Asian Association of Open Universities Journal

Evaluating mobile language learning apps for distance education: insights into design, content and pedagogy

Ka Pik Sun and Billy T.M. Wong
Institute for Research in Open and Innovative Education, Hong Kong Metropolitan University, Hong Kong, China
Kam Cheong Li
Guangdong Open University, Guangzhou, China, and
Manfred M.F. Wu
Institute for Research in Open and Innovative Education, Hong Kong Metropolitan University, Hong Kong, China

Received 11 March 2026
Revised 25 April 2026
3 May 2026
Accepted 14 May 2026

Abstract

Purpose – The proliferation of mobile technologies has transformed language learning in distance education, positioning mobile apps as essential tools for remote and self-directed learners worldwide. This paper presents an evaluation of the feature quality of mobile language learning apps (MLLAs), showing how design, content and pedagogical features of widely used MLLAs can be systematically analysed and compared.

Design/methodology/approach – Eight widely used MLLAs – Duolingo, Memrise, Busuu, Falou, LingoDeer, Beelinguapp, FunEasyLearn and Innovative Language Learning – were evaluated using a structured framework. The framework encompasses four dimensions: app background, design, content and pedagogy.

Findings – The evaluation results highlight the strengths and limitations of each app, as well as the interplay between design, content and pedagogy in supporting learning processes. The results reveal notable disparities in app performance. Busuu emerged as the top-performing app, particularly in content quality and pedagogical features. Beelinguapp showed limited interactivity and applicability. Duolingo, despite its popularity and innovative gamification elements, had shortcomings in content and pedagogy.

Originality/value – This study advances mobile-assisted language learning (MALL) research by situating MLLA evaluation within the context of distance education. It offers a practical, objective and replicable feature-evaluation methodology for selecting apps that align with learner needs and objectives. The results also contribute empirical evidence on app quality disparities, supporting evidence-based tool selection for language learning.

Keywords Language learning, Distance education, Mobile app, Mobile language learning App (MLLA), Mobile-assisted language learning (MALL), Online learning, Evaluation framework

Paper type Research article

Introduction

The widespread adoption of mobile technologies has profoundly influenced language learning and the broader landscape of distance education, with mobile apps becoming essential tools in online and blended language courses as well as in self-directed study (Li *et al*., 2018; Rosell-Aguilar, 2018). In distance learning contexts where learners often have limited face-to-face contact with teachers and peers, mobile language learning apps (MLLAs) are particularly valued for their accessibility across time and place, interactivity and capacity to support personalised, self-paced learning (Kuimova *et al*., 2018; Wong, 2021). They can function both as stand-alone tools for autonomous distance learners and as supplementary resources

Asian Association of Open Universities Journal
Emerald Publishing Limited
e-ISSN: 2414-6994
p-ISSN: 1858-3431
DOI 10.1108/AAOUJ-03-2026-0034

integrated into virtual learning environments. Despite their widespread use, the quality of features that actually support language acquisition remains underexplored (Klímová, 2017). Existing studies relied heavily on surveys and interviews capturing user perceptions (John and Lo, 2024; Raj and Baisel, 2024; Ulfiah *et al*., 2025), which lack the objectivity needed to assess intrinsic design, content and pedagogical features. This study addresses that gap by applying a structured evaluation framework.

The study described in this paper applied a structured evaluation framework to assess MLLAs across four key dimensions: app background, design, content and pedagogy (Li *et al.*, 2025a, b). Unlike prior studies which relied on subjective user feedback (Khlaisang and Sukavatee, 2023; Nami, 2020), this study adopted an objective approach to assess the apps' intrinsic features. It demonstrates their strengths and limitations, offers practical guidance for app developers, educators and learners, and underlines best practices and areas for improvement in technology-enhanced distance language education. This study advances the methodology for assessing the feature quality of mobile-assisted language learning (MALL) tools through rigorous operationalisation and transparency and contributes to evidence-based research that can inform the selection and deployment of MLLAs in distance learning.

Related work

Research on the effectiveness of MLLAs has generally followed two main approaches: subjective studies based on user surveys and interviews (Rosell-Aguilar, 2018; Karjo and Andreani, 2018), and objective evaluations using structured frameworks (Ulfiah *et al*., 2025; Sakalauskė and Leonavičiūtė, 2022). While subjective studies provide insights into user experiences, they often lack generalizability. Objective evaluations, although more structured, may overlook user perspectives. This section reviews studies from both approaches, highlighting their methodologies, findings and implications.

Focusing on user experience, Rosell-Aguilar (2018) assessed Busuu through an in-app survey, examining user demographics, usage frequency and habits and user experiences. While users found the app beneficial for improving their language skills, they expressed dissatisfaction and criticised the lack of real interaction. Rosell-Aguilar (2018) presented detailed user opinions and identified potential areas for improvement but relied solely on subjective feedback and did not employ an evaluation framework, thereby limiting the generalisability of the findings. Karjo and Andreani (2018) evaluated Duolingo and Memrise based on their own user experiences, discussing general advantages and limitations but without an objective, systematic framework, thereby limiting the depth and rigour of their analysis. These studies underscored the importance of a structured framework for conducting comprehensive evaluations of MLLAs. Similarly, Ulfiah *et al*. (2025) focused on the gamified features in Duolingo, ELSA Speak and Cake-Learn. However, their evaluation also relied primarily on user experience data collected through surveys, highlighting the need for a more comprehensive evaluation framework that incorporates gamification as a criterion and enables the systematic analysis of a larger number of apps.

Several researchers have developed their own frameworks for evaluating MLLAs. Sakalauskė and Leonavičiūtė (2022) proposed a framework and evaluated Duolingo by using Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis, Porter's Five Forces analysis and the Value, Rareness, Imitability and Organisation (VRIO) method. Their evaluation emphasised the business and management aspects of the app rather than its effectiveness for language learning, limiting its relevance for analysing the app's feature quality in supporting language acquisition. A framework focused on app design and content is better suited to assessing their practical effectiveness.

Existing MLLA evaluations predominantly emphasise subjective feedback, narrow criteria like gamification, or business models, often lacking scoring transparency or pedagogical grounding. A systematic framework grounded in direct app usage addresses these gaps by providing a transparent, objective basis for evaluating intrinsic feature quality.

Situated within criteria-based evaluation theory and ISO/IEC 25010 software quality standards, this framework maps app design to usability, content to functional accuracy and pedagogy to instructional effectiveness (Sun *et al*., 2026). This alignment ensures a comprehensive assessment that balances technical functionality with educational value. This theoretical grounding distinguishes the framework from purely subjective user surveys and provides a structured, objective basis for evaluating the intrinsic features of MLLAs.

Methodology

App selection criteria

Eight MLLAs, including Duolingo, Memrise, Busuu, Falou, LingoDeer, Beelinguapp, FunEasyLearn and Innovative Language Learning, were selected using the following criteria to ensure relevance, accessibility and methodological consistency:

(1) The apps had to be available on both Google Play Store and Apple App Store to ensure cross-platform accessibility.

(2) Only apps with a rating of 4.5 stars or higher (out of 5) on both platforms were included to ensure quality.

(3) The apps needed to have at least 1 million downloads on Google Play Store (Apple App Store does not disclose download numbers) to ensure substantial user adoption. A substantial download base helps ensure sufficient peer feedback and long-term viability, enhancing the reliability of cross-app comparisons.

(4) The apps had to offer a freemium model, providing free and open access to basic features to ensure fairness and consistency in the analysis. To minimise bias, apps requiring a paid subscription without a free mode or trial period were excluded.

(5) Selected apps were required to support self-directed learning, enabling independent engagement without instructor guidance (Baloh *et al*., 2015). Applications relying on live tutoring were excluded to maintain this autonomous focus.

These criteria ensured the inclusion of widely accessible, free and highly rated language learning apps that supported self-directed learning, minimised bias and promoted consistency in the analysis.

Evaluation framework

The framework adopted in this evaluation was an established systematic model developed by Li *et al*. (2025a, b), as presented in Table 1. The framework synthesises and enhances components from three frameworks proposed in prior studies (Almaiah *et al*., 2022; Essafi *et al*., 2024; Rosell-Aguilar, 2017). It consists of four parts: background, design, content and pedagogy of the app.

A 0–2 scoring rubric was implemented to ensure consistency and clarity in evaluating the quality and effectiveness of the apps. For score aggregation, equal weighting was applied across all 32 criteria to avoid dimensional bias. Final scores for each dimension reflect the number of criteria in each category rather than any hierarchical prioritisation among the dimensions to ensure transparent and feature-level comparability.

Evaluation settings and user profile configuration

To collect data for this study, all eight MLLAs were used under the following settings to ensure comparability and minimise bias:

(1) The default language was Chinese, simulating the experience of a Chinese distance learner studying English. Given the growing focus on English as a global language and

Table 1. Evaluation framework for mobile language learning apps (Li *et al.*, 2025a, b)

Category		Criteria
Background and characteristics		Google Play downloads, Google Play and Apple App Store ratings, year founded, headquarters location, number of language courses offer and learning structure
App design	Multimedia integration	(1) utilises different forms of multimedia (2) utilises multimedia in a didactic and meaningful way
	Offline functionality	(1) functions online and offline (2) offers learning materials for download
	In-app advertising	(1) offers free content (2) does not contain recurring disturbing ads
	App support	(1) provides various channels for support (2) provides instant and personalised responses (3) provides instructions on using (4) gives learners alerts for new notifications
	Interface design	(1) provides good icon design in the user interface (2) allows users to easily identify particular functions
App content	Learning objectives	(1) clear and aligned with the course items (2) achievable and measurable
	Learning content	(1) logically built and structured (2) presents, explains, or models the languages and does not only test it (3) includes background information about the customs and traditions of the language (or the region where the language is spoken) (4) error-free
	Learning activities	(1) align with the Present-Practice-Test (PPT) model (2) varied and interesting (3) offer structured revision tests
	Targeted skills	(1) effectively teaches the targeted skills (2) meaningfully integrates other skills
	App updates	(1) features and contents are updated regularly
App pedagogy	Personalisation	(1) offers placement tests (2) offers easy access to different difficulty levels
	Gamification	(1) utilises gamified features (e.g. unlock levels, earn points, receive virtual tokens, win virtual currency, compete in leaderboards) (2) gamified features have pedagogical added value
	Scaffolding	(1) users can monitor their learning (2) users are scaffolded and offered instant and detailed feedback
	Interaction	(1) targets culture (knowledge learned can be applied in everyday life, not just for practice, e.g. in travel or the workplace) (2) facilitates collaboration between app users (e.g. content exchange and sharing, discussion boards)

Source(s): Authors' own work

the emphasis on English in many MALL studies (Raj and Baisel, 2024), this project used English as the testing language.

(2) The default language level was "beginner" to ensure consistency across apps, as definitions of proficiency levels such as "Proficient," "Advanced," and "Intermediate" vary. Starting from the most basic level in all apps minimised discrepancies and potential bias.

(3) The learning goal was "academic purposes". This purpose was deemed applicable to a broader range of users and ensured consistency in the evaluation process.

(4) The sampled apps predominantly target general English proficiency rather than English for Specific Purposes or discipline-specific content.

To ensure technical replicability, the evaluation was conducted using the following hardware and software configurations:

(1) Device: iPad Air 4

(2) Operating System: iPadOS 18.2

(3) App Versions: All apps were updated to the latest version available during the evaluation period.

(4) Storefront Region: Apps were downloaded from the Hong Kong Google Play Store and Apple App Store to ensure consistent content availability.

(5) Account Mode: All apps were evaluated using standard free accounts (freemium model) without premium subscriptions, unless otherwise noted in the results. Accounts were created specifically for this study to avoid algorithmic bias from prior usage history. The free tiers were assessed for their capacity to deliver foundational communicative competence, which aligns with the academic-purpose learner profile simulated in this study.

Procedure of app evaluation

The app evaluation was conducted in Spring 2026. Each app was evaluated for 2–3 days, completing approximately 20 one-minute lessons per app.

Data collection was carried out in two phases. The first phase involved assigning scores as a new user, capturing the initial user experience without prior familiarity with the app. The second phase was conducted 1–2 days after app usage to verify whether the initial scores aligned with the user experience once users gained familiarity with the app.

The scoring process was conducted in real-time while using the apps to capture immediate impressions and insights. A scoring scale ranging from 0 to 2 was employed, where 0 indicated poor performance and 2 indicated excellent performance. Each score was accompanied by a written explanation that justified the rating and provided detailed insights into the reasoning behind the assigned score.

To assess scoring consistency and address potential evaluator bias, a second independent researcher with expertise in educational technology evaluated the apps using the same rubric and procedures. Inter-rater agreement was calculated using percentage agreement. The two raters achieved 82.05% exact agreement across the criteria. This result indicates strong consistency in the application of the scoring framework. Minor discrepancies were resolved through discussion and did not affect the overall ranking of apps.

Results and discussion

The evaluation of eight MLLAs is organised into four areas: background, app design, app content and app pedagogy.

Background and characteristics

This section provides background information on the eight MLLAs selected for evaluation. To understand their scope, reach and design, the following key characteristics were examined:

(1) Number of downloads on Google Play

(2) User ratings from Google Play and Apple App Store

(3) Year of establishment

(4) Headquarters location

(5) Number of language courses offered

(6) Structure of the learning system

The eight apps vary considerably in user base and growth. Duolingo leads by a large margin, with over 500 million downloads, far exceeding the others, most of which average from 5 million to 50 million downloads. Innovative Language Learning has the smallest user base, with approximately 1 million downloads. Innovative Language Learning, founded in 2005, is the oldest app but has struggled to gain popularity, while Falou, the youngest app, founded in 2019, has grown rapidly to 10 million downloads. User ratings are generally high across all apps, averaging 4.6–4.7 stars, with Memrise and LingoDeer slightly lower at 4.5 stars.

Duolingo offers the largest number of language courses, providing 40 options, including fictional languages, while the other apps offer fewer languages. The apps' headquarters are geographically dispersed, with two each based in Europe, the UK, the USA and Asia. Despite these differences, all apps follow a broadly similar learning structure, typically comprising four to seven levels and consistently including beginner, intermediate and advanced stages. This shared structure provides a familiar progression pathway for learners across platforms.

App design

App design is crucial for enhancing user experience and supporting effective learning outcomes (Rosell-Aguilar, 2017). It was assessed in terms of how the apps use technology to deliver engaging content, provide accessibility, minimise distractions, support users and create intuitive user interfaces. The assessment comprised 12 criteria, each worth 2 points, for a maximum possible score of 24. Table 2 presents the detailed scores.

Multimedia integration. The integration of multimedia contributes to an engaging and effective learning experience (Solmaz, 2025). Apps utilising diverse multimedia in a didactic and meaningful manner scored higher. Duolingo, Memrise and Busuu achieved the maximum 4 points for their rich and varied use of multimedia. Duolingo stood out for its vibrant animations, illustrations and sound effects, as well as its mascot which frequently motivates and encourages users throughout their learning journey. Memrise and Busuu enhanced engagement through video clips featuring live-action pronunciation, accompanied by images and sound effects, making lessons vivid, engaging and memorable.

FunEasyLearn and Innovative Language Learning scored lowest, with 1 out of 4. FunEasyLearn relied almost exclusively on images for its teaching materials, which may limit opportunities for multimodal engagement. Innovative Language Learning primarily used audio and video as its core teaching tools but did not integrate a wider range of multimedia features, resulting in a less dynamic learning experience.

Offline functionality. Offline functionality is critical for learners who lack constant Internet access (Essafi *et al*., 2024). This dimension considers both the extent to which lessons can be used offline and the availability of downloadable materials.

LingoDeer and FunEasyLearn scored full marks, as all their lessons are fully accessible without an Internet connection. Memrise does not support offline access in the evaluated configuration and therefore scored 0.

Regarding downloadable materials, Innovative Language Learning scored full marks by allowing users to download complete lesson materials. Regular users can download the materials available at their current level, while premium users can download all content. Duolingo, Memrise, Falou and FunEasyLearn do not provide downloadable materials and thus scored 0 on this criterion.

In-app advertising. The availability of free content is an important consideration for users (Rosell-Aguilar, 2017). Duolingo, Busuu and FunEasyLearn scored full marks in this criterion. Duolingo provides access to all lesson content, Busuu to most and FunEasyLearn to all lessons within the user's current level. This allows beginners to build a solid vocabulary foundation without needing to subscribe immediately. LingoDeer scored 0, as it offers free

Table 2. App design evaluation scores

Category	Criteria	Score Duolingo	Memrise	Busuu	Falou	LD[a]	BLA[a]	FEL[a]	ILL[a]
Multimedia integration	(1) utilises different forms of multimedia	2	2	2	1	1	1	0	0
	(2) utilises multimedia in a didactic and meaningful way	2	2	2	2	1	1	1	1
Offline functionality	(1) functions online and offline	1	0	1	1	2	1	2	1
	(2) offers learning materials for download	0	0	1	0	1	1	0	2
In-app advertising	(1) offers free content	2	1	2	1	0	1	2	1
	(2) does not contain recurring disturbing ads	0	1	0	0	1	2	2	2
App support	(1) provides various channels for support	2	1	2	1	1	1	1	1
	(2) provides instant and personalised responses	1	1	1	1	1	1	1	1
	(3) provides instructions on using	1	1	0	1	0	1	1	0
	(4) gives learners alerts for new notifications	2	2	2	2	2	2	2	2
Interface design	(1) provides good icon design in user interface	2	2	2	2	2	2	2	2
	(2) allows users to easily identify particular functions	2	2	2	2	2	2	2	2
Total scores		17	15	17	14	14	16	16	15

Note(s) [a]LD: LingoDeer, BLA: Beelinguapp, FEL: FunEasyLearn, ILL: Innovative Language Learning
Source(s): Authors' own work

access to only the first unit (four lessons), making sustained progress difficult without upgrading to the paid version.

To evaluate the impact of advertising, the duration of each lesson and the duration and frequency of ads were recorded and compared. Innovative Language Learning scored 2 because it contains no ads and lessons last 3–6 min. FunEasyLearn also scored 2, with no ads, followed by its 15–30 s lessons. Duolingo, Busuu and Falou scored 0, as they display non-skippable video ads after each short lesson, which substantially disrupts the learning experience. Falou's advertising model is particularly intrusive: once users reach the daily lesson limit, they must watch non-skippable video ads promoting subscriptions.

App support. An effective help section can reduce user attrition (Solmaz, 2025). High-scoring apps should provide various support channels, timely and personalised responses, clear usage instructions and notifications to users.

Duolingo and Busuu scored 2 on the support-channel criterion due to the availability of discussion forums. The other apps scored 1, offering FAQs and email assistance, but no forum. None of the apps provided real-time support (e.g. live chat or hotline); all relied solely on email and therefore scored 1 for "instant and personalized responses," reflecting limited responsiveness.

For usage instructions, Duolingo, Memrise, Falou, Beelinguapp and FunEasyLearn scored 1, as they offered one-time introductory guidance, usually emphasising gamified features or learning approaches. However, these instructions cannot be revisited once completed. Busuu, LingoDeer and Innovative Language Learning each scored 0 because they do not provide explicit instructions on how to navigate the app. Lastly, all apps scored 2 on the notification criterion, as they send daily email reminders to encourage users to complete lessons.

Interface design. A clear and vibrant layout design makes the learning process more engaging and accessible (Solmaz, 2025). This dimension focused on the quality of icon design in the user interface and the ease with which users can identify specific functions based on those icons.

All apps received a perfect score of 4 in this category. They provided user-friendly designs with clear icons that adequately convey their intended functions. Duolingo and Falou feature vibrant, colourful interfaces with cartoon-style elements, creating an engaging and visually appealing experience. In contrast, the other apps employ minimalist, concise design styles that emphasise clarity and simplicity. Despite these stylistic differences, all interfaces are accessible, intuitive and support seamless navigation.

Section summary. Figure 1 illustrates the apps' performance across design dimensions. Interface design stood out with a perfect mean score of 4.00 (100%) and no variability (S.D. = 0.00), reflecting uniform excellence across all apps. App support and multimedia integration demonstrated moderate performance with moderate variability. In-app advertising (2.25, 56.25%) and offline functionality (1.75, 43.75%) were weaker dimensions, both with moderate variability (S.D. = 1.04), revealing inconsistent performance and opportunities for improvement, particularly in offline accessibility.

App content

App content assesses the quality and depth of an app's material, with a focus on the language and skills it aims to develop in learners (Essafi *et al.*, 2024). It was accessed across aspects of educational value and functionality, ensuring a thorough assessment of the apps' ability to

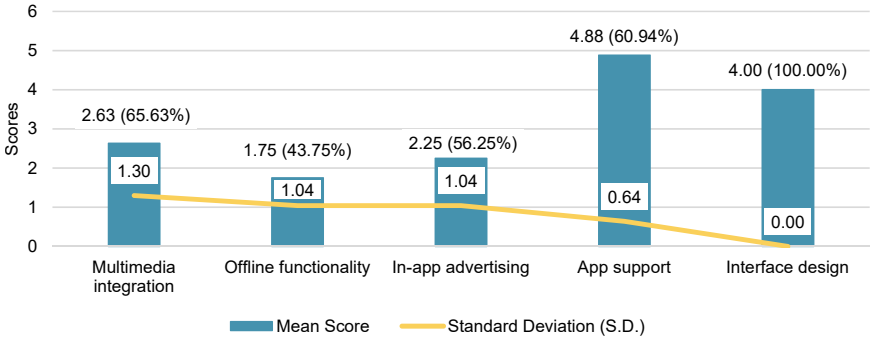

Figure 1. App performance in design. Source: Authors' own work

meet users' learning needs and expectations. Apps were scored out of 24 points across 12 criteria, with 2 points per criterion (see Table 3).

Learning objectives. Apps should have clear and quantifiable goals to guide learners and ensure meaningful improvement (Jeon, 2022; Sewagegn, 2020). This dimension examined whether each app's goals are clear, aligned with its content and achievable. All apps scored 3 out of 4 in this dimension. While their objectives were clearly stated and aligned with the content, most lacked measurable outcomes.

Several apps emphasised making learning enjoyable and engaging (Rosell-Aguilar, 2017). Duolingo aimed to provide fun learning experiences through gamification, offering interactive exercises and rewards. Innovative Language Learning promoted flexible and enjoyable language learning through audio and video content, but lacked interactivity and offered limited engagement. LingoDeer integrated engaging stories into its curriculum to captivate users,

Table 3. App content evaluation scores

Category	Criteria	Score Duolingo	Memrise	Busuu	Falou	LD[a]	BLA[a]	FEL[a]	ILL[a]
Learning objectives	(1) clear and aligned with the course items	2	2	2	2	2	2	2	2
	(2) achievable and measurable	1	1	1	1	1	1	1	1
Learning content	(1) logically built and structured	0	1	2	1	1	0	1	1
	(2) presents, explains, or models the languages and does not only test it	0	1	2	2	2	1	1	1
	(3) includes background information about the customs and traditions of the language (or the region where the language is spoken)	0	0	0	0	1	1	1	2
	(4) error-free	2	2	2	0	1	0	2	2
Learning activities	(1) align with the Present-Practice-Test (PPT) model	1	2	2	2	2	2	2	1
	(2) varied and interesting	1	2	2	1	2	0	2	0
	(3) offers structured revision tests	2	1	2	2	0	0	2	0
Targeted skills	(1) effectively teaches the targeted skills	2	2	2	2	2	2	2	1
	(2) meaningfully integrates other skills	1	1	2	1	1	1	1	0
App updates	(1) features and contents are updated regularly	2	1	2	1	1	2	1	0
Total scores		14	16	21	15	16	12	18	11

Note(s) [a]LD: LingoDeer, BLA: Beelinguapp, FEL: FunEasyLearn, ILL: Innovative Language Learning

Source(s): Authors' own work

though the effectiveness of these narratives varied by individual preference. However, the subjective nature of fun made it challenging to assess how well they achieved this goal.

Tailoring the learning experience to individual users' needs was another core focus for many apps. Duolingo adapted its content based on user progress, delivering personalised education adjusted to individual needs. Memrise offered customised courses and immersive materials suited to users' proficiency levels. Busuu supported independent learning to enable personal growth. Beelinguapp took a unique approach by offering content such as favourite articles, books and music to make language learning more relatable. However, it remained unclear whether users experienced tangible personal or professional growth through these personalised approaches (Diao *et al*., 2018).

Improving accessibility was also a shared objective (Almaiah *et al*., 2022). Falou aimed to make learning more accessible by focusing on practical speaking skills, but its limited free resources and daily restrictions on lessons undermined these claims. FunEasyLearn similarly sought to "make language learning accessible to everyone" by offering diverse language combinations, but its utility for all learners remained subjective. Busuu and Memrise also highlighted accessibility, but the extent to which their platforms remove barriers for all users was inconclusive.

Learning content. Apps should have logically structured and organised content, explain or model the language rather than simply testing users, provide cultural background information about customs and traditions, and be free of errors. These criteria ensure that learners follow a coherent progression, gain a deeper understanding of the language (Sewagegn, 2020) and develop cultural knowledge alongside linguistic skills.

Regarding structure, Busuu scored 2 as its lessons followed the Common European Framework of Reference for Languages and were tailored to learners' proficiency levels (Rosell-Aguilar, 2018). Duolingo and Beelinguapp scored 0 due to a lack of clear progression and discernible organisation, with no evident connection between lessons and levels.

Language learning should emphasise explanation and presentation (Wardoyo, 2018). Busuu, Falou and LingoDeer scored 2. Busuu offered explanations after each question. Falou employed a multi-step teaching approach involving vocabulary practice, simulated conversations and writing exercises. LingoDeer provided grammar tips and vocabulary lists for each unit. Duolingo scored 0 due to its disjointed and ineffective teaching sequence.

Cultural background content can serve as an important source of motivation for learners (McKay, 2003). Innovative Language Learning scored 2 for incorporating cultural insights into its lessons. Its initial audio lesson provided an overview of the cultures and backgrounds of English-speaking countries, with additional lessons that incorporated specific cultural knowledge. Duolingo, Memrise, Busuu and Falou did not provide cultural background information, resulting in a score of 0 on this criterion.

In terms of quality and stability, Duolingo, Memrise, Busuu, FunEasyLearn and Innovative Language Learning scored 2, providing accurate content without noticeable errors, omissions, freezing or crashing. Falou and Beelinguapp scored 0 due to frequent translation errors or synchronisation issues. Falou contained frequent, consistent errors in its Chinese translations of English sentences, undermining the quality of its content. Similarly, Beelinguapp had difficulty aligning sentence highlighting with audio playback.

Learning activities. The learning activities assessed whether the apps provided engaging and effective activities, variety and interest in the exercises and the inclusion of structured revision tests. These factors ensured that users received a well-rounded and dynamic learning experience (Jeon, 2022).

Memrise, Busuu, Falou, LingoDeer, Beelinguapp and FunEasyLearn scored 2 for combining learning materials with quick follow-up review quizzes. Falou, in particular, excelled with its structured approach: after a lesson on pronunciation, users engaged in three repeated simulated conversations, starting with speaking using Chinese translations as hints, followed by writing exercises such as sentence reordering or typing and ending with another speaking activity without hints.

Regarding the exercise variety, Memrise, Busuu, LingoDeer and FunEasyLearn scored 2 for offering a range of activities, e.g. spelling, sentence reordering, word jumbles, fill-in-the-blanks, true-or-false questions and matching exercises. Beelinguapp and Innovative Language Learning scored 0 due to their limited exercise types. Beelinguapp provided only reading comprehension and word-meaning matching exercises, while Innovative Language Learning focused solely on listening to audio and watching videos, offering no interactive tasks.

Duolingo, Busuu, Falou and FunEasyLearn scored 2 for including structured revision tests at the end of each level or unit, helping users consolidate knowledge by reviewing previously learned words and phrases. LingoDeer, Beelinguapp and Innovative Language Learning scored 0 for lacking structured revision tests. LingoDeer included pre-level tests but not revision tests. Beelinguapp did not follow a level-based format and lacked structured revision. Innovative Language Learning did not provide any revision tests.

Targeted skills. The targeted skills dimension evaluated whether the apps effectively teach their primary skills and integrate other language skills into their learning framework. These criteria assessed the breadth and depth of the apps' pedagogical approaches.

Nearly all apps scored 2 for effectively teaching their targeted skills. They demonstrated clear orientation toward their respective focus areas. For example, Duolingo and Busuu emphasised grammar and vocabulary, while Falou focused on speaking. This clarity in instructional goals allowed users to develop the specific skills prioritised by each app. Conversely, Innovative Language Learning scored 1, as its heavy reliance on audio listening made it more akin to a podcast than a structured platform for teaching specific language skills or improving users' English proficiency.

For the integration of other language skills, Busuu scored 2 for covering all basic skills, including reading, writing, listening and speaking. Innovative Language Learning scored 0, as it relied on audio and video content, offered limited interactivity and restricted its vocabulary feature to premium users.

App updates. The app updates dimension assessed the regularity of feature and content updates to determine whether each app remained dynamic, functional and relevant for users (Almaiah *et al*., 2022). Update frequency was based on records from the Apple App Store, as Google Play does not disclose historical update dates. Regular updates indicate an app's commitment to improvement and responsiveness to user needs and technological developments.

Duolingo, Busuu and Beelinguapp scored 2 for frequent updates, typically at least once a week, demonstrating consistent efforts to enhance features and content. Memrise, Falou, LingoDeer and FunEasyLearn scored 1, updating about twice per month. Notably, Falou showed irregular patterns, with months of multiple updates followed by periods without updates. Innovative Language Learning scored 0, as it had not been updated for nearly a year, suggesting a lack of ongoing development and potential stagnation in its features and content.

Section summary. Figure 2 summarises the apps' performance against content-related criteria. Learning objectives showed consistent performance across all apps, with a mean score of 3.00 (75.00%) and no variability (S.D. = 0.00). Learning content averaged 4.13 (51.56%) but exhibited high variability (S.D. = 1.64), reflecting differences in content delivery. Learning activities scored 4.13 (68.75%) but showed greater variability (S.D. = 1.81), indicating varying activity quality. Target skills scored 2.88 (71.88%) with moderate variability (S.D. = 0.83), while app updates had the lowest mean score of 1.25 (62.50%) and moderate variability (S.D. = 0.71), showing uneven performance in addressing skill development and inconsistent update practices across apps.

App pedagogy

Pedagogy refers to the art, science and craft of teaching, encompassing the methodologies and approaches used to facilitate learning (Essafi *et al*., 2024). It shapes how effectively an app engages users and supports their learning process. Apps were scored out of 16 points across 8 categories, with 2 points each (Table 4).

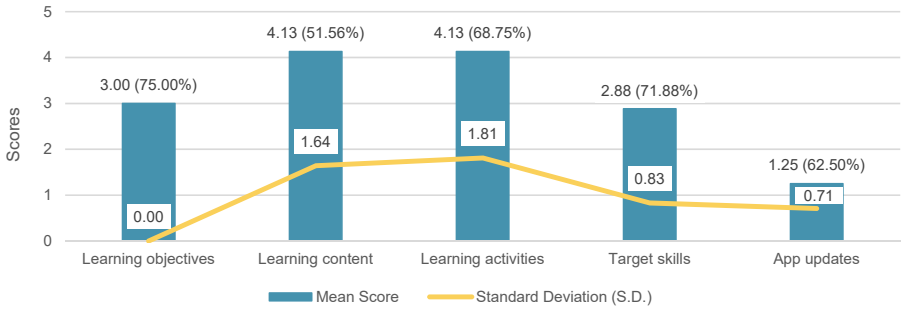

Figure 2. App performance in content. Source: Authors' own work

Table 4. App pedagogy evaluation scores

Category	Criteria	Score Duolingo	Memrise	Busuu	Falou	LD[a]	BLA[a]	FEL[a]	ILL[a]
Personalisation	(1) offers placement tests	2	0	1	0	0	0	0	0
	(2) offers easy access to different difficulty levels	1	0	2	0	0	0	1	2
Gamification	(1) utilises gamified features	2	1	2	1	2	0	2	0
	(2) gamified features have a pedagogical added value	2	1	1	1	1	0	2	0
Scaffolding	(1) users can monitor their learning	2	2	2	2	2	0	2	2
	(2) users are scaffolded and offered instant and detailed feedback	1	2	2	1	1	0	1	0
Interaction	(1) targets culture	1	2	2	2	1	1	1	2
	(2) facilitates collaboration between app users	2	0	2	0	0	0	0	0
Total scores		13	8	14	7	7	1	9	6

Note(s) [a]LD: LingoDeer, BLA: Beelinguapp, FEL: FunEasyLearn, ILL: Innovative Language Learning
Source(s): Authors' own work

Personalisation. Personalised features enhance self-directed learning by allowing learners to tailor content and learning approach to their needs, preferences and interests (Lee and Xiong, 2023; Gunawardena *et al.*, 2024; Walkington and Bernacki, 2020). This dimension

examined whether apps supported placement tests and flexible access to content at different difficulty levels, which can enhance the effectiveness and engagement of the learning process.

Duolingo was the only app with a mandatory placement test at the beginning of the learning process, allowing users to determine whether their chosen starting level was appropriate; it therefore received the highest score of 2. Six of the apps did not include placement tests; users either began at the lowest level by default or selected their starting level without assessment, resulting in a score of 0.

Busuu and Innovative Language Learning scored 2 for allowing unrestricted access to all levels. Four apps scored 0. Memrise and Falou restricted users to their current level, with no option to skip ahead. LingoDeer required users to progress sequentially through the levels, while Beelinguapp limited structured learning to the current level, though it permitted limited access to stories from other levels.

Gamification. Gamification maintains user engagement and fosters a positive, sustained learning experience (Ulfiah *et al*., 2025; Lee and Xiong, 2023). Apps typically incorporate gamified elements such as point systems, rewards, leaderboards and unlockable levels. These elements help motivate learners to stay committed to their language goals, enhance their aspirations and deepen their understanding of the language (Ulfiah *et al*., 2025; Al-Mulhem and Almaiah, 2021).

Duolingo scored 2 for its richest gamified features, including streak maintenance, level unlocking, competitive leaderboards and earning virtual currency, tokens and points. Busuu, LingoDeer and FunEasyLearn also scored 2, providing a rich gamified experience which was slightly less extensive than Duolingo's. Beelinguapp and Innovative Language Learning scored 0, as they were entirely devoid of gamified features.

Gamified features should not be merely superficial but should have pedagogical value by motivating users to engage more deeply with the content. Duolingo and FunEasyLearn scored 2, with FunEasyLearn performing particularly well. It allowed users to unlock full courses using in-app currency earned through lessons, offering an alternative to subscribing to the premium model. This strongly incentivised users to complete more lessons to unlock more advanced content. Language Learning remained at 0 because they lacked gamified systems.

Scaffolding. Instructional scaffolding supports learners in building knowledge and skills incrementally (Wan *et al*., 2024) and is gradually withdrawn as learners achieve greater independence and mastery (Essafi *et al*., 2024). This dimension assessed users' ability to monitor progress and the provision of scaffolded assistance through instant and detailed feedback (Li *et al*., 2025a, b).

All apps except Beelinguapp scored 2, which indicated that they enabled users to identify their current learning level. Beelinguapp, nevertheless, lacked clarity on users' levels. Users selected a starting level and followed "The Journey," where materials were arranged to suit their proficiency, but they could also explore the "Library", which included books from all levels without clear level indicators. This might lead users to select materials beyond their proficiency, causing frustration and hindering progress, therefore Beelinguapp scored 0.

For feedback, Memrise and Busuu scored 2 for explaining incorrect answers and reinforcing understanding through additional questions. Beelinguapp and Innovative Language Learning scored 0. Beelinguapp identified incorrect answers but neither provided the correct answer nor allowed retries. Innovative Language Learning prioritised content delivery over testing understanding, offering minimal opportunities to confirm mastery.

Interaction. Interaction is essential for language acquisition, bridging app-based learning and practical, real-world use (Lee and Xiong, 2023; Behbahani and Karimpour, 2025). This dimension examined whether the app incorporated elements of the target culture, real-world applicability in contexts such as travel or work and user collaboration through content exchange, sharing, or discussion boards, thereby contributing to a more dynamic and engaging learning experience.

Regarding cultural integration and real-life applicability, Memrise, Busuu, Falou and Innovative Language Learning scored 2 for emphasising practical and everyday language use.

Falou stood out for its AI-driven simulated conversations, while Innovative Language Learning offered in-depth explanations of language rules and structures.

For collaboration, only Duolingo and Busuu scored 2, as both featured dedicated discussion forums where users could share content, exchange opinions and engage with other learners. The remaining six apps scored 0 due to their lack of built-in platforms for user collaboration or content exchange.

Section summary. Figure 3 presents the apps' pedagogical performance. Scaffolding had the highest mean of 2.75 (68.75%) with moderate variability (S.D. = 1.28), indicating strong but inconsistent support for learners' progression. Gamification followed with a mean of 2.25 (56.25%) but exhibited high variability (S.D. = 1.58), reflecting differences in how gamified elements were used to enhance engagement. Personalisation and interaction recorded low mean scores (1.13, 28.13%; 2.00, 50%, respectively) with moderate variability (S.D. = 1.36; 1.07), indicating that these features were undervalued or deprioritised by some apps.

Overall app performance

This section analysed the overall performance of the MLLAs. Figure 4 presents the comparison of app performance in each aspect. For app design, the mean score was 15.50 out of 24, corresponding to 64.58% of the total possible marks. The low standard deviation of 1.20 indicated consistent performance among the apps, suggesting relatively similar levels of design quality. App content had a comparable mean score of 15.38 (64.08%), but a higher standard deviation of 2.92, reflecting greater variability in the apps' ability to provide engaging and relevant content. App pedagogy exhibited the lowest mean score of 8.13 out of 16 (50.81% of the total possible marks) and the highest standard deviation (4.09), representing substantial differences in how effectively the apps implemented teaching methods and supported language acquisition.

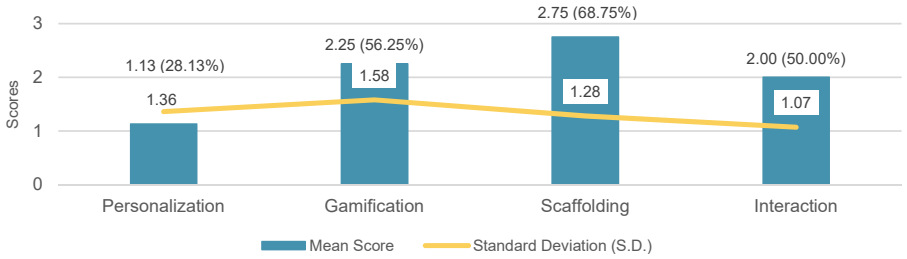

Figure 3. App performance in pedagogy. Source: Authors' own work

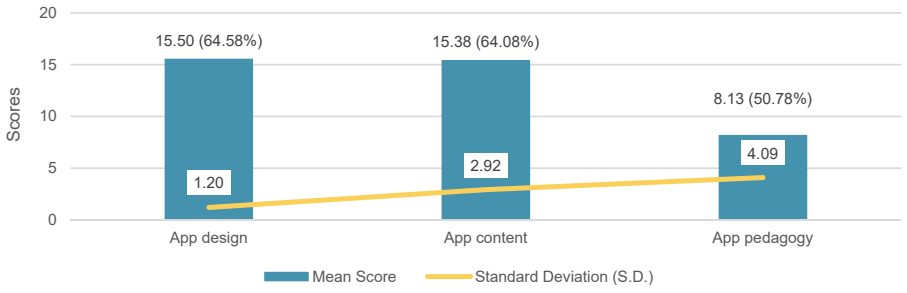

Figure 4. Overall app performance in design, content and pedagogy. Source: Authors' own work

These results suggest that the evaluated apps tend to prioritise design and content, as these visible features help attract users and maintain their engagement, prompting developers to enhance them. The greater variability observed in pedagogy may reflect differing approaches to English language teaching (ELT). For example, Falou and Busuu used communicative language teaching, which is regarded as an effective pedagogy for ELT (McKay, 2003), with a focus on speaking skills. Duolingo adopted a task-based teaching method that emphasises user autonomy over language use and task outcomes and addresses grammar skills during the post-task phase (Coultas and Booth, 2019). These differences illustrate the range of pedagogical orientations across the evaluated apps. Future research could examine how app design choices are related to developers' educational backgrounds and pedagogical beliefs.

The total scores of the apps across design, content and pedagogy reveal substantial variation in overall performance. Busuu achieved the highest total score of 52 out of 64, (81.25% of the total possible marks), excelling particularly in content and pedagogy. Meanwhile, Beelinguapp obtained the lowest total score, 29 out of 64 (45.31%), primarily due to its very low pedagogy score. Figure 5 illustrates the total scores achieved by the apps.

Summary and conclusion

This evaluation provides insights for scholars in the fields of language learning, learning technology and distance education by analysing the strengths and weaknesses of popular MLLAs. It highlights how different design, content and pedagogical approaches impact user experience and learning effectiveness, especially for learners who engage with these apps as primary or supplementary tools in distance learning. The findings can serve as a reference for future research on app-based language learning and its role in second language acquisition at a distance, as well as for app developers to enhance the pedagogical effectiveness, functionality and design of their products. Additionally, the study equips educators and institutions with practical criteria for selecting and integrating MLLAs into distance learning programmes, enabling evidence-based tool recommendations tailored to student needs.

In terms of app design, Duolingo and Busuu scored the highest (17 out of 24), while Falou and LingoDeer scored the lowest (14 out of 24). However, all apps met fundamental usability and navigation standards. For app content, Busuu outperformed the others with a score of 21 out of 24, whereas Innovative Language Learning scored the lowest (11 out of 24 each). This highlighted substantial variation in material depth and structural coherence. With respect to pedagogy, Busuu again scored the highest with 14 out of 16, whereas Beelinguapp performed the worst, with only 1 out of 16, failing essential instructional support. Overall, Busuu emerged as the top-performing app with a total score of 52 out of 64, while Beelinguapp ranked lowest

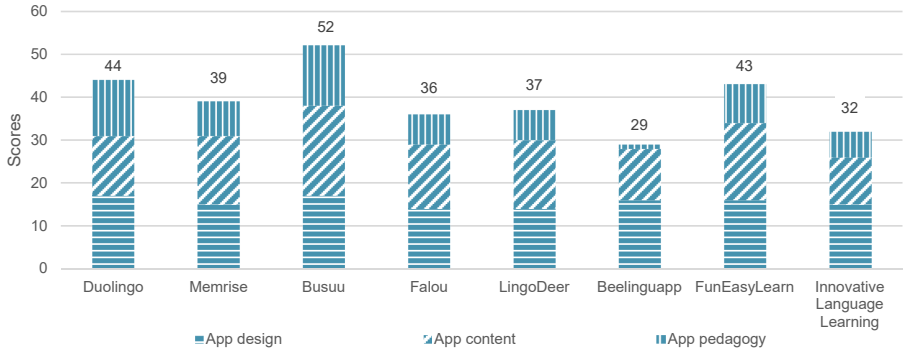

Figure 5. Overall scores of mobile language learning apps. Source: Authors' own work

with 29 out of 64. Nevertheless, Beelinguapp remains a practical choice for learners specifically targeting reading skills, as it provides level-aligned, structured reading passages despite its textbook-like design and limited features.

Despite ranking second overall, Duolingo maintains the largest user base, largely driven by its intensive gamification and humorous mascot, which strongly resonates with learners (Ulfiah *et al*., 2025; Sakalauskė and Leonavičiūtė, 2022). In contrast, the highest-scoring app, Busuu, has only about 10 million users. This discrepancy suggests that app adoption relies heavily on factors beyond intrinsic pedagogical quality, such as gamification intensity or content presentation style. Viewed through the Technology Acceptance Model, while traditional frameworks prioritise perceived usefulness, perceived enjoyment significantly shapes user retention (Abdalla, 2024). Duolingo's strong gamification and intuitive design likely enhance perceived enjoyment, supporting user retention despite its pedagogical limitations. Conversely, Busuu's higher pedagogy score aligns more closely with perceived usefulness but may lack some of the immediate engagement features that drive mass adoption.

This evaluation has limitations. First, the criteria employed may not capture all relevant aspects of MLLAs. Second, while the analysis prioritised objective indicators, some evaluative judgments inevitably involved subjectivity, though explicit rubrics minimised bias. Third, accent variety was excluded. Since apps differ in offering British, American, or multiple accents, future research should examine how accent exposure affects listening skills and learner engagement. Fourth, this evaluation employed a simulated user profile of a Chinese beginner learning English for academic purposes. While it ensured internal consistency, it limits generalisability to other proficiency levels, first languages, or non-academic goals. These findings represent a context-specific comparison rather than a universal ranking. Future research could replicate this framework across diverse user profiles to examine how app performance varies across learner characteristics, instructional designs and distance education settings.

References

Abdalla, R.A.M. (2024), "Examining awareness, social influence, and perceived enjoyment in the TAM framework as determinants of ChatGPT. Personalization as a moderator", *Journal of Open Innovation: Technology, Market, and Complexity*, Vol. 10 No. 3, 100327, doi: 10.1016/j.joitmc.2024.100327.

Al-Mulhem, A. and Almaiah, M.A. (2021), "A conceptual model to investigate the role of mobile game applications in education during the COVID-19 pandemic", *Electronics*, Vol. 10 No. 17, p. 2106, doi: 10.3390/electronics10172106.

Almaiah, M.A., Hajjej, F., Lutfi, A., Al-Khasawneh, A., Alkhdour, T., Almomani, O. and Shehab, R. (2022), "A conceptual framework for determining quality requirements for mobile learning applications using Delphi Method", *Electronics*, Vol. 11 No. 5, p. 788, doi: 10.3390/electronics11050788.

Baloh, M., Zupanc, K., Košir, D., Bosnić, Z. and Scepanovic, S. (2015), "A quality evaluation framework for mobile learning applications", *Proceedings of the 2015 4th Mediterranean Conference on Embedded Computing (MECO)*, Budva, Montenegro, 14-18 June.

Behbahani, H.K. and Karimpour, S. (2025), "Exploring the role of mobile-mediated dynamic assessment in L2 learners' reading comprehension and reading fluency", *Education and Information Technologies*, Vol. 30 No. 5, pp. 6237-6260, doi: 10.1007/s10639-024-12994-4.

Coultas, V. and Booth, P. (2019), "Exploratory talk and task-based learning: a case study of a student's learning journey on an MA (Education) English language teaching course", *Changing English*, Vol. 26 No. 1, pp. 30-47, doi: 10.1080/1358684x.2018.1545109.

Diao, X., Zeng, Q., Duan, H., Lu, F. and Zhou, C. (2018), "Personalized exercise recommendation driven by learning objective within e-learning systems", *International Journal of Performability Engineering*, Vol. 14, pp. 2532-2544.

Essafi, M., Belfakir, L. and Moubtassime, M. (2024), "Investigating mobile-assisted language learning apps: Babbel, Memrise, and Duolingo as a case study", *Journal of Curriculum and Teaching*, Vol. 13 No. 2, pp. 197-215, doi: 10.5430/jct.v13n2p197.

Gunawardena, M., Bishop, P. and Aviruppola, K. (2024), "Personalized learning: the simple, the complicated, the complex and the chaotic", *Teaching and Teacher Education*, Vol. 139, 104429, doi: 10.1016/j.tate.2023.104429.

Jeon, J. (2022), "Exploring a self-directed interactive app for informal EFL learning: a self-determination theory perspective", *Education and Information Technologies*, Vol. 27 No. 4, pp. 5767-5787, doi: 10.1007/s10639-021-10839-y.

John, J. and Lo, Y.Y. (2024), "Improving English speaking proficiency among non-English major learners in Malaysia through mobile language learning applications", *Pertanika Journal of Social Science and Humanities*, Vol. 32, pp. 1599-1618.

Karjo, C.H. and Andreani, W. (2018), "Learning foreign languages with Duolingo and Memrise", *Proceedings of the 2018 International Conference on Distance Education and Learning*, Beijing, China, 26-28 May.

Khlaisang, J. and Sukavatee, P. (2023), "Mobile-assisted language learning to support English language communication among higher education learners in Thailand", *Electronic Journal of e-Learning*, Vol. 21 No. 3, pp. 234-247, doi: 10.34190/ejel.21.3.2974.

Klímová, B. (2017), "Mobile phones and/or smartphones and their apps for teaching English as a foreign language", *Education and Information Technologies*, Vol. 23 No. 3, pp. 1091-1099, doi: 10.1007/s10639-017-9655-5.

Kuimova, M., Burleigh, D., Uzunboylu, H. and Bazhenov, R. (2018), "Positive effects of mobile learning on foreign language learning", *TEM Journal*, Vol. 7, pp. 837-841, doi: 10.18421/tem74-22.

Lee, J.-C. and Xiong, L. (2023), "Exploring learners' continuous usage decisions regarding mobile-assisted language learning applications: a social support theory perspective", *Education and Information Technologies*, Vol. 28 No. 12, pp. 16743-16769, doi: 10.1007/s10639-023-11884-5.

Li, K.C., Lee, L.Y.K., Wong, S.L., Yau, I.S.Y. and Wong, B.T.M. (2018), "Mobile learning in nursing education: catering for students and teachers' needs", *Asian Association of Open Universities Journal*, Vol. 12 No. 2, pp. 171-183, doi: 10.1108/aaouj-04-2017-0027.

Li, K.C., Sun, K.P., Wong, B.T.M. and Wu, M.M.F. (2025a), "Framework development for evaluating the efficacy of mobile language learning apps", *Electronics*, Vol. 14 No. 8, p. 1614, doi: 10.3390/electronics14081614.

Li, K.C., Wong, B.T.M., Kwan, R. and Cheung, S.K.S. (2025b), "Guest editorial: technological mediation in learner-centred pedagogies", *Interactive Technology and Smart Education*, Vol. 22, pp. 1-2, doi: 10.1108/itse-03-2025-253.

McKay, S.L. (2003), "Toward an appropriate EIL pedagogy: Re-examining common ELT assumptions", *International Journal of Applied Linguistics*, Vol. 13, pp. 1-22, doi: 10.1111/1473-4192.00035.

Nami, F. (2020), "Educational smartphone apps for language learning in higher education: students' choices and perceptions", *Australasian Journal of Educational Technology*, Vol. 36, pp. 82-95, doi: 10.14742/ajet.5350.

Raj, K.A. and Baisel, A. (2024), "Empirical study on the influence of mobile apps on improving English speaking skills in school students", *World Journal of English Language*, Vol. 14, pp. 339-348.

Rosell-Aguilar, F. (2017), "State of the app: a taxonomy and framework for evaluating language learning mobile applications", *CALICO Journal*, Vol. 34 No. 2, pp. 243-258, doi: 10.1558/cj.27623.

Rosell-Aguilar, F. (2018), "Autonomous language learning through a mobile application: a user evaluation of the busuu app", *Computer Assisted Language Learning*, Vol. 31 No. 8, pp. 854-881, doi: 10.1080/09588221.2018.1456465.

Sakalauskė, A. and Leonavičiūtė, V. (2022), "Strategic analysis of Duolingo language learning platform", *Science - Future of Lithuania*, Vol. 14, pp. 1-9, doi: 10.3846/mla.2022.17731.

Sewagegn, A.A. (2020), "Learning objective and assessment linkage: its contribution to meaningful student learning", *Universal Journal of Educational Research*, Vol. 8 No. 11, pp. 5044-5052, doi: 10.13189/ujer.2020.081104.

Solmaz, O. (2025), "Impacts of digital applications on emergent multilinguals' language learning experiences: the case of Duolingo", *Education and Information Technologies*, Vol. 30 No. 7, pp. 9185-9214, doi: 10.1007/s10639-024-13185-x.

Sun, X., Ståhl, D., Sandahl, K. and Kessler, C. (2026), "Quality assurance of LLM-generated code: addressing non-functional quality characteristics", *Journal of Systems and Software*, Vol. 238, 112885, doi: 10.1016/j.jss.2026.112885.

Ulfiah, U., Rasyadan, M.F.R., Utami, W.T., Sunardi, S. and Murad, D.F. (2025), "Impact of usability on continuance usage intention in language learning apps with gamification features", *Bulletin of Electrical Engineering and Informatics*, Vol. 14 No. 1, pp. 696-705, doi: 10.11591/eei.v14i1.8023.

Walkington, C. and Bernacki, M.L. (2020), "Appraising research on personalized learning: definitions, theoretical alignment, advancements, and future directions", *Journal of Research on Technology in Education*, Vol. 52 No. 3, pp. 235-252, doi: 10.1080/15391523.2020.1747757.

Wan, H., Zhang, X., Yang, X. and Li, S. (2024), "Which approach is effective: comparing problematization-oriented and structuring-oriented scaffolding in instructional videos for programming education", *Education and Information Technologies*, Vol. 29 No. 14, pp. 17807-17823, doi: 10.1007/s10639-024-12550-0.

Wardoyo, C. (2018), "Using classroom presentation technique in teaching speaking explanation text in senior high school", *Retain: Journal of Research in English Language Teaching*, Vol. 6 No. 1, pp. 34-42.

Wong, B.T.M. (2021), "A survey on the pedagogical features of language MOOCs", *Asian Association of Open Universities Journal*, Vol. 16 No. 1, pp. 116-128, doi: 10.1108/aaouj-03-2021-0028.

Corresponding author
Billy T.M. Wong can be contacted at: tamiwong@hkmu.edu.hk