	
Integrating Artificial Intelligence into CEFR B2 Speaking Interviews in Higher Education: An Hourglass Framework of Human-AI Collaborative Evaluation	187
Nurul Imtiaz Abd Gani; Muhammad Fadhli Mohd Diah; Mohd Zulhusni Ab Raoff	
The World on a Plate (And the Missing Nasi Lemak): A Preliminary Study of a Task-Based Ebook in a Malaysian Vernacular School	207
Woi Min Xuan; Hanushka Anandini A/P Kumaresan; Harwati Hashim; Nur Syafiqah Yaacob; Melor Md Yunus	
Every Voice Matters: Preparing Learners for AI-Supported Speaking Through the Microphone Effect	219
Wan Afiqah Anis binti Wan Ahmad; Mohamad Azrien bin Mohamed Adnan	

AI-Assisted Infographic Poster Preparation in Technical English

2: Students' Perceptions of NotebookLM

Suzianah Sahar, Politeknik Port Dickson, Negeri Sembilan, Malaysia, suzianah@polipd.edu.my

Noor Azua Mohd Bahatera @ Mohd Putera, Politeknik Port Dickson, Negeri Sembilan, Malaysia, noorazua@polipd.edu.my

Diana Ahmad Busra, Politeknik Port Dickson, Negeri Sembilan, Malaysia, diana@polipd.edu.my

Abstract: The integration of artificial intelligence (AI) technologies into higher education offers new opportunities to support information organisation and academic communication. This study investigated engineering students' perceptions of using Google NotebookLM during infographic poster preparation and presentation assessment activities in Technical English 2 at a Malaysian polytechnic. A quantitative descriptive survey design was employed involving 36 Semester 3 engineering students drawn from two existing classes. Data were collected using a post-activity course-evaluation questionnaire comprising closed-ended items and an open-ended question. Descriptive statistics were used to analyse the closed-ended responses, while the open-ended responses were analysed using inductive qualitative content analysis. The study was conceptually informed by the Technology Acceptance Model (TAM) and Cognitive Load Theory (CLT), which were used as interpretive lenses rather than directly measured constructs. Findings indicated generally positive student perceptions of NotebookLM. All 36 respondents (100%) agreed or strongly agreed that NotebookLM was effective for preparing infographic content involving processes, procedures, and instructions, while 35 (97.2%) agreed or strongly agreed that it helped organise key points. All respondents also agreed or strongly agreed that presentations were clearer with NotebookLM. Open-ended responses highlighted perceived benefits related to information organisation, summarisation, ease of use, and assignment preparation. In addition, 35 respondents (97.2%) supported the potential continued use of NotebookLM for presentations and other appropriate assessment contexts. These findings suggest that NotebookLM may have practical potential as a source-grounded support tool for infographic-based Technical English activities. However, the findings are perception-based and context-specific and should not be interpreted as evidence of improved learning outcomes or performance.

Keywords: AI-assisted learning, artificial intelligence in education, Google NotebookLM, , infographic posters, Technical English

Introduction

Artificial intelligence (AI) technologies have increasingly influenced teaching, learning, and assessment practices in higher education. The emergence of generative AI tools such as ChatGPT, Gemini, Claude, and Google NotebookLM has transformed how students access, organise, and synthesise academic information. In educational contexts, AI-assisted learning environments are increasingly recognised for their potential to support information management, academic communication, and learning activities.

Within language and technical communication courses, students are required to process information and communicate technical content systematically. Technical English 2 (TE2) is a compulsory English course offered to engineering students at Politeknik Port Dickson. The course aims to strengthen students' technical communication skills and prepare them for communication in engineering-related workplace contexts. One of the major areas covered in the course involves processes, procedures, and instructions, in which students are required to communicate technical sequences using verbal and visual forms.

As part of the course assessment, students prepare infographic poster presentations related to engineering procedures and technical processes. Such tasks require students to identify relevant technical information, synthesise content from source materials, determine key procedural steps, and present information concisely and coherently. The transformation of detailed technical information into an infographic format may therefore place considerable information-processing demands on students, particularly when information must be selected, summarised, sequenced, and visually organised.

Google NotebookLM provides features that may support these processes. As a source-grounded AI-assisted tool, NotebookLM allows users to upload selected materials and generate summaries, organise notes, retrieve information, and develop structured explanations based on those sources. These features make the platform potentially relevant to infographic preparation activities in which students need to synthesise and organise technical information before presenting it visually.

Although research on generative AI in higher education has expanded considerably, empirical research focusing specifically on Google NotebookLM remains comparatively limited. In particular, there is limited evidence concerning students' perceptions of using NotebookLM to support infographic-based technical communication tasks involving engineering-related processes, procedures, and instructions. This represents a context-specific gap within the emerging literature on AI-assisted language learning and technical communication.

Accordingly, this study investigates engineering students' perceptions of using Google NotebookLM during infographic poster preparation and presentation assessment activities in Technical English 2 at a Malaysian polytechnic. Rather than determining the effectiveness of NotebookLM in improving learning outcomes, the study focuses on students' reported perceptions and experiences of its use within an authentic classroom assessment context.

Problem Statement

Infographic poster preparation in Technical English requires students to transform detailed technical information into concise, logically sequenced, and visually organised content. In process-, procedure-, and instruction-based tasks, students must identify relevant information, synthesise source materials, determine appropriate procedural sequences, and communicate technical content clearly. These requirements can create substantial information-processing demands, particularly when students work with multiple technical sources.

AI-assisted tools such as Google NotebookLM offer potential support for these activities through source-grounded summarisation, information retrieval, note organisation, and synthesis. Such features may assist students in organising technical information before converting it into infographic content. However, the educational value of these features should not be assumed solely from the technological capabilities of the platform.

Although the use of generative AI in higher education has received increasing research attention, empirical evidence concerning Google NotebookLM in technical and language-learning contexts remains limited. More specifically, little is known about how engineering students perceive the use of NotebookLM when preparing infographic presentations involving technical processes, procedures, and instructions. This gap is particularly relevant to Technical English education, where students are required to integrate language, technical information, logical sequencing, and visual communication within a single assessment task.

Furthermore, students' perceptions are important when introducing emerging AI technologies into classroom assessment because perceived usefulness, ease of use, learning support, and experienced challenges may influence how such tools are received and used. At the same time, perception-based evidence should be distinguished from objective evidence of learning effectiveness.

Therefore, this study investigates students' perceptions and experiences of using Google NotebookLM during infographic poster preparation and presentation assessment activities in Technical English 2 at Politeknik Port Dickson. The study provides preliminary classroom-based evidence concerning the perceived benefits and challenges of integrating a source-grounded AI tool into technical communication activities.

In Technical English courses frequently experience difficulties organising technical information for infographic poster presentations. Many students rely on multiple online sources when preparing procedural and instructional content, which may result in information overload and difficulty identifying logical sequencing. Students also encounter challenges converting lengthy textual explanations into concise and visually effective infographic content.

Research Questions

- 1 What are students' perceptions of using Google NotebookLM during infographic poster preparation activities?
- 2 How do students perceive the usefulness of NotebookLM in organizing technical information?
- 3 What benefits do students perceive when using NotebookLM during infographic preparation?

- 4 How do students perceive the role of NotebookLM in supporting presentation preparation activities?

Research Aim

This study aims to investigate students' perceptions of using Google NotebookLM during infographic poster preparation and presentation assessment activities in Technical English 2.

Research Objectives

- 1 To examine students' perceptions of using Google NotebookLM.
- 2 To explore the role of NotebookLM in organising technical information for infographic preparation.
- 3 To identify the perceived benefits and challenges of NotebookLM-assisted infographic presentation activities.

Literature Review

Artificial Intelligence in Higher Education

Artificial intelligence technologies have increasingly transformed teaching and learning practices in higher education. AI-assisted educational tools are integrated into digital learning environments to support academic communication, information organisation, content generation, and student engagement. Recent developments in generative AI technologies have accelerated discussions regarding the role of AI in reshaping educational practices.

Kasneci et al. (2023) argued that generative AI tools may support learning efficiency, information processing, and academic productivity when implemented responsibly. Similarly, Tlili et al. (2023) emphasised that AI-assisted learning environments may create adaptive and student-centred educational experiences through intelligent content support and personalised learning assistance.

Research also indicates that AI technologies may support collaborative learning, academic communication, and student engagement in higher education settings (Crompton & Burke, 2023). (Lo, 2023) further explained that generative AI tools may assist students in brainstorming, summarisation, and academic communication tasks, particularly in language-related learning environments.

Nevertheless, researchers continue to emphasise the importance of ethical AI integration and digital literacy development. Dwivedi et al. (2023) highlighted concerns related to misinformation, hallucinations, overreliance, and academic integrity in generative AI usage. These concerns are particularly relevant in technical and engineering communication contexts where accuracy and procedural clarity are essential.

Cognitive Load Theory and AI-Assisted Learning

Cognitive Load Theory (CLT), introduced by Sweller (1988), explains learning in relation to the limited capacity of working memory during information processing. From a CLT perspective, instructional activities that require learners to process, organise, and integrate multiple sources of information may impose substantial cognitive demands.

In infographic-based technical communication activities, students must understand technical content, identify essential information, determine procedural relationships, summarise detailed explanations, and organise information into a concise visual format. Managing these requirements simultaneously provides a relevant context for considering information-processing demands during infographic preparation.

AI-assisted tools such as Google NotebookLM offer functions including source-grounded summarisation, information retrieval, note organisation, and synthesis. Conceptually, these functions may help students manage information during complex preparation tasks by reducing the need to search repeatedly across multiple source materials and by supporting the organisation of relevant information before visual presentation.

In this study, however, cognitive load was not directly measured using a validated cognitive-load instrument or experimental design. CLT was therefore used as an interpretive lens rather than as an empirically tested construct. Students' reported perceptions that NotebookLM assisted with summarisation or information organisation may be discussed in relation to information-processing demands, but they should not be interpreted as direct evidence that NotebookLM reduced cognitive load. Objective measurement would be required to establish such an effect.

Generative AI Tools and Academic Communication

Generative artificial intelligence (AI) tools have become increasingly integrated into academic communication and language learning environments. Students use AI applications to support brainstorming, summarisation, note organisation, translation, writing, and presentation preparation. These tools have the potential to enhance learning efficiency by assisting students in processing large amounts of information and communicating ideas more effectively (Kohnke et al., 2023; Lo, 2023).

Among recent AI-assisted learning tools, Google NotebookLM has attracted growing educational interest because it employs a Retrieval-Augmented Generation (RAG) approach. Unlike conventional generative AI systems that rely primarily on pretrained knowledge, NotebookLM generates responses based on user-uploaded documents. By grounding its outputs in selected source materials, NotebookLM can produce context-specific summaries, study guides, and explanations while reducing the likelihood of hallucinated or unsupported information. Recent research has identified RAG as a promising approach for educational applications because it improves factual accuracy, enhances transparency, and enables learners to interact more effectively with course-specific materials.

Yeo, Moorhouse, and Wan (2025) reported that NotebookLM supports learner engagement by transforming uploaded academic resources into interactive learning materials, including structured summaries, frequently asked questions, timelines, and AI-generated discussions. These features encourage students to organise information systematically and engage more actively with learning content, particularly during independent study and academic preparation.

Similarly, Tufino (2025) demonstrated that NotebookLM can function as an effective RAG-based educational assistant by grounding responses in instructor-provided materials. The study suggested that NotebookLM supports personalised learning, facilitates information retrieval, and reduces the occurrence of hallucinated responses commonly associated with general-purpose generative AI systems. Such capabilities make NotebookLM particularly relevant for learning activities that require students to synthesise information from multiple technical sources before preparing presentations or other academic tasks.

Despite these advantages, researchers continue to emphasise that AI-generated outputs should not be accepted uncritically. Students remain responsible for verifying information against reliable academic and technical sources, particularly in engineering and technical communication contexts where procedural accuracy is essential (Dwivedi et al., 2023). Therefore, the effective integration of NotebookLM into higher education should be accompanied by AI literacy development and critical evaluation skills to ensure responsible and ethical use.

Infographic Posters in Educational Assessment

Infographic posters are increasingly used as alternative assessment tools in higher education because they combine visual communication, concise textual explanation, and structured information presentation. Infographics encourage students to synthesise information, identify key concepts, and communicate content systematically.

(Smiciklas, 2012) described infographics as visual communication tools that simplify complex information and improve audience engagement. In educational settings, infographic-based learning activities also support visual literacy and information-processing skills.

Recent studies indicate that infographic-supported learning activities may enhance students' critical thinking and communication skills by encouraging concise and systematic information presentation (Ozdamli & Ozdal, 2021). Nevertheless, students may still experience difficulties organising complex information coherently, particularly in technical communication tasks involving process- and procedure-based content.

The integration of AI-assisted learning tools may therefore support infographic preparation by helping students organise information systematically before the visual design process begins.

Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM), introduced by Davis (1989), is widely used to explain individuals' acceptance and use of technology. The model proposes two central constructs: perceived usefulness, the extent to which individuals believe using a particular technology will enhance their performance, and perceived ease of use, the extent to which using the technology is perceived as requiring relatively little effort.

In educational contexts, TAM has frequently been used to examine students' perceptions and acceptance of digital learning technologies. Abdullah and Ward (2016), for example, demonstrated the continuing relevance of perceived usefulness and perceived ease of use in explaining learners' acceptance of e-learning technologies. The framework therefore provides a useful theoretical perspective for considering how students perceive emerging AI-assisted tools within learning environments.

In the present study, however, TAM was not operationalised as a psychometrically measured model, and its constructs were not tested statistically. Instead, TAM served as a conceptual and interpretive lens for understanding students' reported perceptions of NotebookLM, particularly responses concerning its usefulness and ease of use during infographic poster preparation. Consequently, the findings should not be interpreted as validating TAM constructs or demonstrating technology acceptance in the predictive sense proposed by the original model. Rather, TAM provides a theoretical perspective for interpreting students' perception-based responses to using NotebookLM within the Technical English 2 learning context.

Conceptual Framework

This study was conceptually informed by the Technology Acceptance Model (TAM) proposed by Fred Davis (1989) and Cognitive Load Theory (CLT) introduced by John Sweller (1988).

Figure 1. Conceptual Framework of the Study

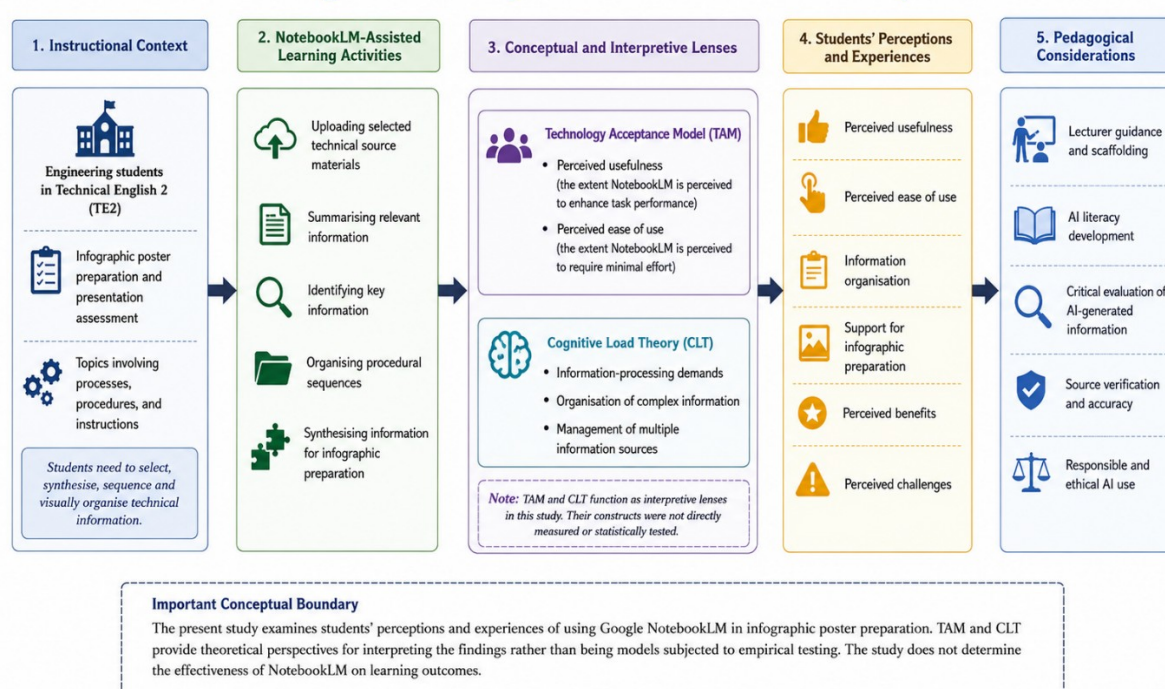


Figure 1 presents the conceptual framework guiding the study. The framework situates the use of Google NotebookLM within the instructional context of infographic poster preparation in Technical English 2 and draws conceptually on the Technology Acceptance Model (TAM) (Davis, 1989) and Cognitive Load Theory (CLT) (Sweller, 1988).

Within the instructional context, students use NotebookLM to support activities such as accessing uploaded source materials, summarising relevant information, identifying key points, organising procedural sequences, and synthesising information before developing their infographic posters. These activities provide the context within which students' perceptions and experiences of the AI-assisted learning activity are examined.

TAM provides an interpretive perspective for considering students' reported perceptions of the usefulness and ease of use of NotebookLM. CLT provides a complementary perspective for considering students' experiences of managing and organising technical information during infographic preparation. However, neither TAM constructs nor cognitive load were directly measured using validated scales in this study. The two theories therefore function as conceptual lenses that support interpretation rather than as models subjected to empirical testing.

The framework consequently focuses on students' perceptions and experiences, including perceived usefulness, ease of use, information organisation, support for infographic preparation, benefits, and challenges. These perceptions are considered within broader pedagogical considerations such as lecturer guidance, AI literacy, source verification, critical evaluation, and responsible AI use.

Methodology

Research Design

This study employed a descriptive survey approach to investigate students' perceptions and experiences of using Google NotebookLM during infographic poster preparation and presentation assessment activities in Technical English 2. The study used closed-ended questionnaire items to collect descriptive quantitative data and supplementary open-ended items to provide further insight into students' experiences using NotebookLM.

A descriptive survey approach was appropriate because the study aimed to describe students' perceptions in an authentic classroom context rather than establish causal relationships or determine NotebookLM's effectiveness in improving learning outcomes (Cohen et al., 2017; Creswell & Creswell, 2018). Quantitative responses were therefore interpreted using descriptive statistics, while open-ended responses provided supplementary contextual explanations of students' perceived benefits and challenges.

Although the study was conceptually informed by the Technology Acceptance Model (TAM) and Cognitive Load Theory (CLT), the questionnaire was not designed as a validated measurement instrument for either framework. TAM and CLT were used as conceptual and interpretive lenses for discussing the findings rather than as theoretical models subjected to statistical testing.

Participants

The participants comprised 36 Semester 3 engineering students enrolled in Technical English 2 at Politeknik Port Dickson, Malaysia, during Academic Session II, 2025/2026. The students were drawn from two existing Technical English 2 classes participating in the same NotebookLM-assisted infographic poster activity. The total enrolment across the two classes was 58 students.

The study employed convenience sampling, as participants were drawn from the two existing classes in which the NotebookLM-assisted learning activity was implemented. Of the 58 students enrolled in the two classes, 36 students completed the course-evaluation questionnaire and constituted the final sample for this study. All respondents had completed the infographic poster preparation and presentation activity involving technical processes, procedures, and instructions before responding to the questionnaire.

The participants were Semester 3 students enrolled in a mechanical engineering diploma programme. Age and gender were not collected as part of the course-evaluation questionnaire. As the study focused on students' perceptions within a specific instructional setting, the participant group represents a bounded classroom context rather than a statistically representative sample of Malaysian polytechnic students.

Instrumentation

Data were collected using a course-evaluation questionnaire administered following the NotebookLM-assisted infographic poster activity. The questionnaire comprised closed-ended items and an open-ended question designed to capture students' evaluations of the assessment activity and their perceptions of using Google NotebookLM.

The closed-ended component consisted of three sections. The first section, Task Instructions and Rubrics, contained two items assessing the clarity of task instructions and the clarity of the assessment rubric. The second section, Contribution to Learning, contained four items addressing students' perceived level of skill or knowledge at the beginning and end of the assessment, the level of skill or knowledge required to complete the assessment, and the perceived contribution of the course to their skill or knowledge. Items in these two sections were rated using a five-point categorical scale ranging from Poor, Fair, Satisfactory, Very Good, to Excellent.

The third section, Use of Google NotebookLM, contained three items addressing students' perceptions of NotebookLM as an effective tool for preparing infographic content involving processes, procedures, and instructions; its usefulness in helping students organise key points; and whether presentations were perceived as clearer with the use of NotebookLM. These items were rated using a four-point agreement scale ranging from Strongly Disagree, Disagree, Agree, to Strongly Agree.

The questionnaire also included one open-ended question: "What aspects of NotebookLM were most useful or valuable?" This item allowed students to describe the features or aspects of NotebookLM that they perceived as beneficial during the activity. Two additional closed-ended questions asked whether students agreed that the use of NotebookLM could enhance presentations in DUE30072 and whether they would recommend applying NotebookLM in other assessments when appropriate.

The questionnaire was originally developed as a course-evaluation instrument rather than as a psychometrically validated scale for measuring TAM or CLT constructs. Prior to administration, colleagues teaching Technical

English reviewed the questionnaire to assess the clarity, relevance, and alignment of the items with the learning activity. Accordingly, the questionnaire findings are interpreted as descriptive, perception-based evidence within this specific instructional context rather than as validated measurements of technology acceptance, cognitive load, or learning effectiveness.

Data Collection Procedure

Students were introduced to Google NotebookLM through lecturer-guided demonstrations before undertaking the infographic poster activity. During the activity, students uploaded selected materials related to engineering processes, procedures, or instructions to NotebookLM and used the platform to support summarisation, identify key points, and organise procedural sequences.

Students subsequently used the organised information to prepare infographic posters for their Technical English 2 presentation assessment. They were reminded to verify AI-generated information against the original technical source materials rather than relying uncritically on AI-generated outputs.

After completing the infographic poster preparation and presentation activities, the 36 students responded to the course-evaluation questionnaire. The questionnaire collected closed-ended responses concerning different aspects of the activity together with supplementary open-ended feedback concerning students' experiences of using NotebookLM.

Participation in the research use of the course-evaluation responses was voluntary, and informed consent was obtained from the participants. Students were informed of the purpose of the study and that their responses would be used for research purposes. No personally identifiable information is reported in the manuscript, and the findings are presented in aggregated or anonymised form. Formal ethical approval was not obtained for this study.

Data Analysis

Closed-ended questionnaire responses were analysed using descriptive statistics. Frequencies and percentages were calculated for each response category to describe patterns in students' evaluations and perceptions of the NotebookLM-assisted infographic activity. Given the descriptive purpose of the study and the relatively small, context-specific participant group, the researchers did not conduct inferential statistical analyses.

Open-ended responses were analysed using inductive qualitative content analysis, in which categories were derived from the students' responses rather than established a priori (Elo & Kyngäs, 2008; Hsieh & Shannon, 2005). The researchers first reviewed all responses to obtain an overall understanding of the data. Recurring ideas were then coded and responses expressing similar meanings were grouped into descriptive categories, such as information organisation, summarisation support, ease of use, assignment preparation, and other benefits or challenges emerging from the data. The categories were subsequently reviewed and refined to ensure that they adequately represented the students' responses.

The open-ended findings were used to supplement and contextualise the quantitative results rather than to constitute an independent qualitative strand. Representative anonymised responses are presented alongside the descriptive categories where appropriate. Frequencies are reported only when they can be verified from the original responses.

Findings

Students' Perceptions of NotebookLM Use

Students reported predominantly positive perceptions of using Google NotebookLM during infographic poster preparation. As shown in Table 1, all 36 respondents (100%) either agreed or strongly agreed that NotebookLM was an effective tool for preparing infographic content involving processes, procedures, and instructions. Specifically, 19 students (52.8%) selected Strongly Agree and 17 (47.2%) selected Agree.

Table 1. Students' Perceptions of Google NotebookLM in Infographic Poster Preparation (n = 36)

Item	Strongly Disagree n (%)	Disagree n (%)	Agree n (%)	Strongly Agree n (%)
NotebookLM was an effective tool for preparing infographic content involving processes, procedures and instructions	0 (0.0)	0 (0.0)	17 (47.2)	19 (52.8)

NotebookLM helped in organising the points	0 (0.0)	1 (2.8)	16 (44.4)	19 (52.8)
Presentation was clearer with NotebookLM	0 (0.0)	0 (0.0)	14 (38.9)	22 (61.1)

Students similarly reported positive perceptions of NotebookLM for information organisation. Nineteen respondents (52.8%) strongly agreed and 16 (44.4%) agreed that NotebookLM helped them organise key points, representing 35 of the 36 respondents (97.2%). One respondent (2.8%) disagreed, while none strongly disagreed.

Perceptions of presentation clarity were also positive. Twenty-two students (61.1%) strongly agreed and 14 (38.9%) agreed that their presentations were clearer with the use of NotebookLM. Thus, all 36 respondents expressed agreement with this item.

Taken together, these findings indicate a high level of perceived usefulness of NotebookLM within the infographic poster activity. From a TAM perspective, these responses may be interpreted as reflecting favourable perceptions of the tool's usefulness in the specific learning context. However, because the questionnaire was not a validated TAM instrument and did not objectively assess presentation quality, the findings should be interpreted as students' self-reported perceptions rather than evidence of improved learning or presentation performance.

Students' Evaluation of Task Instructions and Rubrics

Table 2. Students' Evaluation of Task Instructions and Rubrics (n = 36)

Item	Poor	Fair	Satisfactory	Very Good	Excellent
Clarity of task instructions	0 (0.0%)	3 (8.3%)	13 (36.1%)	7 (19.4%)	13 (36.1%)
Clarity of rubrics	0 (0.0%)	3 (8.3%)	13 (36.1%)	6 (16.7%)	14 (38.9%)

Students generally evaluated the task instructions and rubrics favourably. For clarity of task instructions, 13 students (36.1%) selected Excellent, seven (19.4%) Very Good, 13 (36.1%) Satisfactory, and three (8.3%) Fair. Similarly, for clarity of rubrics, 14 students (38.9%) selected Excellent, six (16.7%) Very Good, 13 (36.1%) Satisfactory, and three (8.3%) Fair. No students rated either item as Poor. Overall, 20 students (55.6%) rated the clarity of task instructions as Very Good or Excellent, while the same proportion, 20 students (55.6%), rated the clarity of rubrics as Very Good or Excellent.

Students' Perceived Contribution to Learning

Table 3: Students' Perceptions of Skills/Knowledge and Contribution to Learning (n=36)

Item	Poor	Fair	Satisfactory	Very Good	Excellent
Skill/knowledge at start of assessment	0 (0.0%)	3 (8.3%)	16 (44.4%)	6 (16.7%)	11 (30.6%)
Skill/knowledge at end of assessment	0 (0.0%)	3 (8.3%)	14 (38.9%)	7 (19.4%)	12 (33.3%)
Skill/knowledge required to complete assessment	0 (0.0%)	2 (5.6%)	14 (38.9%)	8 (22.2%)	12 (33.3%)
Contribution of course to skill/knowledge	0 (0.0%)	3 (8.3%)	12 (33.3%)	8 (22.2%)	13 (36.1%)

Students also evaluated their perceived skill and knowledge in relation to the assessment. At the start of the assessment, 11 students (30.6%) rated their skill/knowledge as Excellent and six (16.7%) as Very Good. At the end of the assessment, 12 students (33.3%) selected Excellent and seven (19.4%) Very Good. Regarding the level of skill/knowledge required to complete the assessment, 12 students (33.3%) selected Excellent and eight (22.2%) Very Good. Thirteen students (36.1%) rated the contribution of the course to their skill/knowledge as Excellent and eight (22.2%) as Very Good.

These ratings represent students' retrospective perceptions collected through the post-activity course-evaluation questionnaire and should not be interpreted as objective evidence of learning gains or as pre-test/post-test measurements.

Perceived Benefits of NotebookLM During Infographic Preparation

Open-ended responses further supported the quantitative findings by revealing several recurring themes related to students' experiences using NotebookLM during infographic preparation activities.

Table 4. Categories Identified from Students' Open-Ended Responses

Theme	Sample Responses
Information Organization	<i>"NotebookLM helped organise information clearly."</i>
Summarization Support	<i>"The summarisation feature is useful."</i>
Ease of Use	<i>"Easy and useful for students."</i>
Assignment Preparation	<i>"It helps my assignments become easier to do."</i>

As shown in Table 4, students highlighted NotebookLM's role in supporting information organisation, summarisation, ease of use, and assignment preparation. Many students perceived the summarisation and note-organisation features as particularly useful during infographic preparation tasks.

The open-ended responses indicate that students valued NotebookLM primarily as a practical support tool during infographic preparation. Students referred to its ability to organise information, summarise source material, facilitate access to relevant content, and make infographic or assignment preparation easier. These responses provide supplementary qualitative support for the positive perceptions reported in the closed-ended questionnaire items.

Perceived Value and Future Use of NotebookLM

Students also expressed positive perceptions regarding the potential continued use of NotebookLM. Thirty-five of the 36 respondents (97.2%) agreed that the use of NotebookLM could enhance presentations in DUE30072, while one respondent (2.8%) disagreed. Similarly, 35 students (97.2%) indicated that they would recommend the use of NotebookLM in other assessments when appropriate, while one respondent (2.8%) did not.

These findings indicate strong student support for the potential continued use of NotebookLM within related assessment contexts. However, the responses represent students' perceptions and preferences rather than objective evidence of the tool's effectiveness.

Discussion

The findings of this study indicate generally positive student perceptions of Google NotebookLM as a support tool for infographic poster preparation in Technical English 2. Students reported favourable perceptions of the tool's usefulness for preparing content involving processes, procedures, and instructions, organising key points, and supporting presentation clarity. The open-ended responses further highlighted information organisation, summarisation, ease of use, and infographic or assignment preparation as perceived benefits. In addition, most students expressed support for the potential continued use of NotebookLM in DUE30072 and other appropriate assessment contexts.

These findings should be interpreted within the descriptive and perception-based nature of the study. The questionnaire measured students' reported experiences and evaluations rather than objective improvements in infographic quality, technical accuracy, presentation performance, or learning outcomes. Accordingly, the discussion considers the findings in relation to the Technology Acceptance Model (TAM), Cognitive Load Theory (CLT), and previous research on AI-assisted learning without treating TAM constructs or cognitive load as directly measured variables.

Perceived Usefulness of NotebookLM for Infographic Preparation

The findings suggest that students perceived NotebookLM as useful within the specific context of infographic poster preparation. All 36 respondents agreed or strongly agreed that NotebookLM was an effective tool for preparing infographic content involving processes, procedures, and instructions. Similarly, all respondents agreed or strongly agreed that their presentations were clearer with NotebookLM, while 35 of the 36 respondents (97.2%) agreed or strongly agreed that the tool helped them organise key points.

From a TAM perspective, these findings can be interpreted as indicating favourable perceptions of the usefulness of NotebookLM in supporting the learning activity. Perceived usefulness is a central construct in TAM and refers to the extent to which users believe that a technology can support or enhance their performance (Davis, 1989).

Previous educational technology research has similarly demonstrated the importance of perceived usefulness in students' acceptance of digital learning technologies (Abdullah & Ward, 2016; Al-Emran et al., 2022).

However, the findings should not be interpreted as a formal measurement of TAM constructs. The questionnaire was developed as a course-evaluation instrument rather than a validated TAM scale. Therefore, the results provide evidence of students' favourable perceptions of usefulness within this particular classroom activity, rather than statistical evidence of technology acceptance as conceptualised in the full TAM model.

Information Organisation and Information-Processing Support

Information organisation emerged as an important aspect of students' experiences with NotebookLM. Quantitatively, 35 students (97.2%) agreed or strongly agreed that NotebookLM helped them organise key points. The open-ended responses complemented this result, with students referring to information organisation, summarisation, ease of use, and support for infographic or assignment preparation.

These findings are relevant to the information-processing demands involved in infographic preparation. Students undertaking process-, procedure-, and instruction-based tasks must identify relevant technical information, distinguish essential from non-essential content, organise procedural sequences, and transform detailed information into concise visual representations. NotebookLM's source-grounded summarisation and information-organisation functions may therefore provide practical support during this preparatory process.

From a CLT perspective, such functions can be considered in relation to the management of information-processing demands. CLT emphasises the limited capacity of working memory when learners engage with complex information (Sweller, 1988). Nevertheless, the present study did not directly measure cognitive load. It would therefore be inappropriate to conclude that NotebookLM reduced students' cognitive load. Rather, students' reported experiences suggest that the tool was perceived as supporting information organisation and summarisation—processes that are conceptually relevant to the management of complex information.

Perceived Value and Continued Use

Students also demonstrated strong support for the potential continued use of NotebookLM. Thirty-five of the 36 respondents (97.2%) indicated that NotebookLM could enhance presentations in DUE30072, and the same proportion indicated that they would recommend its use in other assessments when appropriate.

These findings suggest that students perceived value in extending the use of NotebookLM beyond the immediate infographic activity. This is consistent with the broader TAM proposition that favourable perceptions of a technology's usefulness can contribute to positive attitudes towards its continued use (Davis, 1989). However, willingness to recommend or continue using NotebookLM should not be equated with demonstrated educational effectiveness. The present findings represent students' stated preferences and perceptions within one Technical English course.

The results nevertheless provide useful contextual evidence for Technical English teaching. Rather than positioning AI as a replacement for students' own thinking or technical communication skills, NotebookLM may be more appropriately conceptualised as a support tool during the information-preparation stage, particularly for organising source materials, identifying key information, and preparing content before students construct their own infographic presentations.

Critical Considerations for AI-Assisted Learning

Although students generally reported positive perceptions, the integration of AI-assisted tools into Technical English activities requires appropriate caution. Generative AI systems may produce inaccurate, incomplete, or oversimplified information, particularly when users rely on generated outputs without critically checking the original sources. Concerns regarding hallucination, misinformation, and overreliance have been widely discussed in the literature on generative AI in education (Dwivedi et al., 2023).

These concerns are particularly relevant in technical and engineering communication, where inaccurate procedural information may affect conceptual understanding and the clarity of technical communication. Students should therefore be encouraged to verify AI-assisted outputs against reliable academic or technical sources and to exercise independent judgement when selecting information for infographic presentations.

NotebookLM's source-grounded design may provide an advantage in this respect because students work with uploaded source materials rather than relying exclusively on unrestricted AI-generated information. Nevertheless, source grounding does not eliminate the need for critical evaluation. Lecturer guidance, source verification, AI literacy, and responsible use should therefore accompany the integration of NotebookLM into Technical English activities.

Importantly, these considerations arise from the wider literature and the pedagogical context rather than from participant-reported challenges in the present study. The questionnaire asked students about aspects of NotebookLM they found useful or valuable and did not specifically elicit challenges. Future studies should therefore investigate students' difficulties, concerns, and critical experiences with NotebookLM more directly.

Pedagogical Implications

The findings have several implications for the integration of AI-assisted tools into Technical English teaching. First, NotebookLM may be incorporated as a structured support tool during the preparation stage of infographic-based activities. Students may use the platform to engage with lecturer-selected or verified source materials, identify key information, organise procedural sequences, and summarise technical content before developing their own visual presentations.

Second, AI-assisted activities should be accompanied by explicit AI literacy instruction. Students need guidance not only on how to use NotebookLM but also on how to evaluate AI-assisted outputs, verify information against original sources, identify possible inaccuracies, and make independent decisions about which information is appropriate for technical communication.

Third, lecturers should ensure that AI assistance does not replace the language and communication skills targeted by Technical English assessment. Assessment rubrics may therefore continue to emphasise students' own performance in areas such as logical sequencing, technical clarity, coherence, visual organisation, and oral communication rather than rewarding the use of AI itself.

Finally, the positive student perceptions observed in this study suggest that source-grounded AI tools may have practical value within technical communication activities when their use is carefully scaffolded. However, implementation should remain pedagogically driven, with AI functioning as a support for information preparation and organisation rather than as a substitute for students' critical thinking, technical understanding, or communication skills.

Recommendations

Pedagogical Recommendations

Based on the students' generally positive perceptions, lecturers may consider integrating source-grounded AI tools such as Google NotebookLM into Technical English activities that require students to identify, organise, summarise, and present technical information. In infographic-based activities, NotebookLM may be used as a support tool during the preparation stage, particularly when students work with processes, procedures, and instructions.

However, the use of AI-assisted tools should be accompanied by appropriate lecturer guidance. Students should be encouraged to verify AI-assisted outputs against the original source materials, critically evaluate the accuracy and relevance of generated information, and avoid relying uncritically on AI-generated content. AI literacy and responsible-use principles should therefore be incorporated into the learning activity.

Lecturers may also design assessment rubrics that continue to emphasise students' own technical communication performance, including logical sequencing, technical clarity, coherence, visual organisation, and oral presentation skills. This can help ensure that AI functions as a learning support rather than as a substitute for the skills targeted by the course.

Institutional Recommendations

At the institutional level, professional development may be provided to support lecturers in the pedagogically appropriate use of AI-assisted tools. Such initiatives could include workshops on AI literacy, source verification, responsible AI use, assessment design, and strategies for integrating source-grounded AI into teaching and learning activities.

Institutions may also consider developing clear guidance concerning the acceptable and ethical use of generative AI in coursework and assessment. Such guidance should address issues including transparency of AI use, verification of AI-assisted outputs, academic integrity, data privacy, and the responsibilities of both students and lecturers.

Given the preliminary nature of the present findings, broader implementation of NotebookLM should be accompanied by continued evaluation. Future institutional initiatives could examine its use across different courses, assessment types, and student groups before drawing conclusions regarding its wider educational effectiveness.

Limitations

This study has several limitations that should be considered when interpreting the findings. First, the study involved a relatively small sample of 36 engineering students from two Technical English 2 classes at a single Malaysian polytechnic. The findings are therefore context-specific and should not be generalised to all engineering students, Technical English courses, or Malaysian polytechnics.

Second, the study relied primarily on students' self-reported perceptions collected through a course-evaluation questionnaire. The questionnaire was not developed as a psychometrically validated instrument for measuring Technology Acceptance Model (TAM) constructs or cognitive load. Consequently, the findings should be interpreted as descriptive evidence of students' perceptions rather than validated measurements of technology acceptance or cognitive-load reduction.

Third, the study did not objectively evaluate the quality of students' infographic posters, technical accuracy, presentation performance, or learning outcomes. Although students reported positive perceptions of NotebookLM and presentation clarity, these responses do not demonstrate that the use of NotebookLM objectively improved student performance.

Fourth, students' perceived skill and knowledge at the beginning and end of the assessment were collected retrospectively through the same post-activity course-evaluation questionnaire. These ratings should therefore not be interpreted as pre-test/post-test measurements or evidence of learning gains.

Fifth, the open-ended component was limited to a single question asking students what aspects of NotebookLM they found most useful or valuable. The questionnaire did not explicitly ask students to report challenges or negative experiences. Consequently, the study provides stronger evidence concerning perceived benefits than potential difficulties associated with using NotebookLM.

Finally, the study did not include a comparison or control group undertaking the same infographic activity without NotebookLM. It is therefore not possible to attribute students' perceived experiences or any learning-related outcomes specifically to the use of NotebookLM.

Future studies could address these limitations by employing larger and more diverse samples, validated technology-acceptance and cognitive-load measures, objective assessment rubrics, pre- and post-intervention measures, comparison groups, and more comprehensive qualitative methods such as interviews or focus groups. Future studies should therefore incorporate objective infographic evaluation rubrics, larger participant groups, and comparative analyses involving AI-assisted and non-AI-assisted presentation activities.

Conclusion

This study investigated engineering students' perceptions of using Google NotebookLM during infographic poster preparation and presentation assessment activities in Technical English 2 at a Malaysian polytechnic. The findings revealed generally positive student perceptions of NotebookLM within this specific classroom context. Students particularly valued the tool for organising information, summarising source materials, supporting infographic preparation, and facilitating the organisation of key points. Most participants also expressed support for the potential continued use of NotebookLM in related assessment contexts.

Conceptually, the findings may be interpreted through the Technology Acceptance Model (TAM), particularly in relation to students' perceived usefulness of the tool, and through Cognitive Load Theory (CLT) in relation to the information-processing demands associated with organising and synthesising technical information. However, neither TAM constructs nor cognitive load were directly measured in this study. The findings therefore provide perception-based evidence rather than empirical confirmation of technology acceptance, cognitive-load reduction, or improved learning performance.

The study suggests that NotebookLM may have practical potential as a source-grounded support tool during infographic-based Technical English activities when accompanied by appropriate lecturer guidance, source verification, critical evaluation, and responsible AI-use practices. Nevertheless, the findings should be regarded as preliminary and context-specific because of the small single-institution sample, reliance on self-reported data, and absence of objective learning measures or a comparison group.

Future research should examine NotebookLM using larger and more diverse samples, validated instruments, objective assessment of infographic and presentation quality, and comparative or longitudinal research designs. Such research would provide stronger evidence concerning the role of source-grounded AI tools in Technical English and technical communication education.

References

- Abdullah, F., & Ward, R. (2016). Developing a general extended technology acceptance model for e-learning (GETAMEL) by analysing commonly used external factors. *Computers in Human Behavior*, 56, 238–256. <https://doi.org/10.1016/j.chb.2015.11.036>
- Al-Emran, M., Mezhyuev, V., & Kamaludin, A. (2018). Technology Acceptance Model in M-learning context: A systematic review. *Computers & Education*, 125, 389–412. <https://doi.org/10.1016/j.compedu.2018.06.008>
- Chan, C. K. Y., & Tsi, L. H. Y. (2024). Will generative AI replace teachers in higher education? A study of teacher and student perceptions. *Studies in Educational Evaluation*, 83, 101395. <https://doi.org/10.1016/j.stueduc.2024.101395>
- Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education* (8th ed.). Routledge. <https://doi.org/10.4324/9781315456539>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22. <https://doi.org/10.1186/s41239-023-00392-8>
- 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>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). Opinion paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. <https://doi.org/10.1111/j.1365-2648.2007.04569.x>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>

- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Ozdamli, F., & Ozdal, H. (2021). Developing an instructional design for the design of infographics and evaluation of infographic usage in teaching based on teacher and student opinions. *EURASIA Journal of Mathematics, Science and Technology Education*, 17(3).
- Smiciklas, M. (2012). *The power of infographics: Using pictures to communicate and connect with your audiences*. Que Publishing.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Teo, P. K., Tan, S. M., & Megawati, C. D. (2025). The impact of infographics on learning outcomes in design education. *International Journal of Modern Education*, 7(26), 312–321. <https://doi.org/10.35631/IJMOE.726021>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Tufino, E. (2025). NotebookLM as a Socratic physics tutor: Design and preliminary observations of a retrieval-augmented generation-based learning tool. *The Physics Educator*, 7, 1–9. <https://doi.org/10.1142/S2661339525500143>
- Yeo, M. A., Moorhouse, B. L., & Wan, Y. (2025). From academic text to talk-show: Deepening engagement and understanding with Google NotebookLM. *TESL-EJ*, 28(4), 1–14. <https://doi.org/10.55593/ej.28112int>

Negotiating Digital and Academic Registers: Influence of TikTok Slang on Academic Language Use among University Students

Ahmad Syahmi bin Salehuddin, Faculty of Major Language Studies, Universiti Sains Islam Malaysia, Malaysia, syahmislhdin@gmail.com

Haliza Harun*, Faculty of Major Language Studies, Universiti Sains Islam Malaysia, Malaysia, haliza@usim.edu.my

...

*** Corresponding Author**

Abstract: In technology-enhanced learning environments, social media platforms such as TikTok increasingly shape students' linguistic repertoires, creating tensions between digitally mediated communication practices and formal academic English conventions. This study investigates how TikTok slang influences undergraduate students' academic language use, focusing on how they negotiate language choices across contrasting digital and academic contexts. Combining Halliday's Systemic Functional Linguistics and Gumperz's concept of Contextualization Cues, the study employed an explanatory mixed methods design to examine how TikTok slang shaped students' register awareness, language choices, and academic communication practices. Quantitative results revealed frequent use of digital expressions in students' daily interactions and unintended slips into formal academic writing. In contrast, qualitative analysis further showed that slang served as an interpersonal resource for building peer solidarity and reducing language anxiety among second language learners. Although students demonstrated awareness of appropriate language use by avoiding slang in academic settings, frequent exposure to TikTok slang may unconsciously affect their ability to recall standard academic vocabulary during formal communication. The findings suggest that while students are linguistically competent, they may struggle to shift effectively between informal digital expressions and formal academic communication, highlighting the need for greater critical language awareness in university language instruction.

Keywords: *academic language use; critical language awareness; digital literacy; TikTok slang*

Introduction

Modern communication technologies have accelerated language change, transforming short video platforms such as TikTok to become common platforms for the spread of new slang and expressions (Marzuki, & Mustapha 2023). Among Generation Z undergraduates, digital slang has evolved from a localised subcultural argot, into a widely used part of everyday language across different cultures (Nurhayati & Putri 2024; Tagliamonte, 2016). While this reflects the dynamic nature of language, it also presents challenges in higher education where students are expected to use standard academic English that meets the requirements of academic literacy and professional communication (Li, 2025). This issue is particularly relevant for undergraduate second language learners (L2) who are still developing both their academic literacy and general English proficiency. Empirical evidence from learners' corpus research further demonstrate that L2 writers often struggle to maintain clear boundaries between informal and academic registers, with speech-like and non-academic features appearing in

formal writing due to limited register awareness (Bybee, 2010; Gilquin & Paquot, 2008). This phenomenon is conceptualised in the present study as register leakage, referring to unintentional transfer of informal digital language into formal academic contexts (Gries & Ellis, 2015). While traditional approaches to language teaching often view slang as a form of non-standard language, recent sociolinguistic research recognises it as a meaningful linguistic resource for expressing identity and building social relationships (Ugoala, 2024; Zappavigna, 2012). There is, however, still limited research on how frequent exposure to and use of digital slang influences the way English as a second language university students distinguish between formal and informal language. This gap is particularly evident in the Malaysian higher education context, where these language learners regularly move between digital communication and academic writing. Hence, to address this gap, the study is guided by the following research objectives:

RO1: To identify the most prevalent TikTok slang terms adopted by undergraduates at the Faculty of Major Language Studies, Universiti Sains Islam Malaysia.

RO2: To examine how TikTok usage influences students' language selection across contrasting social and academic contexts.

RO3: To evaluate students' qualitative perceptions of the socio-psychological and cognitive impacts of TikTok slang on their overall English proficiency

Literature Review

Numerous studies have been conducted to examine how social media influences language use across different contexts (Mujeeb et al., 2026; Shofiyah et al. 2026; Ugoala 2024). Many of these studies have highlighted that platforms such as TikTok are not only spaces for entertainment but has also become important medium where new slangs and expressions are created, shared and normalised among young users. Building on this, Ugoala (2024) also emphasises that slang plays a significant role in shaping youth linguistic identity, functioning as a marker of creativity, in-group belonging, and social alignment in informal communication. This perspective is further supported by Tagliamonte (2016), who argues that adolescent language is a key driver of linguistic innovation, with informal and peer-based varieties contributing to broader patterns of language change and the increasing normalisation of non-standard forms in everyday communication. Similarly, Shofiyah et al. (2026). argue that TikTok has become a globalised linguistic space in which slang is rapidly created, circulated, and normalised across different cultural and communicative contexts.

While the influence of TikTok slang is evident in the emergence and normalisation of new linguistic forms, exposure to informal digital language may also extend into students' academic language practices. Harith et al. (2025) for instance, highlighted the influence of social media exposure and informal digital language on students' academic language use. Drawing on a review of existing literature, they found that frequent exposure to slang, abbreviations, and informal language structures contributed to spelling errors, limited vocabulary, repetitive sentence structures, and weaker grammatical accuracy among Malaysian university students. Comparatively, Mohd Zuraini et al. (2024) examined Malaysian ESL learners' perceptions of TikTok as a supplementary tool for English language learning. Using a mixed-methods approach, the findings indicated that students perceived TikTok positively in terms of its potential to enhance vocabulary, pronunciation, communication skills, and exposure to authentic English. However, it is noteworthy to highlight that the study focused primarily on learners' perceptions and the perceived benefits and limitations of TikTok rather than on the linguistic forms to which learners are exposed through the platform. Consequently, it does not address whether exposure to informal linguistic features commonly encountered on TikTok,

particularly slang, may influence students' language choices or lead to the transfer of such forms into formal academic contexts. More broadly, although existing research has increasingly examined the characteristics and functions of digital slang, limited attention has been given to how exposure to and use of such forms may affect ESL undergraduates' ability to distinguish between informal online communication and formal academic English. This distinction is particularly important in higher education, where students are expected to adapt their language appropriately to different communicative purposes, audiences, and academic contexts (Li, 2025)

Conceptual Framework

To address this gap, the present study adopts a conceptual framework informed by Michael Halliday's Systemic Functional Linguistics (SFL) and John Gumperz's Interactional Sociolinguistics. Although both theories view language as a social phenomenon, they offer complementary perspectives on how it is used and interpreted in different communicative contexts. Systemic Functional Linguistics (SFL) explains how speakers make purposeful linguistic choices in response to the social context - Field, Tenor, and Mode (Halliday, 1994; Halliday & Matthiessen, 2014), whereas Interactional Sociolinguistics (IS) focuses on how individuals interpret communicative situations through contextualization cues that shape interaction (Gumperz, 1982). By bringing these perspectives together, the framework provides a comprehensive lens for examining both the observable patterns of TikTok slang use and the socio-cognitive processes that influence students' language choices across digital and academic settings.

Within this framework, Systemic Functional Linguistics (SFL) explains how register varies according to changes in field, tenor and mode, which allows language choice to be examined across different communicative contexts (Halliday & Matthiessen, 2014). More specifically, Systemic Functional Linguistics provides the analytical basis for examining the prevalence of TikTok slang and students' ability to adapt their language across informal and academic contexts through the concepts of register and interpersonal meaning. In contrast, Interactional Sociolinguistics (IS) explains how repeated exposure to digital language can influence the students' interpretation of communicative situations and their perceptions of the cognitive and social influences of TikTok slang (Gumperz, 1982). The integration constructed in the present study will allow examination of digital slang from both functional and interactional perspectives and subsequently provide a stronger conceptual foundation for understanding its influence on Malaysian ESL undergraduates' English language use.

The theoretical integration of this framework also informs the methodological design of the study. The quantitative component is guided by the functional principles of Systemic Functional Linguistics to examine patterns of TikTok slang use and register variation, while the qualitative component draws on Interactional Sociolinguistics to explore students' perceptions of how digital slang influences their language practices. Aligning the theoretical framework with the research design ensures that the data collection and analysis are conceptually grounded and directly address the study's research objectives.

Methodology

Research Design

This study employed a mixed-methods research design to examine both the frequency of TikTok slang usage and the related experiences and perceptions among ESL undergraduates. Mixed-method approaches are widely recognised for enabling both breadth and depth in linguistic research (Braun & Clarke, 2006). Subsequently, the study provides both a broad overview of usage patterns and a more detailed understanding of students' perspectives. This

combination allows the findings to be interpreted more comprehensively, rather than relying solely on numerical trends or subjective accounts.

Participants and Setting

The study involved 73 undergraduate language majors from the Faculty of Major Language Studies at Universiti Sains Islam Malaysia. Participants were selected through purposive sampling, based on the criterion that they used TikTok for at least three hours daily. This requirement ensured that all participants had consistent and meaningful exposure to digital language practices, making them suitable for examining patterns of slang use in authentic contexts. The participants, aged between 19 and 25, were categorised as Generation Z learners, a group whose language practices are strongly shaped by continuous engagement with short-form video platforms. Participation was voluntary, and ethical procedures were strictly followed. Participants were informed that their responses would remain anonymous, and that they had the right to withdraw at any time without penalty.

Research Instrument

Data were collected using a structured online survey designed to address the research objectives through both quantitative and qualitative components. The survey was developed based on the research objectives and the conceptual frameworks identified, with each item aligned with the relevant concepts and research questions. Content validity was established by systematically matching the survey items with the key concepts of TikTok slang use, register variation, language choice, and students' perceptions across digital and academic contexts.

- **Section A: Demographic Information**
This section collected background details such as age, gender, field of study, and self-rated English proficiency. These variables were used to contextualise participant profiles and support interpretation of the findings.
- **Section B: Language Use and Frequency (Quantitative Component)**
This section examined the use of selected TikTok slang terms (e.g., *POV*, *FR*, *Rizz*, *Delulu*, *Slay*) across different communicative contexts, including informal peer interactions and formal academic settings. Participants responded using a five-point Likert scale ranging from *Never* to *Always*. This section provided measurable data on the frequency and context of slang usage, addressing the first and second research objectives.
- **Section C: Open-Ended Responses (Qualitative Component)**
To complement the quantitative data, the survey included open-ended questions that allowed participants to describe their experiences and perceptions of TikTok slang use. These responses explored the reasons behind slang usage, instances where digital language influenced academic writing or speech, and students' views on how social media language may affect their English proficiency. This section provided deeper insight into the personal and cognitive aspects of language use, addressing the third research objective.

Procedures for Data Collection

The survey was administered within a two-week data collection period. Prior to distribution, participants were briefed through institutional digital communication channels on the purpose of the study, its scope, and the requirement for reflective written responses in the qualitative section. The survey was distributed via Google Forms, which included a consent form attached to the opening page to ensure informed participation. Upon completion of the data collection

period, all responses, including both categorical data and open-ended text entries, were exported into a spreadsheet environment for organisation, cleaning, and verification prior to analysis.

Procedures for Data Analysis

Quantitative and qualitative data were analysed separately before being integrated during the final interpretation phase to ensure a comprehensive understanding of the findings.

- **Quantitative Data Analysis:**
Responses from Sections A and B were analysed using descriptive statistical procedures. Frequencies and percentages were computed to identify patterns of TikTok slang usage and to examine its distribution across informal and academic contexts. This analysis provided a numerical overview of slang prevalence and contextual variation among participants.
- **Qualitative Data Analysis:**
Responses from Section C were analysed using reflexive thematic analysis, following the six-phase framework outlined by Braun and Clarke (2006). As the data were collected in written form, analysis began with repeated reading of the responses, followed by systematic line-by-line coding. These codes were then organised into broader themes that reflected participants' experiences, perceptions, and interpretations of digital slang use.

To balance statistical breadth with analytical depth, the descriptive quantitative analysis utilised the full dataset ($N = 73$). In contrast, the qualitative thematic analysis focused on a purposive sub-sample of 15 participants who provided particularly detailed and reflective written responses. This selection prioritised richness of narrative data and depth of meaning over representativeness, in line with established criteria for qualitative rigour (Braun & Clarke, 2006).

Thematic findings were systematically cross-checked against the original data to ensure interpretive consistency and credibility (Braun & Clarke, 2006). These themes were then interpreted through the integrated theoretical framework, where Halliday's Systemic Functional Linguistics, particularly the interpersonal meta function and the register variable of tenor, was used to examine relational meaning and language choice (Halliday, 1994; Halliday & Matthiessen, 2014), while Gumperz's concept of contextualization cues and situational framing (1982) provided explanatory insight into how digital linguistic forms shape interpretation and communicative behaviour.

Findings / Results

The findings are discussed in relation to the research objectives and the theoretical framework. Quantitative results are reported first to outline patterns of TikTok slang usage and contextual variation among undergraduates learning English as a second language, followed by qualitative insights that further explain these patterns. The discussion integrates Systemic Functional Linguistics and Interactional Sociolinguistics to account for both register-based language choices and the socio-cognitive processes underlying students' use and interpretation of digital slang.

Quantitative Findings (RO1 and RO2 – SFL Analysis)

The quantitative findings indicated substantial use of digital-native lexical items within the undergraduate corpus. Rather than forming a uniform or static vocabulary, the data suggested that students selectively and context-sensitively deploy TikTok-derived slang to fulfil specific communicative functions (RO1), consistent with the interpersonal meta function in Systemic Functional Linguistics (Halliday & Matthiessen, 2014).

Table 1. Frequency and Contextual Distribution of Selected TikTok Slang Terms (N = 73)

Slang Term	Linguistic Typology	Primary Functional Role in Discourse	Usage Prevalence (%)	Involuntary Academic Slip (%)
POV (Point of View)	Acronym / Frame Modifier	Establishes a subjective, narrative-first situational frame.	91.8%	64.4%
FR (For Real)	Acronym / Intensifier	Emphasises epistemic sincerity, strong agreement, and validity.	89.0%	52.1%
Rizz (Charisma)	Morphemically Clipped Noun/Verb	Signals social charm, romantic capability, or persuasive attraction.	76.7%	13.7%
Delulu (Delusional)	Reduplicative Clipping / Adjective	Functions as a playful, protective coping mechanism for unrealistic scenarios.	71.2%	19.2%
Slay	Semantic Extension / Verb	Signals enthusiastic approval, excellence, or dominant performance.	68.5%	24.7%

The contextual distribution patterns (RO2) indicated a clear awareness of register appropriateness by the undergraduates, particularly in formal academic setting. In interactions with academic staff, reported use of slang is limited/restricted and suggested that students were able to adjust their vocabulary in response to the formal tenor of the academic setting which is characterised by low familiarity and greater social distance.

However, the findings also indicated partial instability in maintaining these boundaries at the level of automatic language production. A proportion of respondents (64.4%) reported instances in which high-frequency slang terms appeared in formal academic contexts, including written essays, presentations, and continuous assessments. Comparative patterns further suggested variation across lexical forms, with acronyms such as *POV* and *FR* showing higher rates of cross-context transfer than morphologically extended forms such as *Rizz* and *Delulu*.

Qualitative Themes: The Socio-Semantic Paradox (RO3 — Interactional Sociolinguistic Analysis)

The thematic analysis of qualitative responses revealed the socio-cognitive mechanisms underlying both the deliberate use and unintended transfer of TikTok slang among participants. The thematic analysis of qualitative responses from a purposive sub-sample of 15 participants revealed the social and cognitive processes behind both the intentional use and unintended transfer of TikTok slang. The analysis followed a thematic approach to identify common patterns in participants' experiences and views.

Theme 1: Contextualization Cues and Involuntary Mental Slippage

Participants reported difficulties in retrieving formal vocabulary due to frequent exposure to digital slang during academic tasks. Many described a tendency for TikTok slang to appear automatically during academic writing or speaking tasks.

"I think it has a negative effect... it's a struggle sometimes when I need to change to a formal language. I would keep thinking of the TikTok slangs instead of the formal words." (Respondent 4)

This pattern appeared to vary across different slang types. For instance, the relatively high reported transfer of the intensifier *FR* (52.1%) suggested a possible displacement of conventional academic expressions used for emphasis or stance marking, such as *sincerely* or *indeed*. This is further reflected in Respondent 14's account:

"Negative because it makes it harder to improve my vocabulary... I've become so used to online slangs instead of using a wider range of meaningful and more literate words." (Respondent 14)

Theme 2: Interpersonal Meta function for Peer Solidarity

In contrast to the cognitive challenges reported in Theme 1, the qualitative data also highlighted the functional value of TikTok slang within peer interactions. In this instance, participants described the use of terms such as *Rizz*, *Delulu*, and *Slay* as effective linguistic resources for expressing attitudes, managing social closeness, and signalling in-group identity which aligned with the broader patterns of social bonding and stance-taking in digital interaction (Zappavigna, 2012):

"Some of these TikTok slangs could convey my exact messages and reactions the best." (Respondent 5)

"When I'm with my friends... we connect well with just one word." (Respondent 7)

Similarly, *Slay* functions as a concise marker of positive evaluation that reinforces peer alignment and solidarity. This is supported by Respondent 12:

"It makes our speech or writing more expressive, humorous, and relatable." (Respondent 12)

Theme 3: Informal Framing and Reduced Interactional Pressure

Participants reported that the use of familiar digital expressions increases their willingness to communicate in English and reduces hesitation in speaking activities:

"I become more confident and spoke in English more often." (Respondent 13)

"It encourages speakers... to make English a language that is relatable." (Respondent 3)

The finding suggests that TikTok slang contributed to the construction of a more relaxed interactional space in language use.

Discussion

Overall, the findings indicated that TikTok slang is widely used among undergraduates in informal communication, while partial transfer into academic contexts is also reported. Participants demonstrated awareness of register differences, although occasional instability in maintaining strict separation between informal and academic language is evident.

The quantitative results suggested that TikTok-derived slang is not used uniformly across contexts. Rather, it is selectively deployed depending on communicative purpose to fulfil specific communicative functions (RO1), consistent with the interpersonal meta function in Systemic Functional Linguistics (Halliday & Matthiessen, 2014).

This pattern also reflects an alignment with institutional expectations and is consistent with Halliday's account of register sensitivity and boundary control (Halliday, 1994) (RO2). More specifically, it suggested that the students were able to recognise the communicative demands of academic settings and consciously suppress informal linguistic forms when interacting with lecturers or in assessed academic contexts.

The higher frequency of acronyms such as *POV* and *FR*, and their relatively higher rates of reported academic transfer, may be linked to their short form and high frequency in digital

communication environments (Hyland & Jiang, 2021; Mujeeb et al., 2026). This may also indicate that typographically economical and frequently used digital forms are more readily automatised in writing practices, increasing their susceptibility to transfer into academic registers (Al-kadi, 2022). In other words, the frequent use of these shortened forms in fast-paced digital communication may make them more mentally familiar, increasing the chance that they are used unintentionally in formal writing tasks.

The presence of reported “slip” into academic contexts (i.e. register leakage) further indicated that register control was not always fully stable, particularly in spontaneous or less monitored writing tasks. This aligned with corpus-based findings on L2 writers’ challenges in maintaining strict register boundaries (Gilquin & Paquot, 2008). The occurrence of register leakage where informal digital slang occasionally transfers into formal academic contexts may be associated with frequent exposure and increased cognitive automatising of these linguistic forms. However, it did not reflect a lack of linguistic awareness. Instead, participants demonstrated clear sensitivity to contextual appropriateness but experienced occasional difficulty maintaining register boundaries during language production (Androutsopoulos, 2011; Jarvis & Pavlenko, 2008; Schmidt, 1990). The qualitative findings further explained these patterns. From an Interactional Sociolinguistic perspective, repeated exposure may make frequently familiar digital expressions more readily accessible during communication (Gumperz, 1982). This is consistent with evidence of semantic shift and lexical normalisation in short-form digital environments (Mujeeb et al., 2026; Rahman et al., 2022).

The qualitative findings also provided insights into students’ perceptions of the broader socio-psychological and cognitive impacts of TikTok slang use (RO3). While repeated exposure and familiarity increased the automatic retrieval during language production, participants also describes its effects on their social interaction and English language use. From a Systemic Functional Linguistics perspective, the use of slang among peers reflected the emphasis on interpersonal meaning, particularly in peer communication (Halliday & Matthiessen, 2014). Terms like *Delulu* and *Slay* were used not simply as a synonym of lexical items, but as an interpersonal resource that softened critical or unrealistic statements through humour, shared understanding as well as peer solidarity.

Overall, the findings suggested that the use of these slang terms is a purposeful linguistic choice rather than careless language use. Students drew on these resources because they enabled efficient expression, reduced social distance, and strengthened emotional connection within peer communication (Eckert, 2000). Importantly, the findings were consistent with previous studies on youth linguistic identity and digital discourse that position slang as a functional resource for negotiating solidarity and in-group belonging in peer communication contexts (Androutsopoulos, 2011, Shofiyah et al., 2026; Tagg, 2015). However, the present study extends this literature by demonstrating that TikTok slang functions not only as a resource for peer interaction but also as a source of register leakage among L2 undergraduates. The learners recognised contextual language expectations while still experiencing occasional difficulty maintaining boundaries between digital and academic registers. Moreover, repeated exposure to digital expressions may influence formal language choices, highlighting slang as both an interpersonal resource and a challenge to academic register.

Conclusion

This study has demonstrated that TikTok slang is not a temporary or superficial trend, but a socio-semantic resource that influences the communicative practices of university undergraduates. By integrating Systemic Functional Linguistics and Interactional Sociolinguistics, the findings highlight the dual nature of this phenomenon: slang serves as a resource for enhancing interpersonal expression and supporting peer engagement, yet its high

cognitive salience is also associated with instances of unintended transfer into formal academic contexts, as reflected in the reported 64.4% rate of register leakage.

Overall, these findings extend existing research by positioning TikTok slang as both a resource for digital interaction and a factor that may shape register variation and language choices among L2 learners. Consequently, higher education should move beyond prohibition-based approaches and instead adopt Critical Language Awareness frameworks. Such an approach would support students in developing greater register awareness and flexibility, enabling them to navigate effectively between informal digital communication and formal academic discourse.

Recommendations

Implications for Practitioners

The findings suggest that higher education frameworks should move away from restrictive approaches that treat digital slang as a problem by nature. Instead, pedagogical responses should focus on developing students' awareness of how language changes across contexts.

In this regard, the integration of Critical Language Awareness (CLA) into university English curricula is proposed. CLA is grounded in Critical Discourse Analysis (Fairclough, 1992) and supported by pedagogical work in language education (Wallace, 2003). It encourages learners to reflect on how language operates within social contexts and how linguistic choices are shaped by meaning, identity, and communicative purpose. This approach helps students develop awareness of register variation, enabling more appropriate language use across informal and academic settings.

Accordingly, CLA shifts the pedagogical focus from the exclusive teaching of grammar rules to guided reflection on how language is used in real communicative situations. This approach is consistent with recent work highlighting the pedagogical value of analysing authentic digital discourse, where short-form platforms provide rich examples of ongoing semantic variation and language change (Madsen, 2026). More specifically, digital materials such as TikTok content can be used as authentic classroom resources for analysis. Through this, students can examine how Field, Tenor, and Mode influence language choices and how these shifts affect meaning (Halliday & Matthiessen, 2014). For instance, changes in Field involve shifts in the topic or activity being discussed (e.g., casual social interaction versus academic writing), changes in Tenor reflect differences in relationships between speakers (e.g., peer-to-peer communication versus student–lecturer interaction), and changes in Mode refer to the channel of communication (e.g., informal online messaging versus formal written essays). By engaging in systematic comparison between informal digital language and academic language, students may develop greater awareness of register variation and improve their ability to manage language choice across different communicative contexts, thereby reducing the likelihood of unintended transfer into formal writing.

Suggestions for Future Research

The present study focused on undergraduate L2 learners and a selected set of TikTok slang expressions. Future research could examine whether similar patterns of register leakage occur among learners from different educational levels, linguistic backgrounds, or cultural contexts. Comparative studies involving first language (L1) and second language (L2) English speakers may further clarify whether the unintended transfer of digital slang is primarily associated with second language development or reflects a broader consequence of digital language use.

Future studies may also investigate the influence of other digital platforms, such as Instagram, X, or online gaming communities, to determine whether different online environments contribute to similar patterns of language use and register variation. In addition, longitudinal research would provide valuable insights into how sustained exposure to digital slang influences

register awareness and academic writing over time. Intervention-based studies evaluating the effectiveness of Critical Language Awareness (CLA) in improving students' ability to distinguish between informal digital language and academic English would further strengthen the evidence base for pedagogical practice.

Limitations

Several limitations should be considered when interpreting the findings. First, the study focused on undergraduate ESL learners within a single higher education context. As language use is shaped by educational, linguistic, and cultural factors, the findings may not be fully transferable to learners in other institutions or educational settings.

Second, the study examined a selected set of TikTok slang expressions that were commonly used at the time of data collection. Given the dynamic nature of digital language, new slang expressions continue to emerge, and existing expressions may evolve over time. Consequently, the findings should be interpreted as reflecting contemporary patterns of TikTok slang use rather than an exhaustive account of digital slang.

Finally, the study focused specifically on TikTok as the primary source of digital slang. While TikTok has become a prominent platform for the diffusion of emerging language forms, undergraduate learners are also exposed to slang through other digital platforms. Future studies may therefore consider a broader range of social media environments to provide a more comprehensive understanding of digital language use and register variation.

References

- Al-kadi, A. (2022). A cross-sectional study of textese in academic writing: Magnitude of penetration, impacts, and perceptions. *International Journal of Contemporary Educational Research*, 6(1), 29–39. <https://doi.org/10.33200/ijcer.534692>
- Androutsopoulos, J. (2011). Language change and digital media: A review of conceptions and evidence. In T. Kristiansen & N. Coupland (Eds.), *Standard languages and language standards in a changing Europe* (pp. 145–159). Novus Press.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511750526>
- Eckert, P. (2000). *Linguistic variation as a social practice*. Blackwell Publishers.
- Fairclough, N. (1992). *Discourse and social change*. Polity Press.
- Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction*, 1(1), 41–61. <https://doi.org/10.1075/etc.1.1.05gil>
- Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, 65(S1), 228–255. <https://doi.org/10.1111/lang.12119>
- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.
- Harith, N. A. S. B. M., Julindah, N. A. N. B., Rosli, N. A. B., Tahsin, N. Z. B., & Manap, M. R. B. (2025). The negative influence of social media language on academic English writing

- among university students in Malaysia. *International Journal of Research and Innovation in Social Science*, 9(24), 920–924. <https://doi.org/10.47772/IJRIS.2025.924ILEIID00102>
- Hyland, K., & Jiang, F. K. (2021). Academic naming: Changing patterns of noun use in research writing. *Journal of English Linguistics*, 49(3), 255–282. <https://doi.org/10.1177/00754242211019080>
- Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge. <https://doi.org/10.4324/9780203935927>
- Li, H. (2025). Teaching academic English in higher education: Strategies and challenges. *Frontiers in Education*, 10, 1559307. <https://doi.org/10.3389/educ.2025.1559307>
- Madsen, D. O. (2026). Generation Alpha speaks in memes: A conceptual framework for platformed language. *Frontiers in Communication*, 11, Article 1695863. <https://doi.org/10.3389/fcomm.2026.1695863>
- Marzuki, E., & Mustapha, N. S. (2023). Use of TikTok slang among Malaysian Generation Z. *Trends in Undergraduate Research*, 6(2), 12–17. <https://doi.org/10.33736/tur.5516.2023>
- Mohd Zuraini, H. H., Kaur, N., & Kamaruddin, S. F. (2024). Integration of TikTok in language learning: Perceptions of ESL learners. *Journal of Creative Practices in Language Learning and Teaching*, 12(1), 183–205. <https://doi.org/10.24191/cplt.v12i1.25474>
- Mujeeb, N., Lodhi, M. A., Akram, S., & Ali, A. (2026). Beyond cyber slang: Netspeak and academic writing practices in Generation Z higher education. *Journal of Applied Linguistics and TESOL*, 9(2), Article 2247. <https://doi.org/10.63878/jalt2247>
- Nurhayati, I. K., & Putri, A. R. (2024). Exploring cyberculture: A virtual ethnographic perspective on Generation Z slang on TikTok. *Jurnal Sosioteknologi*, 23(3), 400–415. <https://doi.org/10.5614/sostek.itbj.2024.23.3.8>
- Rahman, O. S., Sobarna, C., Sujatna, E. T. S., & Indriyani, L. M. (2022). Semantic shift of English internet slangs used in social media: Morphosemantic analysis. *Research and Innovation in Language Learning*, 5(1), 79–92. <https://doi.org/10.33603/rill.v5i1.6387>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Shofiyah, Q., Aulia, P. N., & Meinawati, E. (2026). Language innovation and identity: The sociolinguistics of Gen Z slang on TikTok. *Humanitatis: Journal of Language and Literature*, 12(2), 169–176. <https://doi.org/10.30812/humanitatis.v12i2.6030>
- Tagg, C. (2015). *Exploring digital communication: Language in action*. Routledge.
- Tagliamonte, S. A. (2016). *Teen talk: The language of adolescents*. Cambridge University Press.
- Ugoala, B. (2024). Generation Z's lingos on TikTok: Analysis of emerging linguistic structures. *Journal of Language and Communication*, 11(2), 211–224. <https://doi.org/10.47836/jlc.11.02.08>
- Wallace, C. (2003). *Critical reading in language education*. Palgrave Macmillan.
- Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web*. Continuum.

AI Tool Perception and Self-Regulated English Learning among Polytechnic Students: Examining Perception Versus Frequency of Use

Sharifah Nozi Syed Hamid*, Politeknik Sultan Azlan Shah, Malaysia,
sharifah@psas.edu.my

* Corresponding Author

Abstract: This study examined the relationship between AI tool perception, frequency of AI tool use, and self-regulated English learning among 90 Malaysian polytechnic students at Politeknik Sultan Azlan Shah. Data were collected via a structured questionnaire and analysed using descriptive statistics, Pearson correlation, and multiple regression in a quantitative cross-sectional design. Students reported positive perceptions of AI tools ($M = 3.94$, $SD = 0.56$), moderate to high self-regulated learning ($M = 3.82$, $SD = 0.50$), and moderate AI tool use ($M = 3.13$, $SD = 0.71$). AI tool perception was significantly associated with self-regulated learning ($r = .597$, $p < .001$) and remained the sole significant predictor in the regression model ($\beta = .578$, $p < .001$, $R^2 = .360$), whereas frequency of use was not ($\beta = .060$, $p = .510$). These findings suggest that AI tool perception, rather than frequency of use alone, plays a more meaningful role in self-regulated English learning, highlighting the need for educators to promote AI literacy and guide students toward critical, purposeful, and independent AI use in polytechnic English classrooms. As the design was cross-sectional, the findings should be interpreted as correlational rather than causal.

Keywords: *AI literacy, artificial intelligence, English language learning, polytechnic students, self-regulated learning*

Introduction

Artificial intelligence (AI) tools have increasingly become part of everyday English learning practices among students. Applications such as ChatGPT, Grammarly, Google Translate, and Duolingo provide immediate feedback, grammar correction, vocabulary support, and opportunities for language practice beyond the classroom (Godwin-Jones, 2022; Kim & Sim, 2024). These tools are especially relevant in polytechnic settings, where English is frequently used for academic tasks, technical communication, and workplace preparation. AI-assisted learning tools are becoming increasingly popular, but their educational usefulness may depend on how students interact with them as well as how frequently they use them. While some students may rely on AI-generated answers passively without meaningful interaction, others may proactively employ AI to edit writing, track mistakes, and consider criticism. This contrast implies that learning behaviors may be significantly influenced by students' opinions about AI technologies.

The ability of students to organize, track, and assess their own learning processes is known as self-regulated learning (SRL) (Zimmerman, 2000). SRL is especially crucial in English language acquisition since language development frequently necessitates individual practice outside of the classroom. In general, students who actively manage their learning are better equipped to edit their work, track their development, and gradually improve their language proficiency. Although AI tools are becoming increasingly common in education, limited research has examined how AI tool perception and frequency of use relate to self-regulated English learning among Malaysian polytechnic students. Most existing studies focus on

university learners or examine technology use in general rather than AI-assisted learning specifically within TVET contexts. Polytechnic students may experience different learning environments, academic expectations, and workplace-oriented communication needs. Recent Malaysian-context research has begun to address this gap: Selvadurai et al. (2025) reported that Malaysian tertiary students used ChatGPT to support self-regulation during the pre-writing stage, while Trinovita et al. (2025) found that generative AI use was associated with stronger self-regulated learning behaviours among EFL learners. However, neither study specifically examined the polytechnic or TVET population, reinforcing the need for the present study.

Another important issue concerns the common assumption that frequent AI use automatically leads to better learning outcomes. In practice, however, students who use AI tools frequently may still engage with them mechanically, while students who perceive AI as meaningful and useful may use them more strategically. Since learner autonomy and self-regulated learning are directly linked to meaningful engagement, it is critical to comprehend this discrepancy. Therefore, this study examines whether AI tool perception and frequency of AI tool use are related to self-regulated English learning among Malaysian polytechnic students. Unlike previous studies that primarily focused on general AI usage or university learners, this study comparatively examines AI perception and frequency of use within Malaysian polytechnic English learning contexts.

Research Problem

Although AI tools are increasingly used in English learning, most studies have not examined how factors such as English proficiency, learning habits, or confidence shape the way TVET students engage with AI tools, nor whether frequent use reliably translates into stronger self-regulated learning. Frequent AI use does not necessarily indicate meaningful engagement: students may rely on AI tools out of convenience or for quick answers without reflecting on or applying the feedback they receive, whereas students who view AI tools positively may be more inclined to use them strategically to organise, monitor, and revise their English learning. This study therefore investigates whether polytechnic students' perceptions of AI tools are a stronger indicator of self-regulated English learning than their frequency of use.

Research Objectives

This study aims to examine the relationship between AI tool perception, frequency of AI tool use,

and self-regulated English learning among polytechnic students. Specifically, it seeks to:

1. identify students' levels of AI tool perception and self-regulated English learning;
2. examine the relationship between AI tool perception and self-regulated English learning; and
3. determine whether AI tool perception and frequency of use predict self-regulated English learning.

Literature Review

Self-Regulated Learning in Language Education

Self-regulated learning (SRL) is crucial to language development because it shows how students actively control their own learning. According to Zimmerman (2000), SRL is a cyclical process that includes monitoring, planning, and reflection. Strong SRL learners are typically

more self-reliant, driven, and determined to advance their language skills. SRL is especially crucial when learning English because language growth frequently necessitates ongoing practice outside of the classroom. Activities such as revising essays, monitoring grammar errors, reviewing vocabulary, and evaluating learning progress require students to regulate their own learning behaviours. Learners who are able to plan and monitor these processes effectively are more likely to achieve meaningful language improvement (Dörnyei & Ryan, 2015). Pintrich (2004) further explains that self-regulated learners actively control their cognition, motivation, and learning strategies in response to academic tasks. In language learning contexts, these behaviours are closely associated with learner autonomy and independent learning.

AI Tools in English Language Learning

The emergence of AI-powered tools has transformed the way students learn English. Applications such as ChatGPT, Grammarly, Google Translate, and Duolingo provide immediate feedback, writing assistance, vocabulary support, and personalised language practice. These features may support SRL by helping learners revise writing, monitor errors, reflect on feedback, and practise independently. Several recent studies suggest that generative AI tools can encourage learner autonomy and reflective learning when used meaningfully (Bahador et al., 2025; Özturan, 2025). Similarly, Kim and Sim (2024) found that AI-assisted learning activities supported autonomous English learning among university students. However, scholars have cautioned that depending too much on AI-generated responses may compromise critical engagement and reduce the development of autonomous language (Liu et al., 2025). In these situations, frequent use of AI might encourage convenience rather than deep learning. This suggests that learners' views of and interactions with AI systems, not accessibility or frequency of use alone, shape the instructional value of these tools. Taken together, these studies show a shift in the literature away from simply documenting how often AI tools are used, toward examining the quality of learner engagement with them. However, few studies have directly compared perception and frequency of use as competing predictors of self-regulated learning outcomes, a gap the present study addresses. Within the Malaysian context specifically, Selvadurai et al. (2025) found that ChatGPT supported self-regulation among tertiary students during the pre-writing stage, but did not examine whether perception or frequency of use was the stronger predictor of this behaviour, a distinction this study is designed to clarify.

Technology Acceptance and AI Tool Perception

According to the Technology Acceptance Model (TAM), people are more inclined to embrace and interact with technology in a meaningful way when they believe it to be practical and simple to use (Davis, 1989). Students may be more inclined to employ AI tools strategically to enhance their performance in classroom settings if they think they are beneficial for language learning. Positive perception may influence the quality of engagement rather than simply the amount of technology use. Students who perceive AI tools positively may use them critically to revise writing, evaluate feedback, and improve language accuracy. By contrast, students who use AI tools mechanically may rely on generated responses without reflecting on learning outcomes. This distinction is important because meaningful engagement is closely associated with self-regulated learning behaviours. The present study therefore applies TAM to examine whether students' perceptions of AI tools contribute more strongly to SRL than frequency of use alone.

AI-Assisted Learning in Malaysian Higher Education

Research on AI-assisted learning in Malaysian higher education is increasing, particularly in university contexts. However, studies focusing on polytechnic or TVET learners remain limited. This gap is important because polytechnic students often experience different learning environments, academic expectations, and workplace-oriented communication needs. English learning in polytechnic settings is closely linked to technical communication, employability, and industry preparation. Therefore, by better understanding how students perceive and utilize AI tools for English learning, educators may develop more effective support for learner autonomy and AI literacy in TVET classrooms. Despite increasing research on AI-assisted language learning, limited studies have comparatively examined whether students' perceptions of AI tools or frequency of use better explain self-regulated English learning, particularly within Malaysian TVET and polytechnic contexts. This study addresses this gap by examining the predictive contribution of both variables within AI-assisted English learning environments. Elsewhere in the EFL literature, Trinovita et al. (2025) similarly found that generative AI use was associated with stronger self-regulated learning behaviours, but their mixed-methods study was conducted outside the Malaysian TVET context, further underscoring why a comparative examination of perception and frequency within Malaysian polytechnic English learning is needed.

Conceptual Framework

This study is guided by Self-Regulated Learning Theory and the Technology Acceptance Model. AI tool perception is treated as a learner belief that shapes meaningful engagement with AI-assisted learning, while frequency of use reflects behavioural exposure to the tools. Self-regulated English learning serves as the dependent variable. The framework assumes that students who perceive AI tools positively are more likely to use them strategically for planning, monitoring, revising, and reflecting on their English learning.

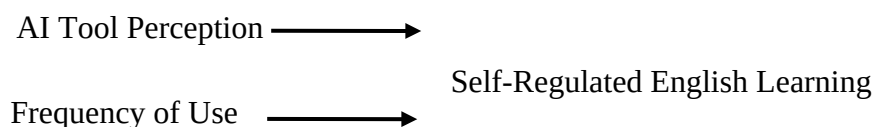


Figure 1: Conceptual Framework of the Study

Methodology

Research Design

This study used a quantitative cross-sectional survey design. Data were collected at one moment in time from a group of polytechnic students to examine the relationships between self-regulated English learning, frequency of AI tool use, and perception of AI tools. A cross-sectional design was considered suitable because the study's objective was to look at correlations between variables rather than monitor changes over time.

Participants

A total of 90 students from Politeknik Sultan Azlan Shah (PSAS) participated in this study. They were selected using purposive sampling and were all enrolled in English language courses at the time of data collection. Purposive sampling was used because it allowed the researcher to specifically target students who were concurrently enrolled in English language courses and therefore had direct, ongoing exposure to the AI-assisted English learning activities relevant to the study's variables. Most of the respondents were male (68.9%), aged between 20 and 21 (50.0%), and in their third year of study (58.9%). In terms of English ability, nearly half of the students (48.9%) rated themselves as having a moderate level, while 36.7% said they were good. For AI tool usage, 55.6% said they use AI tools sometimes, and 30.0% said they use them often. Only a small number said they never use AI tools (2.2%) or rarely use them (12.2%). This sample size was considered appropriate for regression analysis based on the guideline by Hair et al. (2019), which recommends a minimum sample of $50 + 8m$ for regression models.

Research Instrument

A three-part structured questionnaire was used to obtain the data. Section A collected demographic data, including age, gender, year of study, and English self-rating. Using Likert-scale items, the frequency of AI tool use was examined independently as one of the study factors. Section B assessed how students felt about AI tools. The items were modified from the Technology Acceptance Model (Davis, 1989; Venkatesh et al., 2003), emphasizing meaningful involvement with AI-assisted learning and perceived utility. Section C measured self-regulated English learning. The items were modified from well-known frameworks for self-regulated learning and studies on language learning strategies (Oxford, 1990; Pintrich et al., 1991). Every question on the survey employed a five-point Likert scale, with 1 denoting "strongly disagree" and 5 denoting "strongly agree." Prior to the main data collection, a pilot study was conducted with 20 students who shared similar characteristics to the target population (i.e., polytechnic students enrolled in English language courses) but were not included in the main study sample. The pilot testing indicated that the questionnaire items were clear and easy to understand, and no major revisions were required to the instrument.

Reliability Analysis

The questionnaire's reliability was evaluated using Cronbach's Alpha. The AI Tool Perception scale obtained an alpha value of .939 across 10 items, while the Self-Regulated Learning scale recorded an alpha value of .903 across 10 items. These reliability values indicate strong internal consistency among the questionnaire items.

Ethical Considerations

This study received ethical clearance from the Research, Innovation and Commercialisation Unit of Politeknik Sultan Azlan Shah. Prior to data collection, participants were informed of the purpose of the study on the first page of the questionnaire, where consent was implied by their decision to proceed with and complete the survey. Participants were also informed that their responses would remain anonymous, and that participation was voluntary.

Data Analysis

Data were analyzed using IBM SPSS Statistics Version 26. Descriptive statistics summarized self-regulated learning, usage frequency, and perception of AI tools. Multiple regression assessed whether AI tool perception and usage frequency predicted self-regulated English learning, while Pearson correlation examined relationships among variables. Significance was set at $p < .05$. Before interpreting the regression, key assumptions were checked. Multicollinearity was not a concern, as the two predictors were only weakly correlated ($r = .313$, $p < .01$), far below the problematic $r = .80$ – $.90$ threshold (Field, 2013). Standardized residuals ranged from -2.145 to 2.361 ($M = .000$, $SD = .989$), with none exceeding $\pm 3 SD$, indicating approximately normal residuals with no undue outlier influence. Standardized predicted values ranged narrowly from -1.952 to 1.945 , showing no evidence of non-linearity. Overall, the assumptions of normality, linearity, and multicollinearity were reasonably satisfied.

Findings/Results

Demographic Profile of Respondents

The respondents' demographic background is shown in Table 1. Most participants were male (68.9%), in Year 3 of their studies (58.9%), and aged between 20 and 21 years old (50.0%). In terms of English proficiency, the majority of students rated themselves as having moderate (48.9%) or good (36.7%) English ability. Regarding AI tool use, most students reported using AI tools sometimes (55.6%) or often (30.0%), indicating that AI tools are commonly used in their English learning practices.

Table 1: Demographic Profile of Respondents ($N = 90$)

Variable	Category	Frequency	Percentage (%)
Gender	Male	62	68.9
	Female	28	31.1
Age Group	18–19	28	31.1
	20–21	45	50.0
	22 and above	17	18.9
	Year 1	12	13.3
Year of Study	Year 2	25	27.8
	Year 3	53	58.9
	Very Low	1	1.1
English Self-Rating	Low	8	8.9
	Moderate	44	48.9
	Good	33	36.7
	Very Good	4	4.4
Frequency of AI Tool Use	Never	2	2.2
	Rarely	11	12.2
	Sometimes	50	55.6
	Often	27	30.0

Descriptive Statistics

Table 2 presents the mean scores and standard deviations for the three main variables. Students reported a moderately high level of AI tool perception ($M = 3.94$, $SD = 0.56$), indicating generally positive views toward AI-assisted English learning. The mean score for self-regulated learning was also moderately high ($M = 3.82$, $SD = 0.50$), indicating a relatively high level of self-regulated English learning. The mean score for frequency of AI tool use was moderate ($M = 3.13$, $SD = 0.71$), suggesting a moderate level of AI tool engagement among the respondents.

Table 2: Descriptive Statistics

Variable	Mean	SD
AI Tool Perception	3.94	0.56
Self-Regulated Learning (SRL)	3.82	0.50
Frequency of AI Tool Use	3.13	0.71

Correlation Analysis

Table 3 presents the results of the Pearson correlation analysis. AI tool perception showed a substantial positive association with self-regulated learning ($r = .597$, $p < .001$). This suggests that students who were more favourable about AI tools also reported better levels of self-regulated English learning. Frequency of AI tool use had a weak but significant positive relationship with self-regulated learning ($r = .241$, $p < .05$). Frequency of use correlated with perception of AI tools ($r = .313$, $p < .01$). The lower association between frequency of usage and self-regulated learning implies that frequent AI use may not always lead to higher self-regulated learning behaviours.

Table 3: Pearson Correlation Analysis

Variables	SRL	AI Perception	Frequency of Use
SRL	1.000	.597**	.241*
AI Perception	.597**	1.000	.313**
Frequency of Use	.241*	.313**	1.000

Note. ** $p < .01$; * $p < .05$

Multiple Regression Analysis

A multiple regression analysis was performed to investigate whether AI tool perception and frequency of use significantly influenced self-regulated learning. The regression model was statistically significant ($F(2, 87) = 24.433$, $p < .001$), explaining 36% of the variance in self-regulated learning ($R^2 = .360$, Adjusted $R^2 = .345$). Table 4 shows that AI tool perception was a substantial and significant predictor of self-regulated learning ($B = .519$, $\beta = .578$, $p < .001$). The frequency of AI tool use did not substantially predict self-regulated learning ($B = .043$, $\beta = .060$, $p = .510$) when both variables were included in the model. These findings imply that students' perceptions of AI tools may be more closely related to self-regulated English learning than frequency of use alone. In terms of practical significance, the model's R^2 of .360 corresponds to a Cohen's f^2 of approximately .56, which exceeds Cohen's (1988) benchmark for a large effect ($f^2 \geq .35$). This indicates that the predictive contribution of AI tool perception to self-regulated learning is not only statistically significant but also practically meaningful.

Table 4: Multiple Regression Analysis Predicting Self-Regulated Learning

Predictor	B	Std. Error	Beta (β)	Sig.
(Constant)	1.638	.323	—	.000
AI Tool Perception	.519	.081	.578	.000
Frequency of AI Tool Use	.043	.064	.060	.510
Note. $R^2 = .360$; Adjusted $R^2 = .345$; $F(2, 87) = 24.433$, $p < .001$				

Discussion

The findings indicate that students generally held positive perceptions of AI tools and demonstrated moderately high levels of self-regulated English learning. More importantly, AI tool perception showed a stronger relationship with self-regulated learning than frequency of AI tool use. This finding supports the Technology Acceptance Model, which proposes that perceived usefulness and positive attitudes toward technology influence meaningful engagement with digital tools (Davis, 1989). In the present study, students who perceived AI tools positively may have been more likely to use them strategically to revise writing, monitor errors, reflect on feedback, and improve language performance.

These behaviours are consistent with self-regulated learning because they involve planning, monitoring, and evaluating one's own learning processes (Zimmerman, 2000; Dörnyei & Ryan, 2015). The findings also support recent research suggesting that generative AI tools can encourage learner autonomy and reflective learning when used purposefully (Bahador et al., 2025; Özturan, 2025). When AI perception was factored into the regression model, the frequency of AI tool use did not predict self-regulated learning. This shows that regular use alone may not result in significant learning. Students may continually use AI tools for convenience or quick answers rather than critically engaging with the learning process. This interpretation is consistent with Liu et al. (2025), who warn that overreliance on AI-generated solutions may impede critical thinking and individual participation in language learning. The findings also contribute theoretically to AI-assisted language learning research by extending the Technology Acceptance Model within a TVET context. Specifically, the study suggests that positive perceptions of AI tools may influence the quality of learner engagement more strongly than behavioural frequency alone in self-regulated English learning environments.

A plausible explanation for why perception outweighed frequency in this study is that perception functions as a belief-based, evaluative filter through which AI use is translated into regulatory behaviour: students who view AI tools as useful and trustworthy are more likely to interrogate AI-generated feedback, compare it against their own judgement, and use it to plan and revise their work, whereas frequent use without this evaluative stance may reflect habitual or convenience-driven behaviour that bypasses the monitoring and reflection central to SRL. This pattern is broadly consistent with Selvadurai et al. (2025), who found that Malaysian tertiary students' self-regulation during ChatGPT-assisted pre-writing depended on how deliberately they engaged with the tool rather than simply how often they used it. It also extends Trinovita et al.'s (2025) mixed-methods finding that generative AI use supported self-regulated learning among EFL learners, by suggesting that within a Malaysian TVET population, the evaluative dimension of AI use (perception) may matter more than the behavioural dimension (frequency) in predicting self-regulated English learning. Overall, the findings highlight the importance of meaningful engagement, AI literacy, and reflective learning practices in AI-assisted English learning environments.

The findings have several implications for English teaching in polytechnic settings. First, educators should guide students toward strategic rather than passive AI use. Students need support in evaluating AI-generated responses, revising their own work, and using AI feedback to improve learning rather than replace independent thinking. Second, AI literacy should be integrated into English instruction. Classroom activities can encourage students to compare AI-generated feedback with their own ideas, justify revision decisions, and reflect on how AI tools influence their learning processes. Third, lecturers should encourage responsible and reflective AI-assisted learning. AI tools can function as valuable learning supports when used critically and purposefully, but they should not become substitutes for learner effort, reflection, or independent language production. Such approaches may strengthen both language development and learner autonomy among polytechnic students.

Conclusion

This study examined the relationship between AI tool perception, frequency of AI tool use, and self-regulated English learning among students at Politeknik Sultan Azlan Shah. The findings showed that students generally reported positive perceptions of AI tools and moderately high levels of self-regulated English learning. AI tool perception significantly predicted self-regulated learning, whereas frequency of AI tool use did not significantly predict self-regulated learning when both variables were analysed together.

Recommendations

These findings suggest that students' perceptions of AI tools may be more closely associated with self-regulated English learning than simple frequency of use. In other words, the quality of engagement with AI tools appears to matter more than repeated exposure alone. The study contributes to current discussions on AI-assisted language learning by highlighting the importance of AI literacy, meaningful engagement, and reflective learning practices in TVET contexts. For English educators in polytechnic settings, the findings suggest the need to guide students toward critical, purposeful, and independent use of AI tools rather than encouraging frequent but passive use. More specifically, curriculum developers could embed short AI-literacy modules within existing English courses that teach students how to evaluate AI-generated feedback against their own writing goals; lecturers could require brief reflection logs after AI-assisted writing tasks, in which students justify which AI suggestions they accepted, revised, or rejected; and assessment rubrics could allocate explicit marks for evidence of critical, reflective AI use rather than for AI use itself, so that strategic engagement, not frequency, is what is recognised and rewarded.

Limitations

Several limitations should be recognized. First, the study used a cross-sectional design, which limits the causal interpretation of the findings. Second, the study used purposive sampling from a single institution, which limited the generalizability of the results. Third, the data were based on self-reported responses, which could be influenced by social desirability or faulty self-perceptions of AI use and learning habits. To boost generalizability, future study should include bigger sample sizes from multiple polytechnics and universities. Qualitative or longitudinal techniques may potentially provide further information about how students' AI-assisted learning behaviours and self-regulated learning practices change over time.

References

- Bahador, H., Taghipour, M., & Mirzapour Kouhdasht, A. (2025). Artificial Intelligence on Foreign Language Learning Strategies and Self-Regulated Learning: A Qualitative Systematic Review. *Technology Assisted Language Education*, 3(4), 25-41. <https://doi.org/10.22126/tale.2026.13456.1163>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- 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>
- Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage.
- Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, 26(2), 5–24. <https://doi.org/10.12545/llt.26.2.2>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Kim, E., & Sim, J. (2024). Incorporating AI into English language learning: An experimental study focusing on autonomous learning. *English Language Teaching*, 17(10), 82–93. <https://doi.org/10.5539/elt.v17n10p82>
- Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments*, 12, 48. <https://doi.org/10.1186/s40561-025-00406-0>
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Newbury House Publishers.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16, 385–407. <https://doi.org/10.1007/s10648-004-0006-x>
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. University of Michigan.
- Selvadurai, V., Rodrigues, P. D., & Rajamohan, L. P. (2025). Engaging with AI: Exploring ChatGPT's role in self-regulation and pre-writing stage. *VietTESOL International Convention Proceedings*, 5, 211–227.

- Trinovita, D., Nurchurifiani, E., Hastomo, T., Andewi, W., & Hasbi, M. (2025). Exploring the influence of generative AI on self-regulated learning: A mixed-methods study in the EFL context. *Jurnal IQRA*, 10(2), 301–316. <https://doi.org/10.25217/ji.v10i2.6389>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>
- Özturan, T. (2025). Self-regulated language learning and generative AI: A systematic review. *Boğaziçi University Journal of Education*, 42(3), 177–194. <https://doi.org/10.52597/buje.1751534>

Students' Acceptance of CIDOS in Education 5.0-Oriented Blended Communicative English Learning: A Conceptual Framework for Malaysian TVET

Nurfarizah binti Ishak¹, Masniza binti Mansor², Diana binti Ahmad Busra³

¹²³Department of General Studies, Polytechnic Port Dickson, Negeri Sembilan, Malaysia

ABSTRACT

The increasing use of blended learning in Technical and Vocational Education and Training (TVET) has expanded the role of learning management systems in supporting teaching and learning. In Malaysian polytechnics, the Curriculum Information Document Online System (CIDOS) serves as a learning management platform for instructional delivery, communication, assessment-related activities, and access to learning resources. Understanding students' acceptance of CIDOS is particularly relevant in Communicative English courses, where learning involves not only access to digital content but also interaction, participation, and collaborative language-learning activities. Existing technology-acceptance models explain students' perceptions and intentions but may not fully capture the technological and pedagogical conditions surrounding LMS use in communicative learning environments. This conceptual paper therefore proposes an extended Technology Acceptance Model (TAM) framework for understanding students' acceptance of CIDOS in blended Communicative English learning within Malaysian TVET. A focused conceptual review integrated TAM with selected constructs from information systems, technology adoption, and online and collaborative learning literature. The framework retains perceived ease of use, perceived usefulness, behavioural intention, and actual use as its core acceptance mechanism, while incorporating system quality, facilitating conditions, social interaction, and collaborative engagement as contextually relevant antecedents. Education 5.0 provides the framework's broader human-centred educational orientation rather than functioning as an additional causal construct. Eight theoretical propositions explain how technological, institutional, and learning-environment factors may contribute to students' perceptions and subsequent acceptance of CIDOS. The framework provides a context-specific basis for future empirical testing and for examining LMS acceptance in blended Communicative English learning within Malaysian polytechnics.

Keywords: CIDOS, blended learning, Technology Acceptance Model, Communicative English, TVET, Education 5.0, learning management system

1. Introduction

This paper is a conceptual, theory-building study that develops an extended Technology Acceptance Model (TAM) to understand students' acceptance of the Curriculum Information Document Online System (CIDOS) in blended Communicative English learning within the Malaysian TVET context. Rather than empirically testing relationships among variables, the paper synthesises established technology-acceptance theory and relevant blended-learning literature to identify theoretically defensible constructs and relationships for subsequent empirical examination.

The framework retains perceived usefulness and perceived ease of use as the central TAM mechanisms and incorporates four contextually relevant antecedent constructs: system quality, facilitating conditions, social interaction, and collaborative engagement. These constructs are included because accepting an LMS in Communicative English learning involves both technological conditions and pedagogical experiences. In particular, Communicative English requires interaction, participation, collaboration, and meaningful opportunities to use language, while effective LMS use also depends on reliable technological performance and appropriate institutional support.

Education 5.0 serves as the framework's contextual orientation rather than an additional causal construct. It situates CIDOS acceptance within a human-centred educational environment that values meaningful technology use, collaboration, flexibility, and learner participation.

2. Problem Statement

CIDOS has been used as a Learning Management System to support e-learning in Malaysian polytechnics, and earlier research has examined students' acceptance and usability of the platform (Mat Daud et al., 2013). Although its institutional role is established, understanding students' acceptance of CIDOS requires more than examining whether students perceive the platform as useful and easy to use. This is particularly important in Communicative English courses, where learning depends substantially on interaction, communication practice, collaboration, and active participation.

The Technology Acceptance Model provides a well-established theoretical foundation for explaining technology acceptance through perceived usefulness and perceived ease of use. In Malaysian polytechnic English-language education, previous research on LMS-CIDOS has identified issues related to system and internet reliability, content organisation, institutional support, training, communication, and students' digital literacy (Hasnan & Mohin, 2021).

This issue is particularly relevant to blended Communicative English learning. Unlike learning environments in which an LMS primarily distributes content or manages assessment, Communicative English requires learners to communicate, collaborate, exchange ideas, and practise language meaningfully. Consequently, students may evaluate the usefulness of CIDOS partly according to whether the learning environment enables these pedagogical experiences. At the same time, their experience of the platform may depend on system quality and the technological and institutional conditions that enable its use.

However, the literature reviewed in this paper does not yet provide sufficient CIDOS-specific evidence to claim that these relationships have been established in blended Communicative English learning in Malaysian polytechnics. The present study therefore treats this limitation as a theoretical and contextual research opportunity rather than an empirically established deficiency.

Accordingly, this conceptual paper develops an extended TAM framework that integrates technological-environment factors (system quality and facilitating conditions) and learning-environment factors (social interaction and collaborative engagement). The framework provides

theoretically derived propositions for subsequent empirical testing of students' acceptance and use of CIDOS in Malaysian TVET Communicative English education.

3. Research Objectives

This conceptual paper aims:

1. To synthesise TAM and relevant blended-learning literature to establish a theoretical basis for understanding students' acceptance of CIDOS in blended Communicative English learning.
2. To conceptually examine the roles of system quality, facilitating conditions, social interaction, and collaborative engagement as antecedent factors within an extended TAM framework.
3. To develop theoretically grounded propositions explaining the relationships among the proposed constructs, perceived ease of use, perceived usefulness, behavioural intention, and actual CIDOS use.
4. To propose a context-specific conceptual framework for subsequent empirical examination of CIDOS acceptance in Education 5.0-oriented Malaysian TVET Communicative English learning.

4. Research Questions

1. How can TAM provide the theoretical foundation for conceptualising students' acceptance of CIDOS in blended Communicative English learning?
2. How can system quality and facilitating conditions be theoretically integrated into TAM to explain students' perceptions of CIDOS?
3. How can social interaction and collaborative engagement be theoretically integrated into TAM to reflect the pedagogical requirements of Communicative English learning?
4. What theoretically grounded relationships can be proposed among these contextual constructs, perceived ease of use, perceived usefulness, behavioural intention, and actual CIDOS use?
5. How can these relationships be integrated into a context-specific conceptual framework for future empirical investigation in Malaysian TVET institutions?

5. Conceptual Review and Framework-Development Method

5.1 Research Design

This study adopts a conceptual, theory-building approach to develop a context-specific framework for understanding students' acceptance of CIDOS in blended Communicative English learning. Unlike empirical research, the purpose of a conceptual study is not to test relationships among variables but to draw together existing theoretical knowledge to develop new conceptual explanations or models. A clear research design is therefore important to show how relevant

theories are selected, interpreted, and brought together in developing the proposed framework (Jaakkola, 2020).

The Technology Acceptance Model (TAM) provides the framework's theoretical foundation through its core constructs of perceived ease of use, perceived usefulness, behavioural intention, and actual technology use (Davis, 1989). The framework retains TAM as its theoretical core while drawing on related research to incorporate factors that are particularly relevant to LMS use and Communicative English learning. System quality and facilitating conditions represent the technological and institutional conditions surrounding CIDOS use, while social interaction and collaborative engagement represent the pedagogical conditions of the learning environment. Education 5.0 provides the broader human-centred educational context rather than functioning as an additional causal construct.

The purpose of the conceptual analysis is not to demonstrate that the proposed relationships already exist among CIDOS users. Instead, the analysis develops theoretically grounded propositions for future empirical testing in Malaysian TVET contexts.

5.2 Literature Identification

The literature synthesis focuses on five interconnected domains:

- Technology acceptance and TAM;
- LMS adoption and system success;
- blended learning in higher education and TVET;
- social interaction and collaborative learning in online and blended environments; and
- blended Communicative English and language-learning environments.

The literature was identified through a focused conceptual review guided by the purpose of the proposed framework. Rather than attempting an exhaustive or systematic review of all studies on technology acceptance and blended learning, the review concentrated on literature that could help explain the constructs and relationships relevant to CIDOS use in blended Communicative English learning.

The review included seminal theoretical works that established the foundations of TAM and related technology-acceptance constructs, together with more recent studies on LMS use, blended learning, technology-supported language learning, social interaction, collaborative learning, system quality, and facilitating conditions. Particular attention was given to literature from higher education, TVET, and language-learning contexts where it provided useful conceptual or empirical evidence for the proposed framework.

Sources were considered for inclusion when they contributed to at least one of three purposes: defining a construct, explaining its theoretical relevance, or providing evidence for a proposed relationship within the framework. Studies were not treated as evidence that the same relationships necessarily occur among CIDOS users. Instead, the literature was used to develop propositions that remain to be tested empirically in the Malaysian TVET context.

5.3 Analytical Procedure

The development of the proposed framework involved four analytical stages. These stages moved from the established TAM model to selecting additional constructs and, finally, to developing the proposed relationships.

Stage 1: Establishing the theoretical core

TAM was used as the starting point for the framework. The established relationships among perceived ease of use, perceived usefulness, behavioural intention, and actual technology use were retained as the core acceptance mechanism (Davis, 1989). This provided a clear theoretical base before considering additional factors relevant to the CIDOS learning environment.

Stage 2: Identifying contextual limitations of the baseline model

The literature was then examined to identify aspects of blended Communicative English learning that are not fully represented by the core TAM constructs. Particular attention was given to the technological conditions surrounding LMS use and to the interaction and collaboration required in communicative language learning. This stage helped identify areas where additional constructs could improve the framework's explanatory value without changing its TAM foundation.

Stage 3: Selecting and organising the additional constructs

Four additional constructs were selected based on their relevance to the CIDOS context: system quality, facilitating conditions, social interaction, and collaborative engagement. System quality and facilitating conditions were grouped as factors related to the technology and institutional environment, while social interaction and collaborative engagement were grouped as factors related to the learning environment. The constructs were kept conceptually distinct to avoid unnecessary overlap within the framework.

Stage 4: Developing the propositions

The final stage involved linking the additional constructs to the TAM acceptance mechanism. The proposed relationships were developed by considering the theoretical meaning of each construct, evidence reported in relevant literature, and its relevance to blended Communicative English learning. This process resulted in eight propositions (P1–P8). These propositions represent relationships to be examined in future research rather than findings established by the present study.

5.4 Scope of the Conceptual Framework

The proposed framework is intentionally focused rather than comprehensive. It does not attempt to combine all constructs associated with TAM, UTAUT, Information Systems Success, or research on social presence and collaborative learning. Instead, the additional constructs were selected because of their relevance to the specific context of CIDOS use in blended Communicative English learning.

The framework therefore distinguishes between two sets of contextual factors. System quality and facilitating conditions represent the technological and institutional environment surrounding CIDOS use, while social interaction and collaborative engagement represent the learning environment in which Communicative English is taught. These factors are integrated with the core TAM constructs to provide a more context-sensitive explanation of students' acceptance of CIDOS.

Education 5.0 is treated as the framework's broader educational context rather than an independent variable. Its emphasis on learner-driven, flexible, personalised, technology-supported, and collaborative learning helps to situate the proposed framework within contemporary Malaysian higher-education aspirations (Universiti Teknologi MARA [UiTM], 2019). However, this study does not propose a direct causal relationship between Education 5.0 and students' acceptance of CIDOS.

6. Theoretical Architecture of the Proposed Framework

The proposed framework brings together three related layers: the technology environment, the learning environment, and the technology-acceptance mechanism. TAM remains the main theoretical foundation, while concepts from information systems and technology-adoption research, along with research on online and collaborative learning, account for aspects of the CIDOS learning experience not explicitly represented in the original model.

Layer 1: Technology Environment

System quality refers to students' perceptions of CIDOS's technical performance, including its reliability, accessibility, responsiveness, and ease of navigation. The inclusion of system quality is informed by the Information Systems Success Model, which recognises system quality as an important dimension of information-system success (DeLone & McLean, 2003). In the present framework, system quality represents the characteristics of CIDOS itself rather than students' perceptions of the platform's educational value.

Facilitating conditions refer to the resources and support available to students when using CIDOS, such as internet access, suitable devices, technical assistance, institutional guidance, and user support. The construct is informed by UTAUT, which defines facilitating conditions as users' perceptions that the organisational and technical infrastructure needed to support technology use is available (Venkatesh et al., 2003). In an educational setting, these conditions are relevant because students may experience technology differently depending on the support and resources available to them.

Although system quality and facilitating conditions are related, they represent different aspects of the technology environment. System quality concerns the performance and characteristics of CIDOS itself, whereas facilitating conditions concern the external resources and support that enable students to use the platform.

Layer 2: Learning Environment

Social interaction refers to meaningful communication and interpersonal connection among students, lecturers, and peers within CIDOS-supported learning. Its inclusion is informed by research on social presence in online learning, particularly the Community of Inquiry framework, which recognises social presence as an important element of meaningful educational experiences in computer-mediated environments (Garrison et al., 2000). Social interaction is especially relevant to Communicative English because students need opportunities to communicate, exchange ideas, receive feedback, and use language with others.

Collaborative engagement goes beyond interpersonal communication to involve active participation in shared learning activities. In blended language-learning environments, collaborative tasks can give learners opportunities to negotiate meaning, exchange ideas, practise the target language, and work together toward common learning goals. Research in blended foreign-language learning has shown the value of collaborative and authentic activities in supporting communication and language-learning experiences (Avci & Adiguzel, 2017).

Social interaction and collaborative engagement are therefore treated as related but distinct constructs. Students may interact with lecturers or peers without necessarily engaging in sustained collaborative learning. Collaborative engagement, in contrast, involves active participation with others towards a shared learning purpose. Maintaining this distinction helps reduce conceptual overlap between the two constructs.

Layer 3: Technology-Acceptance Mechanism

The third layer retains the central mechanisms of TAM: perceived ease of use, perceived usefulness, behavioural intention, and actual technology use (Davis, 1989). Perceived ease of use reflects the extent to which students expect CIDOS to require little effort, while perceived usefulness reflects the extent to which they believe that using CIDOS supports their learning. These perceptions contribute to students' behavioural intention to use the platform and, ultimately, their actual use.

The additional constructs are not intended to replace TAM or to suggest that the original model is inadequate. Rather, they allow the framework to account for characteristics of the particular learning context examined in this study. The technology environment represents the conditions that may shape students' experience of using CIDOS, while the learning environment represents the interaction and collaboration through which students may experience its educational value. These factors are connected to the TAM mechanism through the propositions developed in the following section.

7. Development of Theoretical Propositions

P1: System Quality → Perceived Ease of Use

System quality concerns the technical characteristics of an information system, such as reliability, accessibility, responsiveness, and ease of navigation. When an LMS operates consistently and allows users to locate and access learning materials without unnecessary technical difficulty, students are likely to experience less effort in using the system. This provides a theoretical basis for linking system quality with perceived ease of use.

Empirical research also supports this relationship. Calisir et al. (2014), in an extended TAM study of a web-based learning system, found that perceived system quality contributed to users' perceived ease of use. Applied to the present context, students who perceive CIDOS as reliable, accessible, and easy to navigate may consequently perceive the platform as requiring less effort to use. The relationship is therefore proposed for future empirical testing among Malaysian TVET students.

P1: System quality positively influences students' perceived ease of use of CIDOS.

P2: Facilitating Conditions → Perceived Ease of Use

Facilitating conditions concern the resources and support available to users when engaging with technology. Although the construct originates from a different technology-acceptance tradition from TAM, it is relevant to the present framework because access to appropriate devices, internet connectivity, technical assistance, and institutional support can reduce the practical difficulties students experience when using an LMS.

Evidence from technology-supported language learning particularly supports this relationship. Ebadi and Raygan (2023) found a significant positive relationship between facilitating conditions and perceived ease of use among EFL learners using mobile-assisted language learning. These findings suggest that when students have sufficient technological resources and support, they may perceive CIDOS as easier to use.

P2: Facilitating conditions positively influence students' perceived ease of use of CIDOS.

P3: Perceived Ease of Use → Perceived Usefulness

Perceived ease of use and perceived usefulness are closely related within TAM. When a technological system requires less effort to operate, users can focus on what the system enables them to accomplish rather than on the difficulties involved in using it. Davis (1989) found that perceived ease of use may act as an antecedent to perceived usefulness, providing the theoretical basis for retaining this relationship in the proposed framework.

In the context of CIDOS, students who find the platform straightforward to navigate and use may be better able to access materials, complete learning activities, communicate, and manage course-related tasks. Ease of use may therefore contribute to students' perceptions that CIDOS supports their Communicative English learning.

P3: Perceived ease of use positively influences students' perceived usefulness of CIDOS.

P4: Social Interaction → Perceived Usefulness

Social interaction represents students' opportunities to communicate with lecturers and peers, exchange ideas, ask questions, receive feedback, and participate in learning-related communication. In Communicative English, these interactions are particularly relevant because communication with others forms an important part of the learning process. When CIDOS supports meaningful interaction, students may perceive the platform as more useful for achieving their learning purposes

However, the relationship between social dimensions of online learning and perceived usefulness should not be treated as firmly established. Research examining social presence alongside TAM has produced mixed findings. For example, Salimon et al. (2021), in a study involving undergraduate students from Malaysian public universities, found that perceived social presence significantly influenced perceived ease of use but did not have a significant direct effect on perceived usefulness. Social presence and social interaction are not identical constructs, but this finding indicates that the relationship between the social experience of online learning and perceived usefulness requires further examination rather than assumption.

For this reason, the proposed relationship is treated as a context-specific theoretical proposition. In blended Communicative English learning, the usefulness students attach to CIDOS may depend partly on whether the platform enables meaningful communication that supports their language-learning activities.

P4: Social interaction positively influences students' perceived usefulness of CIDOS.

P5: Collaborative Engagement → Perceived Usefulness

Collaborative engagement refers to students' active participation with others in shared learning activities. It extends beyond simply communicating with classmates or lecturers by involving participation in tasks that require learners to exchange ideas, solve problems, negotiate meaning, or work towards a common learning outcome.

Research on technology-supported collaborative learning supports the educational relevance of collaboration and engagement. In Malaysian higher education, Al-Rahmi et al. (2018) examined active collaborative learning and engagement through social media among 723 research students from five Malaysian research universities. Their findings showed that technology-supported collaborative learning and engagement enriched learning activities and facilitated group discussion. In language-learning contexts, Avci and Adiguzel (2017) similarly demonstrated the value of collaborative and authentic activities in supporting communication and language-learning experiences.

These studies do not establish that collaborative engagement necessarily causes perceived usefulness in the CIDOS context. Rather, they provide a pedagogical basis for proposing that students may perceive CIDOS as more educationally useful when the platform enables

meaningful participation in collaborative Communicative English activities. The proposed relationship therefore requires empirical testing.

P5: Collaborative engagement positively influences students' perceived usefulness of CIDOS.

P6: Perceived Usefulness → Behavioural Intention

Perceived usefulness is a central determinant of technology acceptance. Users are more likely to intend to use a system when they believe that it helps them perform relevant tasks or achieve desired outcomes. Davis et al. (1989) found that perceived usefulness strongly influenced users' behavioural intentions, providing empirical support for this relationship within the TAM tradition.

In the context of CIDOS, usefulness concerns the extent to which students perceive the platform as supporting their Communicative English learning. This may include accessing course materials, completing learning activities, communicating with lecturers and peers, receiving feedback, and managing course-related tasks. When students see CIDOS as contributing meaningfully to these learning activities, they are more likely to intend to continue using the platform.

P6: Perceived usefulness positively influences students' behavioural intention to use CIDOS.

P7: Perceived Ease of Use → Behavioural Intention

Perceived ease of use may also contribute directly to behavioural intention. A system that is straightforward to learn and operate presents fewer practical barriers to continued use. Davis et al. (1989) found a small but significant effect of perceived ease of use on behavioural intention, although this effect became weaker over time.

This qualification is important to the proposed framework. Ease of use may encourage students' initial willingness to use CIDOS, but its influence may become less important as students gain experience with the platform. The framework nevertheless retains the direct relationship because students who perceive CIDOS as requiring less effort may be more willing to use it for their learning activities.

P7: Perceived ease of use positively influences students' behavioural intention to use CIDOS.

P8: Behavioural Intention → Actual Use

Behavioural intention represents a user's willingness or plan to engage with a technological system and is commonly treated as an important precursor to subsequent technology use. In their longitudinal investigation, Davis et al. (1989) found that behavioural intention was associated with later system use, with the intention–usage relationship becoming stronger when intention was measured closer to the subsequent behaviour.

Applied to CIDOS, students who express stronger intentions to use the platform would be expected to engage with it more consistently for course-related activities. Actual use may include accessing learning materials, submitting tasks, participating in online activities, or using other functions provided through CIDOS. However, actual use may also be affected by factors such as course requirements, lecturer practices, access to technology, and whether particular CIDOS functions are made available to students. The proposed relationship should therefore be examined empirically rather than assumed.

P8: Students' behavioural intention to use CIDOS positively influences their actual use of the platform.

Taken together, the eight propositions position perceived ease of use and perceived usefulness as the main mechanisms through which technological and pedagogical conditions may contribute to students' acceptance of CIDOS. The proposed paths are intentionally selective rather than exhaustive. Alternative relationships may also be plausible; for example, facilitating conditions may influence actual use directly, while social interaction may affect perceived ease of use rather than perceived usefulness. These alternatives are not included in the present framework in order to maintain theoretical parsimony, but they provide useful directions for future empirical testing.

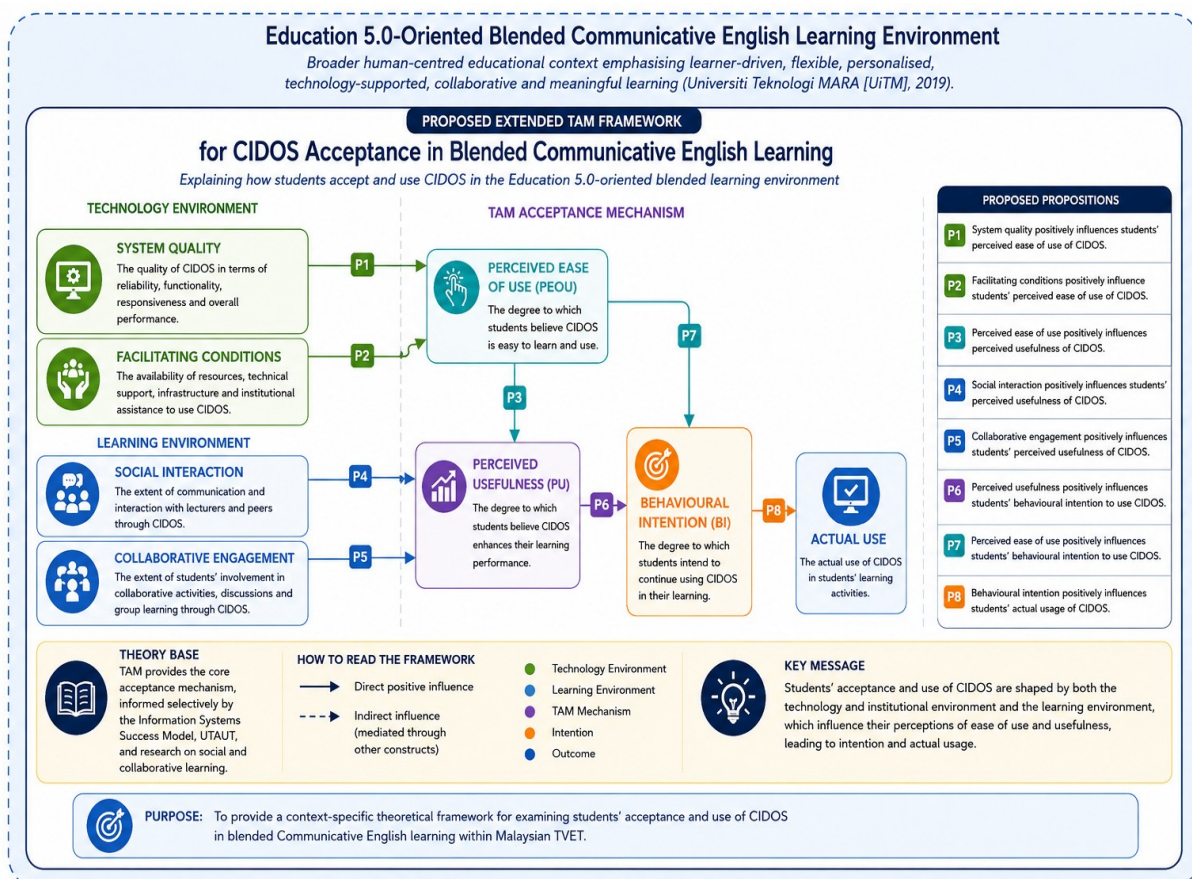


Figure 1. Proposed Extended TAM Framework for CIDOS Acceptance in Blended Communicative English Learning

8. Positioning Against Existing Models

The proposed framework is positioned primarily within TAM while drawing selectively on constructs associated with other established perspectives. TAM explains technology acceptance mainly through perceived ease of use and perceived usefulness and their relationship with subsequent acceptance and use (Davis, 1989; Davis et al., 1989). However, these core constructs do not explicitly represent all the technological, institutional, and pedagogical conditions that may shape students' experiences of an LMS in a communicative language-learning environment.

The Information Systems Success Model offers a complementary perspective by recognising system quality as an important dimension of information-system success (DeLone & McLean, 2003). UTAUT similarly extends the technology-adoption perspective by incorporating facilitating conditions related to the organisational and technical infrastructure surrounding technology use (Venkatesh et al., 2003). In the present framework, these perspectives inform the technology-environment layer rather than replacing the TAM acceptance mechanism.

The learning-environment layer addresses a different aspect of the CIDOS experience. The Community of Inquiry framework, for example, emphasises social presence alongside cognitive and teaching presence in creating meaningful educational experiences in computer-mediated environments (Garrison et al., 2000). The present framework does not reproduce the Community of Inquiry model or treat social interaction as equivalent to social presence. Instead, this literature provides theoretical support for recognising the social dimension of technology-supported learning. Collaborative engagement is similarly included to represent students' active participation in shared learning activities, which is particularly relevant to the communicative orientation of language learning.

The proposed framework therefore does not claim new constructs or suggest that these theoretical perspectives have never been extended or combined. Rather, its contribution lies in purposefully configuring selected constructs around the TAM mechanism for a specific educational setting: CIDOS-supported blended Communicative English learning in Malaysian TVET. System quality and facilitating conditions represent the technology and institutional environment, while social interaction and collaborative engagement represent the learning environment. These factors are linked selectively to perceived ease of use and perceived usefulness before the framework proceeds through behavioural intention to actual use.

This configuration also distinguishes the framework from models that seek to explain technology adoption at a more general level. Its purpose is narrower: to provide a theoretically grounded explanation of how students may come to accept and use CIDOS when the platform functions not only as a technological system but also as part of a communicative learning environment. The proposed framework should therefore be regarded as a context-specific theoretical model that requires empirical validation rather than as a replacement for TAM, UTAUT, the Information Systems Success Model, or the Community of Inquiry framework.

9. Discussion of the Proposed Framework

The proposed framework offers a context-specific explanation of students' acceptance of CIDOS by considering technology acceptance within the broader learning environment in which the platform is used. Its central argument is that students' acceptance of an LMS may not depend solely on whether the technology is perceived as easy to use or useful. These perceptions may themselves be shaped by the quality of the system, the support available to students, and the extent to which the platform enables meaningful interaction and collaborative learning.

The framework distinguishes between two routes through which these contextual factors may operate. System quality and facilitating conditions are proposed to influence perceived ease of use because they relate primarily to the effort required to access and use CIDOS. Social interaction and collaborative engagement, in contrast, are proposed to influence perceived usefulness because their relevance lies primarily in the educational value students may obtain from the platform. This distinction provides a theoretical rationale for the selected paths rather than linking every external construct directly to every TAM variable.

The framework also highlights the particular nature of Communicative English learning. An LMS used in this context is not simply a repository for course materials or a mechanism for submitting assignments. Its educational value may also depend on how effectively it supports communication, feedback, participation, and collaborative language-learning activities. This provides the rationale for including social interaction and collaborative engagement alongside the technological and institutional factors represented by system quality and facilitating conditions.

At the same time, the proposed relationships should not be regarded as fixed or universally applicable. The importance of individual constructs may vary according to how lecturers use CIDOS, the types of activities incorporated into a course, students' previous experience with the platform, access to digital resources, and the extent to which CIDOS functions are used for interaction and collaboration. For example, social interaction may have limited relevance in a course where CIDOS is used mainly to distribute materials, whereas it may become more important when the platform supports discussion, peer feedback, or collaborative tasks. These contextual differences represent important boundary conditions for the framework.

Alternative relationships are also possible. Facilitating conditions, for example, may influence actual use directly rather than operating only through perceived ease of use, while social interaction may contribute to perceived ease of use as well as perceived usefulness. The framework does not exclude these possibilities. Instead, it proposes a parsimonious starting model that future empirical research can compare with alternative specifications.

9.1 Directions for Empirical Validation

The next stage of research should empirically examine the proposed framework among students using CIDOS in Malaysian polytechnics. The eight propositions can be translated into testable hypotheses, and the constructs can be measured using established scales adapted carefully to the CIDOS and Communicative English context. Particular attention should be given to

establishing the conceptual and empirical distinction between system quality and facilitating conditions, as well as between social interaction and collaborative engagement.

Structural equation modelling could then be used to examine the proposed relationships and the model's overall explanatory performance. Future research should also compare the proposed framework with plausible alternative models rather than testing only the hypothesised structure. For example, researchers could examine whether facilitating conditions are directly related to actual use or whether social interaction is more strongly associated with perceived ease of use than perceived usefulness.

Where feasible and ethically permissible, researchers should examine actual CIDOS use using behavioural indicators rather than relying exclusively on students' self-reported usage. Behavioural intention and actual use are related but distinct concepts, and longitudinal research would provide a stronger basis for examining whether students' stated intentions translate into subsequent platform use. The original technology-acceptance literature provides precedent for examining intention and later usage longitudinally (Davis et al., 1989).

Empirical validation should also consider whether the proposed relationships differ across courses, student groups, levels of study, or patterns of CIDOS implementation. Such research would help determine which aspects of the framework are specific to blended Communicative English and which may be applicable more broadly across Malaysian TVET learning environments.

10. Significance and Implications

The proposed framework has theoretical, pedagogical, and practical significance for understanding technology acceptance in blended Communicative English learning within Malaysian TVET.

10.1 Theoretical Significance

Theoretically, the framework extends TAM by positioning students' perceptions of ease of use and usefulness within the technological and pedagogical conditions surrounding CIDOS use. Rather than adding external constructs without a clear conceptual rationale, the framework distinguishes between factors associated with the technology environment—system quality and facilitating conditions—and those associated with the learning environment—social interaction and collaborative engagement. These factors are then connected selectively to the TAM mechanism according to their proposed theoretical roles.

The framework therefore contributes through contextualisation and theoretical configuration rather than by introducing new constructs. This aligns with the purpose of conceptual research, where theoretical contribution may emerge from reorganising and connecting existing theoretical knowledge to provide a clearer explanation of a particular phenomenon (Jaakkola, 2020).

10.2 Pedagogical Implications

Pedagogically, the framework suggests that students' acceptance of CIDOS should not be viewed solely as a matter of technical usability. In Communicative English learning, the platform's educational value may also depend on how it is incorporated into activities that support communication, feedback, participation, and collaboration. Consequently, lecturers may need to consider not only whether students can use CIDOS easily, but also whether its use contributes meaningfully to the course's communicative purposes.

This perspective is particularly relevant when CIDOS is used as part of blended learning. Simply providing learning materials through an LMS may not fully utilise its pedagogical potential. Where appropriate, lecturers could incorporate CIDOS-supported activities that encourage students to communicate, exchange feedback, participate in discussions, or collaborate on language-learning tasks. The framework does not assume that these practices will necessarily increase technology acceptance; rather, it identifies this relationship as an important area for empirical investigation.

10.3 Practical Implications

At the institutional level, the framework highlights both the quality of the technological system and the conditions surrounding its use. Reliable system performance, accessibility, appropriate technical infrastructure, and adequate user support may influence users' experiences of CIDOS, particularly as previous Malaysian polytechnic research has identified technical reliability, internet connectivity, training, and institutional support as important implementation issues (Hasnan & Mohin, 2021). The importance of system quality as a dimension of information-system success is well established in the Information Systems Success literature (DeLone & McLean, 2003).

For Malaysian polytechnics, the framework can therefore provide a basis for examining CIDOS implementation from both technological and educational perspectives. Future findings based on the proposed model could help institutions identify whether students' difficulties arise primarily from the platform itself, the resources and support available to them, or how CIDOS is incorporated into teaching and learning. Such evidence could inform decisions concerning technical support, lecturer development, course design, and the use of LMS functions to support blended Communicative English learning.

11. Conclusion

This conceptual paper proposes an extended TAM framework for understanding students' acceptance of CIDOS in blended Communicative English learning within Malaysian TVET. TAM provides the central acceptance mechanism through perceived ease of use, perceived usefulness, behavioural intention, and actual use, while system quality and facilitating conditions represent the technology and institutional environment, and social interaction and collaborative engagement represent the learning environment.

The framework contributes by bringing these constructs together in a theoretically purposeful configuration for CIDOS-supported Communicative English learning. Rather than proposing new constructs or replacing established technology-acceptance models, it provides a context-specific explanation of how technological, institutional, and pedagogical conditions may contribute to

students' perceptions and subsequent acceptance of the platform. Education 5.0 provides the broader human-centred educational context within which these relationships are situated rather than functioning as an additional causal variable.

Importantly, the eight propositions developed in this paper represent theoretically derived relationships rather than empirical findings. Their validity, relative strength, and applicability therefore remain to be established through empirical research. Future studies should test the proposed model among Malaysian polytechnic students, examine plausible alternative relationships, and determine whether the framework operates differently according to course design, patterns of CIDOS implementation, student characteristics, or other contextual conditions.

By connecting technology acceptance with the technological and communicative learning environments surrounding LMS use, the framework provides a foundation for future research on CIDOS in Malaysian TVET. Empirical validation can subsequently determine whether this configuration offers a useful explanation of students' acceptance and actual use of CIDOS in blended Communicative English learning.

12. References

- Al-Rahmi, W. M., Alias, N., Othman, M. S., Marin, V. I., & Tur, G. (2018). A model of factors affecting learning performance through the use of social media in Malaysian higher education. *Computers & Education*, 121, 59–72. <https://doi.org/10.1016/j.compedu.2018.02.010>
- Avci, H., & Adiguzel, T. (2017). A case study on mobile-blended collaborative learning in an English as a foreign language (EFL) context. *The International Review of Research in Open and Distributed Learning*, 18(7), 45–58. <https://doi.org/10.19173/irrodl.v18i7.3261>
- Calisir, F., Altin Gumussoy, C., Bayraktaroglu, A. E., & Karaali, D. (2014). Predicting the intention to use a web-based learning system: Perceived content quality, anxiety, perceived system quality, image, and the Technology Acceptance Model. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 24(5), 515–531. <https://doi.org/10.1002/hfm.20548>
- 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>
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35(8), 982–1003. <https://doi.org/10.1287/mnsc.35.8.982>
- DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19(4), 9–30. <https://doi.org/10.1080/07421222.2003.11045748>

Ebadi, S., & Raygan, A. (2023). Investigating the facilitating conditions, perceived ease of use and usefulness of mobile-assisted language learning. *Smart Learning Environments*, 10, Article 30. <https://doi.org/10.1186/s40561-023-00250-0>

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2–3), 87–105. [https://doi.org/10.1016/S1096-7516\(00\)00016-6](https://doi.org/10.1016/S1096-7516(00)00016-6)

Hasnan, M., & Mohin, M. (2021). Implementation of LMS-CIDOS in polytechnic English language classroom: Issues and challenges. *Asian Journal of University Education*, 17(4), 527–537. <https://doi.org/10.24191/ajue.v17i4.16253>

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, 10, 18–26. <https://doi.org/10.1007/s13162-020-00161-0>

Mat Daud, K. A., Azizan, A. T., & Deraman, M. S. (2013). Penerimaan dan kebolegunaan rekabentuk e-pembelajaran (CIDOS) sebagai sistem pengurusan pembelajaran di Politeknik Malaysia. *International Journal of Creative Future and Heritage (TENIAT)*, 1(2), 49–58. <https://doi.org/10.47252/teniat.v1i2.110>

Salimon, M. G., Sanuri, S. M. M., Aliyu, O. A., Perumal, S., & Yusr, M. M. (2021). E-learning satisfaction and retention: A concurrent perspective of cognitive absorption, perceived social presence and the Technology Acceptance Model. *Journal of Systems and Information Technology*, 23(1), 109–129. <https://doi.org/10.1108/JSIT-02-2020-0029>

Universiti Teknologi MARA. (2019). *UiTM academic compass: Navigating the future: Education 5.0@UiTM*. UiTM Press.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>

Empowering Responsible Digital Citizens: A Survey of Cyberbullying and Digital Citizenship Among Adolescents in an International School

Submitted to the 7th International Conference on Technology-Enhanced Language Learning (ICTELL 2026)

Track: Digital Pedagogy and Online Learning; Teacher Education and Future Learning Environments

Katherine Chong Pui Kei¹, Shanti Ramalingam², Syeda Rabia Tahir³

Faculty of Education and Humanities, UNITAR International University, Petaling Jaya, Malaysia

¹Corresponding author

TELL023

ABSTRACT

As schools become increasingly reliant on digital platforms for learning and communication, the responsible and humane use of technology has become central to learner wellbeing. This exploratory, single-site mixed-methods study examined cyberbullying involvement, the forms it takes, adolescents' awareness of responsible digital citizenship, and teacher-leaders' school-based responses within an international school context. Quantitative data were gathered through the Adolescent Online Behaviour and Digital Practices Survey administered to 30 students from Years 7 to 11, with descriptive statistics computed in IBM SPSS (v.24). Qualitative data were drawn from semi-structured interviews with two purposively selected teacher-leaders and analysed using Braun and Clarke's six-phase thematic analysis. The findings indicate that cyberbullying is present across victim, perpetrator, and bystander roles, with hostile messaging, rumour-spreading, online exclusion, sharing of private content, and impersonation among the reported forms. Students demonstrated strong awareness of responsible online conduct, yet confidence in reporting incidents to teachers was notably weaker, with the majority of respondents undecided. Teacher-leaders consistently framed cyberbullying as a wellbeing and safeguarding concern, described detection as dependent on student reporting and digital evidence, and identified inconsistent policy implementation as a central challenge. The study contributes contextualised insight into digital citizenship and cyberbullying in international schooling and offers practical implications for humanising technology-rich learning environments through clearer reporting pathways, coordinated whole-school responses, and sustained digital citizenship education. Because collaborative and communicative technology-enhanced learning, including language learning, presupposes psychological safety, the findings are of direct relevance to educators designing digitally mediated classrooms.

Keywords: *cyberbullying; digital citizenship; adolescents; international school; teacher-leader perspectives; whole-school approach*

1. Introduction

Digital technology now occupies a central place in the learning and social lives of adolescents. Rising rates of home internet access and the rapid expansion of social media have made online communication an everyday feature of schooling, exposing students to both enriching and harmful experiences (Seçkin-Kapucu et al., 2021). Within technology-enhanced and increasingly humanised learning environments, students collaborate through digital platforms, sustain peer relationships through messaging, and engage with content well beyond the classroom. As schools deepen their reliance on technology for instruction and communication, they assume a corresponding responsibility to cultivate the safe, respectful, and ethical use of digital tools.

Responsible digital citizenship has become an organising concept for this responsibility. Digital citizenship encompasses not only technical competence but also online conduct that reflects respect, empathy, accountability, and an awareness of the consequences of one's actions. Schools shape digital citizenship by embedding it in the curriculum, establishing behavioural norms, and modelling consistent adult expectations. When these values are not reinforced consistently, harmful online behaviours can become normalised among learners.

Cyberbullying is among the most pressing threats to adolescent wellbeing in connected learning environments. It may involve the spread of false information, exclusion, harassment, or impersonation, and it can originate inside or outside school while its effects frequently re-enter school life through emotional distress, strained peer relationships, and reduced academic engagement. Because the boundary between school and online space is porous, schools cannot treat cyberbullying as a purely external matter. Teachers occupy a pivotal position in prevention and response, yet their actions are shaped by their confidence, the availability of training, the clarity of school policy, and parental expectations (Flôres et al., 2022).

The international school setting offers a distinctive context for this inquiry. International schools typically enrol culturally diverse student populations, producing varied communication styles, help-seeking tendencies, and disciplinary expectations that can complicate how cyberbullying is understood and addressed (Dichek et al., 2021). Much existing research foregrounds student experience while under examining teacher perspectives and the school-based responses that determine whether policy translates into practice. This study addresses that gap by integrating adolescents' self-reported experiences with teacher-leaders' accounts of detection, response, and prevention. In doing so it speaks directly to the conference theme of humanising

technology and empowering minds, since safe and respectful digital environments are a precondition for the kind of confident, collaborative, technology-mediated learning that language and content classrooms increasingly depend upon. In technology-enhanced language classrooms in particular, learners practise and expose their developing language through the same channels in which cyberbullying occurs: class group chats, shared documents, discussion forums, and video-based speaking tasks. Where hostile messaging, exclusion, or the circulation of screenshots goes unaddressed, learners become less willing to take the communicative risks that language development requires. Digital citizenship is therefore not peripheral to digital pedagogy but one of its enabling conditions, and the reporting pathways and whole-school responses examined here form part of the infrastructure on which humane, technology-mediated learning depends.

1.1 Purpose and Research Questions

The purpose of this study is to examine cyberbullying among adolescents in an international school and to analyse how teacher-leaders recognise, respond to, and seek to prevent it. By combining student survey data with teacher-leader interviews, the study aims to identify where students may need stronger guidance in responsible digital behaviour and where teachers may need clearer support in responding to incidents. Four research questions (hereafter RQ1 to RQ4) guided the study:

1. What is the prevalence of cyberbullying involvement among adolescents in an international school, including experiences as victims, perpetrators, and bystanders?
2. What forms of cyberbullying are most commonly experienced by adolescents in the school community?
3. What is the level of adolescents' awareness of responsible digital citizenship practices related to cyberbullying?
4. How do teacher-leaders perceive cyberbullying and respond to it through school-based strategies in the international school context?

Findings for RQ1 to RQ3 are reported in Sections 4.2 to 4.4 and findings for RQ4 in Section 4.5. The two strands are integrated in Section 4.6 before being discussed, question by question, in relation to the literature in Section 5.

2. Literature Review

2.1 Digital Citizenship and Responsible Online Behaviour

Digital citizenship frameworks position responsible online behaviour as a learnable disposition rather than a fixed trait. Programmes that make digital citizenship explicit have been associated with improved understanding of online norms and reduced harmful behaviour among adolescents (Bickham et al., 2021; Capuno et al., 2022). Yuniawati et al. (2024) further argue that strengthening digital citizenship can reduce cyberbullying through its influence on learning outcomes and self-regulation. These accounts converge on a shared premise that empathy, accountability, and awareness of consequences are central to safe participation in digital learning spaces (Silke et al., 2021).

2.2 Cyberbullying Through a Whole-School and Social-Ecological Lens

Cyberbullying is rarely an isolated act between two individuals. Social-ecological and whole-school perspectives situate it within overlapping influences spanning peers, classrooms, school policy, family, and the wider online environment (Ponsford et al., 2022). Whole-school approaches that engage the entire community have repeatedly been identified as more effective than isolated disciplinary measures (Kidwai & Smith, 2024; Bell et al., 2025). Research also documents substantial overlap and role continuity between traditional bullying and cyberbullying, with students frequently moving between victim, perpetrator, and bystander positions across physical and digital settings (Pennell et al., 2020; Sorrentino et al., 2023).

2.3 Prevalence, Forms, and Impact

International evidence points to non-trivial prevalence among adolescents, with some surveys reporting that a majority have encountered cyberbullying in some form (Bali et al., 2023; Zhu et al., 2021). Reported forms commonly include hostile messaging, rumour-spreading, deliberate exclusion, sharing of private content without consent, and impersonation through fake accounts (Malik et al., 2024; Kowalski & Wingate, 2023). The consequences extend to emotional distress, anxiety, reduced confidence, and disengagement from school, with longitudinal work linking victimisation to enduring mental health difficulties (Carvalho et al., 2021; Lee et al., 2025). Within Malaysia, recent work confirms meaningful prevalence among secondary students and situates responses within a developing legal framework (Saruddin et al., 2025; Razali & Nawang, 2022).

2.4 Teacher Perceptions and School-Based Responses

Teachers' willingness and capacity to intervene are shaped by knowledge, self-efficacy, and institutional support. Fischer et al. (2021) report that teacher self-efficacy is a consistent predictor of intervention, while Iorga et al. (2023) document the beliefs and coping strategies that explain why teachers are not always willing to act. Several studies emphasise that detection often depends on student disclosure and digital evidence, and that the publicity and perceived severity of an incident influence how teachers respond (Macaulay et al., 2021; Masoumi et al., 2025). Bystander research highlights that supportive peer responses can interrupt harm, although fear of becoming a target and low confidence in adult follow-through frequently inhibit action (Ferreira et al., 2020; Wong et al., 2021).

2.5 Research Gap and Rationale

Taken together, the literature establishes that effective responses to cyberbullying require alignment between student awareness, peer norms, teacher preparedness, and consistent policy implementation. Yet relatively few studies integrate student and teacher perspectives within a single international school context, where cultural diversity and varied parental expectations add further complexity. The present study addresses this gap by examining both adolescents' experiences and teacher-leaders' school-based responses, thereby generating contextualised insight to inform digital citizenship education and whole-school prevention. The contribution of the study is threefold. First, it extends Malaysian cyberbullying research, which has concentrated on public secondary schools (Saruddin et al., 2025), into the international school sector, where cultural diversity and varied parental expectations shape help-seeking and school response. Second, it positions teacher-leaders as safeguarding gatekeepers whose detection and response practices determine whether digital citizenship policy is enacted, a perspective rarely captured alongside student data in the same setting. Third, by integrating the two strands, it is able to identify where intervention effort is best directed, distinguishing between what students already know and what the school's reporting and response systems enable them to do.

3. Methodology

3.1 Research Design

The study adopted a convergent mixed-methods design, integrating quantitative and qualitative strands to capture both the measurable patterns and the contextual meanings of cyberbullying in the school (Sharma et al., 2023; Skamagki et al., 2024). The quantitative strand documented the prevalence, forms, and digital citizenship awareness reported by students, while the qualitative strand explored how teacher-leaders interpret and respond to cyberbullying. The combination was chosen because cyberbullying is multidimensional and requires both statistical description and situated explanation. The strands were brought together at the interpretation stage, where interview themes were used to help explain survey patterns. The study is exploratory and confined to a single international school; it aims to generate contextualised insight rather than population estimates, and the descriptive statistics reported should be read in that light.

3.2 Participants and Sampling

Purposive sampling was used for both strands. Thirty adolescent students from Years 7 to 11 were recruited for the survey on the basis of their age range and their active use of digital communication platforms, ensuring relevance to the research focus (Ahmad & Wilkins, 2025). Three inclusion criteria were applied: enrolment in Years 7 to 11 at the participating school, an age of between 11 and 16 years, and regular use of at least one digital communication or social media platform. Six students were drawn from each year level so that the sample spanned the secondary age range evenly rather than concentrating on a single cohort, and a target of 30 was set as a manageable size for an exploratory, single-site study using descriptive analysis. The survey invitations were distributed through form tutors and from they also coordinated for the parental consent. To reduce social desirability bias, participation was voluntary and responses were anonymous (Negri-Ribalta, 2022). For the qualitative strand, two teacher-leaders were purposively selected for their combined teaching and leadership roles and their direct involvement in pastoral care, discipline, and digital policy. Two were selected, rather than more, because the aim was to capture the perspective of staff with formal responsibility for policy and response rather than the full range of teacher views. Although small, this sample was appropriate for the aim of depth rather than statistical generalisation (Braun & Clarke, 2021), and its limits are acknowledged in Section 5.2. The participants are referred to throughout as TL1 and TL2 to protect confidentiality.

3.3 Instruments and Data Collection

The quantitative instrument, the Adolescent Online Behaviour and Digital Practices Survey, was developed from established models of cyberbullying and digital citizenship. It comprised three sections. Section A collected demographic and digital-use information (year level, gender, daily time online outside schoolwork, platforms used, and frequency of group-chat activity). Section B measured cyberbullying involvement. Victimization items asked how often respondents had received intentionally hurtful messages, had rumours spread about them, been deliberately excluded from an online group, or had embarrassing or private content shared about them; perpetration items asked how often respondents had sent threatening or aggressive messages, spread rumours online, excluded someone from a group chat, posted embarrassing images or videos, shared screenshots without permission, or created a fake account to impersonate someone; and bystander items asked whether respondents had witnessed someone being bullied online and how they had responded. Involvement items were scored on a four-point frequency scale (0 = never, 1 = once, 2 = a few times, 3 = many times), and “some experience” or “some involvement” in Tables 2 and 3 denotes any response other than 0. Section C measured awareness of responsible digital citizenship and confidence in reporting through statements rated on a five-point Likert scale from strongly disagree to strongly agree (Joshi & Joshi, 2021). The instrument was pilot tested with a small number of students to confirm clarity, and minor wording adjustments were made before full administration (DeVellis & Thorpe, 2021). The survey was completed anonymously during school hours and took approximately fifteen to twenty minutes.

The qualitative data were collected through semi-structured interviews guided by an interview schedule covering teacher knowledge, prevention strategies, reporting and intervention procedures, and policy implementation (Adeoye et al., 2021). Interviews were conducted privately at the school, audio-recorded with consent, and lasted approximately thirty to forty-five minutes.

3.4 Data Analysis

Survey responses were analysed in IBM SPSS (v.24) using descriptive statistics, with frequencies and percentages computed for each item to summarise patterns of involvement, the forms of cyberbullying experienced, and levels of digital citizenship awareness (Weil, 2025). Interview data were analysed thematically following Braun and Clarke's (2006) six-phase procedure: familiarisation, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report (Braun & Clarke, 2021).

3.5 Trustworthiness and Ethical Considerations

Trustworthiness in the qualitative strand was supported through audio-recording for transcription accuracy, the use of illustrative extracts, and cross-case comparison of the two teacher-leaders (Naeem et al., 2023). Ethical approval procedures were observed, including obtaining school administration consent, ensuring voluntary and anonymous student participation, and securing informed consent from interview participants. No personally identifying information is reported, and pseudonymous labels are used for all participants.

4. Findings

4.1 Participant Profile

The 30 respondents were evenly distributed across Years 7 to 11 (six per year level, 20.0% each), with ages ranging from 11 to 16. The sample was balanced by gender, comprising 15 female respondents (50.0%), 14 male respondents (46.7%), and one respondent (3.3%) who preferred not to disclose. Digital exposure was high: half of the respondents (50.0%) reported spending three to four hours online daily outside schoolwork, and a further 16.7% reported more than six hours. WhatsApp (56.7%), Instagram (50.0%), and TikTok (43.3%) were the most frequently used platforms, and 46.7% described themselves as active in group chats many times each day. This profile confirms that the sample inhabits intensive, peer-oriented digital environments relevant to the study.

Table 1

Participant Profile and Digital Engagement (n = 30)

Characteristic	Category	n	%
Year level	Years 7 to 11 (6 each)	30	100.0
Gender	Female	15	50.0
	Male	14	46.7
	Prefer not to say	1	3.3
Daily time online	3 to 4 hours	15	50.0
	1 to 2 hours	8	26.7
	More than 6 hours	5	16.7
	Less than 1 hour	2	6.7
Most used platform	WhatsApp	17	56.7
	Instagram	15	50.0
	TikTok	13	43.3

Note. Platform use was a multiple-response item, so percentages do not sum to 100.

4.2 Research Question 1: Prevalence of Cyberbullying Involvement

Cyberbullying involvement was evident across all three roles. For victimisation, 73.3% of respondents reported receiving intentionally hurtful messages at least once, 50.0% reported being the target of rumours spread through digital platforms, 56.7% reported deliberate exclusion from online groups, and 60.0% reported having embarrassing or private content shared about them. Bystander exposure was also common: most respondents reported having witnessed someone being bullied online. Responses about intervening were mixed, with a substantial proportion reporting that they had not acted and a similar proportion reporting either supportive behaviour or reporting the incident to an adult [authors: insert the exact percentages for witnessing and for each bystander response]. Self-reported perpetration was lower than victimisation and bystander exposure, which is consistent with the under-reporting commonly observed for socially undesirable behaviours (Shahaf-Oren et al., 2021). Taken together, and in answer to RQ1, the pattern indicates that cyberbullying affects a meaningful subgroup of students and is best treated as a wellbeing and safeguarding concern rather than a series of isolated incidents.

Table 2

Self-Reported Victimisation Experiences (n = 30)

Victimisation experience	Some experience (codes 1 to 3) n	%
Received intentionally hurtful messages	22	73.3
Embarrassing or private content shared	18	60.0
Deliberately excluded from an online group	17	56.7
Rumours spread through digital platforms	15	50.0

Note. Code 0 indicates no reported experience; codes 1 to 3 indicate increasing frequency.

4.3 Research Question 2: Forms of Cyberbullying

When asked about engagement in specific forms of cyberbullying, most students selected the lowest code, yet a consistent minority reported involvement across each form. Sharing screenshots without permission and posting embarrassing images or videos emerged among the more frequently reported behaviours, alongside sending threatening or aggressive messages (40.0% reporting some involvement), spreading rumours online (36.7%), and excluding others from group chats (33.3%). Creating fake accounts to impersonate others was the least common form, reported by 16.7% of respondents. The prominence of privacy-related behaviours, such as sharing screenshots and private content, indicates that adolescents may not always classify these acts as cyberbullying, particularly when they are framed as jokes or as group norms (Reneses Botija et al., 2025). In answer to RQ2, and reading Tables 2 and 3 together, the most commonly experienced

forms were hostile or threatening messaging, the sharing of embarrassing or private content, exclusion from online groups, and rumour-spreading, while impersonation through fake accounts was the least common.

Table 3

Reported Involvement in Forms of Cyberbullying (n = 30)

Form of cyberbullying	Some involvement (codes 1 to 3) %
Sending threatening or aggressive messages	40.0
Spreading rumours online	36.7
Excluding someone from a group chat	33.3
Posting embarrassing images or videos	30.0
Sharing screenshots without permission	53.3
Creating fake accounts to impersonate someone	16.7

4.4 Research Question 3: Awareness of Responsible Digital Citizenship

Students demonstrated strong digital citizenship awareness. A large majority understood what cyberbullying means (93.3% agreed or strongly agreed), recognised that sharing private content without permission is wrong (90.0%), acknowledged that online behaviour can affect others emotionally (83.3%), and agreed that people should take responsibility for what they post online (80.0%). Awareness, however, did not extend evenly to reporting confidence. Only 36.7% of respondents felt confident reporting cyberbullying to a teacher, while 53.3% were undecided. Trust that teachers would take reports seriously was moderate, with 60.0% in agreement, 26.7% neutral, and 13.3% in disagreement. In answer to RQ3, awareness of responsible digital citizenship is therefore high, but it coexists with weak reporting confidence. This gap between conceptual awareness and reporting self-efficacy is the most consequential quantitative finding, since strong norms do not automatically translate into help-seeking behaviour.

Table 4

Digital Citizenship Awareness and Reporting Confidence (n = 30)

Item	Agree or strongly agree %	Neutral %
I know what cyberbullying means	93.3	n/a
Sharing private content without permission is wrong	90.0	n/a
Online behaviour can affect others emotionally	83.3	n/a
People should take responsibility for what they post	80.0	n/a
I believe teachers take reports seriously	60.0	26.7
I feel confident reporting cyberbullying to a teacher	36.7	53.3

4.5 Research Question 4: Teacher-Leader Perspectives

Thematic analysis of the two teacher-leader interviews produced six themes that explain how cyberbullying is understood and managed at the school level.

Theme 1: Cyberbullying as a wellbeing and safeguarding concern. Both teacher-leaders framed cyberbullying as a duty-of-care issue rather than merely a disciplinary matter, foregrounding the protection of the affected child. This framing positions emotional impact and student safety at the centre of the school's response.

Theme 2: A blurred boundary between online life and school responsibility. Participants observed that many incidents originate outside school hours on platforms such as WhatsApp, yet still disrupt school relationships, obliging the school to respond to harm it did not directly witness. Frequent platform shifts were seen to make prevention more difficult.

Theme 3: Detection depends on reporting and evidence. Teacher-leaders described student disclosure as the primary entry point for any response and noted that screenshots and other digital evidence are essential for investigation, while private digital spaces limit visibility.

Theme 4: A multi-layered response pathway. Responses combined investigation and verification, proportionate consequences, counselling and pastoral support, and parental engagement to reinforce expectations beyond the school gate.

Theme 5: Policy exists, but consistency is the key challenge. Both participants identified gaps between policy intention and uniform staff implementation, and stressed that policy must keep pace with continuously evolving digital behaviour.

Theme 6: Balancing accountability, learning, and protection. Teacher-leaders sought to correct behaviour while educating and supporting students, escalating responses according to severity rather than defaulting to punishment.

Taken together, these themes answer RQ4. Teacher-leaders perceive cyberbullying as a safeguarding responsibility that extends beyond the school gate, respond through a layered pathway that depends on student disclosure and digital evidence, and identify consistent implementation of existing policy, rather than the absence of policy, as the principal obstacle to effective response.

Table 5

Summary of Teacher-Leader Themes

Theme	What the theme captures
1. Wellbeing and safeguarding	Cyberbullying framed as duty of care, not only discipline
2. Blurred online and school boundary	Off-campus incidents still disrupt school relationships
3. Detection via reporting and evidence	Student disclosure and screenshots drive investigation
4. Multi-layered response pathway	Investigation, consequences, counselling, and parents
5. Policy versus consistency	Gaps between policy intent and uniform staff practice
6. Accountability with education	Correcting behaviour while supporting and protecting students

4.6 Mixed-Methods Integration

The interview themes help to explain the survey patterns. The students' weak reporting confidence aligns with the teacher-leaders' account of detection as dependent on disclosure, suggesting that low confidence directly constrains the school's capacity to detect and respond. Similarly, the high awareness yet uneven labelling of privacy-related behaviours corresponds to the teacher-leaders' emphasis on education alongside accountability. The convergence of the two strands strengthens the interpretation that the principal leverage point is not awareness, which is already strong, but the reporting pathway and the consistency of adult response. This integrated reading of RQ1 to RQ4 frames the discussion that follows.

5. Discussion

In relation to RQ1 and RQ2, the findings indicate that cyberbullying is present in the school across victim, perpetrator, and bystander roles, echoing evidence of role continuity and the porous boundary between physical and digital spaces (Cabrera et al., 2024; Sorrentino et al., 2023). The comparatively high level of bystander exposure is significant, because bystanders are decisive in whether harm escalates or subsides (Ferreira et al., 2020). The mixed willingness to intervene observed here suggests that intervention is constrained by familiar barriers, including fear of becoming a target and uncertainty about adult follow-through (Wong et al., 2021). Lower self-reported perpetration should be interpreted cautiously, since social desirability and the reframing of harmful acts as jokes can mask involvement (Reneses Botija et al., 2025).

In relation to RQ3 and RQ4, the most actionable finding concerns the gap between awareness and reporting confidence. Students clearly understood digital citizenship norms, yet fewer than two in five felt confident reporting to a teacher. This dissociation indicates that knowledge alone is insufficient and that schools must invest in trustworthy, transparent reporting mechanisms and visible follow-through to convert awareness into help-seeking (Khairy et al., 2021). The teacher-

leaders' framing of cyberbullying as a safeguarding concern provides an enabling foundation for this work, but their identification of inconsistent implementation as the central challenge confirms that whole-school coherence is the decisive factor (Kidwai & Smith, 2024; Bell et al., 2025). Beyond the immediate setting, this pattern refines the whole-school literature by showing that, in a context where awareness is already high, the binding constraint on prevention is procedural rather than educational: what students know is less predictive of help-seeking than what they believe will happen after they report.

These results reinforce the conference's call to humanise technology. Empowering responsible digital citizens is not achieved by restricting access or by treating cyberbullying as an external nuisance, but by building learning environments in which students trust that disclosure will be met with consistent, caring, and proportionate adult responses. Such environments are foundational to all technology-enhanced learning, including language learning, where collaborative and communicative digital tasks presuppose psychological safety. In language classrooms specifically, peer feedback on shared documents, discussion-board writing, and recorded speaking tasks require learners to expose imperfect language to their peers; where screenshot-sharing or mockery goes unaddressed, learners withdraw from precisely the risk-taking that communicative competence requires. Digital citizenship education and dependable reporting pathways are therefore part of the pedagogical design of technology-enhanced language learning rather than an adjunct to it.

5.1 Practical Implications

For students, the priority is to strengthen reporting self-efficacy and bystander agency through guided practice in safe disclosure and supportive intervention. In practice, this means scenario-based digital citizenship lessons in which students rehearse concrete steps, such as preserving evidence, muting or blocking, and reporting through a named person or channel, together with peer-support roles that make bystander intervention a recognised and expected behaviour rather than an act of individual courage. For teachers, the findings argue for targeted professional development that builds confidence and consistency, supported by clear protocols for investigation, evidence handling, and escalation. A short written protocol that specifies who receives a report, how digital evidence is stored, how quickly the student will be acknowledged, and the thresholds for escalation to leadership or parents would give staff a common reference point and reduce the inconsistency that the teacher-leaders identified. For school leadership, the

central task is to close the gap between policy and practice through shared procedures, regular review that keeps pace with platform change, and transparent communication of how reports are handled. Concretely, leadership can publish a single anti-cyberbullying procedure in student-friendly language, brief all staff on it at the start of each year, review it annually against the platforms students are actually using, and report back to the school community, in anonymised form, on how incidents have been handled, so that students see that reporting leads to action. For parents, home-school partnership is needed to reinforce expectations across the porous boundary between school and online life (Helfrich et al., 2020). Termly parent briefings covering messaging etiquette, evidence-keeping, and the school's reporting route can align expectations across settings. Finally, for language and other technology-enhanced classrooms, group-chat and collaborative-platform norms can be co-constructed with students when a course begins and revisited whenever a new tool is introduced, embedding digital citizenship in the routines of learning rather than treating it as a separate topic.

5.2 Limitations and Future Research

Several limitations should be acknowledged, and they define the scope within which the findings should be read. This was an exploratory, single-site case study. The quantitative strand drew on a small purposive sample of 30 students ($n = 30$) from one international school, and the qualitative strand on two teacher-leaders ($n = 2$). The findings are therefore descriptive and context-bound: the percentages reported are indicative of patterns within this school and should not be read as prevalence estimates for international schools more broadly, for Malaysian adolescents, or for adolescents in general. With a sample of this size each respondent represents 3.3 percentage points, so small differences between items should not be over-interpreted. The two teacher-leader interviews capture the perspective of staff with formal responsibility for policy and response, not the full range of teacher views, and were analysed for depth rather than for thematic saturation. In addition, the survey instrument was newly developed and analysed descriptively at item level, so scale reliability was not established, and the cross-sectional, self-report design constrains causal inference and may be subject to social desirability bias, particularly for perpetration items. Future research could employ larger and multi-site samples across international schools, incorporate student and parent interviews, and adopt longitudinal designs to examine how reporting confidence and bystander behaviour change as digital citizenship education matures. Comparative work across cultural contexts would further illuminate how diversity shapes help-seeking and school response.

Larger samples would also permit inferential analysis of differences by year level and gender, and psychometric evaluation of the survey instrument.

6. Conclusion

This exploratory, single-site study examined cyberbullying involvement, its forms, adolescents' digital citizenship awareness, and teacher-leaders' school-based responses in an international school. Cyberbullying was present across victim, perpetrator, and bystander roles and took multiple forms, while students demonstrated strong awareness of responsible online conduct that was not matched by confidence in reporting. Teacher-leaders consistently framed cyberbullying as a safeguarding concern and identified inconsistent implementation as the principal obstacle to effective response. The study's contribution lies in integrating student and teacher-leader perspectives within an international school setting and in locating the decisive leverage point in reporting pathways and whole-school consistency rather than in awareness. Its originality lies in extending Malaysian cyberbullying research beyond the public school sector into international schooling, in treating teacher-leaders as safeguarding gatekeepers whose practices determine whether digital citizenship policy is enacted, and in showing that, where awareness is already high, the priority for technology-rich schools is to make reporting trustworthy and adult response consistent. Although the findings cannot be generalised beyond the participating school, they offer a transferable framework and a set of concrete practices that other international and technology-rich schools can adapt. Humanising technology-rich learning, the study concludes, depends on cultivating trust, coherence, and care so that students are genuinely empowered to participate responsibly and safely in digital life.

References

- Adeoye, M. A., et al. (2021). Semi-structured interviewing in qualitative educational research. *Journal of Educational Research and Practice*.
- Ahmad, S., & Wilkins, S. (2025). Purposive sampling in educational research: Principles and applications. *International Journal of Research and Method in Education*.
- Bali, D., Pastore, M., Indrio, F., Giardino, I., Vural, M., Pettoello-Mantovani, C., & Pettoello-Mantovani, M. (2023). Bullying and cyberbullying increasing in preadolescent children. *The Journal of Pediatrics*, 261.

- Bell, M., Keles, S., Furenes Klippen, M. I., Caravita, S. C. S., & Fandrem, H. (2025). Cooperation within the school community to overcome cyberbullying: A systematic scoping review. *Scandinavian Journal of Educational Research*, 1-16.
- Bickham, D. S., Moukalled, S., Inyart, H. K., & Zlokower, R. (2021). Evaluating a middle-school digital citizenship curriculum (Screenshots): Quasi-experimental study. *JMIR Mental Health*, 8(9), e26197.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications.
- Cabrera, M. C., Larrañaga, E., & Yubero, S. (2024). Bullying and cyberbullying in secondary education: A comparison between secondary schools in rural and urban contexts. *Child and Adolescent Social Work Journal*, 41(4), 617-631.
- Capuno, R., Suson, R., Suladay, D., Arnaiz, V., Villarin, I., & Jungoy, E. (2022). Digital citizenship in education and its implication. *World Journal on Educational Technology: Current Issues*, 14(2), 426-437.
- Carvalho, M., Branquinho, C., & de Matos, M. G. (2021). Cyberbullying and bullying: Impact on psychological symptoms and well-being. *Child Indicators Research*, 14(1), 435-452.
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE Publications.
- Dichek, N. P., Antonets, N. B., Shevchenko, S., Kopotun, I. M., & Pohribna, D. V. (2021). Implementation of the cross-cultural approach in the modern school. *International Journal of Learning, Teaching and Educational Research*, 20(4), 210-244.
- Ferreira, P. C., Simão, A. V., Paiva, A., & Ferreira, A. (2020). Responsive bystander behaviour in cyberbullying: A path through self-efficacy. *Behaviour & Information Technology*, 39(5), 511-524.
- Fischer, S. M., John, N., & Bilz, L. (2021). Teachers' self-efficacy in preventing and intervening in school bullying: A systematic review. *International Journal of Bullying Prevention*, 3(3), 196-212.
- Flôres, F. N., Visentini, D. M., Faraj, S. P., & Siqueira, A. C. (2022). Cyberbullying in the school context: Teachers' perceptions. *Psicologia Escolar e Educacional*, 26, e226330.
- Helfrich, E. L., Doty, J. L., Su, Y. W., Yourell, J. L., & Gabrielli, J. (2020). Parental views on preventing and minimizing negative effects of cyberbullying. *Children and Youth Services Review*, 118, 105377.

- Iorga, M., Agati, M., Coutinho, L., Gizir, R., Malinauskienė, D., Milios, P., & Colibaba, C. (2023). Why are teachers not always willing to fight against cyberbullying? *The Medical-Surgical Journal*, 127(1), 99-121.
- Joshi, A., & Joshi, R. (2021). Likert scale: Explored and explained. *Current Journal of Applied Science and Technology*.
- Khairy, M., Mahmoud, T. M., Abd-El-Hafeez, T., & Mahfouz, A. (2021). User awareness of privacy, reporting system and cyberbullying on Facebook. In *International Conference on Advanced Machine Learning Technologies and Applications* (pp. 613-625). Springer.
- Kidwai, I., & Smith, P. K. (2024). A content analysis of school anti-bullying policies in England: Signs of progress. *Educational Psychology in Practice*, 40(1), 1-16.
- Kowalski, R. M., & Wingate, V. S. (2023). Cyberbullying in children and adolescents. *Cyberbullying and Values Education*, 13-29.
- Lee, J., Choo, H., Zhang, Y., Cheung, H. S., Zhang, Q., & Ang, R. P. (2025). Cyberbullying victimization and mental health symptoms among children and adolescents: A meta-analysis of longitudinal studies. *Trauma, Violence, & Abuse*.
- Macaulay, P. J., Betts, L. R., Stiller, J., & Kellezi, B. (2021). The more public it is, the more severe it is: Teachers' perceptions on the roles of publicity and severity in cyberbullying. *Research Papers in Education*, 36(6), 726-753.
- Malik, M., Awang, H., Mansor, N. S., Zolkipli, M. F., Zaini, K. M., & Al-Mashhadani, A. F. S. (2024). Unveiling the shadows: The influence of anonymity and fake accounts on cyberbully intention in social media. *Iraqi Journal for Computer Science and Mathematics*, 5(4), 19.
- Masoumi, D., Bourbour, M., Edling, S., Gill, P., & Francia, G. (2025). School staff strategies for identifying, dealing with and preventing cyberbullying among Swedish primary school pupils. *Computers in the Schools*, 42(4), 366-387.
- Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods*, 22.
- Negri-Ribalta, C. (2022). Anonymity and response authenticity in sensitive survey research. *Survey Research Methods*.
- Pennell, D., Campbell, M., & Tangen, D. (2020). What influences Australian secondary schools in their efforts to prevent and intervene in cyberbullying? *Educational Research*, 62(3), 284-303.

- Ponsford, R., Falconer, J., Melendez-Torres, G. J., & Bonell, C. (2022). Whole-school interventions promoting student commitment to school to prevent substance use and violence: Synthesis of theories of change. *Health Education Journal*, 81(5), 614-637.
- Razali, N. A., & Nawang, N. I. (2022). An overview of the legal framework governing cyberbullying among children in Malaysia. *IIUM Law Journal*, 30, 207.
- Reneses Botija, M., Riberas Gutiérrez, M., & Bueno Guerra, N. (2025). It's just a joke: Gender, sexuality and trivialisation in adolescent online violence such as cyberhate, cyberbullying, and online grooming.
- Saruddin, M. Z., Fattah Azman, A. Z., Mohd Zulkefli, N. A., & Isa, M. R. (2025). Prevalence of cyberbullying and its associated factors among high school students in Klang District, Malaysia. *Malaysian Journal of Medicine & Health Sciences*, 21(3).
- Seçkin-Kapucu, M., Özcan, H., & Karakaya-Özyer, K. (2021). The relationship between middle school students' digital literacy levels, social media usage purposes and cyberbullying threat level. *International Journal of Modern Education Studies*, 5(2), 537-566.
- Shahaf-Oren, B., Madan, I., & Henderson, C. (2021). A lot of medical students, their biggest fear is failing at being seen to be a functional human: Disclosure and help-seeking decisions by medical students with health problems. *BMC Medical Education*, 21(1), 599.
- Sharma, S., et al. (2023). The strengths of mixed-methods research in social and educational inquiry. *International Journal of Social Research Methodology*.
- Silke, C., Brady, B., Boylan, C., & Dolan, P. (2021). Empathy, social responsibility, and civic behavior among Irish adolescents: A socio-contextual approach. *The Journal of Early Adolescence*, 41(7), 996-1019.
- Skamagki, G., King, A., Carpenter, C., & Wåhlin, C. (2024). The concept of integration in mixed methods research: A step-by-step guide using an example study in physiotherapy. *Physiotherapy Theory and Practice*, 40(2), 197-204.
- Sorrentino, A., Esposito, A., Acunzo, D., Santamato, M., & Aquino, A. (2023). Onset risk factors for youth involvement in cyberbullying and cybervictimization: A longitudinal study. *Frontiers in Psychology*, 13, 1090047.
- Weil, J. (2025). Descriptive statistics in educational research: Practical applications. *Educational Measurement and Evaluation Review*.
- Wong, R. Y. M., Cheung, C. M., Xiao, B., & Thatcher, J. B. (2021). Standing up or standing by: Understanding bystanders' proactive reporting responses to social media harassment. *Information Systems Research*, 32(2), 561-581.

- Yuniawati, E. I., Tiatri, S., & Beng, J. T. (2024). Strengthening digital citizenship behavior to reduce cyberbullying through learning outcome mediation. *ENLIGHTEN: Jurnal Bimbingan Konseling Islam*, 7(2), 82-108.
- Zhu, C., Huang, S., Evans, R., & Zhang, W. (2021). Cyberbullying among adolescents and children: A comprehensive review of the global situation, risk factors, and preventive measures. *Frontiers in Public Health*, 9, 634909.

Constructing a Sustainable ICH-STEAM Curriculum Architecture in Higher Education: An Expert Consensus via Delphi Method

Author Sun Yanhu¹, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia .E-mail:p123627@siswa.ukm.edu.my

Author Nuriah Abd Majid^{1*}, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia, E-mail:nuriah@ukm.edu.my

*** Corresponding Author** Nuriah Abd Majid .E-mail:nuriah@ukm.edu.my

Abstract

This study develops a parsimonious curriculum architecture for integrating intangible cultural heritage (ICH), STEAM and digital technologies in higher education without displacing living cultural authority. A three-round Delphi study with 18 China-based experts refined 44 Round 1 items into 53 operational items in Round 2 and a 52-item Round 3 structure. Forty-nine items were directly comparable across Rounds 2 and 3; one was substantially reformulated, three were deleted and two curriculum-design items were introduced only in Round 3 and are therefore treated as provisionally endorsed. The final structure contains seven value-objective, 13 curriculum-design, 16 pedagogical-implementation and 16 structural/ecosystem-barrier indicators. Panel endorsement favoured authentic problems, selective rather than equal STEAM integration, project/task-based learning, co-teaching, professional development and technology use justified by the learning task. Interpreted through safeguarding and critical-heritage scholarship, these findings support cultural accountability as a constraint on curricular and technological choices. The study establishes expert endorsement, not classroom effectiveness, language gains or safeguarding impact. Multimodal language education is proposed only as a testable transfer context.

Keywords: *cultural accountability; Delphi method; higher education; intangible cultural heritage; STEAM education*

1.Introduction

The 2003 UNESCO Convention frames safeguarding as maintaining the viability, transmission and recreation of living heritage with community participation rather than preserving a static cultural object (UNESCO, 2003; Kurin, 2004). Formal education can widen access to cultural knowledge, but it can also decontextualise practice or displace community authority when institutional interpretation becomes dominant (Aral, 2022; Smith, 2006; Waterton & Smith, 2010).

STEAM can organise interdisciplinary inquiry around authentic problems, yet disciplinary breadth alone does not establish integration quality (English, 2016; Perignat & Katz-Buonincontro, 2019; Chappell et al., 2025). Digital technologies likewise expand documentation and representation while selectively rendering embodied and tacit knowledge, so technology is not pedagogically or culturally neutral (Hou et al., 2022).

Language education is a secondary transfer context. Multiliteracies and digital multimodal composing (DMC) show how meaning can be made through linguistic, visual, audio and other

semiotic resources, but multimodal participation cannot be treated as evidence of L2 development without direct measurement (New London Group, 1996; Hafner, 2014; Zhang et al., 2023; Jiang & Hafner, 2024).

In this article, ICH-STEAM denotes higher-education learning in which an ICH practice or safeguarding problem provides the organising cultural content and STEAM disciplines are selectively mobilised as resources rather than treated as a fixed five-subject package. Curriculum architecture refers to the relationship among objectives, curriculum design, pedagogical enactment, assessment and enabling institutional conditions (Biggs, 1996). The primary research question is: What dimensions and indicators should constitute a culturally grounded, technology-enhanced ICH-STEAM curriculum architecture for higher education? A secondary conceptual objective is to identify testable implications for multimodal language education.

2.Literature Review and Conceptual Positioning

2.1 ICH, formal education, and cultural sustainability

Safeguarding and critical-heritage scholarship provide the normative starting point. ICH depends on continued practice, transmission, recreation and participation, while institutional definitions of value can privilege expert authority and misrecognise heterogeneous community claims (UNESCO, 2003; Smith, 2006; Waterton & Smith, 2010). Cultural accountability is therefore used here as an author-level interpretive constraint: curricular decisions should remain answerable to provenance, embodied practice, community-recognised meanings and relevant knowledge holders. It is not treated as a Delphi-rated latent construct.

2.2 STEAM integration, project learning, and curriculum alignment

STEAM research supports problem-responsive interdisciplinarity rather than compulsory equal representation of all five disciplines (English, 2016; Perignat & Katz-Buonincontro, 2019; Chappell et al., 2025). Constructive alignment supplies the minimum curriculum logic: intended outcomes, activities and assessment should reinforce one another (Biggs, 1996). For ICH, disciplinary resources should therefore enter because the cultural problem and learning objective require them, not because every STEAM label must be visible.

2.3 Technology-enhanced learning: technology choice and cultural accountability

TPACK links technology to content and pedagogy, while digital-heritage research shows that digitising embodied ICH is necessarily selective (Mishra & Koehler, 2006; Hou et al., 2022). The decision rule used here is parsimonious: technology should serve the learning task and have defensible implications for representation, attribution and cultural boundaries.

2.4 Multiliteracies, TESL, and digital multimodal composing

Multiliteracies broadens literacy beyond written language, and DMC research shows how language, digital literacy and multimodal design can be combined (New London Group, 1996; Hafner, 2014). Reviews nevertheless emphasise task-specific assessment and direct evidence of language development (Zhang et al., 2023; Jiang & Hafner, 2024). ICH may provide culturally consequential content for such tasks, but cultural material should not become decorative input for practising a dominant language (Davoodi, 2024).

2.5 Conceptual synthesis and study contribution

The study does not propose a new general theory. Its contribution is a parsimonious architecture that places cultural accountability around three familiar curriculum problems:

selecting disciplinary resources, aligning activities and assessment, and creating feasible implementation conditions. The empirical contribution is expert endorsement of the operational domains, not evidence that the architecture improves learning or safeguarding outcomes (Diamond et al., 2014).

2.6 Assumptions and alternative explanation

Three revisable assumptions guide the analysis: formal education supports ICH only under enabling cultural and institutional conditions; good STEAM integration does not require maximal disciplinary presence; and digital representation is selective rather than neutral (Aral, 2022; English, 2016; Hou et al., 2022). A simpler competing explanation is that the final architecture mainly restates good curriculum design. The narrower heritage-specific claim is that familiar mechanisms are constrained by cultural accountability; comparative implementation research is needed to establish any advantage over simpler alternatives.

3. Methodology

3.1 Study design, setting, and scope

A three-round Delphi design was used because the aim was to construct and refine a normative curriculum architecture from distributed expertise rather than estimate population parameters (Beiderbeck et al., 2021; Diamond et al., 2014).

The doctoral research was registered at Universiti Kebangsaan Malaysia (UKM), while all expert participants were based in China and recruited from universities, research institutes and ICH practice/protection organisations.

3.2 Expert selection and recruitment

Purposive sampling, supplemented by expert-network referral, recruited 18 panellists: six STEAM/educational-technology specialists, eight ICH researchers or curriculum-related scholars, and four ICH bearers/practitioners. Eligibility considered topic relevance, professional standing, experience, familiarity with the research problem and capacity to complete repeated rounds (Beiderbeck et al., 2021).

All 18 enrolled experts completed all three rounds. A complete pre-enrolment invitation denominator was not preserved, so no recruitment response rate is reconstructed. Because recruitment targeted people already engaged with ICH, STEAM or educational technology, the panel may have been more receptive to integration than a broader university population; consensus is therefore interpreted as panel endorsement rather than a population estimate.

Table 1. Delphi panel composition and selection rationale (full recruitment and screening audit in Supplementary Table S8)

Expert category	n	Rationale
STEAM / educational technology	6	STEAM education, learning science, maker education, and educational technology; provides technology-integration and interdisciplinary-pedagogy expertise.
ICH researchers / curriculum-related scholars	8	ICH education, heritage protection, cultural theory, curriculum and related research; two had Khoomei-related teaching/research expertise.
ICH bearers / practitioners	4	Recognised practitioners across ICH domains; provides practice-based cultural authority and feasibility judgement.

Expert category	n	Rationale
Total	18	Cross-sectoral panel; all 18 enrolled experts completed all three rounds.

3.3 Instrument development, item generation, and piloting

The Round 1 questionnaire was developed from targeted literature on ICH education, STEAM, curriculum alignment, technology integration and implementation barriers. Three individuals familiar with ICH and STEAM, but not Delphi panellists, reviewed comprehensibility, redundancy, response flow and wording; this preparatory review was not treated as psychometric validation.

Table 2. Initial item-generation traceability (full source and generation audit in Supplementary Table S8)

Initial domain	R1 items	Primary conceptual source	Function in item generation
DIISE: definition	3	Integrated STEM/STEAM definitions; living-heritage safeguarding; project/inquiry learning (UNESCO, 2003; English, 2016; Perignat & Katz-Buonincontro, 2019).	Established the conceptual boundary; treated as a Round 1 anchor rather than a repeatedly rated operational dimension.
IISPVO: value objectives	5	STEAM literacies, interdisciplinary problem solving, and heritage-transmission aims (Pillai & Achilles, 2015; Perignat & Katz-Buonincontro, 2019; Chappell et al., 2025).	Specified intended learner capacities and cultural-educational purposes.
IISCD: curriculum design	10	Interdisciplinary curriculum design, authentic/project learning, and constructive alignment (Biggs, 1996; English, 2016; Chappell et al., 2025).	Specified curriculum structure, task logic, interdisciplinarity, assessment, and cultural content.
IISPI: implementation	15	Project/problem-based and active learning, formative assessment, technology integration, and co-teaching (Black & Wiliam, 1998; Mishra & Koehler, 2006; Perignat & Katz-Buonincontro, 2019).	Specified enactment strategies and teacher/learner roles.

Initial domain	R1 items	Primary conceptual source	Function in item generation
DIISP: barriers	11	ICH formal-education and higher-education implementation constraints (Aral, 2022; Šantek et al., 2025; Zeng et al., 2025).	Specified teacher, curriculum, resource, policy, assessment, and stakeholder constraints.

Round 1 contained 44 rated items: three definitional statements and 41 operational items across value objectives, curriculum design, pedagogical implementation and barriers. Subsequent items were revised, merged, deleted or added through coded comments and controlled feedback.

3.4 Delphi rounds and controlled feedback

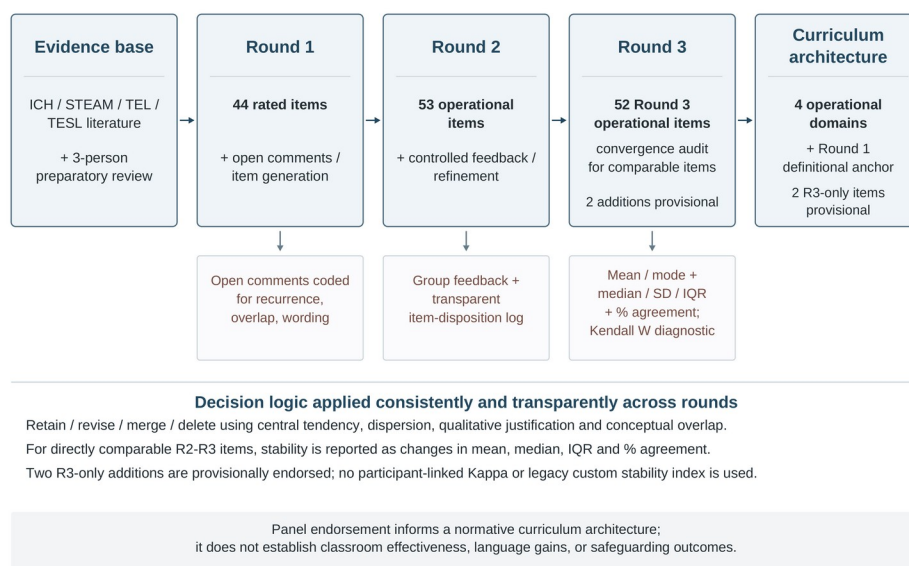
Table 3. Three-round procedure, retention, and item transitions (complete item results in Supplementary Tables S2-S4)

Round	Period	Valid	Items rated	Function
R1	9–31 Oct 2025	18/18	44: DIISE 3; IISPVO 5; IISCD 10; IISPI 15; DIISP 11	Concept clarification, baseline ratings, and open item generation. DIISE established as a definitional anchor.
R2	8–25 Nov 2025	18/18	53: IISPVO 8; IISCD 11; IISPI 17; DIISP 17	Controlled feedback; deletion, merging, and wording revision; recurring suggestions added; single-expert suggestions screened separately.
R3	6–20 Dec 2025	18/18	52: IISPVO 7; IISCD 13; IISPI 16; DIISP 16	Final-round rating; convergence audit for comparable items; two R3-only additions provisionally endorsed.

Open comments were coded by dimension, recurrence, conceptual overlap and wording problem, creating an auditable bridge from qualitative feedback to item decisions. Round 2 contained 53 operational items; Round 3 contained 52.

Later rounds supplied controlled group feedback on prior ratings and decisions. This supports iterative refinement but may also encourage conformity, so later-round agreement is interpreted as endorsement rather than proof that retained items are uniquely optimal (Beiderbeck et al., 2021). The three Round 1 definition statements were not re-rated and are treated only as a conceptual anchor.

Figure 1. Delphi evidence chain, item-generation pathway, and decision logic



3.5 Data analysis, consensus, concordance, and stability

Because Delphi has no universal consensus threshold, item decisions combined central tendency, dispersion, qualitative justification and transition rules (Diamond et al., 2014; von der Gracht, 2012). Mean >3.66 and/or mode = 5 made an item eligible for retention; mean >4.50 was treated only as strong descriptive support. Median, sample SD, IQR and the percentage of ratings at 4-5 were reported for every item.

Kendall's W was used only as a dimension-level rank-concordance diagnostic, not as an item-retention gate. For the 49 items directly comparable between Rounds 2 and 3, stability was described by changes in mean, median, IQR and percentage agreement (Holey et al., 2007).

Participant-linked Kappa and a legacy custom stability index were excluded because the deidentified submission archive did not permit reproducible reconstruction. Reproducibility took precedence over retaining unauditible statistics.

The previously reported Cronbach alpha of .810 described a historical four-attribute expert-screening rubric and was not used as evidence of Delphi consensus, instrument validity or expert quality.

Table 4. Statistical and decision criteria used in the revised Delphi analysis

Indicator	Purpose	Use in the revised article
Mean / mode	Item-level central tendency and study-specific retention rule (Diamond et al., 2014; von der Gracht, 2012).	Mean >3.66 and/or mode=5 made an item eligible for retention; mean >4.50 is a stricter descriptive category, not a universal Delphi rule.
Median / SD / IQR	Distribution and dispersion diagnostics (Holey et al., 2007; von der Gracht, 2012).	Used to identify heterogeneity and wording requiring clarification; not interpreted as interchangeable with consensus.

Indicator	Purpose	Use in the revised article
% agreement (ratings 4–5)	Absolute item-support descriptor used in Delphi reporting (Holey et al., 2007; Diamond et al., 2014).	Reported descriptively for every item in Supplementary Tables S2–S4.
Binary >66.7% support	Study-specific screening rule for single-expert candidates; not a universal cut-off (Diamond et al., 2014).	Eligibility rule for separately presented R2 refinement checks; it did not guarantee retention when overlap/conceptual consolidation applied.
Kendall’s W	Within-round relative rank concordance diagnostic (von der Gracht, 2012; Beiderbeck et al., 2021).	Reported by round/dimension in S7 as a diagnostic only; not an item-retention gate.
R2–R3 distributional stability	Between-round descriptive convergence without respondent linkage (Holey et al., 2007).	For 49 directly comparable items, reports Δ mean, Δ median, Δ IQR, and Δ agreement; full audit in S5.
Kappa / legacy custom stability index	Legacy diagnostics excluded where reproducibility could not be established (Beiderbeck et al., 2021).	Removed from revised inferential/retention claims because they cannot be reproducibly audited from the deidentified submission archive.
Cronbach α =.810	Historical internal-consistency statistic for the screening rubric; not consensus or validity evidence (Tavakol & Dennick, 2011).	Not used to establish expertise, instrument validity, or Delphi consensus.

3.6 Item-transition and reporting audit

All rated items and round-specific statistics are reported in the Supplementary Material. Between Rounds 2 and 3, 49 items were directly comparable, one curriculum-design item was substantially reformulated, three R2-only items were deleted and two curriculum-design items were introduced only in Round 3.

For the 49 directly comparable items, the median absolute mean change was 0.111; 42/49 medians were unchanged; the median absolute IQR change was 0; and 38/49 agreement proportions changed by no more than one expert-equivalent (5.6 percentage points). Because

the two R3-only items had no subsequent re-rating, they are treated as provisionally endorsed rather than longitudinally converged.

3.7 Ethics and transparency

Ethical approval was granted by the UKM Research Ethics Committee (JEP-2025-751; 9 October 2025-8 October 2028). All formal recruitment and data collection occurred after approval. Participants provided informed consent and data were deidentified and stored securely.

4. Results

4.1 Round-by-round development and transition audit

The Delphi process moved from conceptual clarification to operational refinement. Round 1 established a definitional anchor and generated revisions; Round 2 was the principal filtering and expansion stage; and Round 3 produced the reported 52-item operational structure.

The final structure comprises seven value-objective indicators, 13 curriculum-design indicators, 16 pedagogical-implementation indicators and 16 structural/ecosystem-barrier indicators. Two of the 13 curriculum-design indicators were introduced only in Round 3 and remain provisionally endorsed.

Table 5. High-value item transitions across the three Delphi rounds (complete audit in Supplementary Tables S5-S6)

Dimension	Issue / item family	Transition	Interpretation
DIISE	Three definition statements	Synthesised after R1	Retained as conceptual anchor; not re-rated and not described as a three-round consensus dimension.
IISPVO	Real-world problem solving; hands-on skills; knowledge transfer	Added in R2 from recurring comments	Retained in the final value-objective set. Career-interest/further-study item deleted after R2 because of lower salience and conceptual overlap.
IISCD	Equal/balanced integration; integration beyond fixed STEAM domains	Revised, merged, or reformulated	Final design emphasises problem-driven integration, multidimensional assessment, localisation, and interdisciplinary development.
IISPI	Technology-centred wording; standards-based equal-literacy statement	Substantially reframed / deleted	Technology retained only as pedagogically justified mediation; equal-literacy standards item deleted after R2 because of dispersion.

Dimension	Issue / item family	Transition	Interpretation
DIISP	Institutional/ecosystem barriers	Expanded and consolidated	Final set captures stakeholder alignment, social resources, facilities, incentives, assessment, policy, and risk of superficial implementation.

4.2 Final value objectives (IISPVO)

The seven value-objective indicators emphasised transferable interdisciplinary capability, including applying knowledge to real problems, rather than a narrow employability rationale.

4.3 Curriculum design (IISCD)

The 13 curriculum-design indicators supported authentic, student-centred problems, interdisciplinary development, creativity and meaningful assessment. The set favours selecting disciplinary resources according to the cultural problem and learning objective rather than requiring equal presence of all five STEAM domains. Two Round 3-only additions - multidimensional assessment and explicit attention to local/regional/institutional characteristics - require future re-rating.

4.4 Pedagogical implementation (IISPI)

The 16 pedagogical-implementation indicators strongly supported selective STEAM integration, project/task-based learning, learning by doing, cross-disciplinary co-teaching and teacher facilitation. A technology-centred formulation that was weak earlier in the process was retained only after technology was reframed as a pedagogically justified resource rather than an independent curricular end.

4.5 Structural and ecosystem barriers (DIISP)

The 16 barrier indicators identified institutional conditions rather than learner deficits. High-rated concerns included insufficient time, space, facilities and materials; lack of professional development; weak use of social and informal resources; assessment and policy constraints; and risks of superficial or symbolic implementation.

Table 6. Selected final-round item-level evidence (complete Round 3 results in Supplementary Table S4)

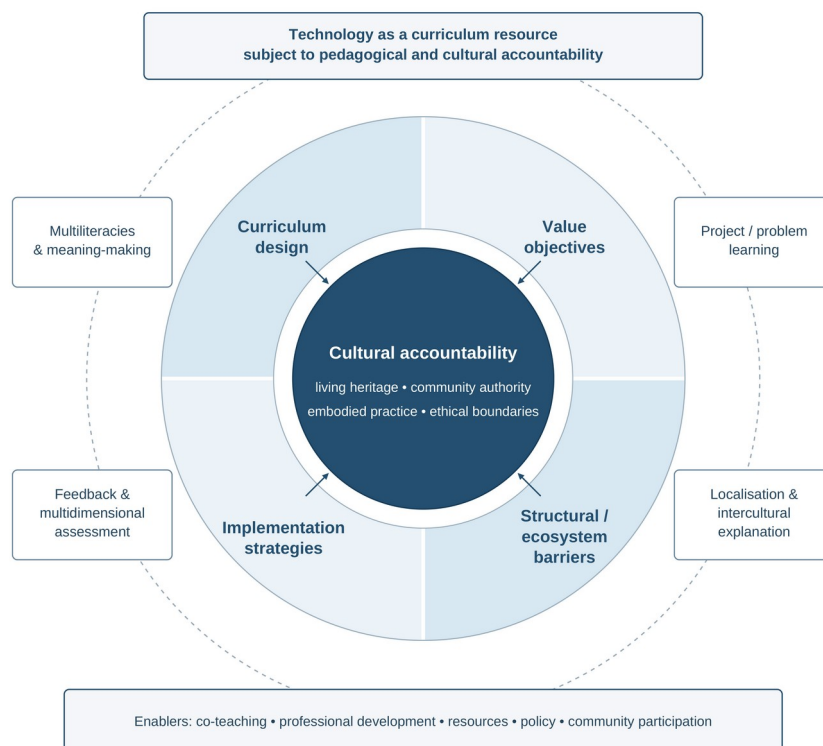
ID	Selected final item	R3 mean	Agree 4–5	Interpretive role
IISPVO-7	Apply interdisciplinary knowledge to real problems	4.78	94.4%	Transferable problem-solving capability.
IISCD-4	Curriculum tasks should be student-centred and oriented toward authentic, real-world problem-solving.	4.72	100.0%	Authentic task as a design principle.
IISCD-9	Learning outcomes should demonstrate design, innovation, cultural interpretation, and meaningful applications.	4.78	94.4%	Aligns creativity with meaningful assessment.

ID	Selected final item	R3 mean	Agree 4 –5	Interpretive role
IISPI-16	Unequal integration of the five STEAM disciplines is acceptable when cultural problems, learning objectives, and available expertise justify the selection of disciplines.	4.89	100.0%	Selective integration; rejects equal-coverage formalism.
IISPI-13	Project-based and task-driven learning should be used to organise extended ICH–STEAM activities.	4.83	100.0%	Project/task structure for enactment.
IISPI-2	Teaching should involve cross-disciplinary collaboration or co-teaching among heritage bearers, subject teachers and technology specialists.	4.83	100.0%	Expertise distributed across teachers and bearers.
DIISP-1 4	The lack of professional development for ICH–STEAM teachers may limit their capacity for interdisciplinary, technological and culturally responsive instruction.	4.89	100.0%	Institutional ecology and social resources matter.
DIISP-4	Time, space, facilities, and material resources may be insufficient to sustain project-based heritage learning in the long term.	4.94	100.0%	Practical/infrastructural capacity is a binding condition.
DIISP-1 2	The lack of localised and culturally appropriate ICH–STEAM teaching materials may constrain implementation.	4.67	100.0%	Localised, culturally appropriate teaching materials are an implementation condition.

4.6 Consensus-derived curriculum architecture

Figure 2 summarises the interpretation adopted here. Cultural accountability is an author-level constraint drawn from safeguarding and critical-heritage scholarship, not a fifth Delphi-derived latent variable. The four operational domains specify intended learner development, curriculum design, pedagogical enactment and implementation conditions. Technology sits within these domains as a resource whose use must be justified by both the learning task and the cultural implications of representation.

Figure 2. Delphi-derived operational domains interpreted through a cultural-accountability constraint



5. Discussion

5.1 What is original about the architecture?

The most economical interpretation is that the panel endorsed a culturally constrained version of familiar good curriculum design rather than a wholly new educational theory. Project learning, curriculum alignment, co-teaching, technology integration and professional development already have established rationales (Biggs, 1996; Mishra & Koehler, 2006; Perignat & Katz-Buonincontro, 2019). The heritage-specific contribution is narrower: these mechanisms are interpreted through cultural accountability so that disciplinary integration, technology, assessment and implementation remain answerable to living practice, provenance and relevant knowledge holders (UNESCO, 2003; Smith, 2006).

This interpretation addresses three recurrent risks: formalisation can displace living authority; STEAM can become disciplinary formalism; and technology can substitute representation for embodied practice (Aral, 2022; English, 2016; Hou et al., 2022). Relevant knowledge holders should not be assumed to represent a homogeneous community position; contested interpretations should remain visible (Waterton & Smith, 2010).

5.2 Technology use under cultural accountability

Technology is neither rejected nor celebrated. It is used when it serves the learning task and when its representation does not obscure provenance, context or authority (Mishra & Koehler, 2006; Hou et al., 2022). This also avoids treating authenticity as a fixed essence: curricular representations may adapt, but their relation to practitioners and cultural context should remain explicit (Smith, 2006).

5.3 Implications for TESL, applied linguistics, and multimodal learning

ICH-centred language tasks could combine disciplinary language, oral histories, performance, video, annotation or multimodal composition, making them plausible contexts for

intercultural explanation and DMC (Hafner, 2014; Jiang et al., 2024). This is a transfer hypothesis, not evidence of language development.

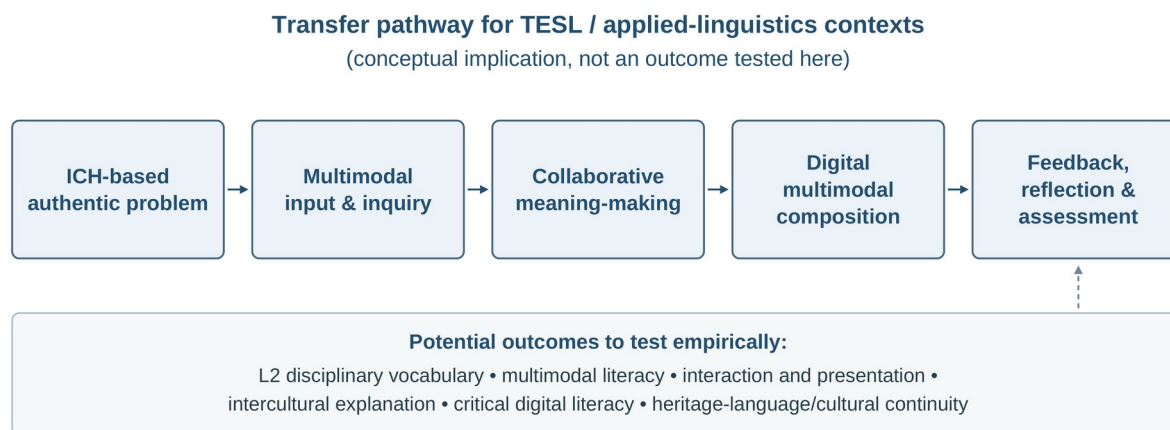
At minimum, cultural material should remain connected to relevant knowledge holders, and assessment should distinguish language/communication, disciplinary reasoning, multimodal design and cultural accountability unless a composite score is validated (Liu & Lam, 2026).

Table 7. Language-education transfer map and falsifiable validation targets

Architecture component	TESL / applied-linguistics design function	Future observable outcome	Status in this study
Cultural accountability + relevant knowledge holders	Select culturally consequential sources and involve people able to contextualise, attribute, or contest their educational use (UNESCO, 2003; Smith, 2006).	Intercultural explanation quality; provenance/attribution awareness; culturally accountable source use.	Theoretical transfer only; not measured.
Problem-driven selective integration	Organise content-based inquiry requiring purposeful disciplinary language and interaction (English, 2016; Chappell et al., 2025).	Disciplinary vocabulary, interactional competence, problem-explanation quality.	Theoretical transfer only; not measured.
Digital/multimodal activity	Use audio, video, visualisation, annotation, or composing tools when the communicative task requires them and examine what the representation foregrounds or obscures (Hou et al., 2022; Jiang & Hafner, 2024).	Multimodal literacy, critical digital literacy, revision quality.	Theoretical transfer only; not measured.
Aligned multidimensional assessment	Keep language/communication, disciplinary reasoning, multimodal design, and cultural accountability analytically distinct unless a composite score is validated (Biggs, 1996; Jiang et al., 2024; Liu & Lam, 2026).	Rubric validity/reliability; feedback uptake; language development; transfer.	Theoretical transfer only; not measured.

The panel was not recruited as a TESL panel and no L2 outcomes were measured. Future language-education studies should therefore pre-specify language and multimodal outcomes and compare ICH-STEAM tasks with credible alternatives (Jiang & Hafner, 2024).

Figure 3. Conceptual transfer pathway to TESL and applied-linguistics contexts; language outcomes require future empirical testing



5.4 Practical and policy implications

For curriculum teams, the architecture implies a simple sequence: identify the heritage practice and relevant knowledge holders; define cultural and disciplinary outcomes; formulate an authentic problem; select only the disciplinary and technological resources required; align activities and assessment; and audit institutional feasibility.

Institutions should treat professional development, timetabling, facilities, assessment policy, incentives and community partnerships as implementation conditions. Community participation should be substantive where decisions concern content, representation, attribution or evaluation (UNESCO, 2003).

5.5 Limitations and future research

The study has important boundaries. Delphi produces structured expert endorsement, not causal evidence or proof of sustainable safeguarding (Diamond et al., 2014). The purposive panel may favour integration-oriented judgements and cannot statistically represent university populations. Four practitioners cannot represent all positions within heterogeneous heritage communities. The Round 1 definition was not re-rated, two curriculum-design items were added only in Round 3, affirmative wording and controlled feedback may have increased agreement, and researcher judgement entered coding, merging, wording and theoretical mapping.

Validation should proceed parsimoniously: re-rate the two Round 3-only items; test feasibility and classroom enactment against credible alternatives; pre-specify L2 outcomes in language-education studies; and require longer-term evidence of continued practice, participation or transmission before making safeguarding claims.

6. Conclusion

A three-round Delphi process produced a 52-item Round 3 ICH-STEAM curriculum structure across value objectives, curriculum design, pedagogical implementation and structural/ecosystem barriers. Fifty indicators had prior-round rating or precursor histories; two curriculum-design additions are provisionally endorsed because they were introduced only in Round 3. The contribution is deliberately parsimonious: familiar curriculum mechanisms are organised under a cultural-accountability constraint that limits how disciplinary integration, technology and assessment are selected and combined. The findings establish expert endorsement within this panel, not classroom effectiveness, unique model

superiority, language gains or sustainable safeguarding. Multimodal language education remains a theoretically grounded transfer context requiring direct empirical validation.

Statements and Declarations

Ethics approval and consent to participate: Ethical approval was granted by the UKM Research Ethics Committee (JEP-2025-751; 9 October 2025–8 October 2028). All formal Delphi participants provided informed consent. Participation was voluntary, and responses were anonymised and handled confidentially.

Funding: The authors report that this study received no external funding.

Competing interests: The authors report no known financial or non-financial competing interests relevant to this work.

Data availability: Participant-level data are not publicly deposited because the Delphi panel participated under confidentiality arrangements. The revised Supplementary Material provides complete round-by-round item-level aggregate statistics, item wordings, and transition/status information. Additional de-identified aggregate materials may be requested from the corresponding author where release is compatible with ethics approval and participant confidentiality.

Author responsibility: The authors are responsible for the accuracy of the analysis, interpretation, source use, and final manuscript.

References

- Aral, A. E. (2022). Seeing the glass half-empty: Implications for safeguarding intangible cultural heritage through formal education in Türkiye. *International Journal of Heritage Studies*, 28(11–12), 1286–1299. <https://doi.org/10.1080/13527258.2022.2153264>
- Beiderbeck, D., Frevel, N., von der Gracht, H. A., Schmidt, S. L., & Schweitzer, V. M. (2021). Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX*, 8, 101401. <https://doi.org/10.1016/j.mex.2021.101401>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347–364. <https://doi.org/10.1007/BF00138871>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Chappell, K., Hetherington, L., Juillard, S., Aguirre, C., & Duca, E. (2025). A framework for effective STEAM education: Pedagogy for responding to wicked problems. *International Journal of Educational Research Open*, 9, 100474. <https://doi.org/10.1016/j.ijedro.2025.100474>
- Davoodi, A. (2024). Digital echoes of heritage: Toward a culturally balanced pedagogy in technology-enhanced bilingual education. *Journal for Multicultural Education*, 18(1/2), 192–205. <https://doi.org/10.1108/JME-10-2023-0107>
- Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, 67(4), 401–409. <https://doi.org/10.1016/j.jclinepi.2013.12.002>

- English, L. D. (2016). STEM education K–12: Perspectives on integration. *International Journal of STEM Education*, 3, 3. <https://doi.org/10.1186/s40594-016-0036-1>
- Hafner, C. A. (2014). Embedding digital literacies in English language teaching: Students' digital video projects as multimodal ensembles. *TESOL Quarterly*, 48(4), 655–685. <https://doi.org/10.1002/tesq.138>
- Holey, E. A., Feeley, J. L., Dixon, J., & Whittaker, V. J. (2007). An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Medical Research Methodology*, 7, 52. <https://doi.org/10.1186/1471-2288-7-52>
- Hou, Y., Kenderdine, S., Picca, D., Egloff, M., & Adamou, A. (2022). Digitizing intangible cultural heritage embodied: State of the art. *Journal on Computing and Cultural Heritage*, 15(3), Article 55, 1–20. <https://doi.org/10.1145/3494837>
- Jiang, L. G., & Hafner, C. A. (2024). Digital multimodal composing in L2 classrooms: A research agenda. *Language Teaching*, 58(4), 528–546. <https://doi.org/10.1017/S0261444824000107>
- Jiang, L., Li, Z. J., & Leung, J. S. C. (2024). Digital multimodal composing as translanguaging assessment in CLIL classrooms. *Learning and Instruction*, 92, 101900. <https://doi.org/10.1016/j.learninstruc.2024.101900>
- Kurin, R. (2004). Safeguarding intangible cultural heritage in the 2003 UNESCO Convention: A critical appraisal. *Museum International*, 56(1–2), 66–77. <https://doi.org/10.1111/j.1350-0775.2004.00459.x>
- Liu, Y., & Lam, R. (2026). Surveying the assessment landscape of digital multimodal composing in L2 contexts: A scoping review. *System*, 140, 104079. <https://doi.org/10.1016/j.system.2026.104079>
- 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>
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60–93. <https://doi.org/10.17763/haer.66.1.17370n67v22j160u>
- Perignat, E., & Katz-Buonincontro, J. (2019). STEAM in practice and research: An integrative literature review. *Thinking Skills and Creativity*, 31, 31–43. <https://doi.org/10.1016/j.tsc.2018.10.002>
- Pillai, J., & Achilles, V. (2015). Learning with intangible heritage for a sustainable future: Guidelines for educators in the Asia-Pacific region. UNESCO Bangkok Office and UNESCO.
- Šantek, G.-P., Kavoura, A., & Briciu, A. (2025). Integrating intangible cultural heritage in higher education—Challenges and obstacles. In A. Kavoura, V. A. Briciu, & A. Briciu (Eds.), *Strategic innovative marketing and tourism* (pp. 929–936). Springer. https://doi.org/10.1007/978-3-031-81962-9_100
- Smith, L. (2006). *Uses of heritage*. Routledge. <https://doi.org/10.4324/9780203602263>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- UNESCO. (2003). *Convention for the safeguarding of the intangible cultural heritage*. UNESCO.

- von der Gracht, H. A. (2012). Consensus measurement in Delphi studies: Review and implications for future quality assurance. *Technological Forecasting and Social Change*, 79(8), 1525–1536. <https://doi.org/10.1016/j.techfore.2012.04.013>
- Waterton, E., & Smith, L. (2010). The recognition and misrecognition of community heritage. *International Journal of Heritage Studies*, 16(1–2), 4–15. <https://doi.org/10.1080/13527250903441671>
- Zeng, Y., Md Noor, H., & Sabri, M. F. (2025). Intangible cultural heritage in fine arts curriculum in Chinese secondary school: A systematic literature review and implications on pre-service teacher education. *Asian Journal of University Education*, 21(1), 135–159. <https://doi.org/10.24191/ajue.v21i1.5563>
- Zhang, M., Akoto, M., & Li, M. (2023). Digital multimodal composing in post-secondary L2 settings: A review of the empirical landscape. *Computer Assisted Language Learning*, 36(4), 694–721. <https://doi.org/10.1080/09588221.2021.1942068>

Navigating the Implementation Constraints of Cultural Heritage Education: A Qualitative Thematic Analysis of University Instructors

Author Sun Yanhu¹, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia .E-mail:p123627@siswa.ukm.edu.my

Author Nuriah Abd Majid^{1*}, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia, E-mail:nuriah@ukm.edu.my

*** Corresponding Author** Nuriah Abd Majid .E-mail:nuriah@ukm.edu.my

Abstract

Intangible cultural heritage (ICH) education increasingly intersects with STEAM, yet universities still face a practical problem: how to combine interdisciplinary expertise with culturally legitimate teaching without allowing technical innovation to displace living transmission. This qualitative study examined that problem through semi-structured interviews with eight university educators and heritage bearers in China. Interviews generated 663 minutes of audio and approximately 110,000 Chinese-language transcript words. Codebook-assisted thematic analysis in NVivo 14 produced five themes: value objectives, course design, pedagogical implementation, faculty requirements, and institutional support. Participants treated ICH transmission as the organising purpose, STEAM disciplines as selectively mobilised resources, and collaboration as necessary because cultural and pedagogical expertise were complementary but not interchangeable. The resulting Dual Mentorship proposition specifies role coordination rather than a universal staffing model. Institutional time, workload recognition, resources, and cross-department coordination enabled this collaboration. The study does not test student learning or the comparative effectiveness of Dual Mentorship; it identifies implementation conditions that warrant direct classroom evaluation.

Keywords: dual mentorship, higher education, intangible cultural heritage, STEAM, thematic analysis

Introduction

The 2003 UNESCO Convention defines safeguarding through continued practice, transmission, and recreation rather than static preservation and assigns communities, groups, and individuals a central role in that process (UNESCO, 2003). Formal education can broaden access to ICH, but curricular inclusion does not by itself secure living transmission; institutionalisation can standardise or decontextualise practice when community participation and culturally grounded judgement are weak (Aral, 2022; Kirshenblatt-Gimblett, 2004; Labrador, 2022; Van Doorsselaere, 2021). Heritage scholarship therefore cautions against treating universities as neutral containers of culture: recognition is negotiated, and expert or institutional authority does not automatically substitute for community-based meaning (Waterton & Smith, 2010). The educational problem is consequently not how much heritage content enters the curriculum, but how formal teaching remains answerable to the practices and people from which that heritage derives.

STEAM can provide useful disciplinary resources for heritage-related inquiry, but placing science, technology, engineering, arts, and mathematics side by side is not meaningful integration. Integrated STEM/STEAM research instead emphasises purposeful connections around authentic problems, with disciplines contributing according to the task rather than in equal proportions (Kelley & Knowles, 2016; Perignat & Katz-Buonincontro, 2019; Quigley et al., 2017). Recent empirical work strengthens this implementation focus. An interview-based C-STEAM study in China found that cultural content must be integrated with disciplinary thinking rather than added decoratively (Chen et al., 2023), while an ethnography of university ICH martial-arts curricula showed that formal curriculum construction must negotiate local knowledge, institutional requirements, and transmission practice (Cheng & Guo, 2024). In parallel, a four-year mixed-methods study found substantial variation in how teachers understand and enact STEAM, indicating that a single definition does not translate automatically into classroom practice (Boice et al., 2024).

Teacher capacity is therefore an implementation condition rather than a background variable. Pedagogical content knowledge distinguishes subject expertise from knowing how to make that expertise teachable, and TPACK shows that technology must be coordinated with content and pedagogy rather than treated as an independent add-on (Mishra & Koehler, 2006; Shulman, 1986). Professional-development research further shows that effective learning for teachers is active, content-focused, coherent, sustained, and collaborative (Garet et al., 2001). Recent STEAM evidence is consistent with this principle: integrated-STEAM professional development can strengthen teachers' PCK (Martins & Baptista, 2024), and partnerships with STEAM experts can expand teachers' capacity to design and implement interdisciplinary lessons (Kessler et al., 2024). These findings make it unnecessary to assume that one instructor should master every cultural, pedagogical, scientific, and technological domain.

The remaining research gap is narrower. Existing scholarship establishes the importance of living-heritage participation, interdisciplinary design, teacher learning, and institutional collaboration, but it specifies less clearly how cultural judgement, curriculum responsibility, and technical expertise should be coordinated when they are distributed across different actors in higher education. Boundary-crossing theory offers a parsimonious explanation: coordination is required because different forms of expertise are complementary but not interchangeable (Akkerman & Bakker, 2011). Accordingly, the study uses one concise conceptual framework to guide interpretation: living-heritage accountability defines the cultural boundary, boundary crossing explains role coordination, and institutional conditions explain whether that coordination can be enacted. PCK/TPACK and professional-development research serve as supporting literatures rather than additional co-equal theories.

This study asks: How do frontline ICH educators and heritage bearers in China understand the constraints of ICH-STEAM implementation, and what pedagogical, professional, and institutional supports do they consider necessary? The aim is not to test the effectiveness of STEAM or to establish a universal staffing model. It is to identify implementation conditions and to examine whether a role-based Dual Mentorship arrangement offers a defensible way to coordinate culturally grounded and pedagogical expertise within formal higher education.

Methodology

Research Design

A qualitative semi-structured interview design was used because the research question concerned experienced practitioners' interpretations of implementation rather than population estimates. The doctoral research was registered at Universiti Kebangsaan Malaysia (UKM), while participants and data collection were located in China. Purposive sampling was used to

identify information-rich participants with direct experience in university ICH teaching, heritage transmission or safeguarding, or ICH-related curriculum development (Palinkas et al., 2015). Sample adequacy was judged by sample specificity, dialogue quality, and thematic sufficiency rather than statistical representativeness (Guest et al., 2020; Malterud et al., 2016).

Sample and Data Collection

Potential participants who met the inclusion criteria were approached directly through formal invitation and consent materials. Selection considered the ICH domain taught or practised, professional role, educational background, teaching or practice experience, institutional context, geographical diversity, and willingness to discuss concrete implementation experience. Eight participants were recruited: university educators, heritage bearers, artists, and practitioner-researchers with 13–43 years of relevant experience. Their fields included folklore/ethnomusicology, Suzhou embroidery, jade carving, Hangzhou embroidery, Shanghai-style lacquerware, Miao batik/ethnic costume, and music aesthetics. The heterogeneous sample was intentional because the study examined cross-domain implementation conditions rather than one craft alone.

Interviews were conducted online or face to face from 10 to 17 January 2026. The guide addressed understandings of ICH–STEAM, course and assessment design, professional preparation, and institutional conditions. The eight interviews totalled 663 minutes (mean approximately 83 minutes) and were recorded with permission and transcribed verbatim in Chinese. Coding was conducted in Chinese; selected quotations were translated into English after theme development. No independent back-translation procedure was documented, so none is claimed. Participants also supplied or identified teaching materials, online resources, presentation materials, and institutional documents. These materials clarified examples and supported contextual triangulation but were not coded as a second or co-equal corpus.

Analyzing of Data

NVivo 14 supported codebook-assisted thematic analysis. Braun and Clarke’s phases were used as a procedural scaffold, but the study is not presented as reflexive thematic analysis because it employed a shared codebook and reliability checks (Braun & Clarke, 2006, 2021). Coding moved from 392 initial concepts and 1,431 coded reference points to 116 preliminary categories, 17 core categories, and five final themes. Before formal coding, the researcher and a graduate assistant completed a three-day calibration exercise using 10% of the interview material. Subsequent coding comparison produced raw section-level agreement from 75.42% to 91.97%; after discrepancy discussion and codebook refinement, the archived final consistency statistic was Cohen’s $\kappa = .83$. Percentage agreement and κ are reported separately because they are different statistics. Three participants reviewed the thematic interpretation for plausibility. The first author’s position as a university arts educator supported domain understanding but could privilege educational interpretations; reflexive notes and supervisory review were used to challenge assumptions and competing explanations.

Ethical approval was granted by the UKM Research Ethics Committee (JEP-2025-751; approval period 9 October 2025–8 October 2028), with China identified as the approved study site. Participation was voluntary; informed consent and recording permission were obtained before interview. Participants were anonymised by code, and identifying details were omitted from reporting.

Findings / Results

Five themes described participants’ accounts of viable ICH–STEAM implementation: value objectives, course design, pedagogical implementation, faculty requirements, and institutional

support. The 17 core categories are consolidated in Table 1 for concise reporting; this compression changes presentation, not the underlying coding hierarchy. The themes are participant perspectives and cross-case interpretations, not measured effects on students.

Table 1. Summary of Themes, Analytic Dimensions, and Coding Evidence

Theme	Underlying core categories	Reference points	Core finding
Value objectives	Development needs; educational value; practical motivation	170	Living transmission remained the organising purpose; innovation was accepted when connected to cultural continuity.
Course design	Course relationship; evaluation; project development; integrated understanding	311	Participants described several integration logics rather than an equal-weight STEAM formula.
Pedagogical implementation	Teacher role; teaching mode; implementation process; reported learning outcomes	547	Hands-on/project/problem work and technology were valued when they supported inquiry, iteration, reflection, and cultural interpretation.
Faculty requirements	Competence; professional development; self-efficacy	215	No single instructor was expected to carry all cultural, pedagogical, and technical expertise; professional learning and collaboration were required.
Institutional support	Implementation resistance; support system; ecosystem collaboration	222	Time, workload recognition, resources, funding, leadership, and cross-department coordination shaped feasibility.

Value objectives placed heritage transmission before disciplinary display. R1 described ICH education as necessary for both “cultural inheritance” and “cultural innovation,” while also framing integrated study as holistic education. Participants did not reject innovation; they judged it by whether it remained connected to the heritage practice. Course design was correspondingly plural: the data included science-centred integration, ICH-as-platform with STEAM-as-tools, engineering-oriented problem solving, and holistic/systemic integration. R4 linked locally grounded content to project design by explaining that the paper-making process contains scientific and technological aspects that can be examined through different disciplinary perspectives.

Pedagogical implementation emphasised structured experience rather than technology for its own sake. Participants valued hands-on work, problem- and project-based activity, collaborative inquiry, and iterative design. Here PBL means problem-based learning and PjBL means project-based learning; the terms are not used interchangeably. Faculty requirements centred on the difficulty of concentrating cultural, pedagogical, and technical expertise in one person. Participants therefore favoured professional learning, co-design, and collaboration rather than the “super-teacher” assumption. Institutional support formed the final theme:

restricted time, unclear curriculum status, resource constraints, funding, workload, and cross-department coordination influenced whether collaborative teaching could occur.

For template-constrained reporting, the five themes are organised into eight reported subthemes in Table 2. Each reported subtheme is anchored by a representative verbatim quotation. The 17 underlying core categories remain part of the coding hierarchy in Table 1 but are not presented as separate formal subthemes.

Table 2. Representative Participant Quotations for the Five Themes and Eight Reported Subthemes

Theme / reported subtheme	Representative quotation	Source status
Value objectives – holistic/cultural purpose	“The key to integrating intangible cultural heritage with STEAM is holistic education.”	R1, verbatim archived quotation
Value objectives – transmission and innovation	“This is not only about cultural inheritance but also about cultural innovation and industrial development.”	R1, verbatim archived quotation
Course design – interdisciplinary integration	“This course provides a platform for interdisciplinary integration, allowing students to learn not only about culture and skills but also about science, technology, engineering, art and mathematics.”	R1, verbatim archived quotation
Course design – local/project contextualisation	“By applying different disciplinary perspectives and tools, we can improve the design and delivery of ICH project-based courses.”	R4, verbatim archived quotation
Pedagogical implementation – practice and cognition	“It is not just about hands-on training; it is also about brain training.”	R1, verbatim archived quotation
Faculty requirements – professional/curriculum development	“The promotion of this curriculum is based on the dual needs of teaching and scientific research.”	R1, verbatim archived quotation
Faculty requirements – implementation burden	“We are building the bridge while walking on it.”	Verbatim teacher quotation; participant code not preserved in archived extract
Institutional support – leadership	“For frontline teachers, the most important factor is whether the school supports you.”	Verbatim participant quotation; participant code not preserved in archived extract

Discussion

The findings support a simple interpretation: ICH–STEAM implementation is primarily a coordination problem. Heritage-education studies already show that formal inclusion does not automatically safeguard living practice (Aral, 2022; Labrador, 2022; Van Doorselaere, 2021), while Chinese empirical studies show that culturally grounded content and local knowledge must be negotiated within formal curricula (Chen et al., 2023; Cheng & Guo, 2024). The present data add a role question to that literature. Participants did not describe cultural judgement, curriculum design, and technical expertise as interchangeable. The contribution is therefore not the general claim that collaboration matters; it is the specification of what must be coordinated and where authority should remain visible.

This interpretation also narrows what “STEAM integration” means. Contemporary STEAM research treats integration as purposeful problem-centred coordination rather than the simultaneous presence of five disciplines (Kelley & Knowles, 2016; Perignat & Katz-Buonincontro, 2019), and recent teacher research shows that enactment varies substantially across educators (Boice et al., 2024). The interview data are consistent with that literature: participants described several integration logics and used science or technology selectively. For ICH, the additional constraint is living-heritage accountability. A technically sophisticated activity is not automatically educationally legitimate if it displaces culturally grounded interpretation. The simplest boundary is therefore heritage-centred integration: disciplines are mobilised when they serve a heritage problem, not to display disciplinary completeness.

Dual Mentorship should consequently be understood as a role arrangement, not as a new general theory or a proven intervention. PCK and TPACK already explain why content, pedagogy, and technology are distinct but related forms of teacher knowledge (Mishra & Koehler, 2006; Shulman, 1986), while boundary crossing explains why coordination is required across complementary expertise (Akkerman & Bakker, 2011). In the present data, heritage bearers contribute embodied and culturally grounded judgement; university educators organise curriculum, scaffolding, assessment, and student support; technical specialists are added when a task requires scientific, digital, or engineering knowledge. This role allocation also avoids assuming that any one bearer represents a homogeneous community. Its practical value is as a design proposition that institutions can test, not as a universal staffing prescription.

The faculty and institutional themes mostly reproduce established implementation findings rather than demanding additional theory. Professional development is more effective when it is active, coherent, content-focused, sustained, and collaborative (Garet et al., 2001), and recent STEAM studies show benefits from PCK-oriented development and partnership with disciplinary experts (Kessler et al., 2024; Martins & Baptista, 2024). Participants’ proposals—co-design workshops, demonstration lessons, case resources, faculty learning communities, and technical clinics—are therefore best interpreted as practical forms of established professional-learning principles. Likewise, planning time, workload recognition, funding, facilities, and coordination are general constraints of interdisciplinary teaching. The ICH-specific remainder is narrower: institutions must enable collaboration without absorbing cultural judgement into university authority.

The evidence also limits the strength of the conclusions. The interviews identify conditions that participants consider workable and culturally accountable; they do not show that ICH–STEAM improves student learning, that Dual Mentorship outperforms single-instructor or alternative team-teaching arrangements, or that the proposed role distinction transfers unchanged across ICH communities. Direct learner evidence, classroom observation, comparative designs, and community-based evaluation are required for those claims. Keeping the contribution at the implementation level preserves explanatory value without converting qualitative perceptions into causal effectiveness claims.

Conclusion

This study identifies a bounded implementation problem in ICH–STEAM higher education: cultural, pedagogical, and technical expertise must collaborate, but they are not interchangeable. Participants treated living transmission as the organising purpose, STEAM disciplines as selectively used resources, and institutional support as the condition that makes coordination feasible. Dual Mentorship captures the central cultural–pedagogical role relationship, with technical expertise added when required by the task. The study therefore contributes an empirically grounded implementation proposition rather than a universal model or evidence of student-learning effectiveness.

Recommendations

Universities developing ICH–STEAM courses should begin with a clearly defined heritage purpose, select only the disciplinary resources required by that purpose, and assign cultural, pedagogical, and technical judgements to actors with relevant expertise. Institutions should protect co-planning time, recognise collaborative teaching in workload systems, provide shared facilities and project resources, and maintain relationships with heritage communities. Professional development should prioritise co-design and practice-based learning. Future evaluation should combine classroom observation, student evidence, and community judgement to test both learning outcomes and the comparative value of Dual Mentorship.

Limitations

The study has several limitations. The sample of eight participants was purposive and heterogeneous, so the findings are analytically informative rather than statistically representative. Evidence came primarily from educator and heritage-bearer accounts; students and wider community members were not directly studied. Reported learning outcomes were anticipated or perceived rather than independently measured. Supplementary materials supported contextual triangulation but were not analysed as a co-equal corpus. English quotations were translated from Chinese after theme development, and no independent back-translation protocol was documented. In addition, Dual Mentorship was derived from participant accounts and role analysis rather than tested through an intervention. These limits restrict claims to implementation perceptions and justify direct classroom and community-based testing.

Ethics and transparency. Ethical approval was granted by the UKM Research Ethics Committee (JEP-2025-751). Participants gave informed consent and recording permission; confidentiality and anonymity were maintained through coded reporting and removal of identifying details. Raw interview transcripts are not publicly deposited because consent and ethical commitments require confidentiality; de-identified coding summaries and the interview guide may be requested from the corresponding author subject to ethical restrictions. The research received no external funding. The authors declare no conflict of interest.

References

- Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research*, 81(2), 132–169. <https://doi.org/10.3102/0034654311404435>
- Aral, A. E. (2022). Seeing the glass half-empty: Implications for safeguarding intangible cultural heritage through formal education in Türkiye. *International Journal of Heritage Studies*, 28(11–12), 1286–1299. <https://doi.org/10.1080/13527258.2022.2153264>
- Boice, K. L., Alemdar, M., Jackson, J. R., Kessler, T. C., Choi, J., Grossman, S., & Usselman, M. (2024). Exploring teachers' understanding and implementation of STEAM: One size

- does not fit all. *Frontiers in Education*, 9, 1401191.
<https://doi.org/10.3389/feduc.2024.1401191>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328–352. <https://doi.org/10.1080/14780887.2020.1769238>
- Chen, Q., Ye, J.-H., & Zheng, C. (2023). Integration of Shangshan culture into the STEAM curriculum and teaching: Results of an interview-based study. *Frontiers in Psychology*, 14, 1251497. <https://doi.org/10.3389/fpsyg.2023.1251497>
- Cheng, Y., & Guo, N. (2024). An ethnography of construction and characteristics of curriculum for inheritance of intangible cultural heritage martial arts in universities. *Frontiers in Sports and Active Living*, 6, 1395128. <https://doi.org/10.3389/fspor.2024.1395128>
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945. <https://doi.org/10.3102/00028312038004915>
- Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PLOS ONE*, 15(5), e0232076. <https://doi.org/10.1371/journal.pone.0232076>
- Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, 3, 11. <https://doi.org/10.1186/s40594-016-0046-z>
- Kessler, T. C., Boice, K. L., Koval, J., Jackson, J. R., Choi, J., Alemdar, M., Grossman, S., Simmons, K., & Usselman, M. (2024). Partnerships in STEAM: How collaborating with STEAM experts impacts K–12 teachers' abilities to implement STEAM lessons in the classroom. *Education Sciences*, 14(6), 666. <https://doi.org/10.3390/educsci14060666>
- Kirshenblatt-Gimblett, B. (2004). Intangible heritage as metacultural production. *Museum International*, 56(1–2), 52–65. <https://doi.org/10.1111/j.1350-0775.2004.00458.x>
- Labrador, A. M. (2022). Integrating ICH and education: A review of converging theories and methods. *International Journal of Intangible Heritage*, 17, 17–36. <https://doi.org/10.35638/ijih.2022.17..001>
- Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13), 1753–1760. <https://doi.org/10.1177/1049732315617444>
- Martins, I., & Baptista, M. (2024). Teacher professional development in integrated STEAM education: A study on its contribution to the development of the PCK of physics teachers. *Education Sciences*, 14(2), 164. <https://doi.org/10.3390/educsci14020164>
- 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>
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>

- Perignat, E., & Katz-Buonincontro, J. (2019). STEAM in practice and research: An integrative literature review. *Thinking Skills and Creativity*, 31, 31–43.
<https://doi.org/10.1016/j.tsc.2018.10.002>
- Quigley, C. F., Herro, D., & Jamil, F. M. (2017). Developing a conceptual model of STEAM teaching practices. *School Science and Mathematics*, 117(1–2), 1–12.
<https://doi.org/10.1111/ssm.12201>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- UNESCO. (2003). Convention for the safeguarding of the intangible cultural heritage. UNESCO. <https://ich.unesco.org/en/convention>
- Van Doorselaere, J. (2021). Safeguarding intangible cultural heritage through formal education in Flanders: A critical analysis of the implementation of the 2003 UNESCO Convention. *International Journal of Intangible Heritage*, 16, 31–44.
<https://doi.org/10.35638/ijih.2021..16.005>
- Waterton, E., & Smith, L. (2010). The recognition and misrecognition of community heritage. *International Journal of Heritage Studies*, 16(1–2), 4–15.
<https://doi.org/10.1080/13527250903441671>

Evaluating Curriculum Priorities in Traditional Vocal Heritage: An Analytic Hierarchy Process (AHP) Approach on Mongolian Khoomei

Author Sun Yanhu¹, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia .E-mail:p123627@siswa.ukm.edu.my

Author Nuriah Abd Majid^{1*}, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia, E-mail:nuriah@ukm.edu.my

*** Corresponding Author** Nuriah Abd Majid .E-mail:nuriah@ukm.edu.my

Abstract

Mongolian Khoomei higher education faces a practical curriculum question: how should digital and STEAM resources be introduced without displacing the embodied and cultural functions that sustain living heritage? This study uses the Analytic Hierarchy Process (AHP) to identify curriculum priorities among ten university-based Khoomei experts in Inner Mongolia, China. Experts compared four criteria and nineteen sub-criteria using Saaty's nine-point scale. All retained aggregate matrices met the conventional consistency requirement of $CR < 0.10$. Cultural Heritage received the highest criterion weight (0.4982), followed by Teaching Implementation (0.3086), Assessment Framework (0.1308), and STEAM Integration (0.0624). Depth of Khoomei techniques, course progression, and traditional cultural content were the three highest global sub-criteria. The results indicate relative priorities within the specified hierarchy rather than attitudes toward technology or evidence of teaching effectiveness. A parsimonious "Heritage-as-Core, STEAM-as-Support" proposition is therefore offered as a conceptual guide for technology-enhanced learning: digital tools should be tested where they support higher-priority heritage and teaching functions, with educational and cultural outcomes evaluated separately.

Keywords: *Analytic Hierarchy Process, digital heritage education, Mongolian Khoomei, STEAM education, technology-enhanced learning.*

Introduction

Intangible cultural heritage (ICH) is sustained through continued practice, transmission, and recreation rather than through documentation alone (UNESCO, 2003; Kirshenblatt-Gimblett, 2004). Formal education can support this process, but institutionalisation may also recontextualise practices through curricular sequencing, assessment, and credentialing (Aral, 2022; Van Doorselaere, 2021). The resulting problem is especially important for performance traditions whose knowledge is partly embodied, tacit, and culturally situated (Pegg, 1992; Jiang & Wu, 2025).

Mongolian Khoomei is a living vocal tradition recognised by UNESCO and characterised by the production of a fundamental tone together with focused overtones (UNESCO, 2009; Bergevin et al., 2020). Contemporary higher education can use acoustic visualisation, recording, digital archives, multimedia production, and other technology-enhanced learning (TEL) resources to make aspects of sound and context more accessible (Tan et al., 2024; Wei

et al., 2026). Yet the educational value of technology depends on how it serves content and pedagogy rather than on tool novelty alone (Mishra & Koehler, 2006).

Recent studies have examined formal ICH education, digital heritage learning, interdisciplinary workshops, and community-engaged digital safeguarding (Aral, 2022; Tan et al., 2024; Xia et al., 2024; Tan et al., 2026). What remains less clear is how university-based heritage specialists prioritise cultural transmission, STEAM integration, teaching implementation, and assessment before a technology intervention is designed. This study addresses that decision problem through the Analytic Hierarchy Process (AHP), which structures expert pairwise judgements into relative priorities within a defined hierarchy (Saaty, 1977, 2008). The research question is: How do university-based Khoomei experts prioritise cultural heritage, STEAM integration, teaching implementation, and assessment within the current curriculum?

Literature Review

The 2003 UNESCO Convention treats safeguarding as a process that includes transmission, education, documentation, and revitalisation, while also emphasising the participation of communities, groups, and individuals who create and transmit ICH (UNESCO, 2003, 2015). This living-heritage perspective means that educational continuity cannot be reduced to preserving fixed representations. For performance heritage, culturally meaningful learning also depends on practice, listening, bodily coordination, and situated judgement (Pegg, 1992; Jiang & Wu, 2025).

Digital heritage education introduces a related but distinct problem. Digitisation can document or disseminate heritage without necessarily improving learning, whereas TEL requires a defined pedagogical function such as explanation, practice, feedback, collaboration, or access (Tan et al., 2024; Wei et al., 2026). Research on digital ICH also warns that technical capability does not by itself secure cultural continuity or community legitimacy (Tan et al., 2026; Wang & Xie, 2026). Accordingly, the present study treats digital tools as potential learning mediators rather than as independent indicators of educational quality.

STEAM similarly has value when disciplines interact around meaningful problems rather than when science, technology, engineering, arts, and mathematics are simply added as separate content areas (Perignat & Katz-Buonincontro, 2019; Chappell et al., 2025). For Khoomei, acoustics and sound analysis can clarify vocal mechanisms, while recording and multimedia production can support reflection and documentation (Bergevin et al., 2020). These possibilities remain pedagogical options, not evidence that more technology is inherently better.

Two mature ideas provide a parsimonious conceptual basis. First, living heritage identifies continued practice and transmission as the substantive aim (UNESCO, 2003). Second, technological pedagogical content knowledge (TPACK) treats technology as educationally useful when it is aligned with content and pedagogy (Mishra & Koehler, 2006). Constructive alignment is used only as a curriculum-design principle connecting intended learning, teaching activity, and assessment (Biggs & Tang, 2011). AHP then serves as the decision method rather than as an additional theory (Saaty, 2008). This combination keeps the framework limited to what is needed to interpret curriculum priorities.

Methodology

Research Design

The study used a quantitative expert-judgement design based on AHP. The bounded case was a university-based Mongolian Khoomei degree programme in Inner Mongolia, China. AHP was

selected because the research question concerns relative priorities among multiple curriculum criteria rather than prevalence, attitudes, or causal effects (Saaty, 1977, 2008).

The hierarchy contained one goal, four criteria, and nineteen sub-criteria. The criteria were Cultural Heritage, STEAM Integration, Teaching Implementation, and Assessment Framework. The sub-criteria covered technique, cultural content, heritage awareness, practice ratio, science and acoustics, digital technology, recording engineering, musical creation, mathematics-related knowledge, theory-practice balance, course progression, teaching methods, staff capability, infrastructure, skills, creativity, interdisciplinary ability, contribution to transmission, and feedback. The instrument was informed by the study objectives and relevant ICH, STEAM, and curriculum literature (UNESCO, 2003; Mishra & Koehler, 2006; Perignat & Katz-Buonincontro, 2019). Three university Khoomei instructors completed a pilot clarity check before formal administration; their feedback was used only to improve wording, interpretability, and formatting and was excluded from the empirical analysis.

Sample and Data Collection

Purposive recruitment targeted experts able to judge university Khoomei curriculum priorities. Eligibility required direct university-level Khoomei teaching or curriculum involvement, substantive familiarity with Khoomei transmission or music pedagogy, capacity to evaluate all four curriculum dimensions, and willingness to complete the pairwise-comparison instrument. Ten eligible instructors were invited and all ten completed the formal survey through Wenjuanxing between 10 and 20 February 2026 after informed consent.

The panel comprised two doctoral-degree holders recognised as higher-education Khoomei inheritors, six master-degree holders with professor titles in Khoomei or music education, and two associate professors engaged in Khoomei-related higher education (Table 1). The panel size was justified by the specialist and bounded decision problem rather than by statistical power, because AHP structures informed judgements and assesses their consistency rather than estimating a population parameter (Saaty & Özdemir, 2014). The sample therefore supports case-bounded analytical interpretation, not population generalisation.

The study was approved by the UKM Research Ethics Committee (protocol JEP-2025-751; approval period 9 October 2025 to 8 October 2028). Formal recruitment and data collection occurred within the approved period. Participation was voluntary, informed consent was obtained, and expert responses were anonymised and reported in aggregate form.

Table 1. Expert Participant Profile for AHP Analysis

Expert category	n	Qualification / role	Relevance to AHP judgement
University-based Khoomei inheritors	2	Doctoral degree holders; recognised higher-education Khoomei inheritors	Embodied vocal transmission, cultural inheritance, and Khoomei pedagogy
Senior Khoomei / music education professors	6	Master-degree holders; professor titles	Curriculum priorities, teaching implementation, and higher music education
Associate professors in Khoomei-related higher education	2	Higher-education qualifications; associate professor titles	Course organisation, assessment, and pedagogical development
Total	10	Senior university-based experts	Institutionally situated curriculum-priority judgement

Analyzing of Data

Experts completed reciprocal pairwise comparisons using Saaty's nine-point scale, where larger values indicate stronger relative importance (Saaty, 1977, 2008). Each individual matrix was checked before aggregation. Corresponding judgements were combined using the geometric mean, a standard approach for aggregating group AHP judgements (Forman & Peniwati, 1998). Priority vectors were then derived for the main criteria and each sub-criteria matrix.

Consistency was assessed through the maximum eigenvalue ($\lambda_{\max}$), consistency index (CI), random index (RI), and consistency ratio (CR). A matrix was retained when $CR < 0.10$, the conventional AHP threshold used for acceptable internal coherence (Saaty, 2008; Ishizaka & Labib, 2011). Cronbach's alpha was not used because pairwise-comparison matrices are not reflective psychometric scales. CI and CR indicate consistency of the judgement structure; they do not establish inter-expert consensus, construct validity, attitude, or teaching effectiveness (Ishizaka & Labib, 2011). Global sub-criterion weights were calculated as criterion weight multiplied by local sub-criterion weight.

Findings / Results

All five aggregate matrices satisfied the stated $CR < 0.10$ requirement (Table 2). The main-criteria matrix had $CR = 0.0243$; the four sub-criteria matrices ranged from 0.0206 to 0.0571. These values support use of the retained priority estimates as internally coherent AHP judgements, without implying that the priorities are objectively true or shared by all possible stakeholders.

Table 2. AHP Matrix Consistency Statistics

Matrix	n	$\lambda_{\max}$	CI	RI	CR	Decision
Main criteria	4	4.0649	0.0216	0.89	0.0243	Retained
A1 Cultural Heritage	4	4.0551	0.0184	0.89	0.0206	Retained
A2 STEAM Integration	5	5.1245	0.0311	1.12	0.0278	Retained
A3 Teaching Implementation	5	5.1988	0.0497	1.12	0.0444	Retained
A4 Assessment Framework	5	5.2557	0.0639	1.12	0.0571	Retained

At criterion level, Cultural Heritage received the largest weight (0.4982), followed by Teaching Implementation (0.3086), Assessment Framework (0.1308), and STEAM Integration (0.0624) (Table 3). The hierarchy therefore places heritage and curriculum organisation above STEAM under the specified pairwise trade-offs. The result is comparative: it does not show that experts reject technology, that STEAM is absent from practice, or that a heritage-centred curriculum is more effective.

Table 3. Criterion Weights and Rankings

Criterion	Weight	Rank	Bounded interpretation
Cultural Heritage	0.4982	1	Highest relative priority: technique, cultural content, heritage awareness, and practice

Criterion	Weight	Rank	Bounded interpretation
Teaching Implementation	0.3086	2	Second priority: progression, theory-practice balance, methods, staff, and infrastructure
Assessment Framework	0.1308	3	Moderate relative priority: skills, creativity, interdisciplinary ability, social contribution, and feedback
STEAM Integration	0.0624	4	Lowest relative priority among the four criteria; not evidence of rejection or non-use

Global sub-criterion priorities provide a more specific picture (Table 4). Depth of Khoomei techniques ranked first (0.2463), course connection/progression second (0.1484), and traditional cultural content third (0.1305). Theory-practice balance ranked fourth (0.0792), followed by ICH inheritance awareness (0.0657) and contribution to transmission (0.0625). Within the STEAM criterion, Art (local weight 0.3925) and Technology (0.3438) were prioritised above Science (0.1120), Mathematics (0.1056), and Engineering (0.0460). These local weights indicate internal ordering within STEAM, whereas their global weights remain comparatively small because the STEAM criterion itself received 0.0624.

Table 4. Global Ranking of the Nineteen Sub-Criteria

Rank	Code and sub-criterion	Local weight	Global weight
1	B1.1 Depth of Khoomei techniques	0.4944	0.2463
2	B3.2 Course connection / progression	0.4808	0.1484
3	B1.2 Traditional cultural content	0.2620	0.1305
4	B3.1 Theory-practice course ratio	0.2566	0.0792
5	B1.3 ICH inheritance awareness	0.1318	0.0657
6	B4.4 Contribution to transmission	0.4778	0.0625
7	B1.4 Ratio of practical courses	0.1118	0.0557
8	B3.5 Hardware and software support	0.1484	0.0458
9	B4.1 Skills assessment	0.2568	0.0336
10	B2.4 Art: traditional music, folk melody, creation	0.3925	0.0245
11	B2.2 Technology: digital piano, MIDI, multimedia score	0.3438	0.0215
12	B3.3 Classroom teaching methods	0.0617	0.0190
13	B4.5 Feedback mechanism	0.1435	0.0188
14	B3.4 STEAM/interdisciplinary staff background	0.0524	0.0162
15	B4.2 Creative ability	0.0800	0.0105
16	B2.1 Science: principles/acoustic analysis	0.1120	0.0070
17	B2.5 Mathematics: melody, harmony, instrumentation	0.1056	0.0066
18	B4.3 Interdisciplinary ability	0.0419	0.0055
19	B2.3 Engineering: recording/audio editing	0.0460	0.0029

Three descriptive conclusions follow. First, embodied technique and cultural content occupy the strongest positions in the observed priority structure. Second, course progression and theory-practice balance show that curriculum organisation is central to the panel's judgement rather than a secondary administrative issue. Third, STEAM elements occupy lower global positions, but technology- and art-related options are the strongest candidates within that domain. AHP weights alone cannot identify resistance, marginalisation, or a conservative pedagogical stance because those claims would require attitudinal, observational, or policy evidence (Ishizaka & Labib, 2011).

Discussion

The strongest result is the priority given to technique and cultural content. This pattern is compatible with living-heritage and embodied-performance accounts that locate transmission in continued practice, culturally situated judgement, and sensorimotor learning (UNESCO, 2003; Pegg, 1992; Jiang & Wu, 2025). It does not prove that heritage-centred teaching is universally optimal. The panel is institutionally situated, and its judgements may reflect university roles as well as broader cultural priorities. The cautious implication is therefore that TEL or STEAM pilots should show how they support high-priority embodied and cultural learning before they displace practice time or practitioner-guided feedback (Mishra & Koehler, 2006; Tan et al., 2024).

The low overall STEAM weight should be interpreted as a relative trade-off, not as an attitude toward innovation. AHP results depend on the hierarchy, criteria, panel, and decision context (Ishizaka & Labib, 2011). A simpler explanation than “resistance” is that experts placed heritage and teaching organisation first within the specific structure they were asked to compare. Within STEAM, Art and Technology received the largest local weights, suggesting that music- and technology-adjacent activities are more plausible starting points than stand-alone scientific or engineering modules. Their effectiveness, however, remains untested.

Teaching Implementation provides the clearest bridge to technology-enhanced learning. Its weight of 0.3086 and the high ranking of course progression indicate that the practical question is not how much technology should be added, but where a tool can support an already valued learning function. This logic is consistent with TPACK, which treats technology as contingent on content-pedagogy fit (Mishra & Koehler, 2006). For example, acoustic visualisation could follow teacher-guided overtone production, while reflective recording could be introduced after foundational technique; such sequences are design propositions grounded in vocal-acoustic and digital-learning literature, not effects demonstrated by the present AHP data (Bergevin et al., 2020; Tan et al., 2024).

Assessment should follow the same principle. Constructive alignment requires that assessment reflect the learning functions a curriculum claims to value (Biggs & Tang, 2011). In the present hierarchy, contribution to transmission and skills assessment outranked interdisciplinary ability (Table 4). Digital recordings or spectrograms may therefore support feedback and evidence, but they should not become the sole definition of successful Khoomei performance because measurable acoustic properties do not exhaust embodied or culturally situated judgement (Jiang & Wu, 2025).

The methodological contribution of AHP is to make curriculum trade-offs explicit before an intervention is designed. It does not identify a universally optimal curriculum or prove which technology improves learning (Saaty, 2008; Ishizaka & Labib, 2011). For this case, the most parsimonious interpretation is “Heritage-as-Core, STEAM-as-Support”: high-priority heritage and teaching functions remain primary, while digital or interdisciplinary tools are introduced when they serve those functions. The proposition is derived from the AHP pattern and

supporting literature; it is not an empirically validated intervention model. Questions of community authority, consent, representation, and cultural acceptability remain external conditions for future testing because they were not measured by the AHP instrument (UNESCO, 2015; Tan et al., 2026).

Conclusion

This study used AHP to identify the relative curriculum priorities of ten university-based Khoomei experts within a four-criterion hierarchy. Cultural Heritage (0.4982) and Teaching Implementation (0.3086) received larger weights than Assessment Framework (0.1308) and STEAM Integration (0.0624), while technique, course progression, and traditional cultural content were the three highest global sub-criteria (Tables 3-4). All retained aggregate matrices met the stated $CR < 0.10$ requirement (Table 2).

The findings support a limited conclusion rather than a broad theory of technology adoption. The defensible TEL implication is that digital and STEAM tools should be tested where they support higher-priority heritage and teaching functions, not assumed to improve learning merely because they are available (Mishra & Koehler, 2006; Tan et al., 2024). “Heritage-as-Core, STEAM-as-Support” is therefore presented as a case-bounded conceptual proposition for pre-intervention curriculum design, not as a validated pedagogy.

Recommendations

Curriculum teams should begin with a defined learning problem and select only the technology needed for that function. Initial pilots could focus on overtone perception, reflective listening, recording literacy, or digital documentation, with clear sequencing in relation to teacher-guided practice (Mishra & Koehler, 2006; Bergevin et al., 2020). Equipment acquisition should be linked to curriculum mapping, staff support, maintenance, and assessment rather than treated as an independent indicator of innovation.

Future research should compare priority structures across university teachers, community-based bearers, performers, and students, because different stakeholder roles may produce different judgements (UNESCO, 2015; Tan et al., 2026). It should then test a small number of discrete TEL interventions using learning outcomes, implementation fidelity, practitioner evaluation, and cultural acceptability as separate measures. Longitudinal research is required before claims are made about sustained transmission beyond the university (UNESCO, 2003; Wei et al., 2026).

Limitations

The study has three main limitations. First, the ten experts represent one university-based case and not the full ecology of Khoomei cultural authority or a probabilistic population sample. Second, AHP priorities are conditional on the researcher-defined hierarchy, criteria, panel, and decision context; a different structure could produce different weights (Ishizaka & Labib, 2011). Third, CI and CR establish internal consistency of pairwise judgements, not consensus, construct validity, actual technology use, attitudes, or causal effectiveness (Saaty, 2008; Ishizaka & Labib, 2011). The conceptual proposition should therefore be tested rather than treated as a validated model.

For transparency, the study received no external funding and the authors declare no conflicts of interest. The hierarchy, aggregate weights, and matrix-level consistency statistics are reported in the article. De-identified instruments and aggregate AHP matrices may be available from the corresponding author on reasonable request, subject to UKM ethics approval and protection of

participant confidentiality. Individual expert judgements are not publicly released because the small specialist panel creates a risk of deductive identification.

References

- Aral, A. E. (2022). Seeing the glass half-empty: Implications for safeguarding intangible cultural heritage through formal education in Türkiye. *International Journal of Heritage Studies*, 28(11-12), 1286-1299. <https://doi.org/10.1080/13527258.2022.2153264>
- Bergevin, C., Narayan, C., Williams, J., Mhatre, H., Steeves, J. G., & Story, B. H. (2020). Overtone focusing in biphonic Tuvan throat singing. *eLife*, 9, e50476. <https://doi.org/10.7554/eLife.50476>
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.
- Chappell, K., Hetherington, L., Juillard, S., Aguirre, C., & Duca, E. (2025). A framework for effective STEAM education: Pedagogy for responding to wicked problems. *International Journal of Educational Research Open*, 9, 100474. <https://doi.org/10.1016/j.ijedro.2025.100474>
- Forman, E., & Peniwati, K. (1998). Aggregating individual judgments and priorities with the analytic hierarchy process. *European Journal of Operational Research*, 108(1), 165-169. [https://doi.org/10.1016/S0377-2217\(97\)00244-0](https://doi.org/10.1016/S0377-2217(97)00244-0)
- Ishizaka, A., & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications*, 38(11), 14336-14345. <https://doi.org/10.1016/j.eswa.2011.04.143>
- Jiang, H., & Wu, W. (2025). How embodied cognition affects vocal performance: An integrative review. *Journal of Voice*. Advance online publication. <https://doi.org/10.1016/j.jvoice.2025.11.005>
- Kirshenblatt-Gimblett, B. (2004). Intangible heritage as metacultural production. *Museum International*, 56(1-2), 52-65. <https://doi.org/10.1111/j.1350-0775.2004.00458.x>
- 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>
- Pegg, C. (1992). Mongolian conceptualizations of overtone singing (xöömii). *British Journal of Ethnomusicology*, 1(1), 31-54. <https://doi.org/10.1080/09681229208567199>
- Perignat, E., & Katz-Buonincontro, J. (2019). STEAM in practice and research: An integrative literature review. *Thinking Skills and Creativity*, 31, 31-43. <https://doi.org/10.1016/j.tsc.2018.10.002>
- Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3), 234-281. [https://doi.org/10.1016/0022-2496\(77\)90033-5](https://doi.org/10.1016/0022-2496(77)90033-5)
- Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1), 83-98. <https://doi.org/10.1504/IJSSCI.2008.017590>
- Saaty, T. L., & Özdemir, M. S. (2014). How many judges should there be in a group? *Annals of Data Science*, 1, 359-368. <https://doi.org/10.1007/s40745-014-0026-4>
- Tan, P., Ren, X., Cheng, Z., & Ji, Y. (2024). A framework for students' digital heritage education in the classroom: A human-engaged computing perspective. *Education and Information Technologies*, 29(8), 10225-10247. <https://doi.org/10.1007/s10639-023-12199-1>

- Tan, Y.-C., Yang, L.-H., & Wang, B. (2026). Community-engaged digital safeguarding of intangible cultural heritage: A review of methods and challenges. *npj Heritage Science*, 14, 184. <https://doi.org/10.1038/s40494-026-02458-y>
- UNESCO. (2003). *Convention for the safeguarding of the intangible cultural heritage*. UNESCO.
- UNESCO. (2009). *Mongolian art of singing, Khoomei*. Representative List of the Intangible Cultural Heritage of Humanity. <https://ich.unesco.org/en/RL/mongolian-art-of-singing-khoomei-00210>
- UNESCO. (2015). *Ethical principles for safeguarding intangible cultural heritage*. <https://ich.unesco.org/en/decisions/10.COM/15.A>
- Van Doorselaere, J. (2021). Safeguarding intangible cultural heritage through formal education in Flanders: A critical analysis of the implementation of the 2003 UNESCO Convention. *International Journal of Intangible Heritage*, 16, 31-44. <https://doi.org/10.35638/ijih.2021..16.005>
- Wang, L., & Xie, X. (2026). Diagnosing the continuation conditions of digital intangible cultural heritage: A socio-technical systems evaluation framework. *npj Heritage Science*. Advance online publication. <https://doi.org/10.1038/s40494-026-02812-0>
- Wei, S., Wang, L., Xu, S., Xu, J., Wang, M., Shi, J., Fan, H., & Gan, F. (2026). How digital technologies empower the safeguarding of intangible cultural heritage: A scoping review of research from 2020 to 2025. *npj Heritage Science*. Advance online publication. <https://doi.org/10.1038/s40494-026-02774-3>
- Xia, J., Yuizono, T., Wang, T.-Y., & Kim, E. (2024). Intangible cultural heritage course design in the digital age and its effects of interdisciplinary workshop. *Education Sciences*, 14(3), 325. <https://doi.org/10.3390/educsci14030325>

Optimizing Student Engagement in Interdisciplinary Heritage Education: A Fuzzy-Set Qualitative Comparative Analysis (fsQCA)

Author Sun Yanhu¹, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia .E-mail: p123627@siswa.ukm.edu.my

Author Nuriah Abd Majid^{1*}, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia, E-mail: nuriah@ukm.edu.my

*** Corresponding Author** Nuriah Abd Majid .E-mail: nuriah@ukm.edu.my

Abstract

Interdisciplinary heritage education can connect living cultural practice with scientific and digital inquiry, but learner support should not be equated with multidimensional engagement or learning effectiveness. This study examined whether support for a STEAM-informed Mongolian Khoomei curriculum depended on one dominant curriculum condition or on alternative combinations. A census of 75 undergraduates at Inner Mongolia Arts Institute completed a 23-item questionnaire covering six content-defined curriculum indices. Necessary Condition Analysis (NCA) and fuzzy-set Qualitative Comparative Analysis (fsQCA) were used for distinct purposes: NCA tested minimum-level constraints and fsQCA tested sufficient combinations. No single condition met the 0.90 set-theoretic necessity benchmark. NCA identified the clearest medium bottleneck tendencies for Future Course Suggestions and STEAM Curriculum Design Preferences. fsQCA identified three high-support configurations (solution consistency = 0.873; coverage = 0.511), with two remaining under a stricter consistency threshold. Because Learning Effect was semantically close to the outcome, it was not interpreted as an independent mechanism. The findings support a bounded conclusion: stronger learner support can arise through multiple combinations of curriculum perceptions rather than one universal recipe.

Keywords: *fsQCA, intangible cultural heritage, learner support, STEAM education*

Introduction

Intangible cultural heritage (ICH) depends on continued practice and transmission, not documentation alone. UNESCO's 2003 Convention therefore treats education as part of safeguarding while requiring the participation of the communities, groups, and individuals who create and transmit heritage (UNESCO, 2003). Mongolian Khoomei is a useful case because it is both an embodied cultural practice and a vocal phenomenon that can be examined through acoustics and physiology (Sundberg et al., 2023). This dual character makes interdisciplinary teaching plausible, but it does not justify assuming that technology or STEAM integration is beneficial by default.

The central outcome in this study is learner support for an interdisciplinary Khoomei curriculum. It is deliberately distinguished from student engagement. Engagement is multidimensional and includes behavioural, emotional, and cognitive participation (Fredricks et al., 2004), whereas the present outcome is an attitudinal judgement about curriculum value. This distinction matters because technology-enhanced learning effects vary with task, pedagogy, disciplinary context, and interaction rather than with technology alone (Bond et al.,

2020; Bedenlier et al., 2020). The study therefore asks how curriculum perceptions combine, not whether one technological feature causes engagement.

A further problem is that heritage education carries cultural constraints absent from generic curriculum models. Formal education can support transmission but can also recontextualise living practice or displace community authority (Aral, 2022; Waterton & Smith, 2010). Digital ICH research has begun to use configurational methods to examine adoption and engagement (Ye et al., 2025; Bao et al., 2026), while higher-education research has used fsQCA to identify alternative pathways to student engagement (Zhao & Zhao, 2026). Less is known about whether support for an embodied heritage curriculum depends on one dominant condition or on alternative combinations of perceptions.

Accordingly, the study asks three questions: (1) Does any curriculum condition operate as a necessary condition or meaningful bottleneck for stronger overall support? (2) Which combinations of conditions are sufficient for stronger overall support? (3) What bounded curriculum implications follow from those patterns? The purpose is analytical diagnosis within one programme, not population estimation or causal proof.

Literature Review

Kahu and Nelson's (2018) educational-interface perspective provides the primary theoretical lens because it locates engagement-related experience at the interaction between learners and educational conditions rather than in one isolated factor. That logic fits a configurational analysis, but the outcome remains curriculum support rather than engagement itself. The study therefore uses engagement research to set a boundary around interpretation, not to relabel the outcome.

STEAM is treated as a curriculum-design approach rather than as an additional causal theory. Reviews describe STEAM as a heterogeneous family of interdisciplinary practices whose value depends on how disciplines and learning purposes are integrated (Perignat & Katz-Buonincontro, 2019). The same caution applies to digital heritage learning. Serious games, immersive tools, and other technologies can support motivation or cultural learning when they fit the task and practice, but evidence does not warrant a general 'more technology is better' assumption (Pang et al., 2024; Bao et al., 2026). The parsimonious proposition is therefore that different combinations of curriculum perceptions may be associated with support.

ICH adds a cultural boundary to that proposition. UNESCO (2003) places community participation within safeguarding, and heritage scholarship warns that institutional recognition can misrepresent or decontextualise community heritage (Waterton & Smith, 2010; Aral, 2022). Student preferences can therefore inform curriculum design without determining cultural legitimacy. The study makes only two interpretive commitments: learner-curriculum interaction is read through the educational-interface perspective, and curriculum implications remain bounded by living-heritage participation. No additional latent mechanism is inferred from the set-theoretic results.

Methodology

Research Design

The study used a quantitative configurational case-study design. The questionnaire contained 23 five-point items grouped into six curriculum-perception domains: Course Content (4 items), Curriculum Implementation (4), Learning Effect (3), Course Promotion (2), Future Course Suggestions (4), and STEAM Curriculum Design Preferences (6). These composites are treated as content-defined curriculum indices rather than as unidimensional reflective psychological scales because their items represent non-interchangeable curriculum features (Diamantopoulos

& Winklhofer, 2001). Outcome Y measured overall support for the contribution of ICH-STEAM course features to Khoomei learning and transmission. It does not measure observed engagement, achievement, performance, or transmission success. Learning Effect (C) is semantically close to Y; it is retained to reproduce the archived analysis but is not interpreted as an independent mechanism.

Content validity was addressed procedurally. Items were mapped to the intended domains, reviewed for relevance, redundancy, wording, and clarity, and piloted with 12 Khoomei undergraduates, three from each academic year. Ambiguous wording and formatting were revised before formal administration. This supports content relevance and comprehensibility but is not a numerical content-validity index. Item-level reliability outputs for the 75-student dataset were unavailable in the analytic record used for this article, so no alpha, omega, or item-total statistic is claimed.

Sample and Data Collection

The participants were all 75 undergraduates enrolled across four academic years in a specialist Mongolian Khoomei programme at the Inner Mongolia Arts Institute, China. The questionnaire was administered through Wenjuanxing between 28 February and 10 March 2026, and all 75 responses were usable. Because this was a programme census rather than a probability sample, the findings are not presented as representative of all Khoomei learners or ICH programmes. Responses were de-identified before analysis.

The study formed part of doctoral research registered at Universiti Kebangsaan Malaysia (UKM) and approved by the UKM Research Ethics Committee (UKM-JEP; JEP-2025-751; approval coverage 9 October 2025–8 October 2028). All participants provided informed consent. Student data were collected in China under this UKM-approved protocol; no separate local Chinese ethics approval is claimed in the available study record. The dataset is not publicly deposited; de-identified analytic data may be requested subject to confidentiality and institutional requirements. No study-specific external funding is reported, and the authors declare no competing interests.

Analyzing of Data

Raw composite scores were directly calibrated into fuzzy memberships using set-specific anchors for full membership (0.95), crossover (0.50), and full non-membership (0.05) (Ragin, 2008). Exact 0.50 memberships were adjusted to 0.501 for truth-table allocation; raw responses and anchors were unchanged. As shown in Table 1, crossover ties were common: 53 of 75 respondents had at least one condition exactly at its crossover before boundary adjustment. This is treated as a calibration boundary, not as evidence of robustness.

Table 1. Calibration Anchors and Crossover Audit

Code	Set	Raw anchors 0.95 / 0.50 / 0.05	Exact crossover n (%)
A	Course Content	4.58 / 4.25 / 3.00	14 (18.7)
B	Curriculum Implementation	4.75 / 4.00 / 3.00	8 (10.7)
C	Learning Effect	5.00 / 4.00 / 3.00	23 (30.7)
D	Course Promotion	5.00 / 4.00 / 3.00	22 (29.3)
E	Future Course Suggestions	5.00 / 4.25 / 3.00	14 (18.7)
F	STEAM Design Preferences	5.00 / 4.50 / 3.23	13 (17.3)
Y	Overall Support	5.00 / 4.00 / 2.70	18 (24.0)

Note. Operational meanings are defined in Research Design. Anchors are raw composite scores mapped to fuzzy memberships of 0.95, 0.50, and 0.05. Across A-F, 94 memberships were at the crossover; including Y, 112 memberships involved 57 respondents.

Necessity and sufficiency were analysed separately. Set-theoretic necessity was checked against the conventional 0.90 consistency benchmark. NCA then estimated ceiling-regression (CR) and ceiling-envelopment (CE) effect sizes, significance, and bottleneck levels; it tests whether a minimum level of X constrains a desired level of Y, not whether X alone produces Y (Dul, 2016; Dul et al., 2023). fsQCA used a frequency cutoff of 1 and a baseline sufficiency-consistency cutoff of 0.844; intermediate solutions were interpreted by consistency and coverage (Pappas & Woodside, 2021). Raising the cutoff to 0.85 provided a limited threshold-sensitivity check. Because the raw case-level matrix was unavailable for alternative-anchor re-estimation, no calibration-robustness claim is made.

Findings / Results

No individual condition reached 0.90 necessity consistency for higher overall support; the highest values were Course Promotion (D = 0.805) and Future Course Suggestions (E = 0.787). NCA provided a different degree-based result. Future Course Suggestions and STEAM Curriculum Design Preferences showed statistically significant medium effects under both ceiling specifications. Course Content, Curriculum Implementation, and Course Promotion showed statistically significant but small effects. Learning Effect was non-significant under both methods and is not treated as necessary.

Table 2. Necessary Condition Analysis Results

Condition	CR d (p)	CE d (p)	Interpretation
Course Content	0.024 (.018)	0.048 (.018)	Small
Curriculum Implementation	0.021 (.011)	0.042 (.011)	Small
Learning Effect	0.067 (.053)	0.133 (.385)	Not significant
Course Promotion	0.042 (.009)	0.083 (.009)	Small
Future Course Suggestions	0.167 (<.001)	0.250 (<.001)	Medium
STEAM Design Preferences	0.139 (.002)	0.278 (.001)	Medium

Note. CR = ceiling regression; CE = ceiling envelopment. At the maximum observed support level, the descriptive bottlenecks for Future Course Suggestions and STEAM Design Preferences were 50.0% and 41.7%, respectively. These values describe constraints, not sufficient causes.

fsQCA identified three intermediate-solution configurations associated with higher overall support (Table 3). Overall solution consistency was 0.872859 and coverage was 0.510953. Coverage refers to membership in the outcome set and should not be interpreted as the percentage of students caused by the configurations. The result is simple: stronger support was compatible with more than one combination of favourable curriculum perceptions.

Table 3. Configurations Associated with Higher Overall Support

Profile	Configuration	Consistency	Raw coverage	Unique coverage
H1	A*C*D*E*~B; F irrelevant	0.8939	0.3914	0.0520
H2	C*D*E*F*~B; A irrelevant	0.8868	0.3682	0.0288
H3	A*C*D*E*F; B irrelevant	0.8670	0.4302	0.0908

Profile	Configuration	Consistency	Raw coverage	Unique coverage
Overall	Three alternative configurations	0.8729	0.5110	-

Note. A = Course Content; B = Curriculum Implementation; C = Learning Effect; D = Course Promotion; E = Future Course Suggestions; F = STEAM Design Preferences. Present and absent refer to fuzzy-set membership. An irrelevant condition is omitted by minimisation and is not educationally unimportant. Because C is semantically close to Y, its recurrence is not treated as an independent mechanism.

H1 had the highest consistency, whereas H3 had the largest raw and unique coverage; H2 made the smallest unique contribution. The absence of high Curriculum Implementation membership in H1 and H2 does not mean that implementation or technology is harmful; it means only that high B membership was not required in those minimised configurations. When the sufficiency-consistency cutoff increased from 0.844 to 0.85, H1 and H3 remained while H2 dropped out. Solution consistency changed from 0.872859 to 0.873327 and coverage from 0.510953 to 0.482185. This supports limited threshold stability of H1 and H3 but not robustness to alternative calibration anchors.

Discussion

The principal finding is equifinality rather than a dominant predictor. No single condition was universally necessary, while three alternative configurations were associated with stronger support. This pattern is consistent with configurational work in digital ICH and higher education, where similar outcomes can arise through different combinations of conditions (Ye et al., 2025; Zhao & Zhao, 2026). The contribution is therefore not a new universal model but evidence that a single-factor explanation is insufficient for this programme.

The simplest competing explanation must also be considered. Students with a generally favourable view of the curriculum may rate several conditions and the outcome highly, and semantic proximity between Learning Effect and Overall Support may further raise consistency. The cross-sectional self-report design cannot separate these possibilities from substantive curriculum mechanisms. For that reason, C is reported for analytical fidelity but given no independent theoretical weight. The results describe patterns of support, not mechanisms of engagement or proof of learning effectiveness.

The curriculum implication is similarly bounded. Reviews of educational technology show that engagement depends on pedagogy and context (Bond et al., 2020; Bedenlier et al., 2020), and embodied heritage studies show that digital affordances are useful only when they serve the cultural practice and learning task (Pang et al., 2024; Bao et al., 2026). The presence of Future Course Suggestions across all high-support configurations and the NCA results for E and F indicate that student preferences are useful evidence for future design. They do not establish that technology should be expanded indiscriminately.

ICH also prevents a purely learner-preference interpretation. UNESCO (2003) and heritage scholarship place communities and practitioners within the authority structure of safeguarding (Waterton & Smith, 2010; Aral, 2022). Curriculum development should therefore combine learner feedback with teacher, practitioner, and community judgement. Methodologically, NCA and fsQCA are useful together only because they answer different questions: NCA identifies possible minimum-level constraints, whereas fsQCA identifies alternative sufficient combinations. Their joint use does not overcome measurement overlap, crossover sensitivity, or the limits of one programme.

Conclusion

This study examined whether stronger support for an interdisciplinary Khoomei curriculum depended on one dominant condition or on combinations of curriculum perceptions. The evidence supports the latter. No individual condition met the set-theoretic necessity benchmark; NCA identified the clearest medium bottleneck tendencies for Future Course Suggestions and STEAM Design Preferences; and fsQCA identified three alternative high-support configurations, with H1 and H3 retained under a slightly stricter consistency threshold. Because Learning Effect is semantically close to the outcome, it is not interpreted as an independent mechanism. The defensible conclusion is therefore limited: different combinations of curriculum perceptions can be associated with stronger support in this bounded programme. The study does not establish multidimensional engagement, objective learning gains, improved Khoomei performance, or successful heritage transmission.

Recommendations

Curriculum designers should prioritise culturally meaningful content and clear learning purposes before adding technological complexity. Student preferences about future course development and STEAM design should inform revision, but they should be considered alongside teacher, practitioner, and community judgement. Future evaluations should measure curriculum support separately from behavioural engagement, learning performance, and heritage-transmission outcomes. Replication should retain item-level data, test measurement structure, compare defensible calibration anchors, and examine whether the main configurations recur across other programmes and ICH practices.

Limitations

Five limitations bound the findings. First, the study is a census of one 75-student programme and supports analytical rather than statistical generalisation. Second, all focal variables were collected in one self-report survey, leaving common-method and general favourable-response explanations open. Third, the outcome measures support rather than direct engagement or learning, and Condition C is semantically close to Y. Fourth, the available analytic record did not contain item-level reliability outputs or a case-level matrix for reconstructed-C and alternative-calibration analyses; many cases also lay at calibration crossovers. Fifth, the cross-sectional design does not establish temporal ordering, intervention effects, performance improvement, community validation, or long-term transmission. These limitations require restrained interpretation and direct testing in future studies.

References

- Aral, A. E. (2022). Seeing the glass half-empty: Implications for safeguarding intangible cultural heritage through formal education in Türkiye. *International Journal of Heritage Studies*, 28(11-12), 1286-1299. <https://doi.org/10.1080/13527258.2022.2153264>
- Bao, Q., Guo, W., & Jung, E. (2026). Playing with the invisible: Unveiling the drivers of user engagement in serious games for intangible cultural heritage. *npj Heritage Science*, 14, 410. <https://doi.org/10.1038/s40494-026-02637-x>
- Bedenlier, S., Bond, M., Buntins, K., Zawacki-Richter, O., & Kerres, M. (2020). Facilitating student engagement through educational technology in higher education: A systematic review in the field of arts and humanities. *Australasian Journal of Educational Technology*, 36(4), 126-150. <https://doi.org/10.14742/ajet.5477>
- Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A

- systematic evidence map. *International Journal of Educational Technology in Higher Education*, 17, 2. <https://doi.org/10.1186/s41239-019-0176-8>
- Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2), 269-277. <https://doi.org/10.1509/JMKR.38.2.269.18845>
- Dul, J. (2016). Necessary Condition Analysis (NCA): Logic and methodology of "necessary but not sufficient" causality. *Organizational Research Methods*, 19(1), 10-52. <https://doi.org/10.1177/1094428115584005>
- Dul, J., Hauff, S., & Bouncken, R. B. (2023). Necessary condition analysis (NCA): Review of research topics and guidelines for good practice. *Review of Managerial Science*, 17, 683-714. <https://doi.org/10.1007/s11846-023-00628-x>
- 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>
- Kahu, E. R., & Nelson, K. (2018). Student engagement in the educational interface: Understanding the mechanisms of student success. *Higher Education Research & Development*, 37(1), 58-71. <https://doi.org/10.1080/07294360.2017.1344197>
- Pang, W. Y. J., Leung, B. W., & Cheng, L. (2024). The motivational effects and educational affordance of serious games on the learning of Cantonese opera movements. *International Journal of Human-Computer Interaction*, 40(6), 1455-1464. <https://doi.org/10.1080/10447318.2022.2112567>
- Pappas, I. O., & Woodside, A. G. (2021). Fuzzy-set Qualitative Comparative Analysis (fsQCA): Guidelines for research practice in information systems and marketing. *International Journal of Information Management*, 58, 102310. <https://doi.org/10.1016/j.ijinfomgt.2021.102310>
- Perignat, E., & Katz-Buonincontro, J. (2019). STEAM in practice and research: An integrative literature review. *Thinking Skills and Creativity*, 31, 31-43. <https://doi.org/10.1016/j.tsc.2018.10.002>
- Ragin, C. C. (2008). Measurement versus calibration: A set-theoretic approach. In J. M. Box-Steffensmeier, H. E. Brady, & D. Collier (Eds.), *The Oxford handbook of political methodology* (pp. 174-198). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199286546.003.0008>
- Sundberg, J., Lindblom, B., & Hefele, A. M. (2023). Voice source, formant frequencies and vocal tract shape in overtone singing: A case study. *Logopedics Phoniatrics Vocology*, 48(2), 75-87. <https://doi.org/10.1080/14015439.2021.1998607>
- UNESCO. (2003). *Convention for the safeguarding of the intangible cultural heritage*. UNESCO.
- Waterton, E., & Smith, L. (2010). The recognition and misrecognition of community heritage. *International Journal of Heritage Studies*, 16(1-2), 4-15. <https://doi.org/10.1080/13527250903441671>
- Ye, X., Ruan, Y., Xia, S., & Gu, L. (2025). Adoption of digital intangible cultural heritage: A configurational study integrating UTAUT2 and immersion theory. *Humanities and Social Sciences Communications*, 12, 23. <https://doi.org/10.1057/s41599-024-04222-8>
- Zhao, Z., & Zhao, M. (2026). Examining Chinese university students' engagement through academic self-efficacy, teacher-student rapport, and flourishing: A PLS-SEM and fsQCA. *European Journal of Education*, 61(3), e70816. <https://doi.org/10.1111/ejed.70816>

Sustaining the Intangible: The Culture-Technology-Ethics Symbiotic Model (CTESM) and Eight-Dimensional Assessment Framework (EDAF) for Heritage Education

Author Sun Yanhu¹, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia .E-mail:p123627@siswa.ukm.edu.my

Author Nuriah Abd Majid^{1*}, Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia, E-mail:nuriah@ukm.edu.my

*** Corresponding Author** Nuriah Abd Majid .E-mail:nuriah@ukm.edu.my

Abstract

The rapid digitalization of cultural heritage in higher education often generates ethical tensions between technocentrism and the commercial commodification of ancient traditions. To address this structural dilemma and support the sustainable future of Intangible Cultural Heritage (ICH), this paper proposes the Culture–Technology–Ethics Symbiotic Model (CTESM). Developed through a multi-method empirical investigation combining Delphi studies, thematic analysis, and quantitative diagnostics of Mongolian Khoomei education, CTESM establishes a tripartite hierarchy: cultural authenticity forms the immutable core; technologies, including AI-based acoustic diagnostics and virtual reality, function as instrumental amplifiers; and an ethical layer safeguards community sovereignty and intellectual property rights. The paper also introduces the Eight-Dimensional Assessment Framework (EDAF) to operationalize this symbiosis. EDAF replaces conventional linear examinations with triangulated, multi-stakeholder evaluation integrating technical precision, heritage mentor review, and social impact indicators. By moving from fragmented teaching toward holistic symbiotic governance, the study offers an ethically bounded policy blueprint for integrating STEAM innovation without eroding indigenous cultural identities. CTESM and EDAF therefore provide a bounded design proposition for future implementation and validation, supporting the living transmission of heritage as dynamic academic and cultural practice rather than static preservation.

Keywords: *assessment, intangible cultural heritage, Mongolian Khoomei, STEAM, technology-enhanced learning*

Introduction

Technology-enhanced learning (TEL) is established across higher education, yet the central question is not whether technology is present but what it improves, for whom, and on what evidence (Kirkwood & Price, 2014). This distinction is especially important for intangible cultural heritage (ICH), where learning includes embodied technique, situated judgement, cultural meaning, and relationships of authority. The UNESCO (2003) Convention therefore treats safeguarding as living transmission with community participation, not as documentation alone. Heritage scholarship likewise shows that institutional selection and representation can reshape what counts as legitimate heritage (Kirshenblatt-Gimblett, 2004; Smith, 2006).

Mongolian Khoomei makes these tensions concrete. Digital recording, acoustic visualisation, immersive media, and AI-assisted feedback can make selected vocal features more visible and repeatable; acoustic research confirms that overtone focusing can be analysed technically

(Bergevin et al., 2020). However, measurable precision cannot by itself establish cultural legitimacy, appropriate reuse, or transmission quality. Critical TEL, TPACK, and ICAP explain how technology can support teacher knowledge and active engagement (Chi & Wylie, 2014; Mishra & Koehler, 2006), while heritage education stresses living practice, cultural authority, and curricular recontextualisation (Aral, 2022; Labrador, 2022). These literatures rarely specify how technical, pedagogical, cultural, and ethical judgements should be ordered when they conflict.

The problem also creates a bounded connection to technology-enhanced language learning (TELL). Khoomei is not a second-language intervention, and this study does not claim language-learning outcomes. The relevant transfer is mechanistic: computer-assisted pronunciation research shows how visual/acoustic feedback can support noticing and practice, while language-learning research increasingly emphasises transparent, human-governed AI (Amrate & Tsai, 2025; Pérez-Paredes, 2026). The same caution applies here: automated feedback is useful only when it remains answerable to pedagogical and culturally authorised human judgement.

Accordingly, the study asks: (1) what curriculum conditions are prioritised for integrated ICH-STEAM education; (2) how should culture, technology, and ethics be related in technology-mediated Khoomei learning; and (3) how can assessment represent technical achievement without reducing heritage learning to machine-measurable performance? The study answers these questions through two linked outputs. CTESM defines culture as the constitutive core, technology as an enabling system, and ethics as a regulatory boundary. EDAF translates that relationship into a multi-source assessment architecture. Neither is presented as a universally validated causal model.

Literature Review

Technology-Enhanced Learning and TELL Boundary

A parsimonious theoretical base is sufficient for the present problem. Critical TEL requires enhancement claims to be tied to learning activity and evidence rather than to the novelty of a tool (Kirkwood & Price, 2014). TPACK clarifies the integration of technology, pedagogy, and content in teacher knowledge (Mishra & Koehler, 2006), while ICAP distinguishes passive technology use from active, constructive, and interactive engagement (Chi & Wylie, 2014). These perspectives support acoustic and AI tools when they improve noticing, practice, explanation, feedback, or reflection; they do not decide who may authorise culturally consequential standards. Recent human-AI work strengthens this boundary by treating agency as distributed across learners, teachers, systems, and institutions rather than located in the AI alone (Cukurova, 2025).

Heritage, Authority, and Ethical Mediation

Heritage education changes the status of knowledge when living practice enters a university curriculum. UNESCO (2003) places communities, groups, and individuals within safeguarding; Smith (2006) treats heritage as a cultural process rather than a self-evident object; and Kirshenblatt-Gimblett (2004) shows that heritage-making is itself a form of cultural production. Formal education can therefore support transmission while also standardising, selecting, or marginalising practice (Aral, 2022; Labrador, 2022). Deng (2022) is useful here because curriculum is not the simple transfer of content: educational knowledge must be selected, transformed, and organised for purposeful learning. For Khoomei, acoustic explanation or digital representation becomes educationally legitimate only when it remains connected to embodied practice and recognised cultural meaning. Digital and immersive heritage studies report gains in access and engagement, but recent reviews also identify limited evidence for sustained learning and long-term cultural outcomes (Zhao et al., 2025; Zhang & Peng, 2026). Human-centred AI guidance adds human oversight, transparency, privacy, fairness, and

contestability (UNESCO, 2021), while CARE emphasises collective benefit and authority to control culturally significant data (Carroll et al., 2020).

Assessment and Model Logic

Assessment must align with educational aims without collapsing heterogeneous forms of judgement into a single metric. Constructive alignment connects outcomes, learning activity, and assessment (Biggs, 1996). Formative assessment and feedback literacy emphasise evidence that learners can interpret and use (Black & Wiliam, 1998; Carless & Boud, 2018), while long-term assessment should develop learners' capacity to make future judgements rather than merely certify short-term performance (Boud & Falchikov, 2006). The resulting gap is relational: existing work explains technology integration, active learning, living heritage, AI ethics, curriculum design, and formative assessment separately, but is less explicit about how these authorities should be ordered when technically measurable performance, pedagogical usefulness, cultural legitimacy, and ethical permission diverge. CTESM addresses that ordering problem; EDAF provides the corresponding assessment logic.

Methodology

Research Design

The study used a sequential four-strand mixed-method design for framework construction rather than intervention-effect testing. Empirical work involved China-based experts, educators, Khoomei instructors, and students; the doctoral study was registered at Universiti Kebangsaan Malaysia (UKM). The strands had non-equivalent roles: Delphi established normative curriculum dimensions; interviews interpreted enactment and ethical boundaries; AHP diagnosed priorities in the existing Khoomei programme; and NCA/fsQCA provided learner-side configurational evidence. Outputs were integrated through a proportional-evidence rule: convergent findings informed core propositions, whereas single-strand or literature-only elements were retained as bounded design propositions. Table 1 summarises the evidence strands and their inferential boundaries.

Sample and Data Collection

The Delphi panel comprised 18 participants: six STEAM/educational-technology specialists, eight ICH/curriculum scholars or policy specialists, and four heritage bearers/practitioners. Eligibility required at least five years of relevant experience, and all panellists completed three rounds. The interview strand purposively selected eight university ICH educators and practitioner-teachers with 13–43 years of relevant experience. Interviews were conducted online and in person in Inner Mongolia, Beijing, and Shanghai on 10–17 January 2026, producing 663 minutes of audio and approximately 110,000 words of verbatim Chinese transcription. Coding was conducted in the source language; selected excerpts were translated after theme development. Ten university-based Khoomei instructors completed the AHP questionnaire containing four criteria and 19 elements. The learner survey invited all 75 undergraduates in the programme across four academic years; 75 valid responses were obtained. A 12-student pilot checked clarity. Ethical approval was granted by the UKM Research Ethics Committee (JEP-2025-751; 9 October 2025–8 October 2028), with China as the approved study site. Participants gave informed consent; data were de-identified and stored under the approved confidentiality conditions.

Instrument Development and Quality Control

Instrument development followed the same staged logic as framework construction. Delphi items were iteratively refined across three rounds using item-transition records, quantitative ratings, and expert feedback. The interview guide then examined enactment, authenticity, technological mediation, assessment, faculty capacity, and institutional support rather than re-testing Delphi ratings. The AHP instrument operationalised four criteria and 19 elements from the existing Khoomei curriculum, while the learner questionnaire translated curriculum

conditions into calibrated perception sets; a 12-student pilot checked wording and clarity before administration. These procedures support content relevance and procedural reliability, but they do not constitute psychometric validation of CTESM or EDAF.

Table 1. Evidence Strands, Quality Procedures, and Inferential Boundaries

Strand	Sample	Analysis / quality	Role / boundary
Delphi	18 experts; 3 rounds	Item-transition audit; central tendency, dispersion, support, concordance; 100% round completion.	Normative curriculum evidence; not intervention-effect validation.
Interviews	8 ICH educators/practitioner-teachers; 663 min	Codebook thematic analysis; 392 concepts -> 116 categories -> 17 intermediate/core categories -> 5 themes; kappa=.83.	Implementation, authenticity, ethics, and support; not prevalence estimates.
AHP	10 Khoomei instructors	4 criteria / 19 elements; individual and aggregate CR<.10; geometric-mean aggregation.	Existing-programme priority diagnosis; not ideal future allocation.
NCA / fsQCA	75 undergraduates; 75 valid responses	12-student clarity pilot; calibrated sets; necessity/sufficiency; .85 threshold sensitivity.	Learner-side perception patterns; not causal learning or transmission effects.

Analyzing of Data

Delphi analysis combined item transitions, central tendency, dispersion, same-direction support, and panel-level concordance rather than treating a single coefficient as proof of consensus. Interview analysis followed a codebook thematic process that remained open to inductive codes and negative cases. It generated 392 initial concepts across 1,431 coded references, consolidated into 116 categories, 17 intermediate/core categories, and five cross-case themes. A 10% calibration subset produced Cohen's kappa=.83; pre-adjudication theme-level agreement ranged from 75.42% to 91.97%, followed by codebook refinement and participant review. AHP matrices were checked individually and after geometric-mean aggregation; all aggregate consistency ratios were below .10 (Saaty, 2008). Student responses were treated as calibrated perception sets rather than latent scales. NCA assessed sample-specific bottlenecks (Dul, 2016); fsQCA tested necessity and sufficient configurations using set-theoretic principles of conjunction, equifinality, and asymmetry (Schneider & Wagemann, 2012). A crossover membership of .50 was adjusted to .501 for truth-table allocation, and robustness was examined at a stricter .85 consistency threshold. These procedures support bounded diagnostic inference, not causal effects or long-term transmission outcomes.

Findings / Results

Cross-Method Evidence

The four strands converged on one central pattern: culture and pedagogical enactment organised the curriculum more strongly than technology alone. Delphi consensus retained cultural grounding, interdisciplinary design, learner-centred pedagogy, project-based implementation, assessment alignment, and implementation barriers as core concerns. Interviews showed that authenticity was not a detachable content item; participants linked it to who teaches, how practice is sequenced, which technologies are appropriate, and whose judgement governs assessment. They also identified faculty capacity, heritage-bearer participation, infrastructure,

and institutional support as enabling conditions. Table 2 summarises the principal quantitative evidence.

AHP and Learner-Side Configurations

AHP quantified the existing programme's priorities. Cultural heritage received the highest criterion weight (.4982), followed by teaching implementation (.3086), assessment (.1308), and STEAM integration (.0624). The highest global element weights were depth of Khoomei technique (.2463), course connection/progression (.1484), and traditional cultural content (.1305). These values diagnose the current programme; they do not prescribe an ideal future allocation. Learner-side analysis reinforced the importance of alignment. NCA identified future course planning and STEAM-compatible design as the largest perceived constraints on high overall course support, with ceiling-envelopment effect sizes of .250 and .278. No single fsQCA condition reached the .90 necessity threshold. Three higher-support configurations were identified: H1 (consistency .8939; raw coverage .3914), H2 (.8868; .3682), and H3 (.8670; .4302). H1 was most consistent; H3 had the highest raw and unique coverage; H2 did not survive the stricter .85 threshold. The stable lower-support profile L1 combined broad weakness in content, implementation, perceived support, promotion, and future planning (consistency .9497; raw coverage .3358). These patterns indicate conjunctural, asymmetric perceptions rather than a universal technological recipe.

Table 2. Key Quantitative Findings and Their Interpretive Boundaries

Evidence	Key result	Interpretation
AHP criteria	Cultural heritage .4982; teaching implementation .3086; assessment .1308; STEAM .0624.	Culture and pedagogical enactment anchor the existing programme; low STEAM weight is diagnostic, not normative.
AHP top elements	Khoomei technique .2463; course progression .1484; traditional cultural content .1305.	Embodied practice, sequencing, and cultural meaning dominate individual priorities.
NCA	Future planning $d=.250$; STEAM-compatible design $d=.278$.	Sample-specific perceived bottlenecks; not treatment effects.
fsQCA - higher support	H1 .8939/.3914; H2 .8868/.3682; H3 .8670/.4302 (consistency/raw coverage).	Multiple sufficient profiles; H2 was threshold-sensitive.
fsQCA - lower support	L1 .9497/.3358.	Broad deficits co-occurred with lower support; asymmetry matters.

The Culture-Technology-Ethics Symbiotic Model

CTESM synthesises the findings as an asymmetric governance model (Figure 1). The core cultural system contains Khoomei technique, traditional cultural knowledge, situated meaning, and embodied mentorship. The technological enabling system includes acoustic analysis, recording, immersive soundscapes, and AI-assisted feedback as optional forms of explanation, practice, and reflection. The ethical regulatory system governs consent, attribution, provenance, intellectual-property sensitivity, cultural authority, and anti-commodification safeguards. Culture therefore defines legitimate educational ends; technology enables bounded mediation; ethics constrains both institutional and digital intervention. 'Symbiotic' denotes interaction, not equal authority.

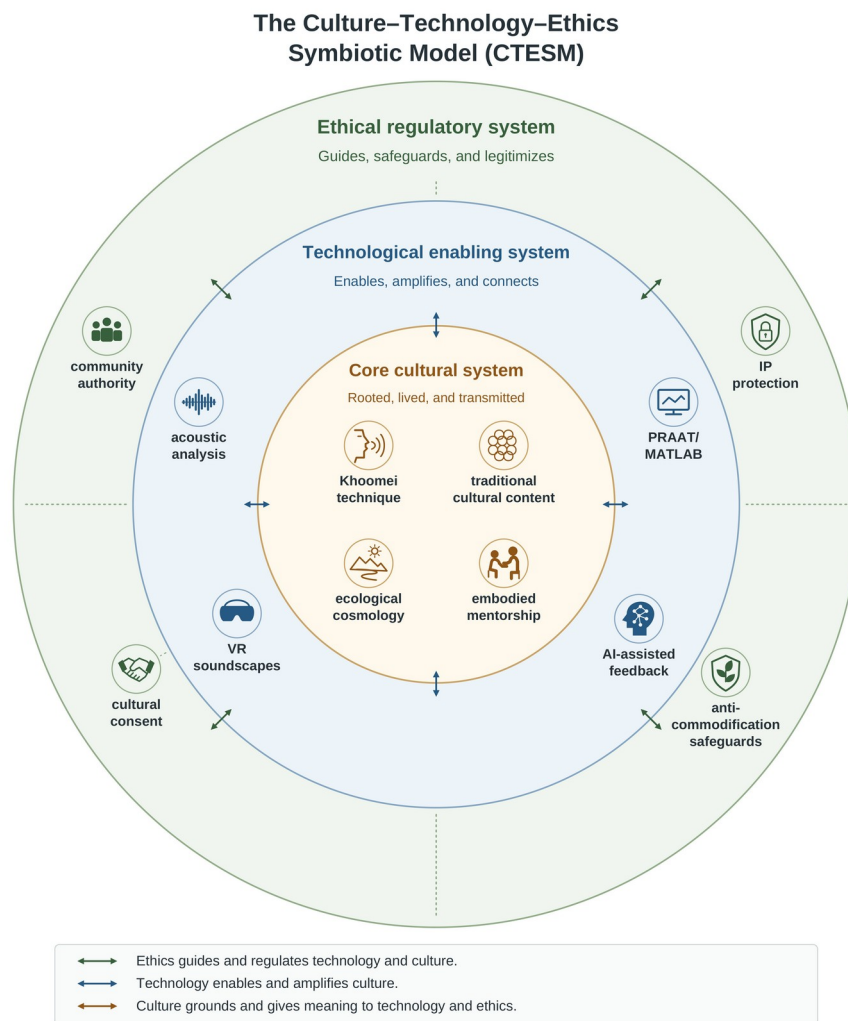


Figure 1. Culture-Technology-Ethics Symbiotic Model (CTESM)

The Eight-Dimensional Assessment Framework

EDAF operationalises the same logic for assessment (Figure 2). Its eight dimensions are technical precision, cultural authenticity, ethical compliance, self and peer evaluation, instructor evaluation, interdisciplinary application, societal impact, and continuous feedback/iterative optimisation. The dimensions are deliberately heterogeneous because they answer different questions. Acoustic measures can strengthen evidence for technical precision, but they cannot independently establish cultural authenticity, ethical legitimacy, or responsible public transmission. EDAF therefore functions as a multi-source assessment architecture rather than a single composite score or psychometrically validated scale.

The Eight-Dimensional Assessment Framework (EDAF)



Figure 2. Eight-Dimensional Assessment Framework (EDAF)

Discussion

The findings support a shift from technology integration to governed mediation. TEL literature already warns that technology does not enhance learning by default (Kirkwood & Price, 2014). CTESM adds a heritage-specific ordering rule: the legitimacy of an educational technology depends not only on whether it improves activity or feedback, but also on whether the learning goal is culturally authorised and the mediation is ethically permissible. This extends TPACK and ICAP without replacing them. Those frameworks remain useful for teacher knowledge and cognitive engagement; CTESM addresses the additional problem of decision authority in culturally consequential learning.

The model also reframes authenticity. Formal education inevitably recontextualises heritage by selecting content, sequencing practice, and making outcomes assessable (Aral, 2022; Deng, 2022). A fixed 'authentic' rubric risks freezing living variation, while unrestricted adaptation risks detaching innovation from recognised transmission. CTESM therefore treats authenticity as situated and reviewable rather than algorithmically settled. University standards should remain transparent, but practitioners and relevant communities need meaningful authority to question representation, reuse, and assessment criteria. This is consistent with living-heritage principles and the view of heritage as an authorised cultural process (UNESCO, 2003; Smith, 2006).

For AI-supported learning, the implication is augmentation rather than substitution. Human-centred AI can extend feedback and analytic visibility, but agency must remain distributed and contestable (Cukurova, 2025; UNESCO, 2021). In Khoomei, a spectrogram can reveal overtone structure and an AI system can flag selected acoustic features, yet neither can determine whether stylistic nuance, context, consent, or reuse are culturally appropriate. The same principle is relevant to TELL: CAPT and AI can support repeated practice and noticing, but the present study does not establish language-learning effects. Its transfer contribution is the governance principle that technically valid feedback should remain subordinate to educational purpose and accountable human judgement.

EDAF contributes by separating measurement from judgement. Constructive alignment requires assessment to reflect intended learning (Biggs, 1996), while formative and long-term assessment require feedback that learners can use and judgement they can carry forward (Black & Wiliam, 1998; Boud & Falchikov, 2006; Carless & Boud, 2018). EDAF applies these principles to ICH by making technical, cultural, ethical, reflective, interdisciplinary, societal, and iterative evidence visible without assuming that they form one latent construct. This reduces the risk that machine-readable signals become proxies for the whole learning outcome.

The practical implication is a sequence of design decisions. Teachers and curriculum teams should first define the cultural and educational purpose, then select a technology for a specific learning problem, specify which judgements remain human or community-governed, and finally align assessment with those boundaries. Institutions should support dual or shared mentorship, faculty development in both digital competence and cultural ethics, and governance for recordings, AI training data, attribution, provenance, reuse, correction, and withdrawal. Policy should therefore evaluate technology-enhanced heritage programmes by the quality of mediated learning and accountable transmission, not by technology quantity or public reach alone.

Conclusion

This study argues that technology-enhanced ICH education is best understood as governed mediation. Across expert consensus, educator interpretation, instructor priorities, and learner-side configurations, the evidence did not support technology as an autonomous driver of educational quality. CTESM therefore orders three functions: cultural authority defines legitimate ends, technology supports bounded mediation, and ethics regulates intervention. EDAF carries this ordering into assessment by preventing technically measurable performance from substituting for cultural, ethical, reflective, interdisciplinary, societal, and feedback evidence. The contribution is a compact design logic for technology-mediated heritage learning, not proof of a universally effective intervention.

Recommendations

Future implementation should co-design CTESM and EDAF with heritage practitioners, teachers, students, and educational technologists before fixing rubrics or automated feedback. Pilot courses should compare technology-supported and lower-technology activities using performance, reflection, cultural-authority review, and learning-process evidence. Institutions should document provenance and permitted uses for recordings and voice data, retain human review for culturally consequential decisions, and establish correction or withdrawal routes. For ICTELL/TELL research, the model should be tested only where a genuine language-learning construct exists; transfer should focus on feedback, agency, multimodal noticing, and ethical mediation rather than treating Khoomei as a language outcome.

Limitations

The study has six main limitations. First, it is a bounded case centred on one university Khoomei programme, so transfer to other ICH domains requires contextual validation. Second, expert

and instructor samples were purposive rather than statistically representative. Third, interview evidence reflects participant interpretation rather than observed classroom behaviour; translated excerpts did not undergo a documented independent back-translation. Fourth, the student survey measured perceptions, and one condition was conceptually close to the overall evaluative outcome; NCA/fsQCA results therefore remain diagnostic rather than causal. Fifth, AHP weights describe perceived priorities in the existing programme, not optimal curriculum proportions. Sixth, CTESM and EDAF have not yet been tested as a completed classroom intervention or validated as a scoring instrument. Longitudinal, cross-site, and co-designed implementation studies are needed before claims about learning gains, transmission outcomes, or general applicability can be made.

Ethics and Transparency

Ethics approval and consent. Ethical approval was granted by the UKM Research Ethics Committee (JEP-2025-751; 9 October 2025–8 October 2028), with China as the approved study site. All participants received study information, participated voluntarily, and provided informed consent; direct identifiers were removed from analytical materials. **Data availability and confidentiality.** Raw interview audio, transcripts, and participant-level specialist data are not publicly deposited because the small cohorts create a risk of deductive identification. De-identified aggregate materials, instruments, and analysis summaries may be requested from the corresponding author where release is compatible with the approved ethics protocol and participant confidentiality. **Funding.** This research received no external funding. **Conflict of interest.** The authors declare no conflict of interest.

References

- Amrate, M., & Tsai, P.-H. (2025). Computer-assisted pronunciation training: A systematic review. *ReCALL*, 37(1), 22–42. <https://doi.org/10.1017/S0958344024000181>
- Aral, A. E. (2022). Seeing the glass half-empty: Implications for safeguarding intangible cultural heritage through formal education in Türkiye. *International Journal of Heritage Studies*, 28(11–12), 1286–1299. <https://doi.org/10.1080/13527258.2022.2153264>
- Bergevin, C., Narayan, C., Williams, J., Mhatre, N., Steeves, J. K. E., Bernstein, J. G. W., & Story, B. (2020). Overtone focusing in biphonic Tuvan throat singing. *eLife*, 9, e50476. <https://doi.org/10.7554/eLife.50476>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347–364. <https://doi.org/10.1007/BF00138871>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399–413. <https://doi.org/10.1080/02602930600679050>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodríguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for Indigenous data governance. *Data Science Journal*, 19, 43. <https://doi.org/10.5334/dsj-2020-043>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>

- Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, 56(2), 469-488. <https://doi.org/10.1111/bjet.13514>
- Deng, Z. (2022). Powerful knowledge, educational potential and knowledge-rich curriculum: Pushing the boundaries. *Journal of Curriculum Studies*, 54(5), 599-617. <https://doi.org/10.1080/00220272.2022.2089538>
- Dul, J. (2016). Necessary condition analysis (NCA): Logic and methodology of 'necessary but not sufficient' causality. *Organizational Research Methods*, 19(1), 10-52. <https://doi.org/10.1177/1094428115584005>
- Kirkwood, A., & Price, L. (2014). Technology-enhanced learning and teaching in higher education: What is 'enhanced' and how do we know? A critical literature review. *Learning, Media and Technology*, 39(1), 6-36. <https://doi.org/10.1080/17439884.2013.770404>
- Kirshenblatt-Gimblett, B. (2004). Intangible heritage as metacultural production. *Museum International*, 56(1-2), 52-65. <https://doi.org/10.1111/j.1350-0775.2004.00458.x>
- Labrador, A. M. (2022). Integrating ICH and education: A review of converging theories and methods. *International Journal of Intangible Heritage*, 17, 17-36. <https://doi.org/10.35638/ijih.2022.17..001>
- 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>
- Pérez-Paredes, P. (2026). Corpus linguistics and AI in the reconfiguration of language learning ecologies. *Language Teaching*, Advance online publication. <https://doi.org/10.1017/S0261444825101109>
- Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1), 83-98. <https://doi.org/10.1504/IJSSCI.2008.017590>
- Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge University Press.
- Smith, L. (2006). *Uses of heritage*. Routledge.
- UNESCO. (2003). *Convention for the safeguarding of the intangible cultural heritage*. <https://ich.unesco.org/en/convention>
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>
- Zhang, J., & Peng, F. (2026). Co-creation of immersive learning for cultural heritage education: A scoping review. *Heritage*, 9(5), 192. <https://doi.org/10.3390/heritage9050192>
- Zhao, Y., Li, Y., Dai, T., Sadini, C., Wu, X., Jiang, W., Li, J., Zhu, K., Zhai, B., Li, M., & Ray, L. C. (2025). Virtual reality in heritage education for enhanced learning experience: A mini-review and design considerations. *Frontiers in Virtual Reality*, 6, 1560594. <https://doi.org/10.3389/frvir.2025.1560594>

Designing CEFR-Aligned Content-Based Speaking Materials with SDG 4 Themes for Islamic Higher Education

Kalsum^{1,2,*}, Rahimah Saimin¹, Engku Muhammad Syafiq Engku Shafruddin¹

¹ Faculty of Major Language Studies, Universiti Sains Islam Malaysia, Malaysia

² English Education Program, Institut Agama Islam Negeri Parepare (IAIN Parepare), Indonesia

* Corresponding author: kalsum@iainpare.ac.id

Abstract. Speaking instruction in Indonesian Islamic higher education (Perguruan Tinggi Keagamaan Islam, PTKI) is often constrained by generic, accuracy-focused materials that offer few opportunities for dialogic interaction, argumentation, and reflection, and that are rarely aligned with an explicit proficiency framework. Although national policy encourages integrating sustainability values into the curriculum, ready-to-use speaking materials linking these values to measurable communicative competence remain scarce. This paper reports the design and development of a CEFR-aligned, content-based speaking module that uses selected SDG 4 (Quality Education) themes—equitable access, relevant skills, and global citizenship (targets 4.3, 4.4, and 4.7)—as meaningful communicative content rather than as content to be tested. Following a Design and Development Research orientation operationalized through the ADDIE model, and grounded in Content-Based Instruction, Communicative Language Teaching, and theories of speaking development, the module organizes three thematic units that progress toward extended interaction and reflective speaking at CEFR B1–B2, with a sub-skill scaffolding scheme, an illustrative prototype unit, and a CEFR-aligned analytic rubric. Designed for technology-enhanced and blended delivery, the module offers a transferable prototype and an assessment architecture on which AI-assisted speaking feedback can be built.

Keywords: CEFR; content-based instruction; SDG 4 Quality Education; speaking materials; technology-enhanced language learning

Abstrak. Pembelajaran berbicara di pendidikan tinggi keislaman (PTKI) di Indonesia kerap terkendala oleh bahan ajar yang umum dan berfokus pada ketepatan linguistik, sehingga sedikit memberi ruang bagi interaksi dialogis, penghujahan, dan refleksi, serta jarang diselaraskan dengan kerangka kompetensi yang eksplisit. Meskipun kebijakan nasional mendorong pengintegrasian nilai kelestarian ke dalam kurikulum, bahan ajar berbicara siap pakai yang menautkan nilai tersebut dengan kompetensi komunikatif yang terukur masih langka. Artikel ini melaporkan perancangan dan pengembangan modul berbicara berbasis kandungan yang selaras dengan CEFR, yang menggunakan tema terpilih SDG 4 (Pendidikan Berkualitas), akses yang setara, keterampilan yang relevan, dan kewarganegaraan global (target 4.3, 4.4, dan 4.7) sebagai kandungan komunikatif yang bermakna, bukan materi yang diujikan. Dengan orientasi Design and Development Research yang dioperasikan melalui model ADDIE, serta berlandaskan Content-Based Instruction, Communicative Language Teaching, dan teori pengembangan keterampilan berbicara, modul ini menata tiga unit tematik menuju interaksi yang diperluas dan berbicara reflektif pada tingkat CEFR B1–B2, dilengkapi skema scaffolding sub-keterampilan, unit prototipe, dan rubrik analitik berbasis CEFR. Dirancang untuk penyampaian berbantuan teknologi dan blended, modul ini menawarkan prototipe yang dapat dialihkan dan arsitektur penilaian yang dapat dikembangkan untuk umpan balik berbicara berbantuan AI.

Kata kunci: bahan ajar berbicara; CEFR; content-based instruction; pembelajaran bahasa berbantuan teknologi; SDG 4 Pendidikan Berkualitas

Introduction

Education is widely recognized as a key driver of sustainable development. Through Sustainable Development Goal 4 (SDG 4), the United Nations (2015) frames inclusive, equitable, and quality education as a foundation for lifelong learning and global citizenship, and UNESCO (2017) underscores the role of higher education in cultivating critical awareness, ethical responsibility, and intercultural competence. In Indonesia, this commitment is reflected in higher-education policies, including Permendikbudristek No. 53/2023 and the Merdeka Belajar–Kampus Merdeka framework, which encourage the integration of sustainability values into university curricula, including in Islamic Higher Education Institutions (PTKI). Universities are increasingly expected to contribute to the sustainable development agenda not only through what they teach, but through the capabilities and dispositions they cultivate (Chankseliani & McCowan, 2021).

Despite these policy directions, the integration of sustainability values into classroom practice remains partial and uneven, particularly in language education (Cordova, 2024; Nasrudin et al., 2025; Sofiana & Haryati, 2025). Read together, these studies converge on a common pattern rather than diverging in their conclusions: quality-education themes are documented as present in curricula and institutional reporting, yet they are seldom carried through into purposeful classroom language use. In tertiary English language teaching, such themes are frequently treated as peripheral topics rather than as meaningful communicative content, and many materials remain generic and decontextualized (Ambariyanto & Utama, 2020; Bui et al., 2024). Reporting at the institutional level has thus tended to outpace pedagogical operationalization in the classroom.

This imbalance is most evident in speaking, where instruction often prioritizes linguistic accuracy and structural control over dialogic interaction, critical discussion, and reflective expression (Robah & Anggrisia, 2023). Within PTKI contexts, including IAIN Parepare, speaking syllabi rarely make explicit connections to SDG 4 educational issues, despite the strategic potential of PTKIs to integrate language learning with ethical and humanitarian values grounded in Islamic educational philosophy (Dariah et al., 2016; Nasrudin et al., 2025). A parallel gap persists in design: issue-based speaking instruction is seldom aligned with a systematic proficiency framework such as the Common European Framework of Reference for Languages (CEFR), so there is limited evidence of how meaningful content contributes to measurable speaking competence (Nahdi et al., 2024). Synthesized across these strands, two recurring limitations stand out: the CEFR appears predominantly as a scoring benchmark rather than a design driver, and sustainability appears as reading or awareness input rather than as communicative content; a systematic review reaches the same converging conclusion, namely that adoption is assessment-driven while sustainability and inclusion remain peripheral rather than structurally embedded (Kalsum et al., in press).

These two limitations are increasingly consequential as speaking instruction moves into technology-enhanced and blended environments, in which recorded, asynchronous, and AI-supported speaking tasks are becoming common. Digital delivery, however, cannot compensate for materials that lack a principled task-and-assessment architecture; without an explicit construct and criterion-referenced descriptors, technology risks reproducing weak, accuracy-focused practice on a scale. Taken together, these observations point to a concrete design need rather than to a further call for awareness: a CEFR-aligned, content-based speaking module in which selected SDG 4 themes function as meaningful communicative content, and whose architecture is ready for technology-enhanced delivery and future AI-assisted feedback. This paper addresses that need. It reports the design and development of such a module for the

English Education Program at IAIN Parepare and presents it as an innovative TESL practice within technology-enhanced language learning and future learning environments.

The module forms part of a larger Design and Development Research study; the present paper concentrates on the analysis, design, and development phases, while the classroom evaluation of effectiveness is reported separately. Conceptually, the study positions SDG 4 themes not as the primary instructional objective but as thematic content through which speaking competence is developed; speaking competence remains the primary outcome. The paper therefore presents the pedagogical principles, the way SDG 4 themes are operationalized as communicative content, the unit architecture and sub-skill scaffolding, an illustrative prototype unit, the CEFR-aligned analytic rubric, and the expert-review and refinement process, so that other PTKIs may adapt the prototype when responding to national sustainability mandates.

Methodology

Research Design

The study adopts Design and Development Research (DDR) as its overarching methodological orientation. DDR is appropriate for studies that aim to design, develop, and evaluate instructional products within authentic educational settings while also generating knowledge about the design process and its pedagogical outcomes (Richey & Klein, 2007; McKenney & Reeves, 2018). Rather than prioritizing statistical generalization, DDR emphasizes contextual relevance, iterative refinement, and practical validity, which makes it well suited to the development of a context-sensitive speaking module. DDR is operationalized through the ADDIE model Analysis, Design, Development, Implementation, and Evaluation (Branch, 2009; Plomp & Nieveen, 2013). For analytical clarity, this paper reports the first three stages (Analysis, Design, and Development), which together yield the prototype module; the Implementation and Evaluation stages, conducted through a quasi-experimental classroom study, are reported elsewhere.

Sample and Data Collection

The analysis stage drew on a needs analysis comprising a documentary review of the compulsory speaking-course syllabi (Rencana Pembelajaran Semester, RPS) and their assessment rubrics, together with consultations with the lecturers who teach these courses in the English Education Program at IAIN Parepare. Because the program has a small, defined teaching team for the speaking courses, the consultation represents the full relevant population of speaking-course lecturers at the research site rather than a sample drawn from a larger pool. This analysis indicated three recurring conditions: thematic content was drawn from every day and institutionally familiar topics rather than from educationally substantive issues; the rubrics evaluated speaking against general criteria without reference to CEFR descriptors of interaction, coherence, or range; and tasks with genuine communicative potential were often realized as written summaries rather than as assessed oral performance. These findings provided empirical grounding for the design that follows.

Ethical considerations. Participation in the lecturer consultations and in the expert-review panel was voluntary and based on informed consent; participants were briefed on the purpose of the study and on how their input would be used and were free to withdraw. No student participants or student data were involved at this design and development stage. Institutional permission for the documentary review of course materials and for the consultations was obtained from the English Education Program, IAIN Parepare.

Analyzing Data

The RPS documents and rubrics were analyzed against CEFR descriptors of spoken performance to establish the extent of proficiency alignment, while the consultation notes were synthesized into a set of design requirements linking the documented needs to concrete task and assessment decisions. The design itself is theoretically grounded rather than intuitive. Content-Based Instruction supplies the rationale for using meaningful content as a vehicle for language development, while Communicative Language Teaching, incorporating task-based principles, explains how learners use language through interaction (Richards, 2006). The design of speaking tasks draws on established accounts of speaking development Levelt's (1989) model of speech production, Long's (1996) interaction hypothesis, Swain's (2005) output hypothesis, and Segalowitz's (2010) account of fluency which explain why conceptually rich content and dialogic, output-oriented tasks advance oral competence. The CEFR Companion Volume provides assessment architecture, supplying analytic criteria for spoken performance (Council of Europe, 2020), and Education for Sustainable Development supplies the rationale for treating quality-education themes as reflective, dialogic content (UNESCO, 2017; Wals, 2015).

Validation. The draft materials, tasks, and assessment instruments were validated through expert review by a panel comprising two internal practitioner-reviewers speaking lecturers experienced in teaching and assessing oral performance in the PTKI context and one external subject-matter expert from a different institution, an arrangement that strengthens content validity while reducing the risk of institution-bound judgement. The review attended to the clarity of task instructions, the appropriateness of scaffolding for B1–B2 learners, the relevance of the themes to speaking objectives, the alignment between tasks and CEFR-based assessment, and the practicality of classroom implementation. Feedback informed iterative revision, and a limited pilot refined prompt comprehensibility, time allocation, and rubric usability before the planned main implementation. In this way, ADDIE functioned not as a rigid linear sequence but as an iterative development cycle in which formative evaluation directly informed design refinement.

Findings / Results

The principal output of the analysis, design, and development phases is the prototype module described in this section. Its design is guided by three interrelated principles that connect the documented needs to concrete decisions: the materials address the specific speaking difficulties identified during analysis particularly students' limited ability to sustain interaction, organize ideas coherently, and articulate viewpoints on social and educational issues; all speaking tasks and assessment criteria are aligned with CEFR B1–B2 descriptors to ensure developmental appropriateness and a clear progression (Council of Europe, 2020); and selected SDG 4 quality-education themes are employed as meaningful communicative content rather than as content to be tested (UNESCO, 2017). The product is therefore best described as CEFR-aligned content-based speaking materials using quality-education themes, because language development, not sustainability knowledge, is the instructional target.

SDG 4 themes as communicative content

Three SDG 4 targets most relevant to higher education and to speaking are selected as thematic content: equitable access to quality tertiary education (4.3), relevant skills for work and life (4.4), and education for sustainable development and global citizenship (4.7). These themes are embedded into speaking tasks not as concepts to be recalled, but as issues that require students to express opinions, explain problems, justify positions, and reflect. A prompt such as "Some students cannot continue to university because of financial, geographic, or social

barriers—what should be done?” illustrates how an educational issue becomes a communicative stimulus that elicits extended speaking, interaction, and reflection. This positioning maintains the linguistic focus of instruction while situating speaking within socially relevant and professionally meaningful content, which is particularly appropriate in PTKI settings where many students will become educators and community leaders.

Unit architecture and thematic organization

The module is organized into three thematic units, each aligned with one SDG 4 target and with an increasingly demanding stage of speaking development. This structure prioritizes depth of engagement and developmental coherence over fragmented topical coverage, allowing students to revisit quality-education themes through progressively more sophisticated language use. Within a single semester, three units provide sufficient instructional depth and repetition to support observable development in fluency, interaction, and coherence while remaining feasible within regular course scheduling. Table 1 summarizes the unit architecture, which progresses from recognition and basic explanation toward positioning and reflective engagement, so that communicative development and language-mediated engagement with the themes advance together without cognitive overload.

Table 1. Unit architecture of the CEFR-aligned content-based speaking module

Unit	SDG 4 target / thematic focus	CEFR speaking focus (B1–B2)	Core speaking tasks	Reflective engagement
1	4.3 — Equitable access to quality tertiary education	Expressing opinions, giving reasons, basic interaction	Opinion sharing; guided discussion; issue-based response	Recognition; understanding of access and inequality
2	4.4 — Relevant skills for work, life, and future challenges	Extended speaking; problem-solving discourse; coherent argumentation	Problem-solving tasks; small-group discussion; proposal presentation	Understanding of skill gaps; positioning on employability
3	4.7 — Global citizenship, equity, and sustainable education	Interaction and turn-taking; justifying viewpoints; reflective speaking	Debate; role play; reflective discussion	Positioning on equity and inclusion; reflection on societal implications

CEFR sub-skill scaffolding

Rather than treating speaking as a single undifferentiated ability, the design scaffolds each CEFR sub-skill through concrete instructional realizations that reduce cognitive load while preserving communicative authenticity, as summarized in Table 2. All five dimensions are retained in scoring to preserve CEFR construct coverage, while fluency, interaction, and coherence receive particular analytical attention because they are most directly aligned with the communicative, dialogic, and reflective nature of the tasks.

Table 2. CEFR sub-skill scaffolding in the speaking tasks

CEFR sub-skill	Instructional realisation
Range	Topic-specific vocabulary and functional expressions for each theme
Accuracy	Guided sentence frames (e.g., “One reason is...”, “I believe that...”)
Fluency	Timed speaking cycles encourage continuous speech
Coherence	Explicit speaking structure: opinion → reason → example
Interaction	Prompted responses, reacting to peers, and turn-taking

Illustrative prototype: Unit 1

To make the design logic concrete, Unit 1 is presented as an illustrative prototype. It focuses on SDG 4.3 (equitable access to quality tertiary education) and targets CEFR B1–B2 speaking through a Communicative Language Teaching-oriented task. The communicative goal is that students can discuss issues of access to higher education and articulate reasoned opinions in spoken English. In the core task a guided small-group discussion students consider: “What barriers prevent students from accessing higher education financial, geographic, or social? Should universities do more to support equal access: why, or why not?” They respond to peers, agree or disagree, and provide brief explanations, prioritizing meaning-making and negotiation of meaning. Scaffolding includes topic vocabulary (access, opportunity, affordability, inclusion), sentence frames for opinion and reasoning, timed turns for fluency, an opinion → reason → example structure for coherence, and response prompts for turn-taking. Performance is scored with the five-dimension CEFR-aligned analytic rubric, with particular attention to fluency, interaction, and coherence. Reflective engagement recognizing unequal access, explaining causes, taking and justifying a position, and connecting the issue to fairness and future opportunity is elicited through the speaking task itself rather than through separate attitude measures.

CEFR-aligned analytic speaking rubric

Speaking performance is assessed through an analytic rubric adapted from the qualitative features of spoken language use in the CEFR (Council of Europe, 2020). The rubric retains five dimensions range, accuracy, fluency, interaction, and coherence to preserve construct alignment, consistent with language-assessment scholarship that treats speaking as a multidimensional construct better captured through explicit criteria than through a single holistic score (Brown, 2004; Fulcher, 2015). For classroom use, each dimension is operationalized into four bands (Below B1, B1, B1+/Approaching B2, and B2). Table 3 summarizes the B1 and B2 anchors for each dimension; the full four-band rubric accompanies the module.

Table 3. Summary anchors of the CEFR-aligned analytic speaking rubric

Dimension	B1 anchor (summary)	B2 anchor (summary)
Range	Enough basic vocabulary and structures for familiar topics; noticeable repetition	Sufficient range to develop viewpoints with reasonable flexibility
Accuracy	Basic patterns with frequent but generally non-impeding errors	Generally good control; errors rarely cause misunderstanding
Fluency	Short stretches of speech; frequent pauses for planning	Relatively even tempo; hesitation does not usually impede communication
Interaction	Responds to simple prompts; brief, support-dependent exchanges	Responds confidently; maintains and builds the exchange with little support
Coherence	Connects simple ideas; organization often uneven or list-like	Organizes and links ideas clearly; develops arguments that are easy to follow

Expert review and refinement

The validation described in the Methodology confirmed the pedagogical soundness, feasibility, and contextual appropriateness of the units, and directly shaped the final prototype. In response to reviewer feedback, task prompts were refined for clarity, linguistic support was adjusted for B1–B2 learners, assessment guidance was clarified, and task sequencing was improved; the limited pilot further refined prompt comprehensibility, time allocation, and rubric usability. These revisions illustrate how formative evaluation was folded back into the design rather than appended to it.

Discussion

The design responds directly to the gap documented in the analysis and in the wider literature. By embedding CEFR descriptors into both task design and the analytic rubric, the module addresses the assessment–pedagogy imbalance in which proficiency frameworks are adopted procedurally for scoring but left out of instructional design (Kalsum et al., in press). Here the CEFR functions as a pedagogical architecture, shaping what is taught and practiced, rather than only as a benchmarking instrument applied at the point of testing. The decision to treat SDG 4 as communicative content rather than as an instructional objective is equally consequential: it keeps the work squarely within English Language Teaching while giving learners something substantive to talk about, which the literature on Content-Based Instruction and on speaking development suggests can stimulate more elaborated and better-organized oral production (Long, 1996; Swain, 2005; Richards, 2006), and it offers a principled answer to a recurrent weakness in sustainability-oriented ELT materials, in which themes appear as reading passages or awareness prompts but are not matched to clear proficiency targets or to assessed oral performance (Cordova, 2024; Sofiana & Haryati, 2025; Yu et al., 2024).

Technology-enhanced delivery and future AI integration

Although the present prototype is not itself a digital artefact, its architecture is deliberately technology-ready, and this is where it connects most directly to technology-enhanced language learning. The task formats that drive the module timed speaking cycles, recorded oral performances, asynchronous small-group discussion, and structured peer interaction map onto the affordances of the blended and online environments now common in Indonesian higher education, and can be delivered through a learning-management system, mobile recording, or videoconferencing with minimal adaptation. More importantly, the CEFR analytic rubric operationalizes fluency, interaction, and coherence as explicit, criterion-referenced descriptors precisely the structure that automated and AI-assisted speaking assessment requires. Temporal fluency measures, automatic speech recognition, and rubric-referenced automated feedback all depend on a principled construct definition to be meaningful; the module supplies that definition. It therefore contributes the pedagogical and assessment architecture on which technology-enhanced, and ultimately AI-enhanced, speaking instruction can be built, guarding against the risk that digital delivery merely reproduces accuracy-focused materials at scale. Positioned within technology-enhanced language learning and future learning environments, the design offers some concrete on-ramp for AI-assisted formative feedback in subsequent iterations.

Originality and significance

It is important to distinguish the two senses in which the module is innovative. Its novelty lies in the integrative combination aligning CEFR descriptors, ADDIE-based design, Content-Based Instruction and Communicative Language Teaching, and SDG 4 themes within a single values-oriented PTKI setting rather than in any single component, each of which is individually well established. Claims about instructional effectiveness are deliberately withheld: the prototype has not yet undergone full classroom evaluation, so its significance at this stage is a design contribution transferable, theoretically grounded architecture whose measured impact is the subject of the planned quasi-experimental study. Understood in this way, the module is best read not as a claim of proven superiority but as a worked, adaptable example of how issue-based, values-oriented content can be aligned with a recognized framework to develop communicative speaking competence, and how a single program can operationalize broad policy directions in specific, assessable instructional terms.

Conclusion

This paper has reported the design and development of a CEFR-aligned, content-based speaking module that uses selected SDG 4 quality-education themes as meaningful communicative content for English language teaching in Islamic higher education. Grounded in a Design and Development Research orientation operationalised through ADDIE, and informed by Content-Based Instruction, Communicative Language Teaching, and theories of speaking development, the module organises three thematic units with sub-skill scaffolding, an illustrative prototype unit, and a CEFR-aligned analytic rubric, refined through expert review and limited piloting. As an innovative TESL practice, it demonstrates how sustainability-oriented content can be aligned with a recognised proficiency framework to support communicative speaking development, and through its criterion-referenced architecture how such instruction can be prepared for technology-enhanced and AI-assisted delivery.

Recommendations

For practitioners, the unit architecture, scaffolding scheme, and analytic rubric can be adapted to other quality-education themes and to other PTKI programs, provided that the linguistic focus is retained and the themes are used as communicative content. For researchers and developers, three directions are recommended: trialling technology-enhanced delivery of the tasks (for example, through recorded and asynchronous speaking on a learning-management system); prototyping AI-assisted formative feedback that draws on the rubric's fluency, interaction, and coherence descriptors; and conducting the planned quasi-experimental evaluation to test the module's effect on speaking competence. Extending the design to additional CEFR bands would further clarify its developmental range.

Limitations

The contribution reported here is a design, and its limitations follow accordingly. The module has not yet been evaluated for instructional effectiveness; it is developed for a particular program and bounded to selected SDG 4 targets and to CEFR B1–B2, so its transferability to other contexts remains to be examined; and the present account reports design decisions rather than learning outcomes. The planned classroom evaluation employs a quasi-experimental pre-test–post-test design to examine the module's effect on speaking competence, with particular attention to fluency, interaction, and coherence, while a complementary qualitative strand examines how students' language-mediated engagement with the themes becomes visible in their discourse.

Acknowledgements

The authors gratefully acknowledge the Faculty of Major Language Studies, Universiti Sains Islam Malaysia, and the State Islamic Institute of Parepare (IAIN Parepare) for the institutional support that enabled this work, and the organizers' of ICTELL 2026 for the opportunity to present and develop it.

Conflict of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Artificial Intelligence (AI) Use Statement

The authors used an AI-based writing assistant for limited language refinement and structural organisation during drafting. The conceptual development, the design of the module, the selection and interpretation of sources, and all argumentations were carried out by the authors, who take full responsibility for the accuracy, originality, and integrity of the work.

References

- Ambariyanto, & Utama, Y. J. (2020). Educating higher education institutions to support SDGs: Indonesian case. *E3S Web of Conferences*, 202, 02015. <https://doi.org/10.1051/e3sconf/202020202015>
- Bialik, M., Bogan, M., Fadel, C., & Horvathova, M. (2016). Character education for the 21st century: What should students learn? Center for Curriculum Redesign.
- Branch, R. M. (2009). Instructional design: The ADDIE approach. Springer. <https://doi.org/10.1007/978-0-387-09506-6>
- Brown, H. D. (2004). Language assessment: Principles and classroom practices. Pearson Education.
- Bui, H. T. M., Bui, T., & Pham, B. T. (2024). The role of higher education in achieving the Sustainable Development Goals: An evaluation of motivation and capacity of Vietnamese institutions. *International Journal of Educational Development*.
- Chankseliani, M., & McCowan, T. (2021). Higher education and the Sustainable Development Goals. *Higher Education*, 81(1), 1–8. <https://doi.org/10.1007/s10734-020-00652-w>
- Cordova, M. (2024). Integrating Sustainable Development Goals in English language and literature teaching. *Frontiers in Education*, 9, 1330034. <https://doi.org/10.3389/feduc.2024.1330034>
- Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment — Companion volume. Council of Europe Publishing.
- Dariah, M., Salleh, M. S., & Shafiai, M. H. M. (2016). A new approach in understanding the Sustainable Development Goals (SDGs) from an Islamic perspective. *International Journal of Islamic Thought*, 10(1), 12–18. <https://doi.org/10.24035/ijit.10.2016.002>
- DeVellis, R. F. (2017). Scale development: Theory and applications (4th ed.). Sage.
- Fulcher, G. (2015). Re-examining language testing: A philosophical and social inquiry (2nd ed.). Routledge. <https://doi.org/10.4324/9781315734889>
- Gericke, N., Boeve-de Pauw, J., Berglund, T., & Olsson, D. (2018). The Sustainability Consciousness Questionnaire: Theoretical development and empirical validation. *Environmental Education Research*, 25(2), 193–210.
- Kalsum, Saimin, R., & Engku Sharfuddin, E. M. S. (in press). CEFR-aligned English-speaking materials in higher education: Teachers' needs, assessment practices, and pedagogical gaps toward SDG 4. *Universiti Sains Islam Malaysia*.
- Levelt, W. J. M. (1989). Speaking: From intention to articulation. MIT Press.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- McKenney, S., & Reeves, T. C. (2018). Conducting educational design research (2nd ed.). Routledge. <https://doi.org/10.4324/9781315105647>
- Nahdi, K., Aisyah, S., Anwar, S., & Rahmawati, D. (2024). Integration of SDGs in language teaching: Assessing the response of Indonesian researchers — A systematic literature review. *Journal of Research and Innovation in Language Learning*, 7(1), 45–61.
- Nasrudin, E., Anwar, S., Suresman, E., & Rahmi, U. Q. (2025). Integration of Sustainable Development Goals in Islamic teaching materials in higher education. *Tafkir: Interdisciplinary Journal of Islamic Education*, 6(1), 17–39. <https://doi.org/10.31538/tijie.v6i1.1266>
- Plomp, T., & Nieveen, N. (Eds.). (2013). Educational design research — Part A: An introduction. Netherlands Institute for Curriculum Development (SLO).
- Richards, J. C. (2006). Communicative language teaching today. Cambridge University Press.
- Richey, R. C., & Klein, J. D. (2007). Design and development research: Methods, strategies, and issues. Routledge.

- Robah, A., & Anggrisia, N. F. (2023). Exploring challenges and strategies in English speaking among Indonesian university students. *Englisia: Journal of Language, Education, and Humanities*, 11(1), 55–74.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
- Sofiana, I., & Haryati, Y. (2025). Incorporating SDG-oriented content into English language instruction in Indonesia. *Proceedings of the ELLiC Conference*, 9(1), 212–221.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Lawrence Erlbaum Associates.
- UNESCO. (2017). *Education for Sustainable Development Goals: Learning objectives*. UNESCO Publishing.
- United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. United Nations.
- Usman, J., Yulia, A., Zakaria, A. R., Muluk, S., Hasan, K. K., & Jarjani, J. S. (2025). Sustainability integration in EFL curricula: Evidence from Indonesian and Malaysian universities. *Studies in English Language and Education*, 12(3), 450–472.
- Wals, A. E. J. (2015). Sustainability-oriented learning in higher education. *International Journal of Sustainability in Higher Education*, 16(2), 152–168. <https://doi.org/10.1108/IJSHE-09-2013-0114>
- Yu, B., Guo, W. Y., & Fu, H. (2024). Sustainability in English language teaching: Strategies for empowering students to achieve the Sustainable Development Goals. *Sustainability*, 16(8), 3325. <https://doi.org/10.3390/su16083325>

Onomatopoeic-Based Approach for Teaching Personification in Poetry Among Malaysian Lower Secondary Proficiency ESL Learners

Pei Yi Tan, The National University of Malaysia, Malaysia, peiyitan29@gmail.com

Nur Izzah Azman, The National University of Malaysia, Malaysia, nurizzahazmn@gmail.com

Nur Haziqah Nadhrah Mohd Fuad, The National University of Malaysia, Malaysia, ieqahaziqah1@gmail.com

Hanita Hanim Ismail, The National University of Malaysia, hanitahanim@ukm.edu.my

Abstract: Lower proficiency ESL learners often face difficulties in understanding poetry, particularly in interpreting figurative language and moving beyond literal meanings. This affects their ability to appreciate and analyse poetry. Therefore, this study presents the development of an onomatopoeic-based approach that integrates sounds as a sensory entry point to enhance learners' understanding of personification in poetry. The approach was developed based on the identified learners' challenges among Malaysian lower secondary level learners, reviewing relevant literature, where a four-stage teaching process was designed involving sound exposure, exploration, and creative application. The effectiveness of the approach is evaluated through pre- and post-test results to examine Malaysian learners' improvement in identifying and applying personification. This approach aims to make poetry learning more engaging, accessible, and meaningful for lower proficiency ESL learners.

Keywords: ESL learners, onomatopoeia, personification, poetry teaching, sensory learning.

Introduction

Literature serves as a cornerstone of English language education by fostering the development of linguistic proficiency, analytical thinking, and appreciation for creative expression. Within the English curriculum, poetry exposes learners to figurative language and encourages them to interpret meanings from the text beyond the literal level. Understanding figurative language is fundamental to comprehending poetry, as it enables learners to recognise the poet's intended ideas, emotions, and messages (Hussein, 2023). Among the variety of poetic devices, personification plays a vital role in helping learners understand imagery, develop interpretative skills, and appreciate poetic expression.

Within the Malaysian secondary context, the teaching of literature has become increasingly challenging following the implementation of the CEFR-aligned curriculum, where the literature component is no longer assessed in the SPM examination (Yee et al., 2023). Although literature remains a part of the English curriculum, Malaysian ESL learners continue to experience difficulties understanding poetry because interpreting figurative language requires them to move beyond literal meanings (Ukat & Ismail, 2022). This challenge is often compounded by a "literal trap", where students are hindered by limited vocabulary knowledge and weak sensory imagination that lead to failure to move beyond literal meanings. According to Mastura (2023), 39% of students view figurative language as the single largest obstacle in poetry

comprehension. To illustrate, when encountering expressions such as “The wind whispered,” many students struggle to read between the lines and ask, “How can wind whisper?” (Umiera & Hamdani, 2025). Therefore, there is an urgency for pedagogical methods to switch from rote memorisation of dictionary definitions to experiential learning (Ukat & Ismail, 2022).

To address these challenges, this study introduces the Onomatopoeic-Based Approach for teaching personification in poetry among Malaysian lower-proficiency ESL learners. By integrating sound-based learning with the teaching of figurative language, this approach seeks to make the concept of personification more concrete, interactive, and meaningful for learners. By guiding learners through a four-stage progression, “Sound → Meaning → Humanise → Personification”, the approach transforms abstract poetic devices into concrete sensory experiences. Ultimately, this study aims to explore the potential of the Onomatopoeic-Based Approach to support learners' understanding and application of personification while enhancing their engagement with and appreciation of poetry. The findings are expected to provide useful insights into the use of an innovative instructional approach that can strengthen literature teaching and learning in Malaysian ESL classrooms.

Literature Review

Teaching Poetry and Personification in ESL Classrooms

Literature, particularly poetry, can provide meaningful opportunities for English language learning by exposing ESL learners to authentic language, vocabulary, creative expression, and interpretation. Poetry can also be integrated with language-focused activities rather than being treated solely as a literary text. Islam (2022), for example, highlights the potential of poetry as authentic material for developing language skills through interactive classroom activities. More recent research has similarly reported that poetry-based activities can support learners' linguistic development, creativity, motivation, and engagement with English (Ismail Shauki & Junit, 2024; Saad et al., 2024). In the Malaysian context, however, literature teaching continues to present challenges. Ukat and Ismail (2022) identify difficulties related to literary texts, cultural differences, and learners' attitudes, while Jais and Ismail (2023) further emphasise the need for alternative and multimodal approaches to make literature learning more accessible to ESL learners.

Poetry may be particularly challenging for lower-proficiency ESL learners because poetic texts frequently rely on figurative expressions, ambiguity, and meanings that extend beyond the literal interpretation of words. Personification is one such device in which human characteristics, actions, or emotions are attributed to non-human entities. Its interpretation, therefore, requires learners to move from a concrete or literal meaning towards an abstract human-related interpretation. Research on personification has demonstrated that it functions as an important mechanism in meaning construction, particularly through the attribution of human characteristics to non-human entities (Fielden-Burns & Piquer-Píriz, 2022). For ESL learners, this abstract process may become more demanding when they have limited vocabulary or limited experience with figurative language. Consequently, teaching personification may require instructional scaffolding that allows learners to develop meaning progressively rather than encountering the literary device only through formal definitions.

Theoretical Foundations

The proposed Onomatopoeia-Based Approach is informed by Dual Coding Theory, the Cognitive Theory of Multimedia Learning, and Constructivism. Dual Coding Theory proposes that verbal and non-verbal representations can work together in learning, providing learners with multiple routes for processing and recalling information (Clark & Paivio, 1991). Similarly, the Cognitive Theory of Multimedia Learning suggests that learning can be supported when information is processed through complementary verbal and visual representations (Mayer, 2024). Constructivist perspectives further emphasise that learners develop knowledge by actively constructing meaning through meaningful experiences rather than simply receiving information from the teacher (Chand, 2023). These perspectives provide a theoretical basis for using sound as an initial stimulus before introducing the more abstract concept of personification.

The combination of these perspectives is particularly relevant to the proposed four-stage framework: Sound → Meaning → Humanise → Personification. In the first stage, learners encounter an authentic or recognisable sound. They then associate the sound with its possible source and meaning before connecting the sound or its associated action to human characteristics. Finally, learners apply this understanding to identify or construct personification. Thus, rather than introducing personification as an isolated literary definition, the approach provides a progression from a more concrete sensory experience towards an abstract figurative concept.

Innovative and Multimodal Approaches to Literature Learning

Recent literature pedagogy has increasingly explored interactive, multimodal, and technology-supported approaches to address the difficulties associated with conventional literature instruction. Jais and Ismail (2023) report that mapping and multimodality can provide alternative ways of supporting learners when engaging with literature in Malaysian ESL classrooms. Similarly, research on poetry-based activities suggests that integrating creative and learner-centred activities can promote greater engagement and meaningful language learning (Saad et al., 2024). Technology-enhanced approaches have also been explored in Malaysian literature classrooms, with recent research examining the use of YouTube to support students' motivation and engagement in learning poetry (Alfaruque et al., 2023). These studies indicate a broader movement away from literature instruction that depends exclusively on textual analysis and teacher explanation towards approaches that incorporate multiple modes of learning.

However, the existing literature largely focuses on multimodality through visual materials, digital platforms, mapping activities, or general poetry-based tasks. Less attention has been given to sound as the primary scaffold for teaching a specific figurative device, particularly personification, among lower-proficiency ESL learners. This distinction is important because sound provides an immediate sensory experience that can be connected to an observable action or source before learners are required to interpret an abstract literary meaning. Therefore, the use of sound may provide a different form of scaffolding from approaches that begin directly with written poetic texts or formal literary terminology.

Research Gap

Recent research has demonstrated growing interest in multimodal and innovative approaches to language and literature teaching. Multimodality allows teachers to combine linguistic, visual, auditory, gestural, and other modes to support meaning-making in language learning (Kessler, 2022). In EFL contexts, multimodal resources have been used to establish cognitive frameworks for new topics and to increase learners' attention and participation (Wu, 2024). In Malaysia, technology-enhanced approaches such as YouTube have also been explored in poetry instruction, reflecting a shift towards more interactive approaches to literature learning (Chung & Ngui, 2025). These studies suggest that literature instruction can benefit from moving beyond conventional text-based explanation and incorporating additional modes of meaning-making.

At the same time, emerging research has examined the relationship between sound and language learning. Zhao (2024) found that explicit instruction on phonesthemes, which are sound sequences associated with particular meanings, supported EFL learners' retention of phonesthemic words. Such findings suggest that attention to sound-meaning relationships may provide useful support for language learning. However, sound-symbolic learning, phonestheme instruction, and onomatopoeia should not be treated as interchangeable concepts. While phonesthemes involve recurring sound patterns associated with meaning, onomatopoeia involves words that imitate or evoke sounds associated with their referents. Therefore, the pedagogical potential of using actual or recognisable sounds as a starting point for teaching figurative language remains insufficiently explored.

Furthermore, existing studies on personification have primarily examined its use and interpretation in discourse rather than the process of teaching personification to lower-proficiency ESL learners. Fielden-Burns and Piquer-Píriz (2022), for instance, demonstrate that personification occurs meaningfully in English-medium academic discourse, but their study does not investigate an instructional framework for helping learners understand or produce personification. Similarly, recent research on figurative language instruction has examined approaches such as songs and explicit instruction, demonstrating the potential of contextualised and active learning for figurative-language development (Maghfira, 2022; Ningrum & Agustina, 2025). However, these approaches do not specifically investigate sound as a scaffold for understanding personification.

Therefore, a specific gap can be identified at the intersection of sound-based learning, multimodal pedagogy, and personification instruction. To the best of the present study's knowledge, limited research has examined a structured approach that begins with an onomatopoeic or recognisable sound and progressively guides ESL learners from sound recognition to meaning-making, the attribution of human characteristics, and finally the interpretation or production of personification. This gap is particularly relevant in the Malaysian ESL context, where poetry remains challenging for learners and where recent research has called for more engaging and alternative approaches to literature instruction (Ukat & Ismail, 2022; Nik Hassim & Hashim, 2024).

In response to this gap, the present study proposes the Onomatopoeia-Based Approach, structured around four stages: Sound → Meaning → Humanise → Personification. The approach uses sound as a concrete entry point before gradually introducing the more abstract concept of personification. By connecting sensory experience with linguistic and figurative meaning, the framework aims to provide a more accessible pathway for lower-proficiency ESL learners to engage with personification in poetry. Rather than claiming that sound alone improves figurative-language learning, the proposed framework provides

a pedagogical model that warrants further empirical investigation in Malaysian ESL classrooms.

Methodology

Research Design

This study employed a pre-experimental one-group pretest-posttest design to examine changes in participants' understanding and application of personification following exposure to the proposed Onomatopoeia-Based Approach. The design was selected because the study aimed to obtain a direct comparison of participants' performance before and after experiencing the proposed instructional approach. The pre-test established participants' baseline understanding, while the post-test examined their performance after the intervention.

The study was conducted during an interactive approach exhibition, where participants first completed a diagnostic pre-test before being introduced to the instructional framework. They then experienced the four-stage instructional intervention, which followed the sequence sound → meaning → humanise → personification, before completing the post-test. This sequence allowed the researcher to observe changes in participants' performance across the learning outcomes targeted by the approach.

The design was appropriate for the exploratory nature of the study because the Onomatopoeia-Based Approach was being presented as a proposed instructional framework. Rather than attempting to establish causal effectiveness, the study focused on identifying observable changes in participants' responses following their exposure to the approach. The comparison specifically examined three intended learning outcomes: (i) associating onomatopoeic sounds with their meanings, (ii) identifying and interpreting personification, and (iii) applying personification through the construction of original figurative expressions.

The pre-test was administered before the intervention to identify participants' existing difficulties with figurative language. Particular attention was given to the tendency to interpret figurative expressions literally, referred to in this study as the "literal trap." For example, when presented with the expression "*The wind whispered*," participants who interpreted "*whispered*" as a literal action rather than recognising it as a human action attributed to the wind demonstrated difficulty with the conceptual basis of personification. The post-test was subsequently used to examine whether participants could demonstrate the targeted concepts after experiencing the four-stage instructional process.

As the study used a one-group design without a control or comparison group, any changes observed between the pre-test and post-test cannot be attributed solely to the intervention. The findings are therefore interpreted as descriptive evidence of changes in participants' performance following exposure to the Onomatopoeia-Based Approach, rather than as definitive evidence of causal effectiveness.

Participants and Sampling

The study employed purposive voluntary sampling. Participants were recruited from visitors who attended the interactive approach exhibition and voluntarily agreed to participate in the study. A total of 17 participants were included in the final analysis.

The intended target population of the proposed approach was lower-proficiency ESL learners who experience difficulty understanding figurative language, particularly personification. However, direct access to lower secondary school learners was not available during the exhibition phase. Consequently, university-level exhibition visitors were recruited as participants. Since the participants did not necessarily correspond directly to the intended school-level population, their suitability for the study was considered based on their baseline characteristics and performance rather than their educational level alone.

Participant characteristics were examined using the information collected through Section A (Demographic Information and Baseline Profile) and Section B (Pre-test Diagnostic). Section A provided information about participants' self-rated English proficiency, confidence in poetry-related tasks, and perceived difficulty in understanding figurative language. Section B provided behavioural evidence of their initial understanding through their responses to diagnostic items on figurative language and personification.

The final sample consisted of participants whose responses reflected characteristics relevant to the learning difficulties targeted by the intervention, including lower self-rated English proficiency, difficulty with figurative language, lower confidence in poetry-related tasks, and a tendency towards literal interpretation in the diagnostic assessment. These characteristics were considered together rather than relying on self-reported proficiency alone.

Purposive voluntary sampling was appropriate for the exhibition setting because participation depended on visitors' willingness to engage with the interactive demonstration. However, the use of a small, non-random sample limits the generalisability of the findings. Therefore, the results are intended to provide preliminary descriptive evidence regarding participants' responses to the proposed approach rather than represent the performance of the wider population of lower-proficiency ESL learners.

Research Instrument

Data were collected using a three-part Google Form assessment instrument administered in conjunction with the interactive intervention. The instrument was designed to establish participants' baseline characteristics, identify their initial difficulties with personification, and assess their performance following exposure to the Onomatopoeia-Based Approach.

Section A: Demographic Information and Baseline Profile

Section A collected participants' basic demographic information and baseline learning profiles. The section included items on self-rated English proficiency, confidence in poetry-related tasks, and perceived difficulty in understanding figurative language.

The information collected in this section served two purposes. First, it provided contextual information about the participants. Second, it helped establish whether participants demonstrated characteristics relevant to the intended target group. However, self-reported information was not treated as the sole measure of participants' proficiency or ability. Instead, it was considered alongside their actual performance in the diagnostic pre-test in Section B.

Section B: Pre-test Diagnostic

Section B consisted of a short diagnostic pre-test administered before the intervention. The purpose of the pre-test was to establish participants' unassisted baseline understanding of the concepts addressed by the Onomatopoeia-Based Approach.

The assessment focused on the learning difficulties that the intervention was designed to address. These included sound–meaning association, interpretation of figurative expressions, recognition of human actions attributed to non-human subjects, and differentiation between literal and figurative meanings.

For example, an item involving the expression “*The wind whispered*” was used to examine whether participants could recognise “*whispered*” as a human action attributed to the non-human subject “*wind*.” A literal interpretation would indicate difficulty in recognising the figurative relationship between the subject and the human action.

The diagnostic pre-test therefore served two related purposes. First, it established participants' baseline performance before exposure to the intervention. Second, it helped identify the specific conceptual difficulties that were relevant to the proposed instructional approach. Participants completed the pre-test without receiving the instructional explanation beforehand so that their initial understanding could be observed.

Section C: Post-test Assessment

Section C was administered immediately after participants completed the four-stage intervention. The purpose of the post-test was to examine participants' performance after exposure to the proposed instructional process.

The post-test addressed the same three broad learning outcomes examined by the intervention: (i) sound–meaning association, (ii) identification and interpretation of personification, and (iii) application of personification. While the pre-test focused primarily on diagnosing participants' initial understanding, the post-test included tasks that required participants to apply the concepts introduced during the intervention.

Participants were required to interpret onomatopoeic words such as “*Whoosh*” and “*Crackle*”, recognise personification in contextualised expressions such as “*The thunder shouted in the sky*,” and demonstrate their understanding of the relationship between a non-human subject and a human action. The post-test also included an application task in which participants were required to construct an original figurative expression using a sound trigger such as “*Boom*.”

The inclusion of an independent production task was important because the study aimed to examine not only whether participants could recognise personification but also whether they could apply the concept in their own figurative expressions. Thus, the post-test provided evidence across different levels of understanding, from recognising sound–meaning relationships to identifying personification and producing an original example.

Intervention

The intervention consisted of a four-stage instructional framework: sound → meaning → humanise → personification. The framework was designed to provide a gradual progression from a concrete sensory concept to the more abstract concept of figurative language.

Stage 1: Sound

In the first stage, participants were introduced to selected onomatopoeic words representing recognisable sounds, such as “*Boom*,” “*Whoosh*,” and “*Crackle*.” The purpose of this stage

was to establish an accessible and concrete starting point for the learning process. Participants were encouraged to recognise the sounds represented by the words and to engage with the auditory quality of the vocabulary.

Stage 2: Meaning

In the second stage, participants were guided to identify the meaning and possible source associated with each sound. For example, “*Whoosh*” could be associated with the sound produced by moving air or wind, while “*Crackle*” could be associated with the sound produced by burning wood or fire.

This stage established a connection between the sound word and the situation or source represented by the sound. Establishing this connection was necessary before participants were introduced to the more abstract process of using the sound as a basis for figurative expression.

Stage 3: Humanise

In the third stage, participants were introduced to the process of assigning human actions or characteristics to non-human subjects. The researcher guided participants to identify an appropriate human action that could be connected to the source or situation established in the previous stage.

For example, after identifying thunder as a source associated with a loud “*Boom*,” participants could associate the sound with a human action such as “*shout*.” This stage served as a conceptual bridge between the concrete sound–meaning relationship and the abstract concept of personification.

Stage 4: Personification

In the final stage, participants were guided to combine a non-human subject with a human action to form a personified expression. For example, the sequence could progress from “*Boom*” to thunder as the sound source, followed by the human action “*shout*,” resulting in the expression “*The thunder shouted*.”

At this stage, the researcher explicitly introduced and reinforced the concept of personification. Participants were guided to recognise that personification occurs when human actions, characteristics, or qualities are attributed to non-human entities.

The four stages therefore provided a structured progression from sound recognition, to meaning construction, to human action, and finally to figurative language. This progression was intended to provide learners with a concrete pathway for approaching personification rather than requiring them to process the figurative concept in isolation.

Research Procedure

The study was conducted in a sequential procedure to ensure that participants’ baseline understanding was established before they were exposed to the intervention.

First, visitors to the interactive exhibition were invited to participate voluntarily in the study. Participants were informed about the activity and completed the assessment instrument through Google Forms. Second, participants completed Section A, which collected their demographic information and baseline learning profile. Third, participants completed Section B (Pre-test Diagnostic) before receiving any explanation of the Onomatopoeia-Based Approach. This ensured that their responses represented their initial

understanding rather than knowledge acquired from the intervention. Fourth, participants experienced the four-stage instructional intervention. The intervention was delivered in the sequence of sound → meaning → humanise → personification, with each stage building upon the concept introduced in the preceding stage. Fifth, immediately after completing the intervention, participants completed Section C (Post-test Assessment). The post-test assessed their ability to recognise, interpret, and apply the concepts introduced during the intervention.

Finally, the completed responses were compiled and analysed using descriptive statistics. The pre-test and post-test results were compared according to the three intended learning outcomes of the intervention. The sequence of the procedure can therefore be summarised as:

Section A: Baseline Profile → Section B: Pre-test → Four-Stage Intervention → Section C: Post-test → Data Analysis

This sequence ensured that the intervention occurred between the two assessment points and allowed participants' responses before and after the instructional experience to be compared descriptively.

Ethical Considerations

Ethical considerations were observed throughout the data collection process. Participation in the study was voluntary, and participants were informed about the purpose and nature of the study before completing the assessment. Participants were informed that the activity involved completing a pre-test, participating in the instructional intervention, and completing a post-test. They were given the opportunity to decide whether to participate and were not required to participate as a condition of attending the exhibition.

Informed consent was obtained from participants before data collection. Participants were informed that the information collected would be used for research purposes and that their participation was voluntary. They were also informed that they could choose not to participate or discontinue their participation without penalty.

Confidentiality and anonymity were maintained throughout the research process. Participants' names and other directly identifying information were not included in the reporting of the findings. The collected responses were treated as research data and analysed in aggregate form. The findings were reported using frequencies and percentages to prevent individual participants from being identified.

Permission to conduct the data collection was obtained from the relevant party responsible for the interactive approach exhibition before participants were recruited. The data collection was conducted within the scope of the exhibition activity and involved only the instructional intervention and assessments described in this study.

The researchers also ensured that the data collected were limited to information relevant to the objectives of the study. The findings were presented in aggregate form, and the collected data were not used for purposes unrelated to the research.

Data Analysis

The quantitative data were analysed using descriptive statistics, specifically frequencies and percentages, to examine participants' responses before and after exposure to the Onomatopoeic-

Based Approach. The analysis was aligned with the three intended learning outcomes of the intervention: (i) sound–meaning association, (ii) identification and interpretation of personification and (iii) application of personification through the construction of figurative expressions.

For each assessment item, the frequency and percentage of correct responses were calculated for the pre-test and post-test. The results were then compared descriptively to identify observable changes in participants' performance following the instructional intervention. Percentage-point differences were used where appropriate to describe the direction and extent of change between the two assessment occasions.

The pre-test results were used to establish participants' baseline difficulties, particularly in distinguishing literal from figurative meanings and associating sounds with appropriate meanings. The post-test results were used to describe participants' performance after completing the four-stage instructional process. The analysis therefore provides descriptive evidence of changes in participants' responses, rather than inferential evidence of statistical significance or causal effectiveness.

Findings / Results

The quantitative findings of this pilot study demonstrate a significant performance shift in lower-proficiency ESL learners' comprehension and application of personification, transitioning from literal reading constraints during the diagnostic pre-test to successful figurative mastery in the post-test.

Participants' Demographic Information

Table 1 presents the demographic information of the participants involved in the study. A total of 17 students from a university in Selangor participated in the Onomatopoeia-Based Approach implementation and completed the pre-test and post-test.

Table 1. Socio-Demographic Characteristics of the Respondents

No	Demographic Characteristics	Frequency (n)	Percentage %
Gender			
1.	Male	5	29.4
	Female	12	70.6
Age Group			
2.	<18	1	5.9
	18 - 25	16	94.1
	26 - 35	0	0
	>35	0	0
Language Background			
3.	English as a Second Language (ESL)	17	100.0
Self-rated English Proficiency			
4.	Level 1	1	5.9
	Level 2	12	70.6
	Level 3	4	23.5

	Level 4	0	0
	Level 5	0	0
Figurative Language Difficulty			
5.	Very Easy	0	0
	Easy	1	5.9
	Moderate	11	64.7
	Difficult	4	23.5
	Very difficult	1	5.9
Personification Understanding in Poetry			
6.	Very confident	0	0
	Confident	1	5.9
	Neutral	10	58.8
	Not confident	4	23.5
	Very not confident	2	11.8

The socio-demographic analysis of the filtered respondents ($n = 17$) confirms that the cohort aligns with the study's target population of lower-proficiency ESL learners. As mentioned in the methodology section, the entire sample (100%) of the pilot test fell within the 18 to 25 age group and identified as ESL learners at university level. This linguistic background is significant as it underpins the foundational vocabulary challenges and the “literal trap” often encountered when processing English poetic devices.

Regarding self-assessed competency, the vast majority of participants (70.6%) rated their English proficiency at Level 2. A small minority of 5.9% and 23.5% identified themselves at Level 1 and Level 3 accordingly, while no participants (0%) rated themselves at the higher mastery levels of 4 or 5.

Baseline Pre-test Diagnostic Analysis: Unpacking the “Literal Trap”

The diagnostic pre-test (Section B) evaluated learners' unassisted baseline comprehension of personification and identified specific cognitive obstacles prior to the sensory intervention. As detailed in Table 2, pre-test errors were classified into three distinct failure modes:

Table 2. Pre-test Diagnostic Baseline Item Analysis (n=17)

Item & Prompt	Construct Measured	Target/ Correct Response	Identified Failure Mode	Correct Rate (n, %)	Diagnostic Error Rate (n, %)
Q1: “ <i>The wind whispered.</i> ” What does this sentence mean?	Non-literal semantic interpretation	B. “ <i>The wind was moving softly.</i> ”	Semantic Failure: Selecting literal options (e.g., “ <i>The wind could speak like a person</i> ” or “ <i>A person was talking outside</i> ”).	8 (47.1%)	9 (52.9%)

Q2: Which sentence is personification?	Contextual identification of figurative human action	C. <i>"The flower danced in the garden."</i>	Concept Blindness: Failing to recognize figurative human action and retreating to literal sentences (e.g., <i>"The boy walked home"</i>).	7 (41.2%)	10 (58.8%)
Q3: Which sound word describes the sound of leaves moving?	Auditory-sensory vocabulary mapping	A. <i>"Rustle"</i>	Sensory Mismatch: Inability to map physical movement to auditory vocabulary (selecting <i>"Boom"</i> or <i>"Crackle"</i>).	13 (76.5%)	4 (23.5%)

When presented with *"The wind whispered"*, 52.9% of participants ($n = 9$) exhibited semantic failure by choosing literal interpretations. Furthermore, when required to identify personification among distractors, 58.8% of the cohort ($n = 10$) demonstrated concept blindness, retreating to literal options such as *"The boy walked home"*. Finally, 23.5% of respondents ($n = 4$) displayed a sensory mismatch, failing to match the movement of leaves to *"Rustle"* and choosing inappropriate sound triggers like *"Boom"* or *"Crackle"*. These results illustrate a clear over-reliance on literal readings and memorised definitions without functional comprehension.

Post-test Performance Metrics: Measuring Phonetic-Sensory Mastery

Following the four-stage sensory intervention (Sound → Meaning → Humanise → Personification), participants completed Section C to evaluate post-intervention learning gains. Table 3 details performance across five post-test evaluation metrics:

Table 3. Post-Test Item Performance and Mastery Analysis ($n = 17$)

Assessment Task / Prompt	Construct Evaluated	Target Response / Applied Rule	Success Rate (n, %)	Error / Incomplete Rate (n, %)
Q1: Choose the meaning of <i>"Whoosh"</i> .	Phonetic-Sensory Decoding	A. <i>"Wind moving quickly"</i>	13 (76.5%)	4 (23.5%)
Q2: Which sentence uses personification?	Contextual Identification	B. <i>"The thunder shouted in the sky."</i>	15 (88.2%)	2 (11.8%)

Q3: Match "Crackle" to its meaning.	Sensory Vocabulary Anchoring	A. "A fire making a small breaking sound"	15 (88.2%)	2 (11.8%)
Q4: Complete: "The wind _____ through the trees."	Instructional Formula Application	A. "whispered" (Sound + Human Action)	14 (82.4%)	3 (17.6%)
Q5: Create your own sentence using "Boom".	Independent Creative Generation	Original Figurative Output (e.g., "My thought boomed in my head")	15 (88.2%)	2 (11.8%)

Following the implementation of the sensory intervention, post-test metrics revealed a quantitative breakthrough across all measured constructs. In phonetic-sensory mapping, 76.5% of learners ($n = 13$) successfully decoded the sound word "Whoosh" as rapid wind movement. Regarding contextual identification, 88.2% of participants ($n = 15$) correctly recognised personification in "The thunder shouted in the sky", while only 11.8% ($n = 2$) remained trapped in literal distractors. Similarly, 88.2% of respondents ($n = 15$) demonstrated accurate vocabulary anchoring by matching "Crackle" to its corresponding sensory description. Furthermore, 82.4% of learners ($n = 14$) successfully executed the instructional framework (Sound + Human Action = Personification) to complete sentence frames such as "The wind whispered through the trees". Finally, 88.2% of participants ($n = 15$) achieved independent creative production, progressing from passive reading to constructing original personified imagery using the auditory trigger "Boom," yielding unique student expressions such as "My thought boomed in my head".

Comparative Synthesis: Pre-Test vs. Post-Test Progression

To measure the overall impact of the framework, baseline accuracy rates from Section B were systematically benchmarked against post-intervention performance metrics in Section C. Table 4 highlights the comparative performance gains across parallel conceptual dimensions:

Table 4. Comparative Progression Between Pre-Test and Post-Test Metrics ($n = 17$)

Learning Dimension / Skill Evaluated	Pre-Test Baseline Accuracy (%)	Post-Test Mastery Accuracy (%)	Net Quantitative Gain (%)	Primary Cognitive Shift Observed
Non-Literal Interpretation / Contextual Identification	47.1%	88.2%	+41.1%	Shifted from Semantic Failure (literal reading) to Embodied Figurative Recognition.
Auditory-Sensory Vocabulary Matching	76.5%	88.2%	+11.7%	Resolved Sensory Mismatch, establishing durable sound-

				meaning links.
Rule Application & Creative Sentence Production	41.2%	88.2%	+47.0%	Transitioned from Rote Memorization to Independent Formula-Driven Output.

The comparative synthesis confirms that the sensory intervention yielded a +41.1% net gain in personification identification and a +47.0% net gain in creative sentence production. These findings demonstrate that utilising sound as an auditory entry point effectively lowers working memory load, bypasses literal reading constraints and enables lower-proficiency ESL learners to achieve creative figurative mastery.

Discussion

The empirical findings of this pilot study provide strong evidence that the Onomatopoeic-Based (Phonetic-Sensory) Approach serves as an effective instructional bridge for lower-proficiency English as a Second Language (ESL) learners. In the context of Malaysian secondary ESL classrooms, where curriculum adjustments under the Common European Framework of Reference for Languages (CEFR) realigned assessment priorities and reduced dedicated focus on literature, learners frequently struggle to interpret poetic devices. By using sound as an accessible auditory entry point, the four-stage framework (Sound → Meaning → Humanise → Personification) transforms abstract figurative concepts into concrete, embodied imagery, enabling students to move beyond literal reading constraints toward creative application.

Overcoming the “Literal Trap” and Semantic Failure through Auditory Scaffolding

The baseline pre-test diagnostic data confirmed a widespread prevalence of the "literal trap", where 52.9% of participants exhibited semantic failure when evaluating *"The wind whispered,"* choosing literal interpretations such as *"The wind could speak like a person"* or *"A person was talking outside"*. Furthermore, 58.8% demonstrated concept blindness, failing to recognise personification in context and retreating to simple literal sentences. These diagnostic baseline results align directly with empirical research by Mastura (2023), who reported that 39% of students view figurative language as the primary obstacle in poetry comprehension, as well as Umiera and Hamdani (2025), who noted that ESL learners frequently struggle to read between the lines, asking literal questions such as *"How can wind whisper?"*.

Following the sensory intervention, post-test performance demonstrated a major quantitative breakthrough, with 88.2% of participants successfully identifying contextual personification in *"The thunder shouted in the sky"*. This sharp increase in identification accuracy indicates that using sound as an initial scaffold allows understanding to precede formal explanation. Rather than forcing lower-proficiency learners to navigate archaic, technical or ambiguous poetic language upfront, the phonetic-sensory hook provides an immediate sensory anchor. Consequently, learners are able to infer underlying mood and human traits without becoming entangled in literal decoding errors.

Addressing the Foundational Vocabulary Deficits via Dual Coding and Realia

The socio-demographic profile of the sample revealed a cohort heavily dominated by Level 2 proficiency learners (70.6%), a group characterised by limited linguistic mastery and lower poetry confidence. Baseline diagnostic testing highlighted this linguistic barrier, as 23.5% of respondents exhibited a sensory mismatch, failing to pair leaf movement with the word "*Rustle*" and choosing inappropriate terms such as "*Boom*" or "*Crackle*". This baseline deficit is consistent with Anaam (2025) and Alkhamash (2022), who observed that 85.5% of lower-proficiency ESL learners lack the foundational vocabulary required to process non-literal and metaphorical meanings.

To bridge this vocabulary gap, Stage 2 of the intervention incorporated realia, such as metal spoons, silk scarves, chip bags and keys, alongside silent video challenges to anchor abstract sound words in physical touch and sight. In the post-test, 88.2% of participants accurately anchored "*Crackle*" to its sensory description. This empirical gain directly supports Dual Coding Theory (Clark & Paivio, 1991; Paivio, 1986), which posits that pairing non-verbal sensory inputs (auditory and tactile experiences) with verbal language representations constructs dual mental traces in memory. By physically experiencing the texture of a sound before learning its formal label, learners transform arbitrary vocabulary words into meaningful mental concepts.

Mitigating Cognitive Load and Building Embodied Imagery in ESL Learners

A significant barrier identified in ESL literature instruction is learners' weak sensory imagination, which prevents them from constructing a "mental movie" or "embodied imagery" while reading. As highlighted by Zhang et al. (2025) and Simina Pitur and Miu (2025), lower-proficiency students often treat poetic text as isolated data rather than an immersive sensory experience. When reading lines like "*The leaves rustled*," students may decode individual words without experiencing the subtle auditory sensation of movement.

The Onomatopoeic-Based Approach addresses this cognitive bottleneck through the principles of the Cognitive Theory of Multimedia Learning / Auditory Sensory Learning (Mayer, 2021, 2024; Sweller, 2011). By engaging the auditory channel first, sound acts as a pattern-recognition trigger that lowers the operational working memory load required to interpret symbolic meaning. The high post-test accuracy across phonetic decoding tasks (76.5% for "*Whoosh*" and 88.2% for "*Crackle*") demonstrates that sound-based scaffolding allows learners to process poetic descriptions as vivid sensory experiences, successfully converting text into embodied mental representations.

Shifting from Rote Memorization to Constructivist Creative Mastery

Traditional literature pedagogy in ESL classrooms often relies on explicit teacher explanations and post-hoc identification of poetic devices after formal definitions are introduced. As documented by Ganesan, Aziz, and Ismail (2025), Cekiso (2025) and Liugan and Lacostales (2026), this definition-first model encourages surface learning and rote memorisation. Students learn dictionary definitions such as memorising personification as "*giving human traits to non-human things*" to pass examinations, yet they remain incapable of identifying or producing figurative language in authentic contexts.

The post-test results demonstrate a clear transition toward Constructivist Learning (Chand, 2023; Piaget, 1973; Vygotsky, 1978), where students actively build knowledge through hands-on exploration and pattern recognition. By introducing the structured formula Sound + Human Action = Personification during Stage 4, 82.4% of participants successfully completed formula-driven sentence frames ("*The wind whispered through the trees*"). More importantly, 88.2% of participants achieved independent creative production, successfully applying auditory triggers to formulate original expressions such as "*My thought boomed in my head*". This progression

from diagnostic failure to creative generation confirms that experiential rule application fosters genuine figurative mastery.

Conclusion

In conclusion, the proposed onomatopoeic-based approach offers an innovative and learner-centred strategy for teaching personification in poetry among Malaysian lower-proficiency ESL learners. By integrating sound-based learning with guided meaning construction, the approach seeks to transform abstract figurative language into a more concrete, engaging, and meaningful learning experience. Grounded in the Dual Coding Theory, Auditory Sensory Learning, and Constructivism, the approach encourages learners to connect auditory experiences with literary interpretation, thereby supporting deeper understanding and active participation (Clark & Paivio, 1991; Chand, 2023; Mayer, 2024).

Furthermore, this study addresses the existing gap in literature instruction by shifting from traditional explanation-based teaching towards a sensory-driven and scaffolded learning approach. It is expected that the onomatopoeic-based approach will enhance learners' understanding and application of personification while fostering greater interest and appreciation of poetry. Ultimately, the findings of this study may provide valuable pedagogical insights for English language teachers and contribute to the development of more effective and engaging literature instruction in Malaysian secondary ESL classrooms.

Recommendations

Future researchers are encouraged to examine the effectiveness of the proposed onomatopoeic-based approach through experimental studies involving larger and more diverse samples of Malaysian ESL learners. Further research may also explore its application in teaching other figurative language devices, such as metaphor, simile, and hyperbole, as well as its integration with digital and multimedia learning tools to enhance literature instruction. For practitioners, English language teachers are encouraged to adopt more creative, learner-centred, and multisensory instructional approaches when teaching poetry. Incorporating sound-based learning into literature lessons may help learners better understand abstract literary concepts, increase classroom engagement, and foster greater appreciation of poetry among Malaysian lower-proficiency ESL learners.

Limitations

Several limitations should be considered when interpreting the findings of this study. First, the study involved a small sample of 17 participants who were recruited through purposive voluntary sampling during an interactive approach exhibition. The participants were university-level exhibition visitors rather than the intended lower secondary school population. Although participants were selected based on characteristics relevant to the learning difficulties addressed by the proposed approach, their educational level and learning context differed from those of the intended target population. Therefore, the findings should not be generalised directly to Malaysian lower secondary ESL learners.

Second, the study employed a one-group pretest-posttest design without a control or comparison group. Although this design allowed the researchers to compare participants' performance before and after the intervention, it does not allow changes in performance to be

attributed solely to the Onomatopoeia-Based Approach. Other factors, such as exposure to the pre-test, the immediate learning experience, or the novelty of the exhibition activity, may have contributed to changes in post-test performance. The findings should therefore be interpreted as descriptive evidence of changes following the intervention rather than as conclusive evidence of causal effectiveness.

Third, the study examined participants' immediate post-intervention performance. The post-test was administered directly after the four-stage instructional process; therefore, the study does not establish whether the observed learning gains would be retained over a longer period. Future studies could incorporate delayed post-tests to examine whether participants retain their understanding and ability to apply personification after the intervention.

Fourth, the study focused specifically on personification and the use of onomatopoeic sounds as a sensory entry point. Consequently, the findings cannot be assumed to apply to other forms of figurative language, such as metaphor, simile, or hyperbole. In addition, participants' familiarity with particular sounds, vocabulary items, and cultural contexts may influence their ability to establish sound–meaning associations.

Fifth, the study relied primarily on frequencies and percentages because of the small sample and exploratory nature of the pilot. The analysis therefore describes patterns in participants' responses but does not establish statistical significance. Further research with a larger sample could employ appropriate inferential analyses to examine changes more rigorously.

Finally, the study was conducted within an interactive exhibition setting rather than a regular classroom environment. The exhibition context may differ from actual classroom conditions in terms of time, learner interaction, teacher support, and environmental distractions. Future research should therefore implement the approach in authentic lower secondary ESL classrooms using larger and more diverse samples, appropriate comparison groups, and follow-up assessments to provide stronger evidence regarding its applicability and longer-term learning outcomes.

Conflict of Interest

The authors declare that there was no conflict of interest that could have influenced the design, data collection, analysis, interpretation, or reporting of the study.

Funding

This study received no external funding. The research and data collection activities were conducted as part of the development and pilot demonstration of the proposed Onomatopoeia-Based Approach.

Data Availability and Transparency

The data supporting the findings of this study consist of responses collected through the study's Google Form assessment. To protect participants' privacy and confidentiality, individual-level responses and identifying information are not publicly disclosed. Anonymised aggregate data may be made available by the authors upon reasonable request, subject to applicable ethical and privacy requirements.

References

- Alkhamash, R. (2022). Processing figurative language: Evidence from native and non-native speakers of English. *Frontiers in Psychology*, 13, 1057662. <https://doi.org/10.3389/fpsyg.2022.1057662>
- Bolhan, N. S., & Ismail, H. H. (2025). Exploring digital game-based learning for teaching poetry in Malaysian primary ESL classrooms: A systematic review. *International Journal of Education, Psychology and Counselling*, 10(61), 363–376. <https://doi.org/10.35631/IJEPC.1061028>
- Chand, S. (2023). Constructivism in education: Exploring the contributions of Piaget, Vygotsky, and Bruner. *International Journal of Science and Research*, 12, 274–278. <https://doi.org/10.21275/SR23630021800>
- Chung, C. E., & Ngui, W. (2025). Blending technology with literature: Leveraging YouTube to teach poetry to secondary school students in Malaysia. *Journal of Research, Policy & Practice of Teachers and Teacher Education*, 15(1), 90–100. <https://doi.org/10.37134/jrpptte.vol15.1.8.2025>
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210. <https://doi.org/10.1007/BF01320076>
- Fielden-Burns, L. V., & Piquer-Píriz, A. M. (2022). Personification and relationships in English as a medium of Instruction business discourse: Crossing paths in metaphorical constructions. *Journal of Pragmatics*, 202, 145–159. <https://doi.org/10.1016/j.pragma.2022.11.002>
- Hussein, O. A. (2023). The significance of figurative language in teaching poetry. *Diyala Journal for Human Research*, 3(97), 492–513. <https://doi.org/10.57592/djhr.v3i97.3603>

Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, 36. <https://doi.org/10.1007/s10648-023-09842-1>

Nik Hassim, N. A. A., & Hashim, F. (2024). Lecturer beliefs and pedagogy: A correlational study of poetry instruction in Malaysian TESL programs. *Malaysian Journal of Social Sciences and Humanities*, 9(10), e003072. <https://doi.org/10.47405/mjssh.v9i10.3072>

Ukat, J. A., & Ismail, H. H. (2022). Teaching literature in Malaysian English classrooms: A review on its challenges and opportunities. *International Journal of Academic Research in Business and Social Sciences*, 12(11), 1866–1878. <https://doi.org/10.46886/IJARBSS/v12-i11/11554>

Yee, B. C., Yusof, N. M., Muniandy, S., Johari, N. D., & Mohammad Yusof, M. (2023). Students' preference for literary texts and literature learning in the Malaysian secondary school. *International Journal of Educational Research on Andragogy and Pedagogy*, 1(1), 71–82. <https://doi.org/10.1234/ijerap.v1i1.3>

Zhao, H. (2024). The role of phonesthemes in EFL learners' word acquisition. *Theory and Practice of Second Language Acquisition*, 10(2), 1–22. <https://doi.org/10.31261/TAPSLA.16086>

Zhou, X., Younas, M., Omar, A., & Guan, L. (2022). Can second language metaphorical competence be taught through instructional intervention? A meta-analysis. *Frontiers in Psychology*, 13, 1065803. <https://doi.org/10.3389/fpsyg.2022.1065803>

Wild Wonders: A CALL-Based Digital Learning Module for Enhancing Reading Comprehension among Year 4 ESL Learners

Reshantini Rajendran*, Universiti Kebangsaan Malaysia, p167176@siswa.ukm.edu.my

Nurul Juni Melissa Muhammad Akif Mark, Universiti Kebangsaan Malaysia, p167580@siswa.ukm.edu.my

Harwati Hashim, Universiti Kebangsaan Malaysia, harwati@ukm.edu.my

Nur Syafiqah Yaccob, Universiti Kebangsaan Malaysia, nursyafiqahyaccob@ukm.edu.my

Hanita Hanim Ismail, Universiti Kebangsaan Malaysia, hanitahanim@ukm.edu.my

* Corresponding Author

Abstract: Computer-Assisted Language Learning (CALL) plays an important role in enhancing ESL learning experiences for primary school learners. This paper presents the design and development of “Wild Wonders”, a CALL-based digital learning module designed to support reading comprehension among Year 4 ESL learners in Malaysian primary schools. The module responds to learning needs identified through a one-hour classroom observation of 35 mixed-proficiency Year 4 pupils, with attention to reading comprehension, writing and learner behaviour. The module is grounded in the CEFR-aligned curriculum and the Malaysian Dokumen Standard Kurikulum dan Pentaksiran (DSKP), and incorporates multimedia learning theory, scaffolding principles and CALL pedagogy. It consists of animal-themed digital content and interactive learning activities designed to provide structured and scaffolded reading experiences. Developed using Google Slides, the module integrates vocabulary development, reading comprehension, written language production and reflection within a single learning sequence. An expert review involving two English teachers was conducted through open-ended feedback to examine the pedagogical features of the module. The feedback highlighted strengths in interaction, feedback, creativity, communication and learner autonomy, while also identifying opportunities to strengthen collaborative learning. Overall, the findings suggest that “Wild Wonders” demonstrates potential as a curriculum-aligned CALL resource for supporting reading comprehension among mixed-proficiency Year 4 ESL learners. The paper contributes to digital pedagogy by illustrating how accessible digital tools can be purposefully integrated with scaffolding and interactive learning activities in the Malaysian primary ESL context.

Keywords: *Computer-Assisted Language Learning (CALL), digital pedagogy, ESL learners, multimedia learning, reading comprehension*

Introduction

Digital technologies have brought an important shift in the field of language education, particularly in ESL classrooms, where the use of Computer-Assisted Language Learning (CALL) has become increasingly significant. In these settings, CALL is often regarded as a means to create an interactive and technology-rich environment that supports learner-centred instruction, learner autonomy and engagement. However, the mere presence of technology does not automatically guarantee meaningful learning outcomes. As stated by Shiu (2025) and Li (2025), digital tools are most effective when they are aligned with instructional design, learner needs and pedagogical objectives. This is especially relevant in reading comprehension, a key component of ESL learning that requires learners to construct meaning from texts and identify

main ideas and supporting details. Some recent studies have shown that digital tools can improve motivation and understanding when they are included in structured learning programmes (Azmi & Zahari, 2024; Wiboolyasarini & Jinowat, 2024). Scaffolding and pedagogical support that encourage active learner participation are important for ensuring the effectiveness of digital tools. Hence, current studies in digital language education increasingly focus on the interaction between technology, pedagogy and learner engagement. Frameworks such as TPACK and SAMR highlight the importance of using technology to transform, rather than merely repeat, traditional instructional practices (Shiu, 2025). Digital pedagogy and technology-enhanced language learning research also emphasise the need for creating learning environments that promote interaction, collaboration and meaningful language use (Li, 2025; Lo, 2025).

Although numerous studies have examined the use of e-books (Azmi & Zahari, 2024), mobile-assisted language learning tools (Wiboolyasarini & Jinowat, 2024), and digital game-based approaches (Gatrell, 2025; Zhou, 2024) to enhance language development, it appears that limited research has conceptualised CALL-based modules that combine reading comprehension instruction, digital media and structured scaffolding in a single learning resource for primary ESL learners. In response to this identified gap, the “Wild Wonders” CALL-based digital learning module was developed to support reading comprehension among Year 4 ESL learners through the integration of interactive animal-themed activities, multimedia resources and guided learning experiences. The module aims to help learners identify main ideas and supporting details, while promoting engagement and learner autonomy. Guided by CALL principles and digital pedagogy, “Wild Wonders” provides a conceptual model for meaningful technology integration into reading instruction in the ESL classroom. Therefore, this paper aims to describe the design and development of the “Wild Wonders” CALL-based digital learning module and examine its pedagogical features through teacher review. The paper addresses the following research questions:

How can a CALL-based digital learning module be designed to support reading comprehension among mixed-proficiency Year 4 ESL learners?

How do English teacher reviewers evaluate the pedagogical features of the “Wild Wonders” module?

Literature Review

TESL Strategies and Reading Comprehension

Reading comprehension is essential in ESL learning because it helps learners construct meaning and identify main ideas and supporting details. Many young ESL learners struggle with limited vocabulary knowledge, a lack of reading strategies and insufficient instructional scaffolding (Li et al., 2022; Abdullah et al., 2026). Li et al. (2022) further emphasise the importance of explicit reading instruction in supporting learners’ comprehension, motivation and self-efficacy, while Abdullah et al. (2026) found that digital scaffolding can support mixed-proficiency learners by guiding them through reading tasks. Together, these studies suggest that structured, step-by-step guidance is important in supporting learners’ reading comprehension, particularly when learners have different levels of proficiency. According to Usman and Salsabila (2025), structured pedagogy in CALL-based instruction enhances reading outcomes, indicating that technology needs to be integrated with explicit instruction rather than used independently. Studies in the Malaysian context have also highlighted the potential of e-books and digital tools to support comprehension while maintaining ESL learners’ engagement (Azmi & Zahari, 2024; Hoe & Mohamad, 2026). Taken together, the literature indicates that effective digital reading

instruction requires a combination of explicit instruction, scaffolding and digital support. These principles informed the reading activities and guided learning tasks embedded in the “Wild Wonders” module.

Technology-Enhanced Language Learning and CALL

Technology-Enhanced Language Learning (TELL) and CALL have been widely implemented in current ESL education due to their ability to support multimodal and learner-centred learning experiences (Zhang & Liu, 2023; Bahari et al., 2025). Zhang and Liu (2023) found that learners’ digital engagement influences their language learning, suggesting that interaction with technology plays an important role in shaping learning experiences. Similarly, digital reading tools have been shown to support comprehension through active engagement with texts (Zhang, 2026), while Wiboolyasarin and Jinowat (2024) found that mobile-assisted learning increases motivation and participation among young learners. Furthermore, CALL projects incorporating AI-driven systems can establish interactive learning environments (Bahari et al., 2025). Collectively, these studies indicate that CALL can provide opportunities for active, multimodal and interactive language learning when digital tools are incorporated into purposeful learning activities. This perspective supports the design of the interactive digital learning activities incorporated into the “Wild Wonders” module.

Digital Pedagogy in ESL Education

Shiu (2025) notes that digital pedagogy involves the use of technology to enhance learning practices in language education. Lo (2025) found that digital pedagogy supports culturally responsive and learner-centred instruction. Various frameworks, such as TPACK and SAMR, focus on the knowledgeable use of technology and the pedagogical design of content and learning strategies to assist learners (Shiu, 2025). These frameworks reinforce the importance of using technology in pedagogically effective ways within the ESL classroom. Rather than treating technology as an isolated tool, these perspectives emphasise the need to align digital resources with instructional purposes and learner needs. Mohammad Shamsul et al. (2024) similarly state that digital tools are most meaningful when used with a clear understanding of instructional goals. Access to digital tools has also been highlighted as a factor contributing to differences in accessibility to technology-based learning environments (Yousofi et al., 2025). Therefore, access and inclusivity should also be considered when applying principles of digital pedagogy. Overall, the literature suggests that effective digital pedagogy requires alignment among instructional goals, technology and learner needs, while also considering accessibility. These considerations are reflected in the instructional design of the “Wild Wonders” module, particularly through its structured and scaffolded digital learning activities.

Learner Engagement in Digital Learning Environments

Learner engagement is an important factor in the learning process, especially in digital environments. According to Krishnan and Sulaiman (2025), digital materials used for reading purposes allow learners to experience a greater level of interest than print-based text materials. Zhou (2024) and Gatrell (2025) consistently reported that gamified learning enhances learners’ motivation, engagement and active participation, particularly among young ESL learners. These studies suggest that engaging digital learning experiences can encourage learners to participate actively in language learning activities and create opportunities for reading development. AI tools such as ChatGPT can also increase learner engagement by providing interactive feedback and quick responses (Jahan & Akbar, 2026). Across these studies, interactivity, gamification and responsiveness emerge as features that can support learner engagement; however, engagement is most meaningful when these features are embedded within activities with clear learning purposes. The “Wild Wonders” module therefore

incorporates interactive tasks and immediate feedback to support learner engagement throughout the learning process.

Educational Technology Integration in ESL Learning

Integration of educational technology in language learning moves beyond the use of individual digital tools to design learning activities that connect content delivery, learner interaction and instructional progression. Usman and Salsabila (2025) suggest that structured digital systems that support coherent learning sequences can be more effective in technology-enhanced learning than isolated activities. Emerging technologies can create adaptive and connected language learning environments when combined with pedagogically meaningful activities (Li, 2025). Thus, technological tools should function as part of an integrated instructional system that connects learning input, practice and output. Moreover, CALL-based environments can provide multiple instructional functions, including input, interactive practice and performance feedback (Bahari et al., 2025). Zhang (2026) demonstrates that structured interaction through computer-assisted reading systems can support learners' reading comprehension. Laksanasut (2026) and Isaee and Barjesteh (2026) highlight the potential of adaptive and AI-supported technologies to personalise learning, although their effectiveness depends on structured instructional design. Taken together, these studies reinforce the importance of sequencing digital learning activities so that learners move from input to guided practice and response with appropriate feedback. This design principle is reflected in the "Wild Wonders" module, where learners progress through structured reading, comprehension and writing response activities supported by digital feedback.

Methodology

Needs Analysis

The needs analysis was conducted through a one-hour classroom observation of 35 mixed-proficiency Year 4 pupils during a reading lesson. A self-developed classroom observation checklist was used to identify the learners' instructional needs. The checklist included criteria related to reading comprehension, sentence construction and learner behaviour, with particular attention to learners' ability to understand key vocabulary, identify main ideas and specific information, answer comprehension questions, and demonstrate engagement and confidence during reading activities. Each criterion was rated using a four-point scale from 0 (rarely demonstrated) to 3 (consistently demonstrated). The observation indicated difficulties in vocabulary understanding, identifying main ideas and specific information, answering comprehension questions, and maintaining engagement and confidence during reading activities. The identified needs were used to inform the development of the "Wild Wonders" module, particularly the inclusion of vocabulary development, structured reading comprehension activities, guided support, immediate feedback and reflective learning.

Design and Development of the Module

A CALL-based module titled "Wild Wonders" was developed to support reading comprehension among Year 4 ESL learners. Based on the needs analysis, the module uses the *Amazing Animals* topic from the *Get Smart Plus 4* textbook because it is a familiar Year 4 CEFR topic that contains rich descriptive vocabulary and informational texts. This topic provides opportunities for learners to develop reading skills while ensuring alignment with the Year 4 CEFR curriculum. The identified learning needs informed the integration of reading comprehension instruction, learner engagement and technology-enhanced learning principles into the module. These considerations led to the development of "Wild Wonders" as a CALL-based learning environment for Year 4 ESL learners. Moreover, the module was developed using Google Slides, while Liveworksheets and Padlet were embedded to support interactive

learning activities. These are accessible digital platforms that can be implemented in primary ESL classrooms. A key feature of “Wild Wonders” is its integration of vocabulary learning, reading comprehension, immediate feedback and reflective learning within a single learning sequence, allowing learners to progress through scaffolded activities independently.

Target Learners and Learning Objectives

The module targets Year 4 ESL learners, as the needs analysis was conducted among learners at this level. The classroom consisted of learners with mixed proficiency levels of English, highlighting the need for differentiated support and scaffolding. Therefore, the “Wild Wonders” module was designed specifically for Year 4 learners to support both guided and independent learning through digital activities. The learning objectives of the module are aligned with the Year 4 CEFR English language curriculum, which focuses on developing learners’ ability to identify main ideas and supporting details in reading texts. By the end of the module, learners should be able to complete an animal-related vocabulary activity, answer comprehension questions based on a reading text and produce a simple written response about their favourite wild animal. These objectives gradually develop vocabulary knowledge, reading comprehension and language production skills among Year 4 learners.

Digital Tools and Instructional Design

The “Wild Wonders” module includes digital elements such as colourful images, puzzles, a reading text and a quiz with immediate feedback. These features allow learners to engage actively with the content and support a learner-centred learning environment. Moreover, the instructional design of “Wild Wonders” reflects scaffolding principles of Vygotsky’s Sociocultural Theory (1978), providing temporary instructional support for learners to complete tasks that they may not be able to complete independently. Scaffolding is provided through the sequencing of activities from vocabulary development to reading comprehension and written language production, allowing learners to become increasingly independent as they progress through the module. According to Abdullah et al. (2026) and Usman and Salsabila (2025), combining structured instructional support with clear learning objectives can support effective digital learning environments. Shiu (2025) and Lo (2025) also emphasise that technology integration should be guided by pedagogical principles rather than technology use alone. Therefore, scaffolding principles are embedded in the design of “Wild Wonders” together with the pedagogically meaningful integration of technology emphasised in the TPACK and SAMR frameworks (Shiu, 2025).

Learning Activities

The activities in the module are designed in a structured sequence that demonstrates progressive learning. Learners begin with the Puzzle Challenge in “Wild Wonders”, which allows them to learn vocabulary related to wild animals by identifying and describing the animal in the puzzle. This game-based learning element activates learners’ prior knowledge and increases their engagement. The Tiger Trail Challenge and Quiz form the second part of the module and focus on reading comprehension. These activities require learners to read a short text and answer comprehension questions to check their understanding of the main ideas and supporting details. The quiz also provides immediate feedback after each response, allowing learners to self-correct. The module then progresses to the My Favourite Wild Animal activity, which requires learners to write a short response based on their knowledge of the topic. Different levels of guidance are provided in this activity to cater to learners with mixed proficiency levels. The final task, My Reflection, requires learners to answer questions to evaluate and reflect on their learning experience. The use of gamified and interactive activities plays an important role in engaging learners and supporting their language development (Zhou, 2024; Gatrell, 2025), which is reflected in the activities designed in the “Wild Wonders” module.

Assessment and Feedback

Formative assessment and final assessment are incorporated into the ‘Wild Wonders’ module as part of the learning activities. The formative assessment is integrated into the quiz activity, allowing learners to assess their understanding of the reading text and receive immediate feedback on their responses. Immediate feedback allows learners to identify errors and engage actively in the learning process. Zhang (2026) emphasises that interactive and immediate feedback can support learners’ engagement and reading comprehension in digital learning environments. A final assessment is also included in the writing part of the module, where learners write a short response about their favourite wild animal. “Wild Wonders” also provides guidance throughout the module, reflecting the importance of scaffolding in facilitating language learning experiences (Abdullah et al., 2026). The module includes a reflection task that functions as a self-assessment, encouraging learners to reflect on their learning progress and identify areas for improvement. Hence, the assessment and feedback features of “Wild Wonders” support learner autonomy and the continuous learning process in addition to assessing learners’ understanding of the content.

Expert Review

Following the development of the “Wild Wonders” module, an expert review was conducted to obtain feedback on its pedagogical features. Two English teachers reviewed the module. Both reviewers hold Master’s degrees in Teaching English as a Second Language (TESL). Teacher 1 has 20 years of English teaching experience and 8 years of experience using ICT in teaching and learning, while Teacher 2 has 18 years of English teaching experience and 11 years of experience using ICT in teaching and learning. The reviewers were provided with the “Wild Wonders” module developed using Google Slides, including its integrated interactive learning activities. Feedback was collected through open-ended comments. The reviewers were invited to comment on the pedagogical features of the module, including interaction, collaboration, feedback, creativity, communication and learner autonomy. The open-ended feedback was analysed by identifying comments that described the strengths of the module and areas for improvement. These comments were then used to describe the pedagogical features of the module and identify aspects that could be strengthened in future development.

Findings

Technology Integration and Expert Review of the “Wild Wonders” Module

The “Wild Wonders” module was developed based on CALL principles to support reading comprehension among Year 4 ESL learners through interactive activities, multimedia components and scaffolded learning experiences. An expert review was conducted to examine the pedagogical features of the module. Two experienced English teachers provided open-ended feedback on six dimensions associated with technology-enhanced language learning: interaction, collaboration, feedback, creativity, communication and learner autonomy (Abdullah et al., 2026). Overall, the reviewers highlighted strengths in the module’s interactive features, feedback, creativity, communication and learner autonomy, while also identifying opportunities to strengthen collaboration.

Interaction through Digital Learning Activities

Interaction is a key component of CALL that provides opportunities for learners to be actively involved with learning materials. The “Wild Wonders” module incorporates interactive activities such as the Puzzle Challenge, Tiger Trail and Quiz to encourage active participation. According to the reviewers, the Puzzle Challenge helps learners engage with vocabulary related

to wild animals by identifying and describing the animals. This allows learners to draw on their existing knowledge while developing key vocabulary for the subsequent reading task. The Tiger Trail and Quiz also allow learners to interact with the reading text by identifying main ideas and supporting details and answering questions to check their understanding. These features were viewed as supporting active learner participation and are consistent with Zhang and Liu's (2023) finding that digital engagement can positively influence language learning.

Opportunities for Collaboration

Although "Wild Wonders" was not designed with embedded collaborative features, the reviewers identified opportunities for teachers to incorporate pair or group work during classroom implementation. The reviewers noted that some activities can be conducted by assigning learners to work together. For example, learners can discuss clues and solve the Puzzle Challenge in pairs or small groups. Similarly, learners can collaborate during the My Favourite Wild Animal activity by sharing their responses and providing feedback to one another. Thus, while collaboration is not directly embedded in the digital format of "Wild Wonders", the module allows teachers to facilitate collaborative learning through classroom implementation. This feedback also highlights collaboration as an area that could be strengthened in future development.

Feedback and Learner Support

Feedback was identified as an important strength of the "Wild Wonders" module, particularly in the Tiger Trail and Quiz activities. Teacher 1 commented on the immediate feedback feature:

"The immediate response eliminates the wait for teacher feedback. Built-in response features allow learners to self-correct and return to the reading text to fix misunderstandings. This feedback feature also provides digital scaffolding for the learner."

The feedback suggests that the immediate response feature allows learners to identify errors, return to the reading text and make corrections with reduced dependence on teacher feedback. This observation is consistent with Abdullah et al. (2026), who highlighted the importance of scaffolding in digital learning environments. Zhang (2026) similarly reported that immediate feedback can support learners' engagement and reading comprehension in computer-assisted reading systems. Therefore, the reviewers viewed the feedback mechanism as a feature that supports learners' understanding, self-correction and independent learning.

Promoting Creativity through Language Production

Creativity is supported through activities that allow learners to produce personalised responses. The "My Favourite Wild Animal" activity requires learners to choose an animal of interest and create their own written descriptions using vocabulary and knowledge acquired throughout the module. The reviewers highlighted this activity as a strength because it allows learners to move beyond comprehension and produce meaningful language. The activity also provides scaffolding while allowing learners to contribute their own ideas based on their preferences. This combination of guidance and personal choice provides support for learners with different proficiency levels. This finding is consistent with Gatrell (2025), who highlighted the potential of engaging digital learning experiences to support creativity in language learning.

Supporting Communication Skills

Communication is incorporated into the "Wild Wonders" module, particularly through the "My Favourite Wild Animal" and "My Reflection" activities. The "My Favourite Wild Animal"

activity allows learners to organise information and communicate ideas through writing, enabling them to apply vocabulary and reading comprehension skills in a meaningful context. The “My Reflection” activity provides learners with an opportunity to communicate their experiences, challenges and perceived learning through responses to the reflection questions. According to the reviewers, these activities provide opportunities for learners to express ideas, share information and reflect on their learning experiences. Thus, the reviewers identified purposeful written expression and reflection as features supporting communication within the module.

Encouraging Learner Autonomy

The structure of “Wild Wonders” supports learner autonomy by allowing learners to progress through the activities with guidance from instructions, scaffolding and timely feedback. The reviewers identified the module’s self-paced format as a significant strength. Learners can navigate the activities at their own pace and monitor their understanding through the assessment and reflection tasks. The “My Reflection” activity also encourages learners to identify their strengths and areas for improvement, which may support metacognitive awareness and self-regulated learning. These features are consistent with Usman and Salsabila (2025), who emphasised the role of well-designed CALL environments in promoting learner autonomy. Overall, the reviewers viewed the self-paced structure, scaffolding and reflection activities as features that support learners’ independent engagement with the module.

Relevance to Malaysian ESL Context

Alignment with Curriculum Needs and Learner Diversity

“Wild Wonders” aligns with the CEFR English language curriculum implemented in Malaysian primary schools. The module is designed to support learners in identifying main ideas and supporting details, addressing key reading comprehension needs identified during the needs analysis. The use of the *Amazing Animals* topic from the *Get Smart Plus 4* textbook further reflects familiar curriculum content and classroom practices. In addition, mixed-proficiency learners are common in Malaysian ESL classrooms, highlighting the need for differentiated support. Therefore, “Wild Wonders” incorporates scaffolded activities, guided support and differentiated levels of guidance to accommodate learners with varying levels of English proficiency. This provides ESL teachers with a curriculum-aligned digital resource that can support reading comprehension instruction in mixed-ability classrooms.

Digital Readiness and Accessibility

The development of “Wild Wonders” reflects the growing use of technology-enhanced learning in ESL education. Research indicates that interactive digital tools can support learners’ engagement and comprehension when integrated with appropriate pedagogical support (Zhang, 2026; Wiboolyasarini & Jinowat, 2024). Recognising that learners may differ in their access to digital resources and digital literacy, “Wild Wonders” was developed using Google Slides and incorporates visual support, clear instructions, guided activities and immediate feedback. These features provide a relatively simple digital learning environment that can support learners as they navigate the activities. Yousofi et al. (2025) also highlight the importance of equitable learning opportunities in digital pedagogy. Therefore, “Wild Wonders” offers a practical approach to technology-enhanced ESL learning through accessible digital tools and structured instructional support.

21st-Century Learning

21st-century learning skills, including communication, critical thinking and learner autonomy, are incorporated into the activities of “Wild Wonders”. For example, the reading

comprehension quiz requires learners to process information, answer both literal and inferential questions, and demonstrate understanding, while the *My Favourite Wild Animal* and *My Reflection* activities provide opportunities to express ideas and reflect on their learning. In addition, gamified and interactive elements can support learner engagement and active participation in language learning (Gatrell, 2025; Zhou, 2024). Immediate feedback and the self-assessment component further encourage learners to monitor their understanding and take greater responsibility for their learning. Therefore, “Wild Wonders” provides Malaysian ESL teachers with a practical digital resource that supports learner-centred instruction while promoting active participation, communication and learner autonomy.

Reflection and Potential Contribution

Strengths of the Module

The learner-centred and scaffolded design of “Wild Wonders” is a key strength of the module. The design provides structured support for learners while encouraging gradual independence. The lesson sequence begins with the Puzzle Challenge, which activates learners’ prior knowledge, followed by reading comprehension, a guided writing task and reflection. This step-by-step sequence provides learners with progressive support as they move from vocabulary development to comprehension and language production. According to Abdullah et al. (2026), digital scaffolding can support learners with diverse proficiency levels. The Tiger Trail and Quiz activities also provide immediate feedback, enabling learners to identify errors and self-correct with reduced dependence on teacher feedback. This supports learner autonomy and is consistent with Zhang’s (2026) findings on the benefits of immediate feedback in computer-assisted reading. Moreover, the use of an animal theme and interactive media provides opportunities to maintain learners’ interest and engagement throughout the module. Gatrell (2025) and Zhou (2024) similarly highlight the potential of interactive and engaging learning activities to support motivation and engagement in language learning. Hence, the learner-centred and scaffolded design of “Wild Wonders” provides a structured approach to support reading comprehension among Year 4 ESL learners.

Limitations of the Module

A limitation of the “Wild Wonders” module is that collaborative learning is not directly embedded within the digital activities. As a result, opportunities for pair or group work depend on teachers’ implementation and adaptation of the activities during classroom use. In addition, the implementation of the module may be influenced by learners’ access to digital devices and stable internet connectivity, which may present challenges in educational settings with limited technological resources. Yousofi et al. (2025) highlighted that unequal access to technology remains a challenge in technology-enhanced learning environments. Thus, even though “Wild Wonders” provides interactive and scaffolded learning experiences, its classroom implementation may require appropriate technological resources and teacher facilitation.

Contribution to TESL Practices, Digital Pedagogy and TELL

The “Wild Wonders” module contributes to TESL practices by demonstrating how a CALL-based digital resource can be designed to support reading comprehension among Year 4 ESL learners. The module integrates vocabulary development, reading comprehension tasks, guided writing and assessment with technology to support learners in identifying main ideas and supporting details. In particular, “Wild Wonders” provides scaffolded guidance for mixed-proficiency learners while also creating opportunities for independent learning. This reflects TESL practices that emphasise differentiated and learner-centred instruction. In the context of digital pedagogy, the module demonstrates how technology can be used purposefully to support

learning objectives rather than functioning as an end in itself. Each activity is designed to support reading comprehension outcomes through scaffolding, multimedia, game-based elements and immediate feedback. This aligns with Shiu (2025) and Lo (2025), who emphasise the importance of aligning technology with pedagogy and content in digital learning environments. Moreover, “Wild Wonders” illustrates how TELL can support vocabulary development, reading comprehension, written responses, reflective learning and immediate feedback within a single learning sequence. These features demonstrate how technology can be meaningfully integrated into language education, consistent with the emphasis on purposeful technology integration in language learning (Usman & Salsabila, 2025; Bahari et al., 2025). Hence, “Wild Wonders” contributes to TESL practices, digital pedagogy and TELL by demonstrating a practical approach to integrating technology, pedagogy and reading instruction for young ESL learners.

Implications for Future Educational Innovation

As educational technology continues to advance, the “Wild Wonders” module could provide a foundation for further innovation in TELL. Future versions of the module could incorporate personalised learning systems, AI-assisted feedback and learning analytics to provide more tailored learning experiences and enable teachers to monitor learners’ progress more effectively. These possibilities are consistent with current developments in educational technology that emphasise flexible and learner-centred learning environments (Li, 2025; Isaee & Barjesteh, 2026). Future development could also incorporate more embedded collaborative features to address the limitation identified through the expert review. Therefore, “Wild Wonders” provides a foundation that could be further developed through emerging technologies and enhanced collaborative features.

Conclusion

There are new opportunities to enhance English language teaching and learning as technology continues to be integrated into educational settings. In particular, technology-enhanced learning can provide opportunities to support reading comprehension among ESL learners. This paper presents “Wild Wonders” as a conceptual CALL module designed for Year 4 ESL learners, which integrates technology with pedagogical principles to promote meaningful learning experiences. The module is structured around key instructional principles, including scaffolding, learner-centred instruction, multimedia integration, game-based learning elements, feedback provision, vocabulary activation, reflection and assessment. Rather than presenting isolated reading activities, “Wild Wonders” offers a sequence of instructional activities that supports learners in gradually developing their reading comprehension skills. Moreover, the module demonstrates that effective technology integration depends on instructional design rather than the use of technology alone. Hence, “Wild Wonders” aligns with curriculum goals, reading comprehension objectives and the diverse learning needs of primary ESL learners, making it relevant to current ESL classrooms. The module also illustrates how CALL can support learner engagement and language learning when technology is integrated in a pedagogically meaningful manner. Overall, “Wild Wonders” provides a practical example of technology integration in primary ESL reading instruction. Although future developments may incorporate more advanced technologies, such as AI-supported features, the module establishes a foundation for further development of TELL in the ESL context.

References

- Abdullah, A. D., Izhar, N. A., & Samsudin, M. A. (2026). Bridging gaps with technology: A systematic review of digital scaffolding in ESL classrooms. *EDUCATUM Journal of Social Sciences*, 12(Special Issue), 72–87. <https://doi.org/10.37134/ejoss.vol12.sp.9.2026>
- Azmi, N. N., & Zahari, Z. (2024). E-books as a tool to improve reading comprehension among Year 4 primary ESL students. *Asian Journal of English Language and Pedagogy*, 12(1), 88–104.
- Bahari, A., Han, F., & Strzelecki, A. (2025). Integrating CALL and AIALL for an interactive pedagogical model of language learning. *Education and Information Technologies*, 30(4), 4123–4142. <https://doi.org/10.1007/s10639-025-13388-w>
- Gatrell, D. (2025). Digital game-based language learning in the young L2 learner classroom. *ELT Journal*, 79(4), 583–596. <https://doi.org/10.1093/elt/ccaf012>
- Hoe, X. Y., & Mohamad, M. (2026). Harnessing digital tools for translanguaging in ESL classrooms: A systematic literature review. *Asian Journal of English Language and Pedagogy*, 14(1), 11–31. <https://doi.org/10.37134/ajelp.vol14.1.2.2026>
- Isaee, H., & Barjesteh, H. (2026). Exploring teachers' and learners' perceptions of AI-supported pedagogical tools in English language teaching. *Discover Artificial Intelligence*, 6(1), Article 118. <https://doi.org/10.1007/s44163-026-00118-2>
- Jahan, F., & Akbar, H. (2026). Exploring ChatGPT in EFL classrooms: Insights into engagement, proficiency and ethics. *Discover Education*, 5(1), Article 44. <https://doi.org/10.1007/s44217-026-00044-7>
- Krishnan, S., & Sulaiman, T. (2025). Comparative effect of digital and print reading materials on engagement and comprehension among young ESL learners. *International Journal of Research and Innovation in Social Science*, 9(3), 112–126.
- Laksanasut, S. (2026). Artificial intelligence in EFL/ESL education: Transforming language learning and teaching through adaptive technology and ethical innovation. *International Journal of Sociologies and Anthropologies Science Reviews*, 6(2), 55–72.
- Li, H., Gan, Z., Leung, S. O., & An, Z. (2022). The impact of reading strategy instruction on reading comprehension, strategy use, motivation and self-efficacy in Chinese university EFL students. *SAGE Open*, 12(3), 1–13. <https://doi.org/10.1177/21582440221086659>
- Li, Z. (2025). Bridging pedagogy and technology: A generative AI and IoT approach to transformative English language education. *Education and Information Technologies*, 30(6), 7781–7798. <https://doi.org/10.1007/s10639-025-13210-6>
- Lo, A. W. T. (2025). A digital pedagogy for transculturing ELT through Global Englishes. *ELT Journal*, 79(3), 428–441. <https://doi.org/10.1093/elt/ccae021>
- Mohamad Shamshul, I. S., Ismail, H. H., & Nordin, N. M. (2024). Using digital technologies in teaching and learning of literature in ESL classrooms: A systematic literature

- review. *International Journal of Learning, Teaching and Educational Research*, 23(5), 112–130.
- Shiu, W. H. C. (2025). Conceptualising the pedagogical purposes of technologies by Technological Pedagogical Content Knowledge (TPACK) and Substitution Augmentation Modification Redefinition (SAMR) in English as a Second Language classrooms. *Education Sciences*, 15(4), Article 411. <https://doi.org/10.3390/educsci15040411>
- Usman, A. S., & Salsabila, H. (2025). The effectiveness of CALL in EFL learning: A synthesis of empirical evidence. *Pedagogy: Journal of English Language Teaching*, 10(1), 1–15.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wiboolyasarini, W., & Jinowati, N. (2024). Revolutionising dual language learning for young learners with the RILCA app: A mobile-assisted approach. *Journal of Computers in Education*, 11(2), 245–262. <https://doi.org/10.1007/s40692-024-00312-8>
- Yousofi, R., Golzar, J., & Yacoub, O. (2025). Digital divides and bridges: Equity implications of EdTech in ESL education. *Frontiers in Education*, 10, Article 1590828. <https://doi.org/10.3389/educ.2025.1590828>
- Zhang, P. (2026). Effects of highlights and annotations on EFL learners' reading comprehension: An application of computer-assisted interactive reading model. *Computer Assisted Language Learning*, 39(1), 1–22. <https://doi.org/10.1080/09588221.2025.2410166>
- Zhang, Y., & Liu, G. (2023). Examining the impacts of learner backgrounds, proficiency level and the use of digital devices on informal digital learning of English: An explanatory mixed-method study. *Computer Assisted Language Learning*, 36(8), 1542–1565. <https://doi.org/10.1080/09588221.2023.2267627>
- Zhou, S. (2024). Gamifying language education: The impact of digital game-based learning on Chinese EFL learners. *Humanities and Social Sciences Communications*, 11(1), Article 215. <https://doi.org/10.1057/s41599-024-02789-3>

Digital Literacy Knowledge and Reported Application of Digital Practices Among Politeknik Nilai Students: A Pilot Study

Arnie Chiah Liling, Fazlind binti Mohd Shuraimi, Nor Farina binti Jamaluddin
Politeknik Nilai
Politeknik Ungku Omar
Politeknik Muadzam Shah

ABSTRACT

The increasing integration of digital technologies and artificial intelligence into higher education has created a growing need for students to locate, evaluate, interpret, and use online information appropriately. This pilot study examined the perceived digital-literacy knowledge and reported academic digital practices of first-semester students enrolled in Communicative English 1 (DUE10012) at Politeknik Nilai. Guided by Eshet's Digital Literacy Framework, the study adopted a quantitative, cross-sectional design using an online questionnaire comprising 20 five-point Likert-scale items. Data were collected from 51 students and analysed using descriptive statistics and Spearman's rank-order correlation. The findings indicated generally positive self-reported perceptions of selected digital-literacy practices. Knowledge-related item means ranged from 3.47 to 3.90, with a composite mean of 3.77, while application-related item means ranged from 3.39 to 3.98, with an overall mean of 3.73. Students reported frequent use of search engines, background research, personal mobile data, and digital resources when completing academic tasks. However, lower or more moderate responses were observed for practices involving independent academic production, language-related digital tools, and direct copying of online material. A strong positive association was observed between the knowledge-related and application-related scores, although this result should be interpreted cautiously because both variables were measured using the same self-reported instrument. The study did not objectively assess students' performance, and the questionnaire had not undergone formal construct validation or internal consistency reliability testing. The findings therefore represent preliminary evidence about students' perceptions and reported practices rather than generalisable estimates of digital-literacy competence. The study suggests that Communicative English instruction could incorporate explicit activities involving source evaluation, paraphrasing, synthesis, citation, academic integrity, and the responsible use of digital and artificial-intelligence tools.

Keywords: *digital literacy, digital literacy knowledge, digital literacy application, Communicative English, polytechnic students, higher education*

1.0 INTRODUCTION

Digital literacy has become an important competence as students increasingly learn, communicate, and produce academic work in technology-mediated environments. In higher education, students are expected not only to access digital information but also to evaluate its credibility, interpret different forms of content, use digital tools appropriately, and produce original academic work. Early definitions described digital literacy primarily as the ability to use digital technologies and access information effectively (Gilster, 1997; Pool, 1997). Later scholars expanded the concept to include technical, cognitive, social, and critical abilities (Bawden, 2008; Lankshear & Knobel, 2015). Similarly, Ferrari (2013) conceptualised digital competence as a combination of knowledge, skills, and attitudes that enables individuals to use digital technologies confidently, critically, and responsibly.

Digital literacy therefore involves more than operating a computer or searching the internet. Students must be able to determine whether online information is accurate, relevant, current, and trustworthy. The Association of College and Research Libraries (ACRL, 2016) explained that information literacy requires learners to recognise an information need, locate appropriate information, evaluate sources critically, and use information ethically. Eshet (2004, 2012) likewise presented digital literacy as a multidimensional competence comprising photo-visual, reproduction, branching, information, socio-emotional, and real-time literacies. These dimensions enable learners to interpret digital content, navigate non-linear information environments, transform existing materials into new meaning, evaluate sources, participate responsibly, and manage digital tasks.

The importance of digital literacy has become even more evident with the development of artificial intelligence (AI) and generative artificial intelligence (GenAI). Recent scholarship distinguishes AI literacy from general digital literacy by emphasising knowledge of AI systems, critical evaluation of AI-generated content, ethical decision-making, and responsible use of AI applications (Ng et al., 2021; Laupichler et al., 2022). Almatrafi et al. (2024) discovered that AI literacy has been conceptualised through several related dimensions, including knowledge, skills, critical thinking, ethics, and practical application. Lintner (2024) similarly reported that existing AI literacy scales measure different aspects of users' understanding and use of AI, which suggests that AI literacy remains a developing and multidimensional field. Hornberger et al. (2023) further highlighted the importance of assessing what tertiary level students comprehend about AI rather than assuming that regular technology use considered to be AI literate.

In English language education, GenAI tools can support idea generation, language feedback, revision, and vocabulary development. However, their use also raises concerns about inaccurate information, overdependence, plagiarism, authorship, bias, and academic integrity. UNESCO (2023) has therefore recommended that educational institutions provide clear guidance on the responsible use of GenAI in education and research (Miao et al., 2023). Castillo Martínez et al. (2024) also reported that AI use in higher education produces both pedagogical opportunities and ethical challenges. In an English language as a foreign language context, Almusharraf et al. (2025) elucidated that students perceived GenAI as useful for writing and reported positive levels of usefulness, self-efficacy, satisfaction, and intention to reuse AI tools. Nevertheless, inclination to use GenAI and familiarity with digital tools do not automatically determine that students can gauge AI-generated output or use such tools properly.

Digital literacy is particularly important in technology-enhanced language learning and English language as a second or foreign language education. Digital tools able to support students' reading, writing, listening, speaking, vocabulary development, and independent learning (Godwin-Jones, 2015, 2016, 2018). Alakrash and Abdul Razak (2021), on the other hand, observed that digital technology and digital literacy are progressively connected to technology-based language education. In the Malaysian ESL context, Li et al. (2018) explained that digital technologies can assist the development of ESL writing among pre-service teachers. However, effective language learning depends on students' ability to select suitable tools, evaluate information, interpret online content, and incorporate information into original written or spoken communication. Access to digital tools alone does not assure the higher-order digital literacy development.

Prior research has also confirmed that digital literacy is not equally developed across all skill areas. Gutiérrez Ángel et al. (2022) reported on differences in how digital literacy is defined, measured, and applied across higher education settings. Tinmaz et al. (2022) similarly justified that digital literacy consists of the ability to access, evaluate,

create, and communicate information through digital technologies. These findings contested the assumption that students are inevitably digitally adept simply because they have grown up using digital devices. Ng (2012) argued that students who are familiar with digital technologies may still need clear instruction in critical evaluation, information management, and responsible digital involvement.

Digital technologies in Malaysia are increasingly integrated into tertiary education, blended learning, online assessment, and academic communication. Malaysian students, therefore, need digital literacy to locate academic sources, evaluate online information, use language-learning applications, and submit coursework through digital platforms. However, digital access does not always lead to applicable academic use. Selwyn and Facer (2013) reasoned that digital disparity extends beyond access to differences in skills, confidence, support, and critical engagement. Starkey (2016), on the other hand, emphasised on the magnitude of an equitable curriculum that prepares students to participate meaningfully in a digital society. Tang and Chaw (2016) further noted that students need more than basic technical skills to gain from digital learning environments.

Communicative English 1 provides an appropriate context for examining these issues among first-semester polytechnic students. The course requires students to locate information, understand online materials, prepare written assignments, deliver presentations, and communicate ideas in English. These activities involved more than language knowledge; in which, students must also identify relevant sources, evaluate credibility, paraphrase and synthesise information, cite sources appropriately, and use digital tools responsibly. The course, subsequently, reflects several dimensions of Eshet's framework, particularly information, branching, reproduction, and socio-emotional literacy (Eshet, 2004, 2012). It also provides a relevant context for considering emerging AI literacy because students may use search engines, online dictionaries, grammar tools, and GenAI applications when completing language-related tasks.

1.1 Background of the Study

Digital literacy is widely understood as a multidimensional competence involving technical, cognitive, and socio-emotional abilities (Gilster, 1997; Ng, 2012; Eshet, 2012). Contemporary interpretations also emphasise critical information evaluation, ethical participation, content creation, safety, and problem-solving. Thus, digital literacy should not be reduced to the frequency with which students use digital devices or applications. Regular technology use may provide familiarity, but it does not necessarily develop the higher-order skills required for academic judgement and knowledge production.

This distinction is especially significant when students use online information for language-learning tasks. Such tasks often require students to read across several sources, determine whether information is credible, identify relevant evidence, paraphrase ideas, and construct an original response. These processes involve more than retrieval. They require critical reading, synthesis, language awareness, and academic integrity. In this respect, digital literacy and English-language competence are interconnected: students must use digital resources to support communication while also maintaining control over the accuracy, meaning, and originality of their work.

The growth of GenAI has made these requirements more complex. AI tools can provide explanations, translations, vocabulary suggestions, writing feedback, and models of academic language. Nevertheless, students must be able to judge whether AI-generated suggestions are accurate and suitable for their task. They must also understand when and how AI use should be acknowledged and how to avoid presenting generated content as their own work. Recent scholarship in EFL education highlights the need for pedagogical approaches that use AI as a scaffold while maintaining learner agency, critical thinking, and ethical responsibility. Eshet's Digital Literacy Framework provides a suitable theoretical foundation for examining these competencies. The framework presents digital literacy as a set of cognitive and practical literacies rather than as simple technical proficiency. Photo-visual literacy involves interpreting visual and multimedia information; branching literacy concerns navigation through non-linear digital environments; information literacy involves searching for and evaluating information; reproduction literacy involves transforming existing digital content into new meaning; socio-emotional literacy concerns ethical and responsible

online interaction; and real-time literacy involves managing multiple digital activities. In the present study, these dimensions are examined through two practical categories: students' knowledge of appropriate digital practices and their reported application of those practices during Communicative English 1 tasks.

1.2 Problem Statements

At the global level, higher education institutions experienced profound shifts as digital tools and generative artificial intelligence became deeply embedded in academic environments, requiring learners to possess advanced competencies beyond basic technological access (Miao et al., 2023; UNESCO, 2023). Despite having regular access to digital technologies, students worldwide continued to experience difficulties when evaluating online information and managing academic tasks critically (Park, 2019). Furthermore, while digital literacy evolved significantly over time, scholars noted that digital familiarity did not automatically result in ethical or critical information use (Eshet Alkalai & Chajut, 2009; Ng, 2012). Consequently, international students often located information quickly but struggled to determine whether a source was credible, biased, current, or suitable for academic assignments, a challenge that intensified with the widespread availability of generative artificial intelligence tools requiring sophisticated evaluation and ethical navigation (Almatrafi et al., 2024; Laupichler et al., 2022).

Following this global trend, the Malaysian higher education landscape experienced rapid technological integration across blended learning platforms, online assessments, and digital communication channels. Within this national context, Malaysian students were similarly expected to locate academic sources, evaluate online materials, and submit coursework through digital portals. However, researchers observed that widespread digital access in Malaysia did not always translate into effective academic application or critical competence, mirroring broader global concerns regarding digital inequalities and skill disparities (Selwyn & Facer, 2013; Tang & Chaw, 2016). Although national educational policies emphasized digital transformation, students frequently encountered difficulties in transforming online information into original academic work, often struggling with paraphrasing, synthesis, and the avoidance of unacknowledged copying (Li et al., 2018).

Narrowing down to the specific local setting, these challenges became particularly pronounced among first-semester students in Technical and Vocational Education and Training institutions, such as those enrolled in Communicative English 1 at Politeknik Nilai. Although previous studies examined general university populations and broad digital access, considerably less attention was directed toward introductory polytechnic students who were required to balance language development in English with complex digital information processing. These learners faced the dual demand of mastering professional communication while navigating online information environments and emerging digital assistance tools without adequate instructional scaffolding. Therefore, this local gap necessitated preliminary empirical investigation into how first-semester polytechnic students understood and applied digital literacy practices when completing authentic language tasks.

1.3 Originality and Contribution of the Study

This pilot study contributed to the existing literature in three distinct ways. First, it extended research on digital literacy beyond conventional university settings by examining first-semester students in a Malaysian polytechnic, which represented a technical and vocational education and training context. Although the study was conducted specifically within the Communicative English 1 course at Politeknik Nilai, the investigated issues possessed broader international relevance. Specifically, global technical and vocational education and training institutions faced similar challenges regarding how students evaluated sources, used information ethically, paraphrased content, synthesised evidence, selected digital tools, and navigated artificial intelligence-driven environments. Furthermore, as Warschauer (2009) observed, digital literacy studies required context-sensitive investigations that accounted for broader institutional and disciplinary demands across international educational landscapes.

Second, the study distinguished between digital literacy knowledge and digital literacy application. This analytical distinction made it possible to examine whether students' theoretical awareness of appropriate digital practices was reflected in their reported academic behaviour. Consequently, the research focused not only on whether students knew about digital literacy, but also on whether they reported applying it when completing English language tasks, thereby addressing the persistent gap between theoretical knowledge and practical execution highlighted by Eshet (2012).

Third, the study provided preliminary implications for curriculum design within digital and artificial intelligence-mediated learning environments. The findings informed the development of targeted classroom activities that explicitly taught learners how to evaluate online and artificial intelligence-generated information, paraphrase and synthesise sources, acknowledge digital assistance, and use language learning tools responsibly, echoing recommendations by Miao et al. (2023) regarding institutional guidance for emerging technologies. As a pilot investigation, the research did not claim to represent all Malaysian polytechnic students or global technical and vocational education and training cohorts. Instead, it provided an initial baseline profile that supported instrument refinement and guided future research involving larger samples, multiple institutions, and performance-based measures of digital and artificial intelligence literacy, directly answering calls by Tang and Chaw (2016) for rigorous foundational methodologies in digital literacy research.

1.3 Research Objectives

- To examine students perceived digital-literacy knowledge when completing Communicative English 1 tasks.
- To examine the relationship between students perceived digital-literacy knowledge and their reported application of selected digital practices.

Research Questions

- What is the level of students perceived digital-literacy knowledge when completing Communicative English 1 tasks?
- What is the relationship between students perceived digital-literacy knowledge and their reported application of selected digital practices?

1.4 Significance of the Study

This pilot study contributes to the understanding of how first-semester polytechnic students perceive and apply digital literacy in technical and vocational education and training (TVET) environments. Although the findings are limited to the scope and context of students enrolled in Communicative English 1 at Politeknik Nilai, they provide useful insights for several key stakeholders. First, the findings offer polytechnic educators a clearer understanding of students' self-reported digital awareness and practices, highlighting critical areas, such as independent academic writing, source evaluation, and ethical AI use, that require closer pedagogical guidance. Second, the results provide curriculum designers and academic developers with a foundational reference for considering how digital literacy, source synthesis, citation practices, and responsible tool integration can be more explicitly embedded into Communicative English curricula and professional development programmes. Furthermore, the study contributes empirical data that informs the refinement of questionnaires used to measure digital literacy, ensuring greater clarity and reliability for future large-scale and multi-institutional studies. Finally, the research establishes a preliminary methodological baseline for investigating digital literacy and technology-enhanced language learning across wider polytechnic settings, supporting future investigations that utilise larger samples and performance-based measures.

1.5 Limitations of the Study

This pilot investigation has several important limitations that shape the interpretation of its findings. First, the study is confined to first-semester students enrolled in Communicative English 1 (DUE10012) at Politeknik Nilai, utilizing a convenience sample of 51 respondents. Consequently, the results cannot be generalized to students in other semesters, programs, or polytechnic institutions across Malaysia. Second, as an exploratory study with a modest sample size, the findings offer preliminary insights rather than definitive conclusions about digital literacy patterns in Malaysian TVET contexts. Tang and Chaw (2016) alerted that small-scale digital literacy studies "provide valuable context-specific insights but require validation through larger, multi-institutional research" (p. 140).

Furthermore, the reliance on self-reported questionnaire data presents inherent limitations regarding participants' perceptions of their competencies rather than observed behaviours. Ng (2012) notes that self-reported digital literacy measures "may overestimate actual performance due to overconfidence among digital natives" (p. 1070). Finally, the study examines digital literacy within a specific course context at one institution, limiting its ability to situate findings within broader Malaysian polytechnic literature. This scarcity of prior research in the local TVET context, as acknowledged by Li et al. (2018), highlight the requirement for following studies to create upon these preliminary findings through expanded samples and diverse institutional settings.

2.0 Literature Review

2.1 Concept of Digital Literacy

Digital literacy evolved considerably from its early conceptualisation. Gilster (1997), as discussed by Pool (1997), initially described digital literacy as "the ability to identify and use information from a variety of digital sources." This early definition focused primarily on the ability to access and understand information through digital technologies. Subsequently, Bawden (2008) traced the digital literacy development from its technical origins to a broader construct that incorporated with information handling, communication, and ethical engagement in digital settings. Later scholars further emphasised on the multidimensional nature of digital literacy. Ng (2012) claimed that digital literacy required "can do skills for effective participation in the digital age" (p. 1066). The claim challenged the conjecture that students who grew up with digital technologies automatically possessed the skills needed for academic and critical engagement. Lankshear and Knobel (2015) distinguished the singular concept of "digital literacy" and the plural concept of "digital literacies." Which explains that digital practices differed according to educational, social, cultural, and disciplinary contexts.

In addition, digital literacy overlapped with related concepts such as digital competence, information literacy, media literacy, and digital citizenship. Ferrari (2013) conceptualised digital competence as the combination of knowledge, skills, and attitudes required for confident, critical, and responsible use of digital technologies. The Association of College and Research Libraries (ACRL, 2016) focused more specifically on information literacy and described it as a process through which learners identified information needs, located appropriate sources, evaluated information critically, and used information ethically. Ribble (2015), meanwhile, emphasised responsible and appropriate behaviour in digital environments. Recent reviews further confirmed that digital literacy was a broad and multidimensional concept. Gutiérrez Ángel et al. (2022) reviewed empirical studies of digital literacy in university settings and found that the concept included technical, cognitive, information-related, and critical dimensions. Their review also indicated that regular use of digital devices did not necessarily demonstrate strong academic digital literacy. Similarly, Tinmaz et al. (2022) conducted a systematic review and found that digital literacy research commonly addressed information evaluation, digital communication, content creation, technical competence, and responsible participation.

However, terminological overlap continued to create challenges for researchers. Scholars did not always use digital literacy, digital competence, information literacy, and media literacy in the same way. Tang and Chaw (2016) observed that this lack of conceptual agreement contributed to "inconsistent measurement approaches and mixed empirical findings across studies". Therefore, the present study used digital literacy as an umbrella concept referring

to students' knowledge and application of appropriate practices when they located, evaluated, transformed, and used online information for academic tasks.

2.2 Digital Literacy in Higher Education

Digital literacy became increasingly important in tertiary education as blended learning, online assessment, digital communication, and electronic learning platforms became more common. Tang and Chaw (2016) emphasised its role in enabling students to “navigate complex digital learning environments effectively”. This ability was particularly relevant to polytechnic students, who were expected to develop both technical knowledge and communication skills for future professional environments. First, students needed to use digital platforms to locate course materials, communicate with lecturers, submit assignments, and collaborate with peers. Second, they needed to evaluate the reliability and relevance of information obtained from online sources. Third, they needed to apply information appropriately in academic tasks while observing ethical standards such as citation, paraphrasing, and respect for intellectual property. Consequently, digital literacy involved more than the technical operation of devices. Alakrash and Abdul Razak (2021) examined digital technology and digital literacy in technology-based language learning. Their study indicated that digital literacy supported students' engagement with technology-enhanced language learning environments. However, students still required guidance to select suitable tools, evaluate online information, and use digital resources purposefully. Eshet Alkalai and Chajut (2009) similarly reported that students able to develop basic digital operations while remaining less competent in higher-order skills such as information verification and critical evaluation.

The assumption that frequent technology use automatically produced digital competence was therefore problematic. Ng (2012) found that familiarity with digital devices did not guarantee academic skills such as source evaluation, ethical content creation, or critical information management. Selwyn and Facer (2013) further argued that digital inequality extended beyond access to technology and included differences in competence, confidence, institutional support, and critical engagement. Starkey (2016) also emphasised the need for an equitable curriculum that prepared students to participate meaningfully in a digital society. Recent research on artificial intelligence extended these concerns. Laupichler et al. (2022) explained that AI literacy involved understanding AI systems, using AI tools, evaluating their outputs, and considering their social and ethical implications. Hornberger et al. (2023) developed and validated an AI literacy test for university students, demonstrating the importance of assessing students' actual understanding rather than having an impression that frequent technology use represented as AI literacy. Almatrafi et al. (2024) later reviewed AI literacy research and identified six related constructs: recognising AI, knowing and understanding AI, using and applying AI, evaluating AI, creating with AI, and navigating AI ethically.

The measurement of AI literacy also remained challenging. Lintner (2024) reviewed AI literacy scales and identified considerable variation in the dimensions assessed by existing instruments. Their findings suggested that researchers needed to distinguish between students' perceived confidence, conceptual knowledge, practical skill, critical evaluation, and ethical judgement. This concern was relevant to the present pilot study because the questionnaire measured students' self-reported digital literacy rather than their performance on an objective information evaluation or AI literacy task.

2.3 Digital Literacy and Language Learning

Digital tools provided important opportunities for language learning when students used them with appropriate guidance and critical awareness. Godwin Jones (2015) identified contributing, creating, and curating content as important digital literacies for language learners. These practices required students not only to consume online materials but also to participate in digital communication, produce original content, and evaluate information before sharing or incorporating it into academic work. Godwin Jones (2016) reviewed the development of technology in language learning and showed that digital tools had expanded opportunities for interaction, independent learning, and access to authentic language resources. Godwin Jones (2018) further demonstrated on how mobile technologies supported “language skills and cultural understanding” through purposeful digital practices (p. 5). These tools included

online dictionaries, grammar applications, translation systems, pronunciation platforms, learning management systems, podcasts, videos, and collaborative writing applications.

In ESL contexts, Li et al. (2018) reviewed the role of digital technologies in supporting ESL writing among Malaysian TESL pre-service teachers. Their review indicated that digital technologies could support writing development, access to language resources, drafting, revision, and collaboration. However, effective digital writing required students to evaluate sources critically and produce original synthesis rather than reproduce online content. Pegrum (2016) therefore argued that language learners needed “multiliteracies” to navigate digital content effectively. These multiliteracies included linguistic, visual, information, technological, and social competencies. The emergence of generative AI further complicated digital literacy in language learning. Students could use AI tools to generate ideas, correct grammar, translate passages, summarise information, and provide feedback on writing. These functions could support language development, but they could also encourage dependence, inaccurate information use, plagiarism, and reduced student involvement in the writing process. Miao et al. (2023) recommended that educational institutions guide the responsible use of generative AI in education and research.

Castillo Martínez et al. (2024) reviewed artificial intelligence in higher education and identified both pedagogical benefits and concerns relating to ethics, privacy, accuracy, academic integrity, and implementation. In EFL writing, Almusharraf et al. (2025) examined students’ perceptions of generative AI and found that students viewed AI positively in relation to writing strategies, self-efficacy, satisfaction, and intention to use the tools. Nevertheless, positive perceptions did not necessarily indicate that students could evaluate AI-generated content or use it ethically.

Liu et al. (2025) conducted a systematic review of generative AI tools and large language models in EFL education. Their review indicated that these tools could support critical thinking when teachers provided structured guidance, and students were encouraged to question and evaluate AI-generated responses. However, uncritical reliance on AI could reduce independent engagement and weaken students’ responsibility for their own learning. These findings demonstrated that technology-enhanced language learning required both access to digital tools and the literacy needed to use them critically.

2.4 Recent Developments in Technology Enhanced Language Learning

Recent research showed that technology enhanced language learning had expanded beyond the use of online dictionaries, learning management systems, and mobile applications. Digital platforms and generative artificial intelligence increasingly supported brainstorming, drafting, translation, feedback, revision, vocabulary development, and independent language practice. However, these developments also created new demands for digital literacy because students needed to evaluate the accuracy, relevance, authorship, and ethical status of digital and AI generated content. Xiao et al. (2025) examined that ChatGPT supported writing instruction in English as a foreign language education and reported that generative AI could support writing development through idea generation, drafting, feedback, and revision. Nevertheless, their review also highlighted concerns related to academic integrity, learner dependence, teacher preparation, and the need for appropriate pedagogical guidance. These findings were relevant to the present study because students’ ability to use digital tools effectively depended not only on access but also on their ability to evaluate information and transform it into original academic communication.

Equally, Mali (2025) reviewed the use of ChatGPT in EFL and ESL writing classrooms. The review demonstrated that students used ChatGPT for both academic and non-academic writing tasks, but it also identified concerns about plagiarism and unethical use. The study recommended institutional policies, process-based writing instruction, innovative assessment design, and ethical co regulation between teachers and students. This evidence supported the need to teach students how to use digital and AI tools transparently while retaining responsibility for the content and language of their academic work. Feng et al. (2025) reviewed research on artificial intelligence and second language writing between 2014 and 2024. Their review demonstrated that AI tools could support feedback, revision, and language development, but it also showed that questions remained concerning learner autonomy, overreliance, writing quality, and academic integrity. In a related systematic review, Limiya and Kumar (2025)

examined 55 studies on generative AI and English language development. Their findings indicated that AI tools supported listening, speaking, reading, and writing development, although effective implementation depended on teacher guidance, appropriate task design, and students' ability to use technology critically.

Recent research also emphasised the role of critical thinking in AI supported language learning. Liu et al. (2025) reviewed the influence of generative AI tools and large language models on critical thinking in EFL education. Their findings suggested that generative AI could support critical engagement when students were required to question, compare, verify, and revise AI generated responses. However, uncritical dependence on AI could reduce independent thinking and weaken students' responsibility for their own learning. Therefore, the educational value of generative AI depended largely on how it was integrated into teaching and assessment. The recent literature also identified institutional and pedagogical barriers to technology enhanced language learning. Sudirman (2025) reviewed digital literacy integration in EFL teaching and reported that digital tools could improve student engagement and language learning, but inadequate infrastructure, unequal access, and limited teacher preparation remained significant challenges. These findings were relevant to Malaysian polytechnic education because students' digital literacy development could be influenced by institutional resources, course expectations, and the extent to which lecturers explicitly taught digital and AI related practices.

Recent research suggested that technology enhanced language learning required more than the provision of digital tools. Students needed to develop information evaluation, source transformation, ethical awareness, AI related judgement, and responsible communication. These requirements were consistent with Eshet's information, branching, reproduction, and socio emotional literacies. They also supported the present study's focus on the distinction between digital literacy knowledge and digital literacy application.

2.5 Methodological Issues in Digital Literacy Research

Methodological differences continued to affect how digital literacy was measured and interpreted. Self-reported questionnaires remained common because they were relatively easy to administer to large groups. However, self-reported confidence did not always correspond with actual performance. Recent AI literacy research raised a similar concern. Hornberger et al. (2023) demonstrated the value of objective AI literacy testing, while Lintner (2024) reported that many AI literacy instruments relied on self-reported perceptions. These findings suggested that students' reported confidence should not be interpreted as direct evidence of actual competence. Instead, self-report results should be understood as indicators of students' perceived knowledge, perceived preparedness, and reported behaviour.

Measurement was also complicated by the multidimensional nature of digital literacy. Some questionnaires combined technical skills, information evaluation, ethical behaviour, communication, and content creation into a single score. Other instruments focused only on computer use or internet access. Gutiérrez Ángel et al. (2022) found that digital literacy studies differed in their definitions, instruments, populations, and measurement procedures. Tinmaz et al. (2022) similarly reported that inconsistent conceptual and methodological approaches made it difficult to compare findings across studies. Context specificity created another methodological issue. General university studies did not always capture the digital literacy demands of programmes or disciplines. Alakrash and Abdul Razak (2021) emphasised that technology-focused programmes required assessments that reflected vocational and academic tasks rather than generic computer skills. In a Communicative English course, students needed to search for online information, evaluate sources, paraphrase content, synthesise ideas, and communicate information in English. Therefore, a context-sensitive approach was necessary.

Because the present study was a pilot investigation with a small convenience sample, it focused on preliminary item-level patterns and students' self-reported responses. The findings were not intended to validate a general digital literacy scale or represent all Malaysian polytechnic students. Instead, the results were intended to identify areas that required further investigation and to inform refinement of the questionnaire for future research.

2.6 Knowledge and Application Gap

A persistent issue in digital literacy research concerned the difference between knowing appropriate digital practices and applying them in academic situations. Eshet (2012) conceptualised digital literacy as a combination of cognitive understanding and practical execution across six domains. From this perspective, students needed to understand why information should be evaluated and then apply source evaluation when completing an actual assignment. Ng (2012) reported that students' familiarity with digital technologies did not automatically lead to effective academic practice. Park (2019) identified several barriers to application, including information overload, difficulty judging source credibility, limited confidence, and weak synthesis skills. These barriers suggested that students could understand digital literacy principles but still have trouble applying them consistently. The distinction between knowledge and application could also be interpreted through Bloom's revised taxonomy. Anderson and Krathwohl (2001) placed knowledge and understanding among the foundational cognitive processes, while application and analysis required learners to use knowledge in particular situations. In the context of digital literacy, students might know that plagiarism was unacceptable but still struggle to paraphrase a source. They might know that sources needed to be evaluated but still select information based on convenience or search engine ranking. They might also know that AI-generated responses could contain errors but still accept those responses without verification.

This gap was particularly important in language learning. Li et al. (2018) showed that digital technologies could support ESL writing, but students still needed to evaluate online information and develop original written responses. Godwin Jones (2015) similarly suggested that language learners needed to move beyond consuming digital content and participate in creating and curating information. Therefore, the ability to use a search engine or language application did not necessarily indicate that students could transform online information into original academic writing. Recent research on AI literacy supported this distinction. Almatrafi et al. (2024) separated knowledge and understanding from using, applying, evaluating, creating, and navigating AI ethically. Liu et al. (2025) also showed that generative AI could support EFL learners' critical thinking only when students were guided to question, evaluate, and revise AI-generated responses. These findings reinforced the importance of examining both awareness and reported practice in technology-enhanced language learning.

2.7 Research Gap

Despite the growth of digital literacy research, several gaps remained relevant to Malaysian polytechnic education. First, many existing studies examined general university populations rather than TVET students preparing for technical and professional careers. Although Gutiérrez Ángel et al. (2022) reviewed digital literacy research in university settings, their findings also demonstrated the diversity of populations, measures, and contexts represented in the literature. Second, course-specific studies remained limited, particularly studies focusing on foundational English communication courses in polytechnic institutions. Communicative English courses required students to combine language ability with information evaluation, digital navigation, paraphrasing, synthesis, and ethical source use. These demands differed from the general technology use examined in many digital literacy studies.

Third, relatively few studies distinguished between students' knowledge of digital literacy and their reported application of digital literacy in academic tasks. Recent AI literacy research showed that knowledge, use, evaluation, creation, and ethical navigation represented related but distinct dimensions (Almatrafi et al., 2024). This distinction was also relevant to digital literacy in ESL and TELL contexts, where students could demonstrate confidence in using digital tools while still experiencing difficulty producing original and well-supported academic work.

Fourth, the growth of generative AI created a new need to reconsider digital literacy instruction. Recent studies indicated that students valued generative AI for writing support, but they still needed guidance to evaluate its outputs, avoid overreliance, and preserve academic integrity (Almusharraf et al., 2025; Liu et al., 2025). This issue was relevant to polytechnic students because their academic and future workplace communication increasingly involved digital and AI-mediated environments.

This pilot study addressed these gaps by examining first-semester students' self-reported digital literacy knowledge and application in a Malaysian polytechnic Communicative English course. The study contributed preliminary evidence from a TVET setting and provided a context-sensitive basis for discussing digital information evaluation, ethical source use, reproduction literacy, and emerging AI literacy. The findings were intended to inform curriculum development and targeted digital literacy instruction in Malaysian polytechnic institutions, while also guiding future research involving larger samples, multiple institutions, expert-validated instruments, and performance-based measures.

3.0 Methodology

3.1 Theoretical Framework

This study was guided by Eshet's Digital Literacy Framework (2004, 2012). The framework conceptualised digital literacy as a multidimensional competence involving cognitive, technical, evaluative, and social abilities. It included photo-visual literacy, reproduction literacy, branching literacy, information literacy, socio-emotional literacy, and real-time literacy. Photo-visual literacy referred to the ability to interpret visual signs, icons, images, and multimedia content. Reproduction literacy referred to the ability to transform existing digital materials into new meaning through paraphrasing, summarising, and creating original content. Branching literacy referred to the ability to navigate non-linear digital environments and move between connected sources. Information literacy involved locating, evaluating, and validating digital information. Socio-emotional literacy involved ethical, safe, and responsible participation in digital environments. Real-time literacy involved managing multiple digital tasks and responding appropriately in rapidly changing digital contexts (Eshet, 2004, 2012). The framework was relevant to this study because Communicative English 1 tasks involved activities such as searching for information, evaluating online sources, using digital language tools, transforming information into original English-language outputs, and communicating responsibly in digital environments. However, the questionnaire was not designed to measure all six literacy domains as separate constructs.

Instead, the study focused on two analytical categories: perceived digital-literacy knowledge and reported application of selected digital practices. Perceived digital-literacy knowledge referred to students' self-reported awareness of practices such as online safety, source checking, plagiarism avoidance, ethical information use, and familiarity with digital tools. Reported application referred to students' self-reported use of selected digital practices when completing Communicative English 1 tasks, including searching for information, using keywords, conducting background research, and using digital language tools. These categories were treated as exploratory indicators rather than fully validated constructs. The questionnaire therefore provided a preliminary assessment of selected digital practices relevant to the course rather than a comprehensive measurement of Eshet's six literacy domains. The study did not objectively assess students' ability to evaluate sources, paraphrase and synthesise information, cite evidence, or evaluate artificial-intelligence-generated content.

3.2 Research Design

This pilot study adopted a quantitative, cross-sectional, descriptive correlational design. Data were collected at one point in time, and no intervention, experimental treatment, or manipulation of variables was introduced. The descriptive component was used to examine students' responses to the digital literacy items, while the correlational component was intended to explore the possible association between digital literacy knowledge and application. The study was exploratory rather than confirmatory. Its purpose was to provide preliminary evidence about students' reported digital literacy practices and to identify possible areas for future instrument development. This approach was appropriate because previous research showed that digital literacy was a multidimensional construct and that its measurement varied across educational contexts (Gutiérrez Ángel et al., 2022). In addition, research on AI literacy indicated that knowledge, practical use, evaluation, creation, and ethical navigation represented related but distinct abilities (Almatrafi et al., 2024; Knoth et al., 2024).

Due to the a small convenience sample used in this study and a preliminary questionnaire, the findings were interpreted cautiously. The study did not aim to establish causal relationships, validate a measurement scale, or generalise the findings to all Malaysian polytechnic students.

3.3 Participants and Sampling

The participants comprised 51 Semester 1 students enrolled in Communicative English 1 at Politeknik Nilai. Convenience sampling was used because the participants were accessible to the researchers and were directly involved in the course activities examined in the study. The participants were considered relevant to the pilot study because they were at the early stage of tertiary education and were required to complete academic tasks involving online information and digital tools. The sample was used to generate preliminary findings within a specific polytechnic context and to identify possible improvements to the questionnaire and research procedures. The sample was not intended to represent all Malaysian polytechnic students. Therefore, the findings were interpreted as context-specific and were not generalised beyond the participating cohort.

3.4 Instrument Development

Before administration, the questionnaire underwent a preliminary review to assess the clarity, relevance, and suitability of its items for first-semester polytechnic students enrolled in Communicative English 1. The items were developed with reference to the objectives of the study and selected concepts from Eshet's Digital Literacy Framework, particularly information literacy, reproduction literacy, branching literacy, and socio-emotional literacy. The wording was reviewed to ensure that the statements were understandable and relevant to the students' academic and digital-learning context. A preliminary language review was also conducted to identify unclear expressions, grammatical problems, and possible ambiguities. In addition, a small-scale pilot evaluation was carried out to identify difficulties that participants might encounter when interpreting or responding to the items. Feedback from these preliminary processes was used to improve the wording, clarity, and presentation of the questionnaire before the main data collection.

The questionnaire was divided into two sections. Items 1–10 addressed students' perceived knowledge of selected digital practices, including online safety, source checking, plagiarism awareness, and familiarity with digital tools. Items 11–20 addressed students' reported use of selected digital practices during academic tasks, including searching for information, using keywords, conducting background research, and using digital language tools. These two sections were treated as exploratory analytical categories rather than as fully validated constructs. As this study was designed as a pilot investigation, the questionnaire was used as a preliminary research instrument rather than as a validated measure of digital-literacy competence. Formal expert-based content validation, construct validation through factor analysis, and internal-consistency reliability testing were not conducted in the present study. Instead, the preliminary review, language review, and small-scale pilot evaluation were used to improve the clarity, relevance, and presentation of the items before the main data collection. Accordingly, the findings are interpreted as preliminary evidence of students' perceived digital-literacy knowledge and reported digital practices, rather than as definitive measures of their actual competence.

The instrument also relied on self-reported responses and did not objectively assess students' ability to evaluate sources, paraphrase or synthesise information, cite evidence, or verify artificial-intelligence-generated content. Some items may therefore require further refinement to strengthen their alignment with the study objectives. Before the questionnaire is used in a larger-scale study, the revised instrument should undergo expert review by specialists in digital literacy, language education, and educational technology, followed by further pilot testing to examine item performance and establish appropriate validity and reliability evidence. These considerations are acknowledged as limitations of the present pilot study.

One questionnaire item, Q3, stated, "I am aware that everything I search on the internet is being watched." Although the item was retained in its original form for the present analysis because the data had already been collected using the administered questionnaire, its wording may be considered broad and open to different interpretations. In particular, the terms "everything" and "being watched" do not specify the nature of the monitoring or the parties involved. Therefore, Q3 should be reviewed and refined in future versions of the questionnaire to more clearly represent a specific digital-literacy knowledge practice. For the present study, responses to Q3 are interpreted as students' self-reported awareness that internet activity may be monitored rather than as evidence of knowledge about a specific form of online monitoring or surveillance.

3.5 Research Procedure

Before data collection, permission to distribute the questionnaire was obtained from the relevant academic authority at Politeknik Nilai. The participants were then, provided with an information statement explaining the purpose of the study, the voluntary nature of participation, the expected completion time, and the procedures used to protect their responses. The questionnaire was distributed through Google Forms. Before accessing the questionnaire, participants were asked to indicate their informed consent. They were informed that participation was voluntary, that they could decline to participate, and that they could withdraw without academic penalty. Participants completed the questionnaire individually and were not required to provide their names or other directly identifying information. After the data collection period ended, the responses were downloaded from Google Forms and checked for completeness. Responses were then organised according to the two conceptual sections of the questionnaire. Items 1 to 10 represented digital literacy knowledge, while Items 11 to 20 represented digital literacy application. The cleaned data were subsequently prepared for descriptive and exploratory correlational analysis.

3.6 Ethical Considerations

Institutional permission was obtained before the questionnaire was administered to the participants. An information statement was provided at the beginning of the Google Forms questionnaire to explain the purpose of the study, the voluntary nature of participation, and how the information collected would be used. Participants provided informed consent before proceeding with the questionnaire and were informed that they could decline to participate or withdraw without academic penalty.

No names or other direct personal identifiers were collected. The responses were treated confidentially and were used only for the purposes of this study. The collected data were stored securely and were accessible only to the researchers. The study did not involve any intervention or experimental manipulation, and participation was not expected to cause more than minimal risk to participants. Formal institutional research ethics committee approval was not obtained for this pilot study. The study was conducted with institutional permission and in accordance with the applicable institutional requirements for student research. The authors declare that there is no conflict of interest associated with this study. This study received no external funding. The research was conducted using the authors' own resources.

3.7 Data Analysis

The collected data were analysed using descriptive and exploratory inferential procedures. Descriptive statistics, including means and standard deviations, were calculated for each questionnaire item to describe patterns in students' perceived digital-literacy knowledge and reported application of selected digital practices. Standard errors of the mean were calculated for statistical reporting where required but were not used to interpret students' digital-literacy levels.

For the first research objective, Items 1–10 were examined to describe students' perceived knowledge of selected digital practices. For the second objective, Items 11–20 were examined to describe the frequency with which students reported applying selected digital practices during academic tasks. Preliminary composite scores were also calculated for the two item groups to summarise the overall response patterns within the participating cohort. For the application items, responses were retained in their original coding, and the application composite was calculated using the ten application items (Q11–Q20). Q14 was a negatively worded item concerning the practice of copying and pasting content from relevant articles and was therefore interpreted separately when discussing the application-item results. The application composite was treated as a summary of the original responses across the ten items and was not interpreted as a uniformly positive or desirable measure of digital practice. The composite scores were treated as preliminary analytical summaries rather than validated measures of separate digital-literacy constructs due to the questionnaire had not undergone formal construct validation or internal-consistency reliability testing.

Before the correlation analysis was conducted, the distributions of the two composite scores were assessed using the Shapiro–Wilk test. The perceived-knowledge score produced a statistically significant result, $W = .930$, $p = .005$, while the reported-application score also produced a statistically significant result, $W = .922$, $p = .003$. Both results were significant at the .05 level, indicating departures from normality. Accordingly, Spearman's rank-order

correlation was selected to examine the direction and strength of the association between students' perceived digital-literacy knowledge and reported application of selected digital practices. This non-parametric procedure was considered appropriate for the exploratory analysis of the ordinal-scale questionnaire responses and the observed distributions of the composite scores.

The correlation was interpreted as an association between students' perceived digital-literacy knowledge and their reported application of selected digital practices. It was not interpreted as evidence of causation or as proof that students' perceived knowledge resulted in objectively demonstrated behaviour. Because both variables were obtained from the same self-report questionnaire and contained conceptually related content, the possibility that shared measurement method and item similarity contributed to the strength of the observed association was also recognised. Independent-samples t-tests and one-sample t-tests were not conducted because the study did not involve comparisons between independent groups, pre-test and post-test measurements, or hypothesis testing against a predetermined population value. The analysis therefore focused on item-level response patterns, preliminary composite summaries, and the exploratory relationship between students' perceived knowledge and reported application. This analytical approach was consistent with the pilot and descriptive-correlational nature of the study.

4.0 Data Analysis

Data were analysed using descriptive and correlational statistical techniques appropriate for exploratory pilot research (Tang & Chaw, 2016). To address the first research objective, descriptive statistics, including mean, standard deviation, and standard error, were computed to describe students' perceived digital-literacy knowledge and reported application of selected digital practices. Preliminary composite scores were calculated for the knowledge-related item group (Items 1–10) and the application-related item group (Items 11–20) to summarise the overall response patterns within the participating cohort. The responses for Items 11–20 were retained in their original coding, and the application composite was calculated using the ten application items (Q11–Q20), including Q14 without reverse-coding. Q14 was a negatively worded item concerning the reported practice of copying and pasting content from a relevant article and was therefore considered separately when interpreting the item-level findings. Consequently, the application composite was treated as a summary of the original responses across the ten items rather than as a uniformly positive or desirable measure of digital practice. Because the questionnaire had not undergone formal construct validation or internal-consistency reliability testing, the composite scores were treated as preliminary analytical summaries rather than validated measures of separate digital-literacy constructs.

For the second research objective, the relationship between the preliminary knowledge-related and application-related composite scores was examined using Spearman's rank-order correlation. Before the correlation analysis was conducted, the distributions of the two composite scores were assessed using the Shapiro–Wilk test. The perceived-knowledge score produced a statistically significant result, $W = .930$, $p = .005$, while the reported-application score also produced a statistically significant result, $W = .922$, $p = .003$. Both results were significant at the .05 level, indicating departures from normality. Accordingly, Spearman's rank-order correlation was selected to examine the direction and strength of the association between students' perceived digital-literacy knowledge and their reported application of selected digital practices. This non-parametric procedure was considered appropriate for the exploratory analysis of the questionnaire responses and the observed distributions of the composite scores.

The correlation was interpreted as an association between students' perceived digital-literacy knowledge and their reported application of selected digital practices. It was not interpreted as evidence of causation or as proof that students' perceived knowledge resulted in objectively demonstrated behaviour. Because both variables were obtained from the same self-report questionnaire and contained conceptually related content, the possibility that shared measurement method and item similarity contributed to the observed association was also recognised.

Independent-samples t-tests and one-sample t-tests were not conducted because the study did not involve comparisons between independent groups, pre-test and post-test measurements, or hypothesis testing against a predetermined population value. The analysis therefore focused on item-level response patterns, preliminary composite summaries, distributional characteristics, and the exploratory relationship between students' perceived

knowledge and reported application. This analytical approach was consistent with the pilot and descriptive-correlational nature of the study.

4.1 Results and Discussion

This section presents the preliminary findings of the pilot study. The results were interpreted in relation to Eshet's Digital Literacy Framework and relevant research on digital literacy, technology-enhanced language learning, and artificial intelligence. Because the questionnaire had not undergone formal construct validation or reliability analysis, the findings are presented primarily as students' self-reported perceptions and reported practices. They should not be interpreted as direct evidence of students' actual digital-literacy performance.

4.1.1 Digital Literacy Knowledge among Polytechnic Students

Item	Statement	Mean	SD	SEM
Q1	I am aware of the appropriate conduct to safeguard personal information online.	3.88	0.95	0.13
Q2	I can check multiple sources to verify online information.	3.84	0.92	0.13
Q3	I am aware that everything I search on the internet is being watched	3.90	0.94	0.13
Q4	I am aware that copying and pasting habit may cause serious consequences	3.80	0.94	0.13
Q5	I comprehend the definition of words given by my lecturer after browsing	3.86	0.92	0.13
Q6	I am familiar of the use of both hardware and software to complete my assessment	3.86	0.89	0.13
Q7	I am capable of performing written tasks without referring to any internet resources	3.47	1.06	0.15
Q8	I am proficient of utilising digital technologies for everyday tasks	3.75	0.93	0.13
Q9	I can bookmark or add a web page to my favourites	3.75	0.98	0.14
Q10	I am familiar and utilise Artificial Intelligence in helping me accomplish tasks	3.55	1.08	0.15
Overall Mean		3.77		

Table 1 : Descriptive Statistics for Digital Literacy Knowledge Items (Q1 to Q10)

Note. N = 51 for all items.

Table 1 presents the descriptive statistics for the ten knowledge-related items. The item means ranged from 3.47 to 3.90, while the overall mean was 3.77. The highest mean was recorded for Q3, concerning students' awareness that internet activity could be monitored ($M = 3.90$, $SD = 0.94$). This was followed by Q1, concerning the protection of personal information online ($M = 3.88$, $SD = 0.95$), and Q5 and Q6, both of which recorded a mean of 3.86. The lowest mean was recorded for Q7, concerning the ability to complete written tasks without referring to internet resources ($M = 3.47$, $SD = 1.06$).

Overall, the response pattern indicated that students perceived themselves as relatively aware of online safety, digital tools, and selected digital practices. The lower response for Q7 suggested that students reported less confidence in completing written tasks independently of internet resources. This finding may indicate that online information had become an important part of students' academic work; however, the questionnaire did not assess the quality or originality of their written work. Students also reported relatively high familiarity with the hardware and software used for assessments ($M = 3.86$, $SD = 0.89$). This result indicates perceived familiarity with commonly used digital tools, but it should not be interpreted as evidence of advanced digital-literacy competence. Similarly, Q10, concerning familiarity with artificial intelligence for completing tasks, recorded a moderate mean of 3.55 ($SD = 1.08$). This item measured perceived familiarity or use of AI; it did not assess students' conceptual understanding of AI, ability to verify AI-generated information, or ability to use AI ethically.

Overall, the knowledge-related responses provide a preliminary description of students' perceived awareness of selected digital practices. They also suggest possible areas for instruction, including independent academic writing, source evaluation, and responsible use of emerging AI tools.

4.1.2 Application of Digital Literacy Skills in Academic Tasks

Item	Statement	Mean	SD	SEM
Q11	I am familiar and use Translation software/programs/websites to accomplish my task	3.76	0.93	0.13
Q12	I tend to research thoroughly on the topic given by my lecturer by reading articles extensively before completing the assessment	3.73	0.94	0.13
Q13	I used relevant keywords on the topic given by my lecturer to look for related information of the written assessment	3.80	0.89	0.13
Q14	I often copy and paste the content of the relevant article to complete my written assessment	3.39	1.02	0.14
Q15	I usually use my own mobile data to browse for information in order to complete my assignment	3.98	1.01	0.14
Q16	I don't use polytechnic's own WIFI to search for information out of fear in violating my privacy	3.57	1.10	0.15
Q17	I am aware that a strong password is required when I sign up for websites that contain the information I need	3.88	1.05	0.15
Q18	I am familiar of using online thesaurus to help me complete my assignment	3.45	1.06	0.15
Q19	I would use the search engine to look up for extra information on the values of the given topics in the assignment	3.88	1.01	0.14
Q20	I would do a background research/reading before attempting to complete my assignment	3.88	0.99	0.14
Overall Mean		3.73		

Table 2 : Descriptive Statistics for Digital Literacy Application Items (Q11 to Q20)

Note. N = 51 for all items.

Table 2 presents the descriptive statistics for the ten application-related items. The item means ranged from 3.39 to 3.98, while the overall mean was 3.73. These results describe students' reported frequency of selected practices; they do not represent observations of students' behaviour during actual assignments.

The highest mean was recorded for Q15, concerning the use of personal mobile data to browse for information when completing assignments ($M = 3.98, SD = 1.01$). Relatively high means were also recorded for Q17, concerning strong passwords ($M = 3.88, SD = 1.05$), Q19, concerning the use of search engines to obtain additional information ($M = 3.88, SD = 1.01$), and Q20, concerning background research before completing an assignment ($M = 3.88, SD = 0.99$). Students also reported using relevant keywords when searching for information ($M = 3.80, SD = 0.89$). These responses indicate that students reported frequent use of digital connectivity, search engines, keywords, and background reading to support academic tasks. They primarily describe access and reported use, however, and do not establish whether the information retrieved was accurate, credible, or appropriately incorporated into students' assignments.

Q14 recorded the lowest mean among the application-related items ($M = 3.39, SD = 1.02$). Because the item concerned copying and pasting content from an article, a higher score would represent more frequent endorsement of a potentially problematic practice rather than stronger digital-literacy application. The item should therefore be interpreted separately and should not be treated as a positive indicator in the application score unless it has been reverse-coded and this procedure is explicitly justified.

The result for Q18, concerning the use of an online thesaurus, was also relatively low ($M = 3.45, SD = 1.06$). This may indicate that students reported using digital resources more frequently for information retrieval than for vocabulary development. Nevertheless, the item does not establish why students used or did not use the tool.

Taken together, the responses indicate frequent self-reported use of digital resources for searching and accessing information, alongside less consistent endorsement of selected practices related to independent writing, language refinement, and direct copying. These patterns support the need for explicit instruction in source evaluation, paraphrasing, synthesis, citation, academic integrity, and responsible use of digital and AI tools. They should not, however, be interpreted as direct evidence of students' demonstrated competence in these areas.

4.1.3 Relationship between Digital Literacy Knowledge and Application

For the second research objective, the relationship between students' perceived digital-literacy knowledge and their reported application of selected digital practices was examined using Spearman's rank-order correlation. The analysis used preliminary composite scores derived from Items 1–10 for perceived digital-literacy knowledge and Items 11–20 for reported application. The application composite was calculated from the original responses to Items 11–20, including Q14 without reverse-coding. Therefore, the composite represented the overall response pattern across the application items and was not interpreted as a uniformly positive or desirable measure of digital practice.

Construct	Mean	SD	1	2
1. Knowledge (Q1 to Q10)	3.77	0.78	1.000	
2. Application (Q11 to Q20)	3.73	0.81	.925**	1.000

Table 3 : Spearman's Rank-Order Correlation between Digital Literacy Knowledge and Application
($N = 51$)

Note. ** $p < .001$ (two-tailed).

The Shapiro–Wilk test indicated that the distributions of both composite scores departed significantly from normality. The perceived digital-literacy knowledge composite produced a statistically significant result, $W = .930, p = .005$, while the reported application composite also produced a statistically significant result, $W = .922, p = .003$. Accordingly, Spearman's rank-order correlation was used to examine the association between the two composite scores. The analysis indicated a strong positive association between students' perceived digital-literacy knowledge and their reported application of selected digital practices, $\rho = .925$. This result indicated that students with higher perceived digital-literacy knowledge scores also tended to report higher scores across the original application items. However, the finding should be interpreted as an association between two self-reported measures rather than as evidence that digital-literacy knowledge caused greater application or as evidence of objectively demonstrated digital behaviour.

Q14 was considered separately in the interpretation of the item-level application findings because it described the reported practice of copying and pasting content from a relevant article. As Q14 was retained in its original coding, the strong correlation should be understood in relation to the overall response pattern across Items 11–20 rather than as evidence that higher application scores consistently represented more desirable digital practices. In addition, because both composite scores were obtained from the same self-report questionnaire and contained conceptually related content, shared measurement method and item similarity may have contributed to the strength of the observed association.

4.2 Discussion and Implications

The findings indicate that first-semester students reported generally positive perceptions of selected digital-literacy practices. Their reported application varied across items, particularly those involving independent academic production, language-tool use, and direct copying. Because the data were obtained through self-report, these patterns

should be understood as indicators of perceived awareness and reported behaviour rather than as evidence of objectively demonstrated competence. The findings are therefore most useful for identifying possible instructional priorities. Communicative English lecturers may consider incorporating structured activities in which students compare sources, identify evidence and bias, paraphrase and synthesise information, cite sources, and evaluate AI-generated content. These recommendations are pedagogical implications rather than conclusions that students currently lack or possess specific skills, because those skills were not directly tested in the present study.

Furthermore, when comparing these findings with recent scholarship on second-language writing and emerging technologies, important parallels emerged. Recent systematic reviews and empirical investigations have demonstrated that learners frequently utilise generative artificial intelligence and digital language models to support writing tasks, while concerns remain regarding critical thinking, independent source evaluation, and responsible technology use (Almusharraf et al., 2025; Feng et al., 2025; Liu et al., 2025). For instance, Almusharraf et al. (2025) found that students expressed satisfaction and behavioural intention to use generative artificial intelligence tools in English as a foreign language writing, whereas Liu et al. (2025) cautioned that over-reliance on such technologies could hinder learners' independent critical thinking unless appropriate pedagogical scaffolding was provided. In this context, the present findings reinforce the importance of examining not only students' awareness of digital-literacy principles but also how they report applying those principles in academic contexts. Foundational digital knowledge alone should therefore not be assumed to guarantee competent and ethical application within AI-mediated learning environments (Hornberger et al., 2023; Knoth et al., 2024).

The relationship between perceived digital-literacy knowledge and reported application provides further insight into the findings. The Spearman's rank-order correlation showed a strong positive association between students' perceived digital-literacy knowledge and their reported application of digital practices, $\rho = .925$, $p < .001$, $N = 51$. This indicates that students who reported higher levels of perceived digital-literacy knowledge also tended to report more desirable and frequent application of the measured digital practices. In interpreting the Application score, higher composite scores generally represented more desirable reported digital practices, Q14 requires consideration because it described the reported practice of copying and pasting content from a relevant article. The item was retained in its original coding and was not reverse coded in the present analysis. Consequently, the application composite should be understood as a summary of the original responses across Items 11–20 rather than as a measure in which higher scores uniformly represent more desirable digital practices. The wording and treatment of Q14 also highlight the need for further refinement of the application items in future versions of the questionnaire, particularly where positively and negatively valenced digital practices are included within the same item group. However, the strength of the association should be interpreted cautiously. The correlation does not establish that digital-literacy knowledge causes students to apply these practices because both dimensions were measured through the same self-report questionnaire and consisted of conceptually related aspects of digital literacy. Shared measurement method and item similarity may therefore have contributed to the strength of the observed association. The finding should consequently be interpreted as evidence of a strong relationship between students' perceived knowledge and their reported application, rather than as evidence of a causal relationship or objectively demonstrated competence.

The strong association between knowledge and reported application also has implications for instructional practice. Rather than suggesting that students necessarily experience a substantial separation between what they know and what they do, the findings indicate that perceived knowledge and reported application were closely related within the present sample. Nevertheless, the self-reported nature of the measures means that students' perceptions of their knowledge and practices may not fully correspond to their actual performance. This highlights the value of providing students with opportunities to demonstrate digital-literacy practices through authentic academic tasks. As emphasised by Miao et al. (2023) and the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2023), educational institutions need to move beyond passive technology adoption and provide explicit instructional guidance. Within technical and vocational education and training contexts, language educators should not assume that students automatically know how to evaluate AI-generated outputs, avoid unacknowledged text recycling, or integrate digital tools responsibly (Laupichler et al., 2022). Instead, curriculum designers and English language instructors should intentionally provide structured opportunities for students to apply digital-literacy principles through activities

involving critical source evaluation, ethical artificial-intelligence use, paraphrasing and synthesis, appropriate citation, and reflective academic writing.

Overall, the findings suggest that digital literacy should be approached as both a knowledge-oriented and application-oriented dimension of students' academic practices. The strong positive association observed in this study provides evidence that students' perceived knowledge and reported application were closely connected, while the limitations of self-report measurement caution against interpreting this relationship as proof of actual competence or causation. Future instructional efforts may therefore benefit from combining explicit digital-literacy instruction with performance-based activities that allow students to demonstrate how they evaluate, use, and communicate digital information in authentic academic contexts.

5.0 Conclusion

This pilot study examined first-semester students' perceived digital-literacy knowledge and their reported application of selected digital practices within a Communicative English 1 course at Politeknik Nilai. The findings indicated generally positive self-reported perceptions of digital awareness alongside varied frequencies in application-related practices, particularly concerning independent academic production and language-tool usage. Furthermore, Spearman's rank-order correlation revealed a strong positive association between students' perceived knowledge and reported application ($p = .925$, $p < .001$, $N = 51$), though this relationship must be interpreted cautiously due to the self-reported nature of the instrument and shared measurement characteristics.

While limited by a small convenience sample and the absence of formal construct validation, this pilot investigation provides preliminary empirical insights into students' digital habits in a Malaysian TVET context. The findings suggest that educational institutions should not assume digital natives inherently possess critical information evaluation skills or ethical AI navigation abilities. Consequently, curriculum designers and language instructors are encouraged to integrate explicit instructional scaffolding that guides students from passive technology use to critical source evaluation, paraphrasing, synthesis, and responsible academic communication. Future research should build upon these findings through multi-institutional studies, larger sample sizes, and objective performance assessments to better support digital literacy development in higher education.

REFERENCES

- ACRL. (2016). *Framework for information literacy for higher education*. Association of College & Research Libraries.
- Alakrash, H. M., & Abdul Razak, N. (2021). Technology-based language learning: Investigation of digital technology and digital literacy. *Sustainability*, 13(21), 12304. <https://doi.org/10.3390/su132112304>
- Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education: Open*, 6, 100173. <https://doi.org/10.1016/j.caeo.2024.100173>
- Almusharraf, A., Bailey, D., Almusharraf, N., & Alotaibi, T. (2025). Students' perceptions of generative AI in EFL writing: Strategies, self-efficacy, satisfaction and behavioural intention. *Australasian Journal of Educational Technology*, 41(5), 18-36.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

- Bawden, D. (2008). Origins and concepts of digital literacy. In L. M. C. Martin & P. A. Combes (Eds.), *Digital literacies: Concepts, policies and practices* (pp. 17–32). Peter Lang.
- Castillo-Martínez, I. M., Flores-Bueno, D., Gómez-Puente, S. M., & Vite-León, V. O. (2024). AI in higher education: A systematic literature review. *Frontiers in Education*, 9, 1391485. <https://doi.org/10.3389/feduc.2024.1391485>
- Eshet, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, 13(1), 93–106.
- Eshet, Y. (2012). Thinking in the digital era: A revised model for digital literacy. *Issues in Informing Science and Information Technology*, 9, 267–276. <https://doi.org/10.28945/1621>
- Eshet-Alkalai, Y., & Chajut, E. (2009). Changes over time in digital literacy. *CyberPsychology & Behavior*, 12(6), 713–715.
- Feng, H., Li, K., & Zhang, L. J. (2025). What does AI bring to second language writing? A systematic review (2014–2024). *Language Learning & Technology*, 29(1), 1–27. <https://doi.org/10.64152/10125/73629>
- Ferrari, A. (2013). *DIGCOMP: A framework for developing and understanding digital competence in Europe*. European Commission Joint Research Centre.
- Gilster, P. (1997). *Digital literacy*. Wiley.
- Godwin-Jones, R. (2015). Contributing, creating, curating: Digital literacies for language learners. *Language Learning & Technology*, 19(3), 8–20. <https://doi.org/10.64152/10125/44427>
- Godwin-Jones, R. (2016). Looking back and ahead: 20 years of technologies for language learning. *Language Learning & Technology*, 20(2), 5–12. <https://doi.org/10.64152/10125/44457>
- Godwin-Jones, R. (2018). Using mobile technology to develop language skills and cultural understanding. *Language Learning & Technology*, 22(3), 3–17.
- Gutiérrez-Ángel, N., Sánchez-García, J.-N., Mercader-Rubio, I., García-Martín, J., & Brito-Costa, S. (2022). Digital literacy in the university setting: A literature review of empirical studies between 2010 and 2021. *Frontiers in Psychology*, 13, 896800. <https://doi.org/10.3389/fpsyg.2022.896800>
- Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about artificial intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence*, 5, 100165. <https://doi.org/10.1016/j.caeai.2023.100165>
- Knoth, N., Decker, M., Laupichler, M. C., Pinski, M., Buchholtz, N., Bata, K., & Schultz, B. (2024). Developing a holistic AI literacy assessment matrix–Bridging generic, domain-specific, and ethical competencies. *Computers and Education Open*, 6, 100177.
- Lankshear, C., & Knobel, M. (Eds.). (2008). *Digital literacies: Concepts, policies and practices* (Vol. 30). Peter Lang.
- Laupichler, M. C., Aster, A., Perschewski, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3, 100101. <https://doi.org/10.1016/j.caeai.2022.100101>

- Li, K. L., Razali, A. B., Noordin, N., & Abd Samad, A. (2018). The role of digital technologies in facilitating the learning of ESL writing among TESL pre-service teachers in Malaysia: A review of the literature. *Journal of Asia TEFL*, 15(4), 1139–1156. <https://doi.org/10.18823/asiatefl.2018.15.4.10.1139>
- Limiya, A. V. G., & Kumar, P. K. N. S. (2025). Generative artificial intelligence in enhancing English language skills: A systematic review. *Studies in Media and Communication*, 13(4), 256–267. <https://doi.org/10.11114/smc.v13i4.7763>
- Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning*, 9, 50.
- Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence tools and large language models influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments*. <https://doi.org/10.1186/s40561-025-00406-0>
- Mali, Y. C. G. (2025). Exploring the use of ChatGPT in EFL/ESL writing classrooms: A systematic literature review. *Journal of Language and Education*, 11(2), 137–156. <https://doi.org/10.17323/jle.2025.21793>
- Miao, F., Holmes, W., Huang, R., & Zhang, H. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education*, 59(3), 1065–1078. <https://doi.org/10.1016/j.compedu.2012.04.016>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Park, S. (2019). Digital literacy and information evaluation skills among university students. *Information Research*, 24(1), 1–13.
- Pegrum, M. (2016). Languages and literacies for digital lives. In B. Tomlinson (Ed.), *Technology-enhanced language learning for specialized domains* (pp. 9–22). Routledge.
- Pool, C. R. (1997). A new digital literacy: A conversation with Paul Gilster. *Educational Leadership*, 55(3), 6–11.
- Ribble, M. (2015). *Digital citizenship in schools: Nine elements all students should know* (3rd ed.). International Society for Technology in Education.
- Selwyn, N., & Facer, K. (2013). Beyond digital divide: Toward an agenda for change. In L. A. Tomei (Ed.), *Digital literacy: Concepts, methodologies, tools, and applications* (pp. 1678–1696). IGI Global.
- Sudirman, A. (2025). Integrating digital literacy in EFL teaching and learning contexts: A systematic literature review. *Jejak Digital: Jurnal Ilmiah Multidisiplin*, 1(4), 1401–1417. <https://doi.org/10.63822/b2sdt853>
- Starkey, L. (2016). An equitable curriculum for a digital age. *Curriculum Matters*, 12(12), 29–45.
- Tang, C. M., & Chaw, L. Y. (2016). Digital Literacy: A Prerequisite for Effective Learning in a Blended Learning Environment?. *Electronic Journal of E-learning*, 14(1), 54–65.

Tinmaz, H., Lee, Y.-T., Fanea-Ivanovici, M., & Baber, H. (2022). A systematic review on digital literacy. *Smart Learning Environments*, 9, 21.

UNESCO. (2023). *Global education monitoring report 2023: Technology in education: A tool on whose terms?* UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000385723>

Warschauer, M. (2009). Digital literacy studies: Progress and prospects. *The future of literacy studies*, 123-140.

Xiao, F., Zhu, S., & Xin, W. (2026). Exploring the landscape of generative AI (ChatGPT)-powered writing instruction in English as a foreign language education: A scoping review. *ECNU Review of Education*, 9(1), 20965311241310881.

Integrating Artificial Intelligence into CEFR B2 Speaking Interviews in Higher Education: An Hourglass Framework of Human–AI Collaborative Evaluation

Nurul Imtiaz Abd Gani*, School of Languages, Civilisation and Philosophy, Universiti Utara Malaysia, Sintok, Kedah, Malaysia, nurulimtiaz@uum.edu.my

Muhammad Fadhli Mohd Diah, School of Languages, Civilisation and Philosophy, Universiti Utara Malaysia, Sintok, Kedah, Malaysia, md.fadhli@uum.edu.my

Mohd Zulhusni Ab Raoff, School of Languages, Civilisation and Philosophy, Universiti Utara Malaysia, Sintok, Kedah, Malaysia, zulhusni@uum.edu.my

...

*** Corresponding Author**

Abstract: The expanding recognition of Artificial Intelligence (AI) in higher education has opened new opportunities to improve language evaluation, particularly in speaking proficiency. Although the potential of generative AI for automated scoring and feedback has been demonstrated, CEFR B2 speaking interview assessment remains highly challenging due to its reliance on subjective human judgement, multidimensional criteria, growing learner cohorts, and the substantial time required for scoring and feedback. Limited research has explored how to integrate these technologies systematically within a coherent educational structure. To address this gap, this conceptual paper proposes the Hourglass Framework of Human–AI Collaborative Evaluation as a structured model for English proficiency courses in higher education. This framework conceptualises AI-supported speaking evaluation as a progressive, three-stage process that funnels macro-level educational challenges down to micro-level collaborative evaluations, which then expand back out into macro-actionable learning analytics. Grounded in a robust thematic synthesis of 54 peer-reviewed sources, the study proposes the Hourglass Framework as a foundation for enhancing the evaluation of CEFR-aligned speaking assessments while promoting educational integrity, personalised learner development, and future innovations in evaluation practices.

Keywords: *Artificial Intelligence, CEFR B2 Speaking Interview Assessment, English Proficiency Learners, Hourglass Framework.*

Introduction

In modern higher education, speaking interview assessment represents one of the most complex domains of language evaluation. Traditional face-to-face speaking interviews require learners to demonstrate multiple dimensions of communicative competence simultaneously. These include fluency and coherence, interactional competence, vocabulary range, and grammatical accuracy. Evaluating these multidimensional facets in real-time requires considerable time, professional expertise, and consistent judgement from instructors. As learner enrolments expand in higher education institutions, language faculties face a compounding logistical challenge in maintaining quality, reliability, and standardisation of speaking assessments amid high-volume grading workloads. While the global adoption of the Common European Framework of Reference for Languages (CEFR) has standardised proficiency benchmarks, university language programmes require effective evaluation approaches to reconcile rigorous assessment criteria with the practical limitations of the institutional context. Focusing specifically on the B2 ‘independent’ user threshold, reliably evaluating spontaneous, multidimensional spoken interaction at scale introduces enormous logistical and evaluative burdens on language instructors and faculties.

To address these issues, recent advances in Artificial Intelligence (AI) have offered feasible means of assistance and mitigated these challenges. Technological advancement has transformed the field of higher education by influencing teaching, learning, assessment and evaluation approaches. It has emerged as one of the most popular innovations by creating new options to enhance educational practices and supporting data-driven decision-making (Holmes et al., 2019). AI technologies such as ChatGPT, Gemini, and Claude have received much attention for their ability to generate natural language, human-like responses, and personalised feedback, as well as to perform other educational tasks. As a result, AI is being explored in higher education to improve assessment efficiency, consistency, and learner support (Kohnke et al., 2023), including in speaking interview assessment. Despite these advancements, the use of AI in speaking interview assessment remains a complex issue due to concerns related to reliability, transparency, fairness, ethics, and pedagogy. Several researchers have studied AI-assisted scoring, feedback generation, chatbot-supported learning, and conversational AI applications. However, these studies have focused only on individual AI functions, and much of this literature is about technological performance rather than educational integration (Long and Magerko, 2020). Moreover, there is a lack of a comprehensive conceptual framework to explain how AI can be systematically integrated into CEFR-aligned speaking interview assessment without compromising the evaluation's validity, reliability, and integrity.

This gap is evident in the CEFR B2 speaking interview, where evaluation involves complex human judgement, multidimensional communicative ability, interactional skills, and contextual explanation. These criteria extend beyond simple linguistic analysis. Therefore, instead of explaining what AI technologies can do, higher education institutions need a framework for integrating AI into the evaluation of CEFR B2 speaking interviews to enhance reliability, feedback, analytics, and outcomes. To bridge this gap, this paper proposes the Hourglass Framework of Human-AI Collaborative Evaluation as a conceptual framework for integrating AI into CEFR B2 speaking interview assessments among English learners in higher education. In this conceptual framework, AI is integrated into a developmental process that connects multiple educational challenges, Human-AI collaborative evaluation practices, and broader educational outcomes. This concept treats AI as a partner for instructors rather than a replacement. To establish the theoretical foundation for the proposed Hourglass Framework, the following sections review four interrelated areas: CEFR B2 communicative competence and speaking assessment, Human-AI collaborative evaluation, language assessment quality, and learning analytics. These four theoretical foundations explain what is assessed, how Human

–AI collaboration can support the evaluation process, how assessment quality is safeguarded, and how assessment evidence can be transformed into actionable insights to enhance learning and speaking development.

CEFR B2 Communicative Competence and Speaking Assessment in Higher Education

The Common European Framework of Reference (CEFR) stands as the baseline for language teaching, curriculum development, and competency assessment. The CEFR provides a standard reference framework for evaluating learners' language competences and ensures consistency with worldwide language standard measurement of learners' ability to speak, write, listen to, and read a language. The CEFR advocates an action-oriented approach that shifts the educational paradigm from a linear structural progression to a focus on learners' capacity to use language successfully. The CEFR consists of six levels, from A1 (beginning) to C2 (expert). However, this study focuses on the B2 level only. According to the CEFR, B2-level learners are independent users who can interact with a certain degree of fluency and spontaneity. They can present and justify viewpoints and participate effectively in discussions on common and a range of topics (Council of Europe, 2020). This CEFR communicative orientation draws on the communicative competence proposed by Canale and Swain (1980) and later expanded by Canale (1983). Communicative competence is conceptualised as a multidimensional construct which consists of grammatical, sociolinguistic, discourse and strategic competences. These dimensions collectively enable learners to communicate effectively and appropriately in an authentic context. These multidimensional perspectives make communicative competence an important conceptual foundation for understanding speaking proficiency in CEFR-aligned assessments. Therefore, the CEFR B2 interview assessment extends beyond evaluating isolated linguistic knowledge and focuses on learners' overall communicative ability. This includes the learners' ability to produce coherent discourse, interact appropriately in different social contexts, and strategise communication whenever needed.

Speaking performance in the context of CEFR-aligned assessment is measured using dimensions such as range, accuracy, fluency, interaction, coherence, and phonology. However, specific evaluation criteria may vary from one institution to another. To establish construct validity and ensure alignment with both global assessment benchmarks and national quality standards, the qualitative criteria used in the institutional interview assessments are mapped directly onto the official CEFR B2 illustrative scales. The speaking rubric analysed in this study is adapted from the validated blueprints of the National English Language Higher Education Assessment in Malaysia (ELHAM) E-Book, Volume 2: Speaking (Zuraidah Mohd Don et al., 2025), which serves as the local operationalisation of the CEFR. Rather than operating disconnected benchmarks, the local five-point rubric constructs comprising Fluency and Coherence, Interaction, Vocabulary, and Grammar are strategically designed to operationalise the broader CEFR B2 communicative 'can-do' descriptors into practically assessable classroom metrics (Council of Europe, 2020). These dimensions reflect the multidimensional nature of communicative competence and require an evaluation approach that assesses linguistic accuracy and communicative effectiveness. To illustrate this alignment, Table 1 provides a systematic mapping of the Universiti Utara Malaysia (UUM) English Language Enrichment II (ELE II) speaking interview criteria to the official CEFR B2 illustrative scales and their corresponding descriptors.

Table 1. Mapping and Construct Alignment of the Institutional Speaking Rubric with CEFR B2 Illustrative Descriptors

Rubric Criterion	Institutional Band 5 Descriptors	Mapped CEFR B2 Scale(s) and Descriptors (Council of Europe, 2020)
Fluency & Coherence	Maintains a generally smooth flow of speech and develops and connects ideas clearly and appropriately throughout the interaction. Hesitation rarely affects communication and is mainly used to organise ideas.	● Fluency: Can produce stretches of language with a uniform tempo; although they can be hesitant when searching for patterns and expressions, with few noticeably long pauses. ● Coherence: Can use a limited number of cohesive devices to link utterances into clear, coherent discourse.
Interaction	Responds appropriately and sustains natural interaction throughout the interview with minimal or no prompting.	● Interaction: Can take the initiative in an interview, expanding and developing ideas with little help or prompt from an interviewer.
Vocabulary	Uses vocabulary appropriately and with sufficient range to express and develop ideas clearly. Paraphrases effectively when needed to maintain communication.	● Vocabulary Range: Can use a sufficient range of language to give clear descriptions and express viewpoints on most general topics, with little noticeable searching for words.
Grammar	Uses grammatical structures appropriately to communicate ideas clearly with generally good control. Occasional errors rarely affect communication.	● Grammatical Accuracy: Can maintain a relatively high degree of grammatical control. Does not make errors that cause misunderstanding and can correct most of the mistakes.

While the official CEFR framework includes a distinct illustrative scale for Phonological Control (specifically focusing on pronunciation, intonation patterns, and accent intelligibility), this speaking interview rubric deliberately subsumes phonological elements within the Fluency and Coherence criteria rather than treating ‘Pronunciation’ as an isolated, standalone scoring category. By incorporating pronunciation as an integral component of a learner's ability to maintain a smooth flow of speech or sustain natural interaction, the rubric streamlines the assessor's focus on the overall communicative outcome of the utterance rather than on formalistic micro-linguistic details.

Human–AI Collaborative Evaluation in Speaking Assessment

Contemporary discussion on AI in education increasingly advocates Human-AI collaborative evaluation rather than full automation. This perspective holds that human assessors and AI systems have distinct strengths that complement one another and should work collaboratively throughout the evaluation process (Baker, 2021, p.51). Human assessors contribute to contextual interpretation, pedagogical judgement, ethical decision-making, and final assessment decisions. AI, at the same time, contributes to the task of analysing performance, including scoring support, consistency, scalability, generating feedback and rapid analysis of learners’ performance. By combining all these complementary capabilities, Human-AI collaborative evaluation has the potential to improve evaluation quality while at the same time preserving educational integrity and professional judgement.

Language Assessment Quality in AI-Supported Evaluation

Artificial Intelligence (AI) has been widely used in educational evaluation due to its ability to analyse language, identify patterns of performance, and generate personalised feedback (Wollny et al., 2021; Kohnke et al., 2023). Early AI applications focused on automatic scoring systems for objective evaluation. Recently, advances in generative AI have expanded its potential to support language evaluation through performance analysis, automatic scoring, and diagnostic feedback (Kasneci et al., 2023; Tlili et al., 2023). Even though AI has the potential to support evaluation efficiency and provide timely feedback on learners' performance, it continues to require human interpretation and professional judgement. As AI becomes more integrated into language assessment, it is crucial to uphold the quality of evaluation. Bachman and Palmer (1996) stated that effective language assessments should be driven by the principles of reliability, validity, fairness, and usefulness to ensure that assessment decisions are consistent, accurate, and appropriate for their intended purposes. These principles remain relevant in AI-supported evaluation, where automated analysis should complement and not replace human judgement. Therefore, AI integration into speaking skills evaluation should follow a framework that preserves assessment integrity while addressing concerns about reliability, transparency, fairness, and ethical implementation. This approach would also enable AI to support formative and summative assessments without compromising the quality and credibility of evaluation practices.

Learning Analytics for Speaking Development

Learning Analytics has emerged as a prominent domain of educational research focusing on the measurement, collection, analysis, and interpretation of educational data to better understand and optimise learning processes (Siemens, 2013). Initially, learning analytics was primarily used to monitor learner engagement through behavioural indicators such as Learning Management System (LMS) access logs, participation records, and assessment outcomes. However, recent advances in AI have significantly expanded its capabilities, enabling more sophisticated analyses of learner performance, the generation of automated personalised diagnostic feedback, and evidence-based instructional decision-making (Ifenthaler & Yau, 2020). In higher education, learning analytics extends the role of assessment beyond mere grading by transforming assessment data into meaningful insights that support teaching and learning. Within language education specifically, learning analytics enables instructors to map and identify learners' strengths and weaknesses, monitor the development of speaking proficiency, provide personalised feedback, and implement targeted curricular enhancements. Consequently, integrating AI-supported evaluation with learning analytics offers a transformative opportunity to elevate speaking assessments from simple summative measures into dynamic, information-rich systems that promote continuous language development and raise the standard of evaluation in higher education.

Conceptual Integration

The proposed Hourglass Framework is grounded in four theoretical foundations: communicative competence, Human–AI collaborative evaluation, language assessment quality, and learning analytics. Although each foundation is well established in its respective literature, their systematic integration within a unified CEFR B2 speaking assessment framework remains insufficiently developed. The proposed Hourglass Framework addresses this gap by conceptualising these foundations as interrelated elements of a continuous assessment-to-

learning process rather than viewing them as independent or isolated domains. Communicative competence defines what should be assessed in CEFR B2 speaking performance. It determines the aspects of linguistic and pragmatic constructs that must be measured. Meanwhile, Human–AI collaborative evaluation serves as the operational mechanism, describing how human and AI-supported capabilities can complement one another during the evaluation process. Language assessment quality establishes the principles through which the validity, reliability, fairness, and alignment of the collaborative execution are safeguarded. Finally, learning analytics explains how structured assessment evidence can be transformed into actionable insights for learner development and pedagogical decision-making.

Fundamental Assumptions and Boundary Conditions

The Hourglass Framework is grounded in a fundamental assumption concerning complementary Human–AI capabilities and is delimited by a boundary condition specifying its intended proficiency level and assessment context. Together, these establish the conceptual scope within which the framework and its propositions should be interpreted.

1. Fundamental Assumption: Complementary Human–AI Capabilities

Human assessors and AI systems have asymmetric, non-overlapping strengths in language assessment (Baker, 2021). AI excels at rapid, large-scale pattern recognition of micro-linguistic features (e.g., speech rate, pauses, lexical characteristics, grammatical structures, and other measurable aspects of spoken performance), while human assessors excel at macro-communicative features (e.g., contextual, sociolinguistic appropriateness, turn-taking dynamics, co-constructed interactional meaning and pragmatic aspects of communication). Therefore, AI is conceptualised as a supportive element that provides additional evidence, while human assessors remain responsible for interpreting and validating this evidence.

2. Boundary Condition: Target Proficiency and Assessment Context

The framework is contextualised for independent language learners (CEFR B2) in face-to-face speaking interviews at higher education institution. The main emphasis is on a structured speaking assessment in which learners are expected to communicate independently and relatively spontaneously. The framework is therefore not designed to represent all forms of language assessment nor all levels of proficiency. It is not designed to model basic formulaic production (A1/A2) or highly abstract academic discourse (C1/C2). Its application to other CEFR levels, assessment formats, or educational contexts would require further theoretical and empirical investigation.

Theoretical Propositions of the Hourglass Framework

The four theoretical propositions articulate the sequential relationships through which the theoretical foundations of the proposed Hourglass Framework are connected to speaking development. These propositions present the conceptual connection of communicative competence, Human–AI collaborative evaluation, assessment-quality assurance, learning analytics, and CEFR B2 speaking development. Figure 1 illustrates these proposed relationships.

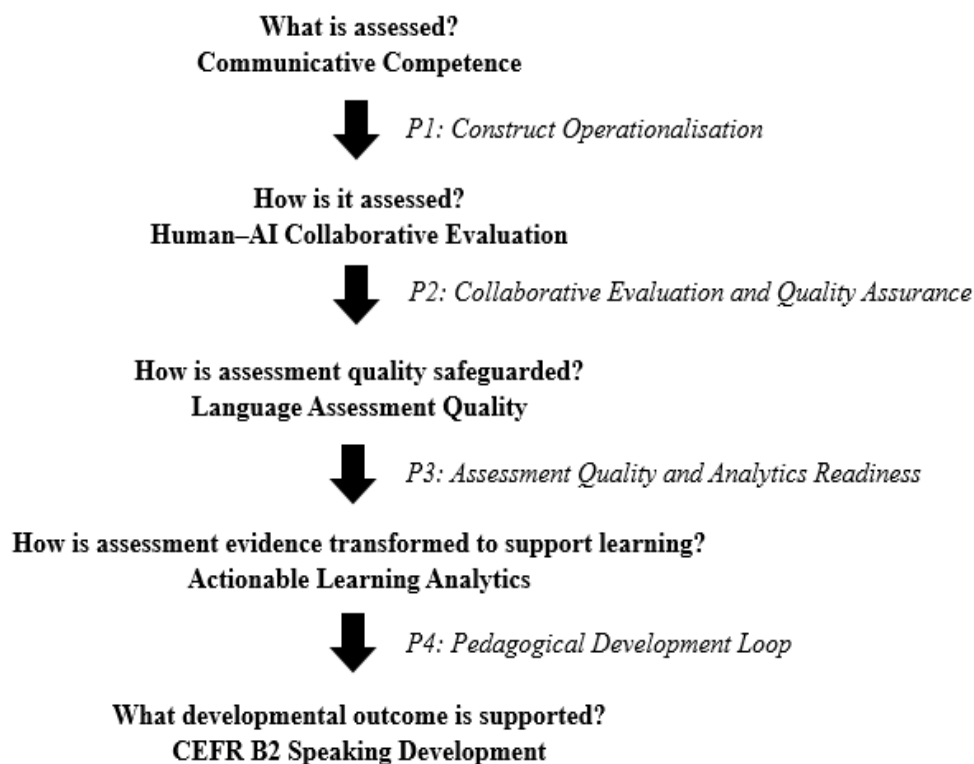


Figure 1. Theoretical Propositions of the Hourglass Framework

Proposition 1: Construct Operationalisation

P1: CEFR-aligned communicative competence can be operationalised through Human–AI collaborative evaluation, in which AI-supported technologies assist in analysing observable linguistic and speech-related features while human assessors evaluate and validate broader dimensions of communicative performance.

This proposition is based on the multidimensional nature of speaking competence. Complete automation (AI-only) or traditional face-to-face grading (human-only) insufficiently represents the CEFR B2 speaking performance. AI-supported analysis may supply structured evidence for observable aspects of performance, while human assessors may interpret these features within the broader communicative context. Effective evaluation of independent spoken interaction is maximised only when micro-linguistic dimensions (grammar and vocabulary) are pre-analysed via AI-powered language analysis, while macro-communicative dimensions (interaction and coherence) are validated and moderated by human assessors. The framework therefore proposes that combining these complementary capabilities can provide a more comprehensive basis for CEFR-aligned speaking evaluation.

Proposition 2: Collaborative Evaluation and Quality Assurance

P2: Human–AI collaborative evaluation can support language assessment quality by combining the complementary capabilities of human assessors and AI-supported technologies, while managing the limitations and potential sources of bias associated with human-only and AI-supported assessment.

Limitations may exist for both human and AI assessment. Human assessors may experience workload, cognitive fatigue, or inconsistency, particularly when evaluating large numbers of speaking performances. AI systems may also generate inaccurate or inappropriate outputs due to shortcomings in voice recognition, accent variation, algorithmic bias, or insufficient contextual understanding. The framework therefore does not presume that human or AI assessment is intrinsically better. Instead, it proposes that collaboration enables human assessors to review, analyse and authenticate evidence generated by AI.

Proposition 3: Assessment Quality and Analytics Readiness

P3: Language assessment quality supports actionable learning analytics by ensuring that assessment data are well-structured, rubric-aligned, appropriately moderated, and sufficiently trustworthy for meaningful interpretation and analysis.

Actionable learning analytics in language education cannot be generated from raw, unmoderated AI scores or subjective human grades. Instead, the structured, rubric-aligned diagnostic feedback and scores produced by moderated Human–AI collaborative evaluation serve as the necessary high-fidelity data input required to feed predictive and descriptive learning analytics models. Learning analytics is only as effective as the quality of the data on which it is built. A single total score in speaking provides limited information about why a learner performed at a particular level. In contrast, structured assessment data across domains covering fluency, interaction, vocabulary, grammar, coherence, and intelligibility can provide more meaningful information about learners' strengths, weaknesses, and areas for improvement. The framework, therefore, proposes that Human–AI collaborative evaluation should produce structured, diagnostically meaningful assessment evidence before any evidence is used in learning analytics. This information can then be utilised to discover individual, classroom and programme-level patterns in speaking performance.

Proposition 4: The Pedagogical Loop

P4: Through the transformation of suitably controlled assessment data into diagnostic feedback and targeted learning recommendations, learning analytics can enable a continuous cycle of reflection, practice and reassessment for speaking development.

This proposition places assessment as part of learning rather than the end of education. When learners receive diagnostic feedback and personalised learning paths derived from Human–AI collaborative evaluation analytics, they may engage in more focused self-reflection and iterative practice in their CEFR B2 spoken proficiency compared to traditional summative grading. For example, a learner may have sufficient grammatical accuracy and vocabulary but may have difficulty in sustaining fluency or responding appropriately in interaction. Rather than simply reporting the learner's score, learning analytics can help identify specific areas of difficulty and inform targeted learning activities, such as turn-taking practice, follow-up questions, fluency-building exercises, or simulated speaking interactions. The subsequent learning activities can then lead to further speaking practice and reassessment. New assessment data can provide additional evidence of progress and identify areas that still require attention. Thus, the framework considers speaking assessment as a continuing process in which this cycle links assessment directly to the growth of communicative competence. The Hourglass Theoretical Framework consequently moves beyond viewing AI as an automated scoring technology. Instead, it positions AI as part of a bigger, quality-orientated and human-supervised evaluation ecosystem. The framework links what is assessed, how it is assessed, how assessment quality

is maintained, how assessment data are interpreted, and how the resulting information can contribute to learning and teaching. Importantly, the framework is conceptual and therefore does not claim that Human–AI collaborative evaluation will automatically improve reliability, validity, fairness, or learning outcomes. Rather, it provides a theoretically informed framework through which these relationships can be investigated empirically in future research.

The authors of this study set research objectives as follows:

1. To develop the Hourglass Framework of Human–AI Collaborative Evaluation by combining related conceptual perspectives.
2. To examine how AI can be systematically integrated into the evaluation of CEFR B2 speaking interview assessments for English language learners.
3. To explore the implications of the proposed framework for learning analytics, speaking skills development, evaluation practices, and curriculum improvement in higher education.

Methodology

This study employs a qualitative conceptual research design to develop a theoretically grounded framework that integrates AI into the CEFR B2 interview speaking assessment. The development of the proposed Hourglass Framework of Human–AI Collaborative Evaluation was carried out in three sequential stages: (1) literature identification and narrative synthesis, (2) conceptual integration, and (3) framework development and contextualisation.

Stage 1: Literature Identification and Narrative Synthesis

The initial phase encompassed a focused literature review involving narrative and thematic synthesis of current scholarly works derived from secondary documentary sources. A two-week literature search was conducted across five electronic databases, Scopus, Web of Science, ERIC, ScienceDirect, and Google Scholar, to establish a systematic and transparent evidence base. The search was organised around four main thematic domains that correspond to the theoretical foundations of the proposed framework: (i) CEFR B2 communicative competence and speaking assessment, (ii) Human–AI collaborative evaluation in speaking assessment, (iii) language assessment quality in AI-supported evaluation, and (iv) learning analytics for speaking development. During the screening process, duplicate records and studies that did not align with the educational focus of tertiary spoken interaction were removed. A total of 54 peer-reviewed sources were included in the final synthesis corpus. These selected articles were reviewed, compared, categorised, and qualitatively synthesised to map conceptual relationships, assessment-related challenges, and key research gaps. The resulting literature base is summarised in Table 2 according to these four thematic domains and subsequently integrated to establish the four theoretical foundations of the Hourglass Framework.

Table 2. Summary of Literature Synthesized for the Hourglass Framework

Thematic Domain	No. of Sources	Key Representative Literature	Core Contribution to the Hourglass Framework
Domain 1: CEFR B2 Communicative Competence and Speaking Assessment in Higher Education	14	Canale & Swain (1980); Canale (1983); Council of Europe (2020); Bachman & Palmer (1996); Hymes (1971); Taylor (2011); Weir (2005); Fulcher (2010); Galaczi & Taylor (2018); O'Sullivan (2011); Mohd Don et al. (2025); Nakatsuhara et al. (2023); Seed (2024); Green (2021)	Establishes the CEFR B2 speaking construct and its validation basis, covering communicative competence, assessment usefulness and validity, interactive speaking constructs, contextual task parameters, and construct-to-descriptor alignment.
Domain 2: Human-AI Collaborative Evaluation in Speaking Assessment	13	Baker (2021); Luckin (2018); Zawacki-Richter et al. (2019); Pardo & Siemens (2014); Selwyn (2019); Bond (2023); Chounta et al. (2022); Holstein et al. (2019); Lodge et al. (2023); Rets & Herodotou (2024); Seufert et al. (2021); Cope et al. (2025); Yan et al. (2026)	Establishes a human-in-the-loop approach in which AI supports scalable linguistic analysis, dashboards, and assessment workflows while human educators retain professional and pedagogical judgment, reducing fatigue and mitigating automation risks.
Domain 3: Language Assessment Quality in AI-Supported Evaluation	14	Chuang & Yan (2025); Holmes et al. (2019); Kasneci et al. (2023); Kohnke et al. (2023); Long & Magerko (2020); Wollny et al. (2021); Tlili et al. (2023); Attali & Burstein (2006); Chen & Feng (2022); Crossley (2024); Jiang & Yan (2023); Lee & Lee (2024); Yan & Chuang (2024); Zou et al. (2025)	Establishes the role of AI and generative AI in language assessment, including automated linguistic analysis, scoring, feedback, and speech processing, while identifying reliability, construct-under-representation, and automation limitations.
Domain 4: Learning Analytics for Speaking Development	13	Siemens (2013); Ifenthaler & Yau (2020); Ferguson (2012); Gašević et al. (2015); Tempelaar et al. (2015); Wise & Vytasek (2017); Rienties et al. (2022); Tsai et al. (2023); Wong et al. (2024); Winne (2025); Knight & Buckingham Shum (2017); Lang et al. (2017); Macfadyen & Dawson (2010)	Establishes learning analytics as the mechanism for transforming assessment and learning data into actionable diagnostic information, supporting feedback, self-regulated learning, institutional implementation, and curricular alignment.

Stage 2: Conceptual Integration

The second stage was to integrate the four selected conceptual underpinnings identified during the literature synthesis: communicative competence, Human–AI collaborative evaluation, language-assessment quality, and learning analytics. Within this context, the structural and pedagogical relationships between these foundations were investigated to understand their collective role in explaining what should be assessed, the complementary roles of human assessors and artificial intelligence systems, how assessment quality is maintained, and how the evaluation outcome contributes to learners’ developmental pathways. This conceptual integration provided the theoretical framework and formal propositions of the proposed Hourglass Framework of Human–AI Collaborative Evaluation.

Stage 3: Framework Development and Contextualisation

The final stage involved translating the integrated conceptual foundations into the proposed Hourglass Framework of Human–AI Collaborative Evaluation. These integrated concepts were organised into a coherent, visual model comprising interconnected components that illustrate the systematic progression of AI integration in CEFR B2 speaking assessments. This framework was subsequently contextualised for the English Language Enrichment II (ELE II) course at UUM, demonstrate its potential application in CEFR B2 speaking interview evaluation in higher education. This implementation illustrates how Human–AI collaborative evaluation could be executed while preserving educational integrity, pedagogical value, learning analytics, and learner outcomes.

Findings: The Hourglass Framework of Human–AI Collaborative Evaluation

The main contribution of this conceptual research is the development of the proposed Hourglass Framework of Human-AI Collaborative Evaluation. This is a conceptual framework for integrating AI into the CEFR B2 speaking interview assessment. This development addresses the first research objective by integrating communicative competence, Human–AI collaborative evaluation, language assessment quality, and learning analytics into a unified framework. This framework uses the hourglass structure to illustrate the progression from broad educational challenges at the upper macro level, to a focused evaluation process at the micro level, and, later, to broader educational outcomes and learner development at the lower macro level. Figure 2 displays the Hourglass Framework of Human–AI Collaborative Evaluation for CEFR B2 speaking interview assessment in higher education.

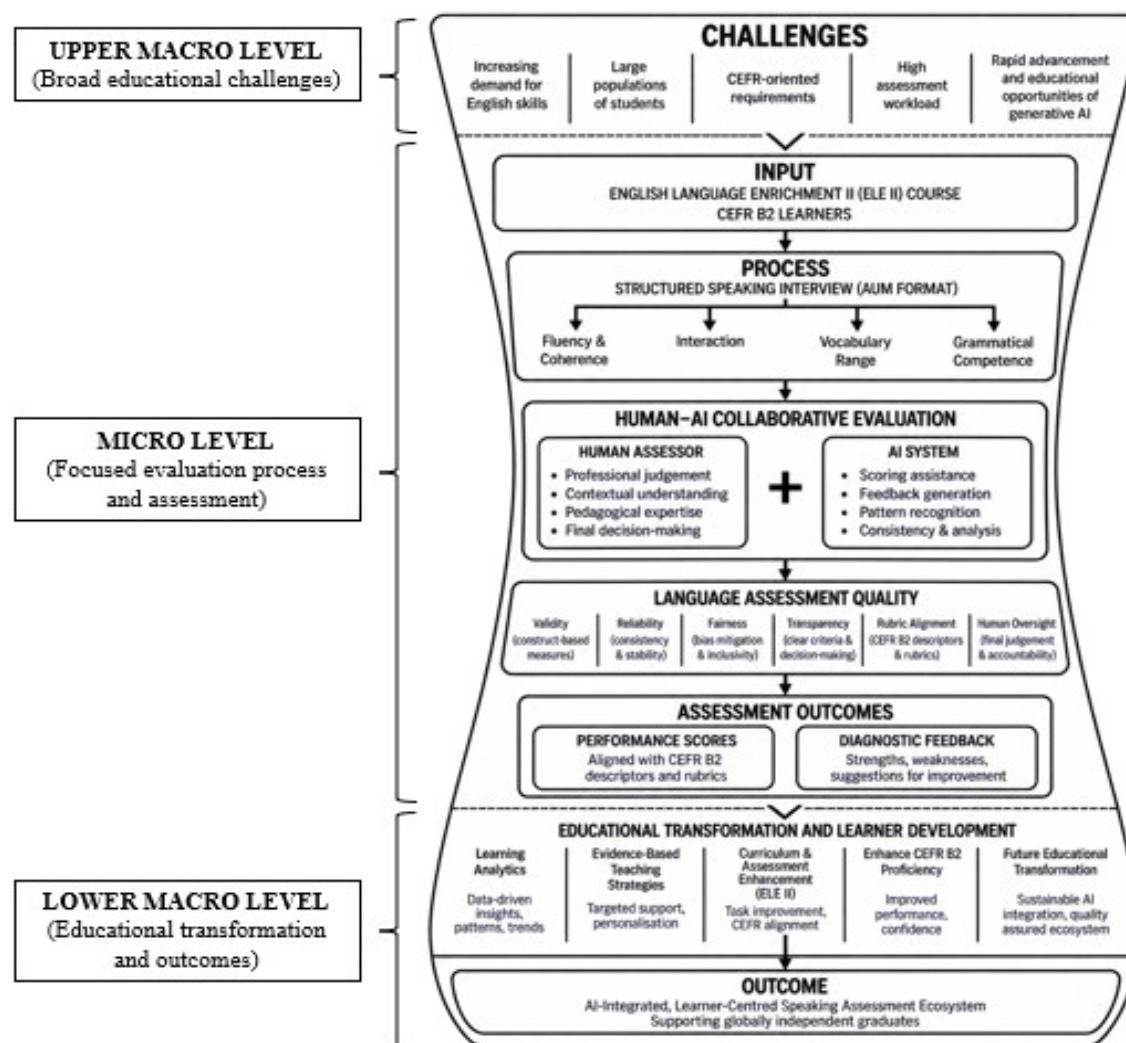


Figure 2. The Hourglass Framework of Human–AI Collaborative Evaluation for CEFR B2 Speaking Interview Assessment in Higher Education

Upper Macro Level: Educational Challenges and the Need for AI Integration

The upper chamber of the Hourglass Framework of Human–AI Collaborative Evaluation represents a broader educational landscape that requires the integration of AI into language evaluation. Higher education institutions increasingly face expectations to produce more graduates with strong English-language communication skills who can effectively function in academic, professional, and global contexts. In the meantime, universities face growing pressure to uphold assessment quality, ensure consistency in evaluation, and provide meaningful feedback while managing large learner populations. Within CEFR-aligned language courses, speaking interview assessments present challenges because they require a multidimensional evaluation of communicative competence, including fluency and coherence, interaction, vocabulary, and grammar. These dimensions demand a lot of time, expertise, and professional judgement from the instructors. Hence, the traditional evaluation approach seems difficult to sustain as the evaluation workload increases. The emergence of generative AI is a potential means of supporting these evaluation challenges by providing adaptive, personalised learning and assessment (Chuang & Yan, 2025). AI systems can process language data efficiently, including analysing learners' performance, generating diagnostic feedback, and

supporting evidence-based decision-making. Therefore, this framework identifies AI as not only a technological innovation but also a strategic educational tool that would enhance the efficiency and consistency in assessment and learner support in higher education.

Micro Level: Human–AI Collaborative Evaluation in CEFR B2 Speaking Interviews

The narrow section of the hourglass represents the core component of this framework and functions as the focal point of AI integration. At this stage, the framework addresses the second research objective by illustrating how AI can be systematically integrated into CEFR B2 speaking interview assessment while maintaining educational integrity and supporting language learners' development. This level specifically addresses the CEFR B2 speaking interview assessment for English-proficiency learners enrolled in the English Language Enrichment II (ELE II) course at Universiti Utara Malaysia. Within this ELE II course, the framework can support rubric-aligned analysis of CEFR B2 speaking performance across four evaluation criteria: fluency and coherence, interaction, vocabulary and grammar. The framework proposes a Human–AI collaborative evaluation process, where instructors and AI-supported technology work together in the evaluation of learners' CEFR B2 speaking performance. Within this collaborative process, instructors contribute to pedagogical expertise, contextual understanding, ethical and professional judgement and final decision-making. Meanwhile, AI contributes towards analytical capabilities by providing preliminary scoring, automated speech analysis, pattern recognition across performances, and feedback generation. Together, these complementary roles may support the quality of assessment when AI-generated outputs are appropriately reviewed, validated, and interpreted by qualified human assessors.

Besides that, the implementation of Human–AI collaborative evaluation should also be guided by appropriate ethical and governance principles. The use of AI-supported speaking assessment should be underpinned by clear protocols on informed consent, transparency, data privacy, accountability, and secure handling of learners' speech data. Such governance mechanisms are important for supporting an educationally responsible, ethically appropriate, and institutionally and legally compliant Human–AI collaborative evaluation process. Language assessment quality functions as a key safeguard within the Human–AI collaborative evaluation process. In the proposed framework, the integration of human and AI capabilities should be guided by principles of validity, reliability, fairness, transparency, rubric alignment, and human oversight. These quality concerns are designed to provide a way to effectively analyse, moderate and align AI-generated evidence with CEFR B2 speaking assessment criteria before contributing to assessment outcomes. The integration of these AI-supported functions generates two key assessment outputs:

1. Performance Scores

Performance scores provide structured evidence of learners' performance across speaking interview assessment criteria. AI-assisted scoring would support consistency in the evaluation process by identifying performance patterns across fluency and coherence, interaction, vocabulary, and grammar criteria aligned with CEFR B2 descriptors and rubrics. However, the score generated by AI should be reviewed, interpreted and validated by instructors, who are responsible for the final assessment decision based on professional considerations and contextual understanding.

2. Diagnostic Feedback

Diagnostic feedback is a vital component of learner-centred assessment. AI-supported feedback generation can provide learners and instructors with timely information regarding strengths, areas for improvement, and specific aspects of speaking

performance that require further development. Such feedback may encourage learner reflection and support targeted improvement in speaking performance.

Together, performance scores and diagnostic feedback constitute the key assessment outputs of the Micro Level. These outputs are not the end products of assessment, but rather they are structured evidence that may be turned by means of learning analytics into actionable information for learner development, instructional improvement, and broader educational decision-making.

Lower Macro Level: Learning Analytics, Educational Transformation, and Learner Development

The lower chamber of the hourglass represents broader educational outcomes that emerge from Human-AI collaborative evaluation practice.

This stage addresses the third research objective by highlighting the implications of the proposed framework for learning analytics, speaking skills development, evaluation practices, and curriculum improvement across three distinct levels of educational practice: learners, instructors, and institutions. The framework proposes that the performance score and diagnostic feedback should not remain isolated evaluation products but be transformed into actionable educational intelligence through learning analytics. When supported by sufficient longitudinal data and appropriate analytical procedures, learning analytics would help instructors and institutions identify learners' performance trends, monitor their progress, and develop targeted interventions. At the learner level, learning analytics facilitates a personalised learning path that helps learners understand their performance patterns and target specific areas for improvement. This process directly contributes to the development of CEFR B2 communicative competence by encouraging continuous reflection, practice, and improvement. Meanwhile, at the instructional level, instructors can use evaluation data to refine their teaching strategies, develop targeted learning activities, and provide support for learners' diverse needs. Lastly, at the institutional level, aggregated assessment data can enhance curriculum development, programme evaluation, and quality assurance initiatives. This framework, therefore, conceptualises Human-AI collaborative evaluation as a catalyst for educational transformation, rather than merely as an evaluative tool. By linking assessment outcomes to learning analytics and teaching improvement, this framework may create a continuous cycle of evaluation, feedback, and learner development.

Discussion

The findings of this conceptual research demonstrate the value of Artificial Intelligence (AI), especially in supporting the evaluation of CEFR B2 speaking interview assessments in higher education. The proposed Hourglass Framework of Human-AI Collaborative Evaluation provides a comprehensive conceptual explanation for systematically integrating AI into CEFR B2 speaking interview assessments in higher education while preserving educational integrity and pedagogical relevance. Building on existing research, this framework offers an integrated perspective that bridges evaluation processes, learning analytics, and learners' development within a unified CEFR B2 speaking assessment system. The framework starts with broader educational challenges that drive the application of AI technology. Then it narrows to the implementation of Human-AI collaborative evaluation in CEFR B2 speaking interviews, before expanding to broader educational outcomes through learning analytics, instructional improvement, and learner development. Within the CEFR B2 interview assessment, instructors provide contextual understanding, pedagogical judgement and ethical oversight and decision-

making. While AI may contribute to assessment reliability and consistency, it also provides immediate analytical support, generates diagnostic feedback, and conducts structured performance analysis. By connecting the evaluation process to learning analytics and educational transformation, the proposed Hourglass Framework expands the role of assessment beyond mere measurement and grading. Instead, it positions evaluation practices as a continuous process that supports enhancement in speaking proficiency, curriculum improvement, and evidence-based pedagogical approaches.

The main contribution of this framework lies in its emphasis on Human-AI collaborative evaluation. The findings show that the future of interview-based assessment should not replace human assessors with AI; instead, AI should function as a complementary assessment partner that elevates the role of the instructor rather than diminishes it. This perspective is in tandem with Human-AI collaborative evaluation, which emphasises the complementary roles of human expertise and technological capabilities. The integration of these capabilities has the potential to enhance the quality of evaluation while preserving the human-centred nature of language learning. However, AI evaluation alone could introduce risks of algorithmic bias, inconsistent performance across groups, data privacy concerns, and limited transparency in automated decision-making. Therefore, AI-generated output must be considered complementary evidence supporting professional decisions and does not serve as the sole basis for evaluation. The framework also continues to conceptualise performance scores and diagnostic feedback as two complementary outputs of Human-AI collaborative interview evaluation. This proposed framework reconceptualises the interview evaluation as a development process that yields valuable insights to inform pedagogical strategy, improve the curriculum, and support learners. This is in line with learning analytics theory, which supports the use of data in education to improve learning outcomes and instructional decision-making. In the context of the ELE II course at UUM, this framework offers practical relevance for CEFR B2 interview assessment. The integration of AI-supported scoring, immediate feedback generation, and performance analysis would enable instructors to provide consistent yet personalised support while also reducing assessors' workloads. Moreover, this framework emphasises the importance of leveraging assessment data to identify learners' needs to refine pedagogical approaches and support the continuous development of speaking skills.

Overall, these research findings show that the proposed framework needs to be seen not only as a mere technological innovation but as a structured approach to support the quality, consistency, and pedagogical usefulness of CEFR B2 speaking interview evaluation. The proposed Hourglass Framework of Human-AI Collaborative Evaluation contributes to some important literature. Firstly, it introduces a new conceptual framework that integrates AI into the CEFR B2 interview evaluation. Secondly, it broadens the scope of Human-AI collaborative evaluation in CEFR-aligned language assessments, emphasising the complementary use of AI and instructors' expertise in assessment decision-making. Thirdly, this framework integrates learning analytics with interview assessment, demonstrating how assessment outcomes can be transformed into actionable insights for learners' development. Finally, this framework reconceptualises AI as a developmental assessment tool, rather than merely a scoring or grading mechanism. This contribution, taken together, lays the groundwork for future empirical research and further theoretical development in AI-assisted language evaluation.

Conclusion

To sum up, this paper proposes the Hourglass Framework of Human-AI Collaborative Evaluation as a conceptual framework for English learners in higher education. This framework explains how AI can be systematically integrated into a CEFR B2 interview assessment by connecting Human-AI collaborative evaluation and learner-centred learning outcomes. The

framework is grounded in four interconnected theoretical foundations: communicative competence, Human–AI collaborative evaluation, language assessment quality, and learning analytics. Together, these foundations explain what constitutes CEFR B2 speaking performance, how human and AI capabilities can be integrated during evaluation, how assessment quality can be safeguarded, and how assessment evidence can be transformed into actionable insights for speaking development. The importance of this framework lies in the AI conceptualisation as a source of supplementary evaluation evidence rather than replacing instructors. By combining human expertise and technological capabilities, this framework encourages a balanced approach that enhances the effectiveness of evaluation while preserving professional consideration and ethical responsibility. Moreover, this framework broadens the role of evaluation beyond grading by showing how evaluation results support evidence-based teaching, improve the curriculum, and provide continuous speaking development. As AI continues to evolve, this framework offers a theoretically grounded and pedagogically meaningful approach to integrating AI into language evaluation. This framework contributes to ongoing discussion of the future of education-based AI and provides a foundation for empirical research to validate and refine its application in higher education.

Recommendations

Based on the proposed framework, this paper offers a few recommendations for higher education institutions, language instructors, and future researchers. First, universities have to accept AI as a supporting tool within speaking evaluation process. The implementation of collaborative Human-AI evaluation must ensure that instructors continue to play an important role in making the final assessment. Secondly, institutions have to strengthen assessment and AI literacy among instructors. They need to acquire the necessary knowledge and skills to critically interpret the scores, feedback, and analytical output that AI provides. Hence, institutions should prioritise a professional development programme that focuses on the responsible use of AI in assessments. Thirdly, universities must establish a clear and well-defined framework, along with ethical guidelines, for AI-based assessments. Higher education institutions must diligently address issues of accountability, transparency, data privacy, and bias to ensure the responsible use of AI. Fourthly, a system supported by AI can use the assessment data it produces to support learning analytics and make evidence-based decisions. The data is essential for improving the curriculum, providing learning interventions, and supporting continuous quality improvement in language learning and teaching. Lastly, future research should examine the empirical effectiveness of the proposed framework across educational contexts, competence levels, and AI platforms. Emphasis should be given to the validity of AI-based assessments, the quality of AI-generated feedback, learners' perceptions of Human-AI collaborative evaluation, and comparisons of assessments from emerging AI generative technologies.

Limitations

Despite its theoretical contributions, this study has a few limitations. Firstly, the Hourglass Framework for Human-AI Collaborative Evaluation is a conceptual model and has not yet been empirically validated in classroom or experimental research. Hence, the proposed relationship among Human-AI collaborative evaluation, learning analytics, and learners' development remains theoretical and requires empirical testing. Secondly, this framework is specifically tailored for interview assessments of CEFR B2 English language skills in higher education. As a result, the applicability of this framework to other language skills, educational environments, and disciplinary contexts remains uncertain. Thirdly, this framework emphasises the benefit of

integrating AI into education and does not comprehensively research the limitations of the technology, algorithmic bias, or institutional obstacles that would affect the implementation.

References

- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater v2.0. *The Journal of Technology, Learning, and Assessment*, 4(3), 1–30.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Baker, R. (2021). Artificial intelligence in education: Bringing it all together. In *OECD Digital Education Outlook 2021: Pushing the frontiers with artificial intelligence, blockchain and robots* (pp. 43–56). OECD Publishing. <https://doi.org/10.1787/f54ea644-en>
- Bond, M. (2023). Human-in-the-loop AI-enhanced assessment in higher education: A scoping review. *Educational Technology Research and Development*, 71(3), 903–928. <https://doi.org/10.1007/s11423-023-10214-w>
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2–27). Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/applin/I.1.1>
- Chen, L., & Feng, G. (2022). Automated speaking assessment in the wild: Acoustic and linguistic feature analysis. *Computer Assisted Language Learning*, 35(5-6), 1145–1172. <https://doi.org/10.1080/09588221.2020.1788579>
- Chounta, I. A., Bardone, E., Raudsep, A., & Pedaste, M. (2022). Designing Human–AI collaboration in language education: A co-design study with language teachers. *Computers & Education: Artificial Intelligence*, 3, 100089. <https://doi.org/10.1016/j.caeai.2022.100089>
- Chuang, P.-L., & Yan, X. (2025). Language assessment in the era of generative artificial intelligence: Opportunities, challenges, and future directions. *System*, 134, 103846. <https://doi.org/10.1016/j.system.2025.103846>
- Cope, B., Kalantzis, M., & Searsmith, D. (2025). The human-in-the-loop: Reimagining assessment and feedback in the age of generative AI. *Journal of Educational Technology in Higher Education*, 22(1), 18. <https://doi.org/10.1186/s41239-025-00465-3>
- Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment. Companion Volume*. Council of Europe Publishing. <https://www.coe.int/en/web/common-european-framework-reference-languages>
- Crossley, S. A. (2024). Linguistic analysis of second language speaking and writing using natural language processing tools. *Language Assessment Quarterly*, 21(1), 89–105. <https://doi.org/10.1080/15434303.2024.2309112>

- Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5-6), 304–317. <https://doi.org/10.1504/IJTEL.2012.051816>
- Fulcher, G. (2010). *Practical language testing*. Routledge.
- Galaczi, E. D., & Taylor, L. (2018). Trends in second-language speaking assessment. *Language Assessment Quarterly*, 15(4), 406–423. <https://doi.org/10.1080/15434303.2018.1537559>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. <https://doi.org/10.1007/s11528-014-0822-x>
- Green, A. (2021). *Linking language assessments to the Common European Framework of Reference for Languages*. Cambridge University Press.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promise and implications for teaching and learning*. Center for Curriculum Redesign.
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing teacher-facing AI dashboards for live classrooms: A case study. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3290605.3300821>
- Hymes, D. (1971). *On communicative competence*. University of Pennsylvania Press.
- Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics for study success. *Educational Technology Research and Development*, 68(4), 1961–1964. <https://doi.org/10.1007/s11423-020-09788-z>
- Jiang, Y., & Yan, X. (2023). ChatGPT as an automated rater for language examinations: Reliability, agreement, and feedback quality. *Assessing Writing*, 58, 100780. <https://doi.org/10.1016/j.asw.2023.100780>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasparetti, D., Gnest, S., Gokhale, A., Hauswirth, K., Hüllermeier, E., Zacharias, S., Kern-Isberner, G., Koedinger, K., Lipizzi, C., Loll, F., Pinkwart, N., Pontoldof, L., Scholes, O., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Knight, S., & Buckingham Shum, S. (2017). Theory and infrastructure for learning analytics. *Journal of Learning Analytics*, 4(2), 17–29. <https://doi.org/10.18608/jla.2017.42.3>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 242–250. <https://doi.org/10.1177/00336882231162868>
- Lang, C., Siemens, G., Wise, A., & Gašević, D. (Eds.). (2017). *Handbook of learning analytics*. Society for Learning Analytics Research.
- Lee, H., & Lee, J. H. (2024). The use of generative AI in second language oral assessment: An exploration of scoring accuracy and construct representation. *ReCALL*, 36(1), 112–128. <https://doi.org/10.1017/S095834402300021X>

- Lodge, J. M., de Barba, P., & Broadbent, J. (2023). Learning in the age of generative AI: The role of human agency and cognitive offloading in learner evaluation. *Computers and Education*, 205, 104889. <https://doi.org/10.1016/j.compedu.2023.104889>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3313831.3376727>
- Luckin, R. (2018). *Machine learning and human intelligence: The future of education in the 21st century*. UCL IOE Press.
- Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an early warning system for educators: A proof of concept. *Computers & Education*, 54(2), 580–590. <https://doi.org/10.1016/j.compedu.2009.09.008>
- Mohd Don, Z., Abdul Raof, A. H., Zainal Abidin, S. A., Md. Rashid, S., Kassim, R., & Mohd Shamsuddin, C. (2025). Implementing the CEFR-aligned English language education reform: Assessing speaking in universities (Vol. 2). Department of Higher Education, Ministry of Higher Education Malaysia.
- Nakatsuhara, F., Inoue, C., & Taylor, L. (2023). Double-testing or co-constructing? Re-examining spoken interaction constructs in paired speaking tests. *Language Testing*, 40(2), 241–267. <https://doi.org/10.1177/0265532222111244>
- O'Sullivan, B. (2011). Language testing and validation. In B. O'Sullivan (Ed.), *Language testing: Theory and practice* (pp. 13–32). Palgrave Macmillan.
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. <https://doi.org/10.1111/bjet.12152>
- Rets, I., & Herodotou, O. (2024). Orchestrating Human-AI collaborative assessment: Teacher perspectives on workflow integration. *British Journal of Educational Technology*, 55(2), 654–672. <https://doi.org/10.1111/bjet.13388>
- Rienties, B., Boroowa, A., Cross, S., Kubiak, C., & Mayles, K. (2022). Scaling up learning analytics-informed interventions: A case study of institutional adoption. *Journal of Learning Analytics*, 9(1), 114–129. <https://doi.org/10.18608/jla.2022.7317>
- Seed, G. (2024). Aligning university-level placement tests with the CEFR: Pragmatic and structural challenges in spoken assessment. *System*, 122, 103289. <https://doi.org/10.1016/j.system.2024.103289>
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.
- Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-enhanced language learning under the lens of the socio-technical systems theory: A collaborative human-AI perspective. *Education and Information Technologies*, 26(4), 4811–4832. <https://doi.org/10.1007/s10639-021-10502-3>
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400. <https://doi.org/10.1177/0002764213498851>

- Taylor, L. (Ed.). (2011). *Examining speaking: Research and practice in assessing second language speaking*. Cambridge University Press.
- Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search of the most informative data for feedback: LMS logs or learning analytics. *Computers in Human Behavior*, 47, 157–167. <https://doi.org/10.1016/j.chb.2014.07.054>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Huang, R., & Chang, T.-W. (2023). What if the devil is my guardian angel: ChatGPT as a case study of generative artificial intelligence in education. *Smart Learning Environments*, 10(1), 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Tsai, Y. S., Whitelock-Wainwright, A., & Gašević, D. (2023). Moving beyond the pilot: Implementing learning analytics at scale in higher education. *Higher Education Research & Development*, 42(3), 702–718. <https://doi.org/10.1080/07294360.2022.2082390>
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave Macmillan.
- Winne, P. H. (2025). Learning analytics as a catalyst for self-regulated language learning. *Educational Psychologist*, 60(1), 12–28. <https://doi.org/10.1080/00461520.2024.2339103>
- Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation in higher education: Institutional challenges and systemic solutions. *Journal of Learning Analytics*, 4(2), 6–21. <https://doi.org/10.18608/jla.2017.42.2>
- Wollny, S., Schneider, J., Di Mitri, D., Rittercov, A., & Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, 654924. <https://doi.org/10.3389/frai.2021.654924>
- Wong, J., Baars, M., de Koning, B., & Paas, F. (2024). Actionable learning analytics in online self-regulated language learning. *Computers & Education*, 210, 104961. <https://doi.org/10.1016/j.compedu.2023.104961>
- Yan, X., & Chuang, P. L. (2024). Generative artificial intelligence and language assessment: A critical review. *Language Assessment Quarterly*, 21(3), 203–221. <https://doi.org/10.1080/15434303.2024.2351112>
- Yan, X., Jiang, Y., & Chuang, P. L. (2026). Human-AI collaborative grading in second language oral interviews: Reliability, efficiency, and examiner cognitive load. *System*, 138, 104112. <https://doi.org/10.1016/j.system.2026.104112>
- 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 domains of AI? *International Journal of Educational Technology in Higher Education*, 16(1), 1–27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zou, D., Kohnke, L., & Moorhouse, B. L. (2025). Generative AI for formative speaking assessment: A pedagogical framework for higher education. *System*, 131, 103567. <https://doi.org/10.1016/j.system.2024.103567>

The World on a Plate (And the Missing Nasi Lemak): A Preliminary Study of a Task-Based Ebook in a Malaysian Vernacular School

Woi Min Xuan, Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia,
p162813@siswa.ukm.edu.my

Hanushka Anandini A/P Kumaresan, Faculty of Education, Universiti Kebangsaan
Malaysia, Malaysia, p165597@siswa.ukm.edu.my

Harwati Hashim, Nur Syafiqah Yaacob & Melor Md Yunus, Faculty of Education,
Universiti Kebangsaan Malaysia, Malaysia

Abstract: The adoption of Common European Framework of Reference (CEFR) aligned textbooks in Malaysian vernacular schools has introduced notable instructional challenges, as Western-centric materials often increase cognitive load for students learning English as a third language (L3). This pilot study examines the efficacy of a culturally responsive, multimodal interactive ebook—*The World on a Plate (And the Missing Nasi Lemak)*—developed using the ASSURE model to scaffold grammar acquisition and reading comprehension for Year 5 learners. Grounded in Culturally Responsive Pedagogy, Schema Theory, and Mayer’s Cognitive Theory of Multimedia Learning, the intervention was evaluated using a pre-test and post-test design with a convenience sample of 13 Tamil vernacular school (SJKT) students. Quantitative findings revealed marked academic improvement: the cohort's mean score rose from 51.5% to 95.4%, achieving 100% discrete grammar mastery. These results indicate that localized narrative content and guided multimedia significantly reduce extraneous cognitive load. While constrained by a small sample size and potential digital infrastructure limitations in rural settings, the findings underscore the potential of culturally situated digital tools to foster inclusive L3 acquisition in vernacular classrooms.

Keywords: Cognitive load, Culturally responsive pedagogy, Multimedia learning, Vernacular education

Introduction

Background of Study

To ensure local graduates are prepared to participate effectively in global socio-economic domains, the Malaysian Ministry of Education has undertaken extensive curricular reforms over the past decade. A primary initiative within this nationwide agenda was the implementation of the Standards-Based English Language Curriculum (SBELC), systematically aligned with the Common European Framework of Reference for Languages (CEFR). A central operational dimension of this reform entailed the nationwide deployment of commercial, internationally published CEFR-aligned learning packages—including Super Minds, Get Smart Plus 3, and Academy Stars 6—across primary institutions nationwide (Takal et al., 2021; Ya Shak et al., 2021). The foundational objective was to benchmark English language communicative competencies against standardized global metrics, modernizing instructional frameworks and classroom assessments.

Challenges

Despite their structured pedagogical design, these imported instructional packages have created considerable instructional friction, particularly within National-Type Tamil Schools (Sekolah Jenis Kebangsaan Tamil, SJKT). Critical analyses of imported CEFR textbooks consistently demonstrate a pronounced cultural imbalance, where textual scenarios, illustrative imagery, and communicative activities are heavily anchored in Euro-American contexts (Ya Shak et al., 2021). Learners are frequently required to process cultural concepts, unfamiliar geographical references, and European culinary traditions that bear little resemblance to the daily lived experiences of Malaysian children.

In the SJKT vernacular environment, this contextual mismatch creates severe learning hurdles. From the onset of formal schooling, SJKT learners operate within a complex trilingual dynamic: Tamil functions as their primary mother tongue (L1) and medium of classroom instruction; Bahasa Melayu serves as the mandatory national second language (L2); and English is acquired strictly as a third language (L3). Navigating this tripartite linguistic ecosystem imposes a substantial 'trilingual burden' on young learners, depleting essential working memory resources on phonetic decoding, translation, and syntactic mapping (Sani & Ismail, 2021). When L3 English lessons rely on materials depicting foreign schemas—such as European seasonal transitions or unfamiliar Western dining conventions—the input becomes pedagogically opaque. Rather than supporting intuitive acquisition, these culturally detached materials impose an overwhelming extraneous cognitive load, forcing learners to decode unfamiliar cultural schemas while simultaneously grappling with new linguistic structures (Sweller et al., 2011).

Aim of study

To resolve this pedagogical mismatch, educators require localized, multimodal instructional materials grounded in Culturally Responsive Pedagogy (CRP) (Sendera Mohd. Yakin et al., 2025). Rather than presenting cognitive barriers, instructional resources should leverage students' existing cultural funds of knowledge to scaffold target language acquisition. This preliminary investigation evaluates the instructional design, implementation, and descriptive outcomes of an interactive digital ebook created via BookCreator, titled *The World on a Plate (And the Missing Nasi Lemak)*, mapped explicitly to Unit 5 (Food) of the Year 5 Malaysian ESL syllabus. Integrating Schema Theory, Task-Based Language Teaching (TBLT), and

Mayer's Cognitive Theory of Multimedia Learning (CTML), the intervention is designed to support the cognitive and linguistic profiles of L3 vernacular learners.

The study addresses two central research questions:

1. How does the inclusion of Culturally Responsive Pedagogy within an interactive ebook influence the cognitive processing and reading comprehension of Year 5 SJKT learners?
2. How does a Task-Based Language Teaching (TBLT) digital ebook facilitate short-term discrete grammar acquisition and procedural language mastery among Year 5 SJKT learners?

Literature Review

Cognitive Overload and the Trilingual Dynamics in SJKT Settings

Malaysia's vernacular education system is designed to preserve cultural heritage while developing national and international linguistic competence. However, instructional approaches that treat third-language (L3) acquisition identically to first- or second-language learning often trigger severe cognitive strain (Biang, 2025; Hasin & Nasir, 2021). In vernacular classrooms where English instruction has historically emphasized structural drills, pupils frequently revert to Tamil or Malay as a natural psychological coping mechanism when English demands exceed their working memory capacity (Azizah et al., 2024).

Cognitive Load Theory posits that working memory is strictly limited in capacity and duration when processing novel information (Sweller et al., 2011). When instructional materials are detached from learners' lived experiences, extraneous cognitive load increases, diverting mental resources away from productive schema construction (Biang, 2025). Embedding structured pedagogical scaffolding and multimodal cues eases this trilingual strain, enabling learners to allocate working memory toward target language acquisition (Sami et al., 2025; Zhou, 2025).

Western-Centric Materials and Schema Theory

From an interactive-constructivist perspective, reading comprehension in a second or third language is not a passive reception of text; rather, meaning is actively constructed as the reader connects textual information with existing background knowledge structures, termed schemata (Anderson & Pearson, 1984; Carrell & Eisterhold, 1983). Schema theory establishes that readers rely on internal cognitive databases of culturally specific scripts, settings, and social conventions to interpret textual input (Bukroni et al., 2026; Khan & Iqbal, 2025).

When imported, CEFR-aligned textbooks present culturally unfamiliar contexts, L3 learners lack the requisite cognitive frameworks to interpret texts effectively, resulting in comprehension breakdowns (Khan & Iqbal, 2025; Radzi & Radzi, 2024). In contrast, anchoring instruction in familiar cultural touchstones—such as local markets, traditional foods, and community settings—bridges these cognitive gaps and reinforces cultural identity (Sudarsi et al., 2025). By activating familiar schemas, learners avoid the mental exertion of decoding alien cultural paradigms, freeing working memory to focus on syntactic decoding and lexical acquisition (Xu, 2025).

Culturally Responsive Pedagogy (CRP) and World Englishes

Culturally Responsive Pedagogy (CRP) extends beyond superficial tokenism by using the cultural knowledge, prior experiences, and frames of reference of diverse students to make

learning more relevant and effective (Gay, 2013; Ladson-Billings, 1995). Within multicultural ESL environments, CRP fosters classroom inclusivity, reduces the affective filter, and boosts learner motivation (Matiso, 2024; Putera & Akid, 2022).

Furthermore, contemporary CRP aligns with World Englishes and plurilingual education frameworks, which challenge monolithic native-speaker norms and validate linguistic diversity (Azlan & Kaur, 2023; Boonsuk, 2024; Chaweewan & Boonsuk, 2025). Incorporating diverse, intelligible phonetic varieties and strategic translanguaging within digital tools validates learners' emerging identities, lowers language anxiety, and supports listening comprehension (Krashen, 1982; Suzuki et al., 2025).

Task-Based Language Teaching (TBLT) and Focus on Form Shifting Beyond Mechanical Drills

To move away from traditional grammar-translation methods, this study utilizes Task-Based Language Teaching (TBLT), structuring instruction around goal-oriented communicative tasks (Ellis, 2003). To maintain linguistic precision, the intervention incorporates Focus on Form (FonF), which draws explicit attention to linguistic elements as they arise during meaningful communication (Long, 1991).

Recent scholarship demonstrates that integrating FonF within task-based environments enhances grammatical precision and communicative competence in multilingual contexts (Sami et al., 2025; Xin et al., 2024). Contextualized tasks—such as preparing a recipe or assembling a virtual meal tray—turn discrete grammar constructs, like countable/uncountable nouns (much vs. many) and imperative sequencers, into functional tools required to complete the task, accelerating acquisition without overloading working memory (Xin et al., 2024).

Mayer's Cognitive Theory of Multimedia Learning (CTML)

Digital learning environments offer rich multimodal input, but poorly structured media can trigger severe cognitive overload in L3 learners. The structural architecture of the digital intervention is grounded in Richard Mayer's Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2002; Mayer & Moreno, 2003). Research confirms that multimedia materials designed around CTML principles improve retention and comprehension through coordinated dual-channel processing (Februati & Dewa, 2025; Zhou, 2025).

To support L3 vernacular learners, the ebook incorporates three core CTML principles:

1. The Coherence Principle: Extraneous visual elements, background audio, and decorative distractions were eliminated to maintain focus on learning objectives (Mayer & Moreno, 2003).
2. The Spatial Contiguity Principle: Explanatory text and corresponding illustrations were placed side-by-side on the screen to prevent split-attention fatigue (Mayer & Moreno, 2003; Prasetya, 2024).
3. The Segmenting Principle: Instructional content and narrative sequences were organized into learner-paced, digestible modules (Maamuujav, 2025; Mayer & Moreno, 2003).

Methodology

Research Design

This study utilized a quantitative, quasi-experimental single-group pre-test and post-test design. Conceived as a small-scale preliminary investigation, the primary objective was to assess pedagogical feasibility, evaluate classroom implementation, and identify descriptive learning trends. Given the focused sample size ($N = 13$) and the one-week duration, inferential statistical models were omitted in favor of detailed descriptive statistics, focusing on mean scores, accuracy rates, and percentage shifts.

Participants

A convenience sample of 13 Year 5 pupils ($N = 13$) was selected from a rural SJKT school in Malaysia. All participants share Tamil as their native language (L1), with foundational proficiencies in Bahasa Melayu (L2) and English (L3). The rural school setting contends with a recognized digital divide, characterized by limited home computer access and variable internet connectivity (Ravendaran & Nasri, 2025). To maintain consistency, the intervention was conducted entirely within the school's computer laboratory under teacher supervision across a five-day instructional week.

The Interactive Ebook and the ASSURE framework

The digital intervention, *The World on a Plate (And the Missing Nasi Lemak)*, was developed on the BookCreator platform using the ASSURE instructional design framework (Adedapo & Opoola, 2021; Eliana et al., 2024):

- **Analyze Learners:** The cohort comprised 13 trilingual L3 learners susceptible to cognitive overload when encountering foreign materials (Sani & Ismail, 2021). The ebook utilized familiar Malaysian cultural schemas to remove cultural decoding barriers.
- **State Standards and Objectives:** Objectives were aligned with Unit 5 (Food) of the Year 5 SBELC syllabus, targeting: (a) expressing dietary preferences ('I love...', 'I dislike...'), (b) distinguishing countable and uncountable nouns (much vs. many), (c) applying procedural sequencers (first, next, then, finally), and (d) demonstrating literal and inferential reading comprehension.
- **Select Strategies, Technology, Media, and Materials:** BookCreator was selected for its multimodal capabilities (Muhisom et al., 2023). AI-generated global voices provided clear, diverse accents, while a 360-degree Virtual Reality (VR) widget offered spatial immersion in local food culture (Cheng et al., 2025).
- **Utilize Technology, Media, and Materials:** Interface layouts adhered to CTML principles. Tactical translanguaging supported comprehension (Nur et al., 2022), and the VR module was isolated on a standalone page to prevent split-attention fatigue during reading tasks.
- **Require Learner Participation:** Interactive Wordwall mini-games, response polls, and a drag-and-drop lunch tray activity engaged learners actively (Alghasab & Handley, 2023). Pupils recorded a short spoken video explaining their meal selections using the target structures (Ngoi et al., 2024; Zhussupova & Shadiev, 2023).
- **Evaluate and Revise:** Learning outcomes were measured through embedded pre- and post-tests, supplemented by student feedback (Ibrahim et al., 2025).

Assessment Instruments, Validity, and Reliability

Data was collected using a uniform digital Google Forms assessment administered on Day 1 (pre-test) and Day 5 (post-test). The instrument comprised two sections: (1) Reading Comprehension and Vocabulary (5 multiple-choice items), and (2) Discrete Grammar and

Sequencing (10 items assessing much/many usage, preference sentence patterns, and procedural imperative sequencing). Content validity was established through alignment with the Malaysian Year 5 SBELC syllabus and confirmed via expert review by a senior TESL lecturer. Internal consistency testing of the pre-test instrument yielded a Cronbach's Alpha of $\alpha = 0.528$, which is deemed acceptable for a preliminary pilot study with a small sample size ($N = 13$) and concise item set.

Findings / Results

Descriptive analyses revealed substantial performance improvements across all assessed domains following the intervention. The cohort's overall composite mean score increased from a pre-test baseline of 5.15 out of 10 (51.5%) to 9.54 out of 10 (95.4%) on the post-test, representing an absolute gain of 43.9%. Table 1 provides a detailed summary of domain-specific gains.

Assessment Construct	Pre-test Mean Score	Pre-test Accuracy (%)	Post-test Mean Score	Post-test Accuracy (%)	Absolute Accuracy Shift (%)
Reading Comprehension & Vocabulary	3.15 (Max 5)	63.0%	3.77 (Max 4)	94.2%	+ 31.2%
Discrete Grammar (Countable / Uncountable)	0.85 (Max 2)	42.5%	2.00 (Max 2)	100.0%	+ 57.5%
Imperatives & Sequencing	1.15 (Max 3)	38.3%	3.77 (Max 4)	94.2%	+ 55.9%
Total Composite Score	5.15 (Max 10)	51.5%	9.54 (Max 10)	95.4%	+ 43.9%

Table 1: Comparative Analysis of Pre-Test and Post-Test Means (N=13)

Findings Related to Cognitive Processing

In reading comprehension, the cohort recorded a baseline pre-test mean of 3.15 out of 5 (63.0%). Following interaction with the ebook, post-test scores increased to a mean of 3.77 out of 4 (94.2%), reflecting a 31.2% net gain in accuracy. Replacing unfamiliar Western scenarios with relatable Malaysian cultural touchstones (e.g., school canteens and traditional markets) enabled pupils to comprehend the text without experiencing cultural decoding strain.

Findings Related to Short-Term Language Acquisition

The most substantial academic gains were observed in discrete grammar. For countable and uncountable nouns (much vs. many), the cohort's baseline accuracy rose from 42.5% (mean = 0.85/2) to 100% (mean = 2.00/2), marking a 57.5% improvement. Competence in procedural imperatives and chronological sequencers advanced from 38.3% (mean = 1.15/3) to 94.2% (mean = 3.77/4), a gain of 55.9%. Overall, 11 of the 13 participants (84.6%) achieved a perfect score of 10/10 on the post-test assessment.

Discussion

The empirical findings demonstrate that integrating Culturally Responsive Pedagogy with interactive multimedia design effectively mitigates the instructional friction associated with standard CEFR-aligned textbooks. The cohort's composite score growth (51.5% to 95.4%) indicates that supporting trilingual L3 learners requires both linguistic scaffolding and purposeful management of cognitive load.

Cognitive Load Mitigation through Cultural Schema Activation

Trilingual SJKT learners often experience working memory exhaustion when learning English, leading to L1 or L2 translation as a coping strategy (Azizah et al., 2024; Hasin & Nasir, 2021). Imported textbooks exacerbate this burden by presenting unfamiliar cultural settings (Sweller et al., 2011). The 31.2% improvement in reading comprehension demonstrates the benefit of culturally familiar contexts. Consistent with Schema Theory (Anderson & Pearson, 1984; Carrell & Eisterhold, 1983), the ebook's narrative activated learners' existing knowledge (Khan & Iqbal, 2025). Removing extraneous cultural decoding enabled pupils to dedicate working memory directly to syntactic parsing and vocabulary acquisition (Sudarsi et al., 2025; Xu, 2025).

Grammar Acquisition through Contextualized TBLT and Focus on Form

Achieving 100% mastery in countable and uncountable nouns supports the effectiveness of Task-Based Language Teaching combined with Focus on Form (Ellis, 2003; Long, 1991). Rather than presenting grammatical rules through isolated drills, the ebook embedded structures within meaningful activities, such as selecting ingredients for Nasi Lemak and assembling a lunch tray. Abstract grammar rules (much vs. many) became functional communicative assets necessary to complete the task, accelerating acquisition without overloading working memory (Sami et al., 2025; Xin et al., 2024).

Multimedia Scaffolding and Affective Engagement

The 55.9% gain in procedural imperatives aligns with Mayer's Cognitive Theory of Multimedia Learning (Mayer, 2002; Mayer & Moreno, 2003). Positioning text alongside step-by-step

cooking visuals (Spatial Contiguity) and segmenting steps into user-paced modules prevented split-attention fatigue (Februati & Dewa, 2025; Zhou, 2025). Additionally, 11 of the 13 students achieved perfect post-test scores, highlighting the positive affective influence of culturally responsive learning (Gay, 2013; Ladson-Billings, 1995). Incorporating tactical translanguaging and diverse accent varieties lowered the affective filter and encouraged task engagement (Boonsuk, 2024; Chaweewan & Boonsuk, 2025; Putera & Akid, 2022).

Contributions, Limitations, and Future Directions

Key Pedagogical Contributions

- **Instructional Scaffolding:** Demonstrates how localized digital narratives reduce extraneous cognitive load for trilingual L3 learners.
- **Form-Meaning Integration:** Illustrates the value of TBLT and Focus on Form in contextualizing abstract grammar rules within interactive tasks.
- **Practical Scalability:** Highlights cloud-based ebook platforms (such as BookCreator) as flexible, cost-effective alternatives to static imported textbooks (Lee & Ompok, 2025).

Limitations of the Study

- **Sample Size and Scope:** The small convenience sample (N = 13) from a single rural SJKT school restricts broad generalization across all Malaysian ESL contexts.
- **Assessment Horizon:** The evaluation measured immediate short-term recall; longitudinal follow-up is needed to verify long-term retention.
- **Potential Novelty Effect:** High engagement may partly reflect initial novelty toward interactive digital tools (Jeno et al., 2019).
- **Digital Divide:** The intervention required stable computer laboratory infrastructure, which may vary across rural schools (Ravendaran & Nasri, 2025).

Recommendations for Future Research and Practice

Future studies should expand to larger, multi-site cohorts across diverse vernacular (SJKT/SJKC) and national (SK) schools. Incorporating delayed post-tests will help assess long-term retention. Additionally, developing low-bandwidth and offline-accessible versions of digital learning modules will be critical to ensuring equitable access for resource-constrained schools.

Conclusion

This preliminary study highlights the instructional and cognitive challenges created when trilingual vernacular primary pupils are taught using culturally unfamiliar, standardized ESL textbooks. Implementing *The World on a Plate* (And the Missing Nasi Lemak) demonstrates that integrating Culturally Responsive Pedagogy with Task-Based Language Teaching and CTML-guided multimedia effectively mitigates extraneous cognitive load. The intervention produced an overall accuracy gain of 43.9% and complete mastery of target grammar forms. While sustainable scaling requires addressing digital infrastructure disparities, these findings underscore the value of localized, multimodal learning tools in supporting inclusive English language acquisition in Malaysian vernacular classrooms.

References

- Adedapo, A., & Opoola, B. T. (2021). Levels of integrating the ASSURE model in lesson delivery of selected primary school teachers in Nigeria. *Journal of Language Teaching and Research*, 12(1), 177–187. <https://doi.org/10.17507/jltr.1201.19>
- Alghasab, M., & Handley, Z. (2023). The role of task-based interaction in digital language learning environments: Fostering grammatical accuracy and learner autonomy. *System*, 114, 103012. <https://doi.org/10.1016/j.system.2023.103012>
- Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255–291). Longman.
- Azizah, N., Karim, A., & Ahmad, S. (2024). First language interference and translanguaging strategies among young multilingual learners. *Asian Journal of English Language Teaching*, 34(1), 45–63.
- Azlan, N. D. A., & Kaur, J. (2023). Malaysian ESL teachers' perspectives on curriculum shift to Global Englishes language teaching. *The English Teacher*, 52(3), 97–109. <https://doi.org/10.52696/mdub4967>
- Biang, W. S. (2025). Improving multilingual education in Malaysia: Tailoring language teaching for SJKC schools. *International Journal of Research and Innovation in Social Science*, 9(4), 679–694. <https://doi.org/10.47772/IJRISS.2025.90400052>
- Boonsuk, Y. (2024). “Teacher, I don’t know how to describe ‘Tom Yum Kung’ in English, but I know ‘hamburger’ and ‘pizza’:” Will Global Englishes language teaching be a beacon of hope or a recipe for disaster in Thai ELT? *International Journal of Applied Linguistics*, 35(2), 885–898. <https://doi.org/10.1111/ijal.12667>
- Bukroni, A., Rozelin, D., & Tartila, T. (2026). Schema activation in EFL reading: Implementation and perspectives from students and teacher. *Eltin: Journal of English Language Teaching in Indonesia*, 14(1), 13–28. <https://doi.org/10.22460/eltin.v14i1.p13-28>
- Carrell, P. L., & Eisterhold, J. C. (1983). Schema theory and ESL reading pedagogy. *TESOL Quarterly*, 17(4), 553–573. <https://doi.org/10.2307/3586613>
- Chaweewan, T., & Boonsuk, Y. (2025). When English is no longer singular: Asian students' journey to English plurality in Thai higher education. *Language, Culture and Curriculum*, 38(3), 361–379. <https://doi.org/10.1080/07908318.2025.2484734>
- Cheng, L., Hidayat, R., & Mahat, J. (2025). Immersive and interactive learning: A systematic review of virtual reality in primary education. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(16), 163–188. <https://doi.org/10.3991/ijim.v19i16.54901>
- Eliana, N., Wati, U. A., & Rahmadona, S. (2024). Leveraging the ASSURE model for optimized information technology-based learning media. *AL-ISHLAH: Jurnal Pendidikan*, 16(3), 3120–3131. <https://doi.org/10.35445/alishlah.v16i3.5639>

- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Februati, B. M. N., & Dewa, D. M. R. T. (2025). Case study: Multimodal approach as an alternative for effective English learning and teaching in Polytechnic of Energy and Mineral. *Jurnal Locus Penelitian Dan Pengabdian*, 4(10), 9625–9638. <https://doi.org/10.58344/locus.v4i10.4922>
- Gay, G. (2013). Teaching to and through cultural diversity. *Curriculum Inquiry*, 43(1), 48–70. <https://doi.org/10.1111/curi.12002>
- Hasin, I., & Nasir, M. K. M. (2021). The effectiveness of the use of information and communication technology (ICT) in rural secondary schools in Malaysia. *Journal of Education and E-Learning Research*, 8(1), 59–64. <https://doi.org/10.20448/journal.509.2021.81.59.64>
- Ibrahim, N. A., Harun, R. N. S. R., & Yunus, M. M. (2025). Gamified AI for EFL: A model to support engagement, formative evaluation, and communicative competence. *Higher Education Studies*, 15(4), 551–564. <https://doi.org/10.5539/hes.v15n4p551>
- Jeno, L. M., Vandvik, V., Eliassen, S., & Grytnes, J.-A. (2019). Testing the novelty effect of an m-learning tool on internalization and achievement: A Self-Determination Theory approach. *Computers & Education*, 128, 398–413. <https://doi.org/10.1016/j.compedu.2018.10.008>
- Khan, Z., & Iqbal, M. (2025). Cultural schemata in EFL reading: Instructor perspectives and pedagogical implications. *Journal of Applied Linguistics and TESOL (JALT)*, 8(3), 836–852. <https://doi.org/10.63878/jalt1048>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- 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>
- Lee, B. N., & Ompok, C. (2025). *Cultural relevance in digital learning materials: Scaffolding early language development*. Universiti Malaysia Sabah Press.
- Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspectives* (pp. 39–52). John Benjamins Publishing.
- Maamuujav, U. (2025). Multimodal writing in multilingual space. *Education Sciences*, 15(11), 1446. <https://doi.org/10.3390/educsci15111446>
- Matiso, N. H. (2024). Optimising culturally responsive pedagogies in multicultural English second language classrooms. *International Journal of Learning, Teaching and Educational Research*, 23(11), 384–401. <https://doi.org/10.26803/ijlter.23.11.20>
- Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation*, 41, 85–139. [https://doi.org/10.1016/S0079-7421\(02\)80005-6](https://doi.org/10.1016/S0079-7421(02)80005-6)
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52. https://doi.org/10.1207/s15326985ep3801_6

- Muhisom, Destiani, & Pangestu, D. (2023). Utilizing Book Creator as a digital module to improve student listening skills in elementary school. *Advances in Social Science, Education and Humanities Research*, 764, 355–361. https://doi.org/10.2991/978-2-38476-046-6_35
- Ngoi, S., Tan, K. H., Alias, J., & Mat, N. (2024). Digital storytelling to improve English narrative writing skills. *International Journal of Academic Research in Business and Social Sciences*, 14(4), 112–126.
- Nur, S., Yusri, H., Huzaimi, A., & Sulaiman, N. (2022). Translanguaging pedagogies in primary ESL environments. *International Journal of Academic Research in Business and Social Sciences*, 12(9), 607–619. <https://doi.org/10.6007/IJARBS/v12-i9/14542>
- Prasetya, R. E. (2024). Competency-based English language teaching through Microsoft Teams: A framework for adaptation and integration. *English Learning Innovation*, 5(2), 298–315. <https://doi.org/10.22219/englie.v5i2.33323>
- Putera, Z. F., & Akid, A. (2022). Teaching culturally responsive pedagogy in multilingual classrooms: Valuing linguistic and cultural diversity. *Transformational Language Literature and Technology Overview in Learning (Transtool)*, 1(4), 19–24. <https://doi.org/10.55047/transtool.v1i4.1366>
- Radzi, A. H. M., & Radzi, A. H. M. (2024). The influence of content schemata on ESL learners' comprehension of 'The Hunted Fox' and 'The Breakup': ODL experience. *Asian Journal of University Education*, 20(2), 180–194.
- Ravendaran, S., & Nasri, N. M. (2025). Challenges in the implementation of digital learning in primary school. *International Journal of Research and Innovation in Social Science*, 9(9), 504–516.
- Sami, A., Raza, S., & Chen, Y. (2025). Implications of task-based learning and cognitive theory on language learning. *Journal of Applied Linguistics and TESOL (JALT)*, 8(2), 1446–1459.
- Sani, S., & Ismail, H. H. (2021). Assessing the use of learning strategies among young Malaysian English as second language learners. *Creative Education*, 12(9), 2124–2145. <https://doi.org/10.4236/ce.2021.129163>
- Sendera Mohd. Yakin, H., Abu Bakar, A. L., & Kassim, N. (2025). The model of cultivation, innovation, and marketing module development: A culturally responsive framework. *Global Research and Development Services Journal*, 8(1), 14–28.
- Sudarsi, E. T., Mangoki, D., Wahyuni, S., Allo, M. D. G., & Irianto, T. U. (2025). Integrating local culture-based learning media in multilingual classrooms: An analytical study of implementation and educational impact. *Eduvelop: Journal of English Education and Development*, 9(1), 61–78.
- Suzuki, S., Hato, Y., & Miller, J. (2025). Incorporating diverse accent varieties in computer-assisted language learning. *Language Teaching Research*, 29(2), 210–231.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer Science & Business Media. <https://doi.org/10.1007/978-1-4419-8126-4>

- Takal, G. M., Planisamay, T., Ibrahim Gani, N. A. B., Peter, C. V., & Rahmat, A. (2021). An evaluation of 'Get Smart' English language textbook from the perspectives of English teachers. *Aksara: Jurnal Ilmu Pendidikan Nonformal*, 7(2), 727–746. <https://doi.org/10.37905/aksara.7.2.727-746.2021>
- Xin, W., Zhang, Q., & Luo, X. (2024). Promoting language acquisition and cross-cultural competence in multilingual environments: Effective teaching strategies and educational support. *Advances in Humanities Research*, 9(1), 49–53. <https://doi.org/10.54254/2753-7080/9/2024107>
- Xu, Y. (2025). Cognitive strategies for developing intercultural competence in English language reading: Challenges and interventions. *Cognitive Strategies in Study*, 1(2), 1–12. <https://doi.org/10.63808/css.v1i2.139>
- Ya Shak, M. S., Mohd Naw, A. H., Mohd Radzuan, N. R., Che Wan Ahmad, C. N. F., & Mohd Tahir, M. (2021). Revisiting postmethod pedagogy: Analyzing Kumaravadivelu's macrostrategies in the Super Minds Year 1 & 2 English textbook. *International Journal of Academic Research in Business and Social Sciences*, 11(9), 1675–1689.
- Zhou, Y. (2025). Enhancing listening comprehension: A multimodal approach grounded in cognitive load and coding theories. *Lecture Notes in Education Psychology and Public Media*, 111(1), 1–7. <https://doi.org/10.54254/2753-7048/2025.nd26257>
- Zhussupova, R., & Shadiev, R. (2023). Digital storytelling to facilitate academic public speaking skills: Case study in culturally diverse multilingual classroom. *Journal of Computers in Education*, 10(3), 499–526. <https://doi.org/10.1007/s40692-023-00259-x>

Every Voice Matters: Preparing Learners for AI-Supported Speaking Through the Microphone Effect

Dr. Wan Afiqah Anis binti Wan Ahmad *, Universiti Malaya, Malaysia,
afiqahanis@um.edu.my

Dr. Mohamad Azrien bin Mohamed Adnan, Universiti Malaya, Malaysia,
mdazrien@um.edu.my

Abstract: Artificial intelligence (AI) can provide language learners with abundant, patient, and ostensibly judgement-free opportunities to practise speaking. Yet access does not guarantee speech. This conceptual and practice-based paper examines a pedagogical problem observed among foundation students at Universiti Malaya Education Centre, Bachok where learners who could comprehend and write in English still froze when required to speak into a microphone or to an AI system. The paper introduces the Microphone Effect, defined as the psychological threshold between passive language knowledge and audible participation, and proposes a graduated intervention comprising mic drop, mic mates and mic on. Across these phases, generative AI supports idea generation and task planning, while spontaneous delivery, peer interaction, drama and reflection remain human-led. Illustrative learner reflections suggest a possible movement from perfection-oriented silence towards expression, audience awareness, and ownership of voice. These observations are synthesised in the V.O.I.C.E. framework: Validate anxiety, Orient learners to the AI environment, use Incremental speaking tasks, provide Conversational roleplay and Evaluate and reflect. The paper argues that effective AI-enhanced speaking pedagogy should prepare learners psychologically and communicatively before expecting independent human-AI oral interaction.

Keywords: artificial intelligence, language anxiety, oral communication, speaking pedagogy, voice

Introduction

Generative artificial intelligence (GenAI) has altered the practical conditions of second-language learning. A learner with an internet-connected device can now request explanations, examples, prompts, feedback, roleplays, and conversational practice at almost any time. For speaking instruction in particular, this availability appears to resolve a long-standing logistical problem: in a large class, no teacher can provide every learner with sustained one-to-one oral practice. AI-supported tools, by contrast, offer repeated interaction without fatigue and can respond immediately. Recent reviews associate chatbot-supported language learning with cognitive outcomes such as communicative competence and with emotional outcomes such as confidence, willingness to communicate, enjoyment, and reduced nervousness (Li et al., 2025; Lo et al., 2024).

However, technological availability should not be confused with learner readiness. A system may be ready to listen while a learner remains unable to speak. This distinction is especially important for learners whose receptive knowledge or written English masks substantial verbal anxiety. In the context presented in this paper, foundation students at Universiti Malaya Education Centre (UMEC), Bachok, during the 2025/2026 session achieved an average Malaysian University English Test (MUET) Band of 3.5. They were described as capable in comprehension and writing but heavily restricted by verbal anxiety. One learner captured the discrepancy directly: “I am one of the students that excel in English, well in WRITING via text message or letter... but when it was time to speak about it, I froze.” The

problem, therefore, was not simply a lack of linguistic content. It was a barrier between knowing and vocalising.

Traditional classroom speaking and AI-mediated speaking create overlapping but different pressures. In human-to-human interaction, learners may fear peer judgement, teacher correction, interruption, or loss of face. AI removes much of this social exposure, but speaking to a machine may produce a different form of scrutiny. The microphone, recording indicator, or silent interface can heighten awareness of one's own voice, pronunciation, pauses, and errors. What feels technically safe may nevertheless feel psychologically sterile and permanent. Anxiety can shift from social judgement to performance perfectionism: the learner is no longer primarily asking, "What will others think of me?" but "How do I sound, and must this recording be perfect?"

This paper calls that threshold the Microphone Effect: a proposed psychological barrier between available language knowledge and audible participation, particularly when microphone use or an AI interface heightens self-monitoring. It is a descriptive conceptual lens, not a clinically or psychometrically validated construct. The psychological phenomenon and the pedagogical response are distinct: the MIC Series, MIC DROP, MIC MATES and MIC ON that provides a graduated instructional response, supported by the V.O.I.C.E. framework. AI scaffolds preparation while learners retain responsibility for speaking and teachers retain responsibility for pedagogical judgement.

Accordingly, the paper has three purposes. First, it explains why judgement-free AI access may still fail to elicit speech. Second, it articulates the design logic of the three-phase MIC Series. Third, it synthesises the design into an actionable framework for English language educators. The guiding questions are: (1) How does the Microphone Effect explain silence during AI-supported speaking? (2) How do the three MIC phases scaffold movement from preparation to spontaneous and embodied speech? and (3) What principles should guide psychologically responsive AI-enhanced speaking instruction?

Literature Review

Foreign language speaking anxiety is not reducible to limited grammar or vocabulary. Oral production unfolds in real time and exposes learners to evaluation while they must retrieve language, organise meaning, monitor accuracy, and respond to an interlocutor. These concurrent demands help explain why learners who perform adequately in receptive or written tasks may struggle to speak. Anxiety can consume attentional resources and intensify monitoring at precisely the point when automaticity and message-focused interaction are needed. Within the presentation underpinning this paper, the resulting pattern is represented as a "barrier of silence": comprehension and writing competence remain available, but audible participation is interrupted.

AI-supported conversation appears well positioned to reduce this barrier because it offers privacy, repetition, immediate response, and freedom from direct human ridicule. Research increasingly supports its value. Fathi et al. (2024) reported improvement in speaking skills and willingness to communicate through AI-mediated interaction. Wang et al. (2024) examined conversational GenAI chatbots in relation to willingness to communicate, foreign language speaking anxiety, and perceived communicative competence. A systematic review by Huang et al. (2022) similarly found that chatbots can extend language practice, although their pedagogical usefulness depends on how learners and teachers use them. More recent evidence has strengthened the design argument: Li et al.'s (2026) meta-analysis of 25 GenAI-supported task-based studies found a positive overall effect on language performance and identified AI literacy, task planning, and the assessor as meaningful moderators.

These findings challenge a technologically deterministic assumption that providing a tool will itself produce interaction. ChatGPT and comparable systems are general-purpose

applications rather than language pedagogies. Their responses may be too complex, inaccurate, generic, or poorly calibrated to a learner's proficiency. Learners also need to know how to initiate, sustain, redirect, and critically evaluate interaction. Reviews therefore emphasise the continued importance of teachers, instructional rules, peer collaboration, and task design (Ji et al., 2022; Li et al., 2024; Li et al., 2025). In activity-theoretical terms, the tool is only one component of an instructional system; outcomes emerge from the relationship among the learner, task, community, rules, and division of labour.

The affective promise of AI also requires qualification. Removing human judgement does not remove self-judgement. In human conversation, pacing is fluid, repair can be negotiated, and feedback may be conveyed through nuanced verbal and non-verbal cues. Human-AI interaction can instead feel disjointed, turn-based, and explicit. A learner may interpret automated transcription, corrective feedback, or the recording interface as a clinical assessment rather than a conversation. The presentation describes a cognitive-load trajectory in which anticipation rises towards a "recording spike," followed by an initial AI response and, only with acclimation, a more fluent state. Although this trajectory remains a conceptual representation rather than a measured curve, it highlights a plausible instructional issue: the first moments of speaking to AI may be the most demanding and may require deliberate preparation.

Task planning provides one such preparation mechanism. GenAI can help learners generate keywords, organise content, model an outline, simplify vocabulary, and rehearse possible turns. Used well, this support can reduce the burden of inventing content while speaking. Used poorly, it can encourage learners to read AI-generated scripts verbatim, outsource decisions, or imitate language they do not own. The relevant pedagogical distinction is therefore between AI-supported preparation and AI-substituted performance. The former preserves the learner as speaker; the latter may produce polished text without communicative agency. Kohnke et al. (2023) similarly argue that ChatGPT's value in language teaching depends on principled integration and critical awareness rather than novelty alone.

Peer interaction supplies another missing dimension. Much chatbot research concentrates on individual learner-tool exchanges, although collaboration can strengthen emotional, behavioural, and cognitive outcomes (Li et al., 2025). A known partner can provide reassurance, co-construct meaning, model repair, and make mistakes feel socially survivable. Structured progression from script to bullet points to unscripted conversation also resembles graduated scaffolding: support is strongest at the beginning and deliberately withdrawn as control transfers to learners. Roleplay and drama extend this logic by allowing learners to speak through a character. The temporary distance between self and role can reduce self-consciousness while requiring embodied choices about timing, expression, and audience.

The literature thus supports a human-centred position on AI-enhanced language education. AI can expand practice and preparation, but emotional safety, task sequencing, collaboration, reflection, and teacher mediation determine whether availability becomes participation. UNESCO (2023) likewise recommends a human-centred approach to GenAI in education, including protection of human agency, critical use, and pedagogical appropriateness. The Microphone Effect and MIC Series contribute to this discussion by focusing on the moment immediately before and during vocal interaction, the point at which a technically capable learner may still remain silent.

Methodology

Research Design

This is a conceptual, practice-based pedagogical paper rather than an experimental effectiveness study. It develops an explanatory construct and instructional framework from the classroom problem, intervention sequence, and learner evidence documented in the conference presentation "Every Voice Matters: Preparing Learners for AI-Supported Speaking Through

the Microphone Effect.” The presentation was treated as the primary evidential record. Claims not contained in that record such as participant number, demographic distribution, instrument reliability, statistical significance, or causal effect were not inferred.

Context and Evidence Base

The pedagogical context comprised foundation students at UMEC Bachok in the 2025/2026 academic session. The presentation reports an average MUET Band of 3.5 and characterises learners as capable in comprehension and writing while constrained by verbal anxiety. The evidence base consists of the stated classroom problem, a comparison of human-to-human and human-to-AI speaking, the three instructional phases, the V.O.I.C.E. synthesis, and four learner reflection excerpts reproduced in the presentation. Because the slides do not document a formal sampling or elicitation procedure, the excerpts are treated as illustrative evidence of learner experience, not as a representative qualitative dataset.

Conceptual Development and Analysis

The analysis proceeded through pedagogical synthesis. First, the documented barrier was defined by contrasting linguistic capability with vocal participation. Second, the presentation’s human-to-human and human-to-AI comparison was examined to identify the changing source of anxiety. Third, the MIC phases were mapped according to the type of support provided, the learner’s degree of spontaneity, the social arrangement, the role of AI, and the intended confidence transition. Fourth, the learner excerpts were read for convergent meanings rather than coded as a formal thematic corpus. The recurring meanings, freezing despite competence, prioritising expression over perfection, increased audience awareness, and ownership of voice were then related to the intervention logic. Finally, these elements were synthesised into the V.O.I.C.E. framework and interpreted against recent scholarship on AI chatbots, task planning, speaking anxiety, collaboration, and human agency.

The mapping separates three levels of interpretation: the classroom descriptions and quotations constitute the source material; the Microphone Effect offers a proposed explanation of the participation barrier; and MIC Series and V.O.I.C.E. specify an instructional response. The four excerpts are considered separately as illustrations of freezing, tolerance of imperfection, audience awareness and ownership of voice. Their alignment with the framework is interpretive and does not establish frequency, temporal change or agreement across the cohort. Alternative explanations include task familiarity, peer support, proficiency differences and the novelty of the activity. The proposed implementation and evaluation below extend the design argument; they are not additional procedures or outcomes reported in the original presentation.

Ethical and Interpretive Boundaries

No personally identifying learner information appears in the presentation or this paper. Quotations are anonymous and reproduced only as presented. The analysis does not claim clinical diagnosis, population prevalence, measured anxiety reduction, or statistically established learning gain. Its contribution is design-oriented: it makes explicit the pedagogical reasoning embedded in the MIC Series and derives propositions that can be tested in subsequent research.

Conceptual Findings

The pedagogical synthesis yields four interrelated propositions and a framework synthesis. These organise the practice account and illustrative reflections; they do not estimate intervention effects.

The Microphone Effect Frames a Gap Between Knowledge and Vocal Participation

The first proposition is that silence may persist when learners have usable English knowledge. The reported cohort average of MUET Band 3.5 provides context, but an aggregate

score cannot establish any individual learner’s speaking competence or rule out proficiency-related difficulty. The learner who described success in writing but “froze” when speaking illustrates a perceived gap between available knowledge and vocal participation. In the proposed account, a microphone may make speech feel inspectable and potentially permanent. Tolerating imperfect, audible production is therefore a plausible instructional target, alongside continued development of language proficiency.

AI Changes Rather Than Automatically Removes Anxiety

The second proposition concerns the potentially mixed psychological experience of speaking to a machine. AI may reduce exposure to peer embarrassment and direct teacher correction, while recording, transcription and automated feedback may amplify self-monitoring. These pressures can coexist rather than replace one another. The presentation’s cognitive-load trajectory depicts a hypothesised recording spike before acclimation; it does not measure when pressure peaks or show that every learner experiences such a spike. This account offers one possible explanation for difficulty initiating AI-supported speech, to be tested against alternatives such as unfamiliarity with the interface or difficulty formulating content.

Three Graduated Phases Structure the Transition From Preparation to Voice

The third proposition concerns instructional sequence. The MIC Series distributes support across three phases rather than demanding immediate independent AI conversation. Table 1 summarises this progression.

Table 1. Instructional Progression Across the MIC Series

Phase	Arrangement	AI contribution	Human performance focus
MIC DROP	Individual	Prompted talking points or outline	Prepared but unscripted delivery
MIC MATES	Pair	Collaborative content and structure	Progression from script to bullet points to free conversation
MIC ON	Group of 6–7	Script refinement	Embodied roleplay, interpretation, interaction, and audience engagement

MIC DROP establishes prepared spontaneity. Learners use a specific GenAI prompt to obtain structured talking points or an outline. They then speak from those points without reading a generated script verbatim. AI reduces the content-planning burden, but the learner remains responsible for selection, wording, and delivery. The aim is a manageable first success: an effective “mic drop” moment that associates preparation with confident expression.

MIC MATES adds peer scaffolding through a scripted-to-spontaneous ladder. A pair may begin with a pre-written AI-assisted text, progress to guided bullet points, and then hold an unscripted conversation. The partner provides familiarity, feedback, and mutual support, while AI assists collaborative content generation and organisation. The phase is not merely more practice; it changes the social conditions of risk by making spontaneity shared and supported.

MIC ON moves to larger-group drama involving approximately six or seven learners. AI may help refine a script, but human interpretation, characterisation, interaction, and performance dominate. Adopting a role—for example, a familiar fictional character allows learners to step temporarily outside their habitual self-consciousness. The phase integrates individual preparation and paired interaction into an embodied, memorable communicative event.

Learner Reflections Illustrate a Possible Reorientation From Perfection to Expression

The fourth proposition concerns a possible attitudinal reorientation illustrated by the reflections. One learner stated that being brave enough to speak “broken English” was more valuable for personal growth than remaining silent to appear perfect. Another declared,

“Finding my ‘voice’ is way more important than being perfect.” A third reported increased awareness of planning, delivery, and interaction with the audience to make a presentation memorable. Read together, the excerpts foreground three dimensions: tolerance of linguistic imperfection, ownership of voice, and attention to communicative impact. These are not formal outcome measures, but they align with the intervention’s intended shift from self-protection to expression.

The V.O.I.C.E. Framework Synthesises the Intervention

The phases are consolidated in V.O.I.C.E.: Validate the anxiety; Orient learners to the AI environment; use Incremental speaking tasks; provide Conversational roleplay; and Evaluate and reflect. Validation prevents silence from being misread as laziness or inability. Orientation demystifies the interface, modelling prompts, turn-taking, privacy choices, and the fallibility of AI feedback. Incremental tasks lower the entry demand and gradually withdraw scripts. Conversational roleplay connects machine-supported preparation to meaningful human interaction. Evaluation and reflection return AI experiences to the classroom, where learners and teachers examine what helped, what felt difficult, and what should change. The framework’s central result is thus a design proposition: psychological preparation should be designed into AI speaking practice.

Practical Implementation Example

The following is a proposed adaptation for three 50-minute lessons on “Helping new foundation students settle into campus life”; these timings and prompts are illustrative, not a reconstruction of the original implementation. Before the sequence, the teacher models one imperfect but intelligible response, explains permitted AI use, demonstrates microphone controls and checks access. Learners should use institutionally permitted tools without entering personal information. A learner who needs more support may begin with private rehearsal and repeat a step before progressing.

Lesson 1—MIC DROP: Allocate 10 minutes to orientation, 10 to planning, 20 to short speaking turns and 10 to reflection. A sample planning prompt is: “Give me three short talking points in accessible English about helping a new student feel welcome. Do not write a complete speech.” Each learner checks and adapts the suggestions, selects up to five keywords, and delivers a 60–90-second response to a partner. The partner asks one follow-up question. Feedback identifies one clear message and one next step; a second attempt can follow.

Lesson 2—MIC MATES: Use 5 minutes to review the goal, 10 to prepare a short exchange, 25 for paired practice and 10 for feedback and reflection. One learner plays a new student seeking advice and the other a student mentor. Pairs first rehearse an AI-assisted model that they have checked, then retain keywords only, and finally repeat the exchange without a script. They swap roles and introduce one unprepared question. A useful repair phrase is “Could you explain what you mean?” The teacher checks that both partners contribute and allows another keyword-supported attempt where needed.

Lesson 3—MIC ON: Allocate 10 minutes to group planning, 15 to rehearsal, 20 to short performances in parallel groups or to the class, and 5 to reflection. Groups of six or seven create a campus welcome scenario in which every character has a speaking responsibility. AI may suggest a complication or refine the outline, but learners decide how to enact it. After each performance, an audience member asks an unprepared question answered in character. The final reflection asks what made starting easier or harder, which AI suggestion was changed, and what the learner will try next.

Across the lessons, V.O.I.C.E. guides teacher action: validate discomfort without diagnosing it; orient learners to the tool and task; reduce support incrementally; use conversational roleplay; and evaluate through performance and reflection. Suggested progression checks are completing a brief message with keywords, answering a partner’s follow-up, and contributing a relevant turn in group interaction. These are teaching checks, not

validated cut-offs or conditions for penalising an anxious learner. A simple observation sheet can record task completion, support used, response relevance and one learner-selected next step.

Discussion

The paper's central argument is that the principal barrier in AI-supported speaking may occur before sustained conversation begins. This reframes the common affordance claim that AI provides unlimited, judgement-free practice. Unlimited opportunity has little educational value when learners cannot initiate speech. The Microphone Effect makes initiation itself a legitimate object of pedagogy. It also explains why apparently safe technology can provoke discomfort: the absence of a human evaluator removes one threat, but the interface can intensify self-evaluation.

This interpretation extends affective accounts of language learning by differentiating social and performance-oriented anxiety. In a classroom, the learner may anticipate negative evaluation from peers or teachers. In machine interaction, attention may turn inward towards pronunciation, fluency, and the permanent trace of a recording. The distinction should not be overstated as a universal binary; learners can experience both sources simultaneously. Nevertheless, it offers teachers a sharper diagnostic question. Rather than asking only whether students fear people, teachers can ask which features of the interface—silence, recording, transcription, correction, or unfamiliar turn-taking—make speaking feel evaluative.

The MIC Series addresses this problem through calibrated transfer of responsibility. MIC DROP reduces conceptual load with AI-generated prompts but preserves unscripted delivery. MIC MATES adds relational safety and progressively removes textual support. MIC ON places AI in the background while foregrounding collective, embodied communication. This movement is significant because it prevents AI from becoming the final audience. The educational goal is not fluent conversation with a machine as an end in itself, but stronger participation with people. AI supports rehearsal and organisation; communicative agency remains human.

This division of labour accords with current research. Li et al. (2026) found stronger GenAI-supported language outcomes where learners had AI literacy and task-planning support, and where human teachers or researchers assessed performance. Li et al. (2025) similarly describe chatbot-supported outcomes as products of an activity system rather than of the tool alone. The MIC Series operationalises these insights at lesson level: orientation develops practical AI literacy, prompts structure planning, peers form a community, teachers define rules and evaluate performance, and learners retain responsibility for expression.

The learner reflections also suggest that confidence should be conceptualised as willingness to act under imperfection, not the elimination of error. “Broken English” is not celebrated as an instructional endpoint; rather, temporary imperfection is accepted as the cost of participation and growth. This stance is especially relevant to oral fluency because excessive online monitoring can interrupt message construction. By shifting the success criterion from flawless form to meaningful expression, the intervention may release attention for audience, interaction, and purpose. The reflection about making a presentation memorable indicates precisely this outward movement—from monitoring the self to influencing listeners.

The framework additionally contributes a safeguard against overreliance on generated language. Each phase specifies where AI assistance should stop. In MIC DROP, the learner does not read verbatim. In MIC MATES, text is reduced to bullet points and then removed. In MIC ON, AI may refine a script, but performance requires interpretation and interaction. These boundaries answer an important concern in GenAI-supported language learning: polished output can conceal weak ownership, limited critical evaluation, or reduced spontaneity. The framework treats fading support as a design requirement rather than an optional extension.

For language assessment, the paper implies that educators should distinguish support used during preparation from performance evidence. If AI generates an outline, assessment can still focus on how learners select ideas, adapt wording, respond, deliver, and engage an audience. Transparent task rules are essential. Teachers should specify whether AI may generate topics, questions, vocabulary, bullet points, or sample responses; whether scripts are permitted; and which elements must be spontaneous. This preserves construct relevance while recognising that responsible tool use is becoming part of contemporary communicative preparation.

The intervention is especially relevant in large classes where individual speaking time is scarce. AI can distribute preparation and low-stakes rehearsal beyond teacher-led turns, while pairs and groups multiply interaction. Yet scale should not erase care. Validation, orientation, and reflection require a classroom climate in which mistakes are treated as evidence of engagement. The teacher's role changes from sole interlocutor to designer of conditions for voice: selecting attainable tasks, modelling critical AI use, watching for avoidance or dependence, and reconnecting private practice to public communication.

Finally, the V.O.I.C.E. framework advances a human-centred conception of AI-enhanced learning. "Human-centred" does not mean rejecting automation; it means evaluating automation by whether it expands learner agency. Within this framework, AI does not replace the human voice. It gives learners a structured route towards finding, testing, and eventually owning that voice. The intended transformation is therefore not from human instruction to machine instruction, but from barrier to bridge.

Conclusion

AI-supported speaking offers extraordinary access, but access alone cannot solve the silence produced by anxiety, perfectionism, and unfamiliar interaction. The Microphone Effect names the threshold at which learners must convert language knowledge into audible participation. The practice account from UMEC Bachok illustrates why this proposed threshold merits investigation: learners may comprehend and write successfully yet freeze when speaking, and machine interaction may heighten self-monitoring even when direct social judgement is reduced.

The MIC Series responds through a deliberate progression. MIC DROP combines AI-prompted planning with spontaneous individual delivery. MIC MATES uses peer collaboration and fading textual support to bridge scripted and unscripted speech. MIC ON shifts AI to a supporting role within group drama and embodied communication. V.O.I.C.E. makes the pedagogy portable by requiring teachers to Validate anxiety, Orient learners, use Incremental tasks, provide Conversational roleplay, and Evaluate and reflect. The paper proposes that educators build psychological readiness into AI-supported speaking through supported initiation, gradually reduced scaffolding and opportunities for reflection. Whether this sequence improves participation beyond other forms of rehearsal remains an empirical question.

Recommendations

English language educators should begin AI-speaking activities with explicit emotional and technical orientation. Learners should see the interface demonstrated, understand what is recorded or stored, practise stopping and restarting, and be reminded that automated feedback can be incomplete or wrong. Initial prompts should generate keywords or short outlines appropriate to proficiency rather than sophisticated scripts.

Task sequences should fade support visibly. Teachers can move from model text to bullet points, from individual rehearsal to a familiar partner, and from rehearsed turns to unscripted interaction. A short reflection after each phase should ask what reduced anxiety, what increased cognitive load, which AI suggestions were accepted or rejected, and what the learner will attempt next. Assessment rubrics should reward intelligibility, responsiveness, message development, and audience engagement alongside accuracy.

Institutions should provide guidance on ethical and responsible GenAI use, including privacy, disclosure, verification, bias, and academic integrity. Teachers also need professional development in prompt design and in recognising when AI use supports planning versus replacing learning.

Proposed Empirical Evaluation

A future mixed-method pilot could first test feasibility over four teaching weeks, with baseline tasks before the MIC sequence, post-tasks immediately afterwards and a delayed follow-up two weeks later. Recruit consenting learners from comparable foundation classes and document enrolment, prior proficiency, AI experience, attendance and attrition. If feasible, allocate classes to MIC Series with AI-supported preparation, individual AI-supported rehearsal, or human-only rehearsal. Match topic difficulty, practice time and teacher contact. If random allocation is not possible, report baseline differences and retain appropriate limits on causal interpretation.

The primary behavioural outcome could be successful speech initiation, defined in advance as beginning an intelligible, task-relevant utterance within a standardised response window. Record initiation latency from the speaking cue to that utterance, non-initiation separately, task completion, support used and relevant responses to follow-up questions. Technical delays should be recorded separately. Parallel unfamiliar topics can reduce simple memorisation effects. Assess both a brief AI-mediated task and a human-interaction task to distinguish interface-specific readiness from transfer to interpersonal speaking.

To examine the Microphone Effect more specifically, counterbalance comparable tasks involving a familiar partner and an AI voice interface, document recording conditions and obtain a brief task-specific anxiety rating before and after each task. Combine these ratings with validated speaking-anxiety and willingness-to-communicate measures appropriate to the language and population. Interviews should explore whether learners attribute difficulty to evaluation, recording, language formulation or technical unfamiliarity. Any new items about microphone-related self-monitoring would require separate content review, piloting and psychometric evaluation before being treated as a scale. This triangulation could test whether the proposed lens adds explanatory value beyond general speaking anxiety.

An implementation checklist can document whether validation and orientation occurred, keywords replaced scripts, both partners contributed, every group member spoke, and reflection followed each phase. Teacher logs should record deviations, task duration and the AI tool and version used. Two trained raters, unaware of condition where feasible, could assess recordings for intelligibility, message development, responsiveness and audience engagement, with agreement reported. Analyse change and uncertainty using methods suited to repeated observations and class clustering; determine sample size prospectively for the primary outcome. A small pilot should prioritise feasibility and measurement quality over definitive effectiveness claims. Interviews and reflections can help explain contrasting participation patterns, including cases that do not fit the proposed account.

Future data collection should obtain institutional ethics approval and informed consent, separate research participation from course grades, and specify permission, access, retention and deletion arrangements for recordings. An equivalent learning activity should remain available to students who decline research participation. Improved participation without reduced anxiety, gains shared by every condition, or persistent interface-specific difficulty would each require a different interpretation. The study should therefore evaluate the framework rather than assume its benefit.

Limitations

This paper is limited by the evidence available in the conference presentation. The participant number, age and gender distribution, selection process, duration of implementation,

exact AI prompts, data-collection procedure, and ethics approval were not reported. The learner excerpts are anonymous illustrations and cannot establish prevalence, saturation, or causal change. The average MUET Band of 3.5 describes the cohort broadly but does not provide a speaking-specific baseline or distribution. Moreover, the proposed cognitive-load trajectory and Microphone Effect have not yet been psychometrically validated. The framework should therefore be understood as a theoretically informed pedagogical proposition supported by practice evidence, not as proof of intervention effectiveness. Formal mixed-method evaluation is required before claims of generalisability or impact can be made.

References

- Chen, A., Jia, J., Li, Y., & Fu, L. (2025). Investigating the effect of role-play activity with GenAI agent on EFL students' speaking performance. *Journal of Educational Computing Research*, 63(1), 99–125. <https://doi.org/10.1177/07356331241299058>
- Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 103254. <https://doi.org/10.1016/j.system.2024.103254>
- Huang, W. J., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237–257. <https://doi.org/10.1111/jcal.12610>
- Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, 206, 104898. <https://doi.org/10.1016/j.compedu.2023.104898>
- Ji, H., Han, I., & Ko, Y. (2022). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, 55(1), 48–63. <https://doi.org/10.1080/15391523.2022.2142873>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELJ Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- Li, Y., Shadiev, R., & Chiu, T. K. F. (2026). Applying generative artificial intelligence to task-based language teaching and learning: A systematic review and meta-analysis. *TechTrends*, 70, 150–163. <https://doi.org/10.1007/s11528-025-01140-7>
- Li, Y., Zhou, X., & Chiu, T. K. F. (2024). Systematic review on artificial intelligence chatbots and ChatGPT for language learning and research from self-determination theory: What are the roles of teachers? *Interactive Learning Environments*. Advance online publication. <https://doi.org/10.1080/10494820.2024.2400090>
- Li, Y., Zhou, X., Yin, H., & Chiu, T. K. F. (2025). Design language learning with artificial intelligence (AI) chatbots based on activity theory from a systematic review. *Smart Learning Environments*, 12, 24. <https://doi.org/10.1186/s40561-025-00379-0>
- Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S.-Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, 11, 50. <https://doi.org/10.1186/s40561-024-00342-5>
- UNESCO. (2023). Guidance for generative AI in education and research. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, 127, 103533. <https://doi.org/10.1016/j.system.2024.103533>