

Article

Framework Development for Evaluating the Efficacy of Mobile Language Learning Apps

Kam-Cheong Li , Ka-Pik Sun , Billy T. M. Wong  and Manfred M. F. Wu

Institute for Research in Open and Innovative Education, School of Open Learning, Hong Kong Metropolitan University, Homantin, Kowloon, Hong Kong, China; kcli@hkmu.edu.hk (K.-C.L.); tamiwong@hkmu.edu.hk (B.T.M.W.); mmfwu@hkmu.edu.hk (M.M.F.W.)

* Correspondence: kpsun@hkmu.edu.hk

Abstract: Mobile-assisted language learning (MALL) has emerged as a powerful tool for language education, offering flexibility, multimedia integration, and personalized learning experiences. Despite its growing adoption, most studies have focused on user perceptions and learning outcomes, with limited attention given to systematically evaluating the design, content, and pedagogical efficacy of mobile language learning apps (MLLAs). To address this gap in the effective design of MALL tools, this study developed an evaluation framework by integrating and refining elements from three established models in the field. The framework is organized into four dimensions: background and characteristics, app design, app content, and app pedagogy. It incorporates objective criteria alongside a standardized scoring system (0–2) to ensure consistent and systematic evaluations. The resulting framework provides researchers and educators with a tool to analyze and compare MLLAs based on their alignment with effective teaching and learning principles. This study contributes to the advancement of MALL app evaluation, supporting their development and improving teaching practices and learner outcomes.

Keywords: language learning; mobile learning app; mobile-assisted language learning (MALL); evaluation; framework; efficacy



Academic Editor: Christos J. Bouras

Received: 28 February 2025

Revised: 2 April 2025

Accepted: 15 April 2025

Published: 16 April 2025

Citation: Li, K.-C.; Sun, K.-P.; Wong, B.T.M.; Wu, M.M.F. Framework Development for Evaluating the Efficacy of Mobile Language Learning Apps. *Electronics* **2025**, *14*, 1614. <https://doi.org/10.3390/electronics14081614>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mobile learning has become an increasingly prominent focus in educational research, particularly in the growing field of language education [1,2]. Mobile-assisted language learning (MALL) enables flexible and independent access to educational materials and multimedia resources, supports the continuity of study across various devices, ensures adaptability to personal study habits, and is enhanced by teacher guidance and assistance [3,4]. Among the research on MALL, most of these studies have investigated the effectiveness of mobile learning in improving language proficiency [5–7], students' acceptance of mobile technologies for language learning [8–10], and the advantages and challenges associated with adopting m-learning in language education [11–13]. This phenomenon has also been acknowledged by Singh and Suri [14], Kessler et al. [15], Fang and Chew [16], and Moghaddam et al. [17], who highlighted similar findings in their study on research trends in MALL. While these studies have provided valuable insights, they predominantly relied on surveys or interviews to assess user perceptions and learning outcomes [16–20]. This left an important research lacuna in the systematic evaluation of the design, functionality, and pedagogical efficacy of language learning mobile apps themselves.

To address this gap, this paper aims to develop an evaluation framework that can serve as a foundation for analyzing the efficacy of mobile language learning apps (MLLAs).

Unlike previous studies that focused on user experiences and outcomes [7,9,12], this paper proposes a structured evaluation framework, derived from a review of the relevant literature [21–23]. The development process is described step by step. This framework examines the background and characteristics of these apps, including their accessibility and learning pathways. As noted by Almaiah et al. [21], many studies lacked clear and well-defined standards and rubrics for assessing mobile learning apps from a technical perspective. Therefore, the inclusion of design as a key evaluation aspect is essential. Accordingly, this framework analyzes the app design as well, particularly interface design, multimedia integration, user support, and offline functionality. It also evaluates the quality, variety, and relevance of the content offered by these apps and assesses their pedagogical approaches, analyzing their incorporation of effective teaching and learning strategies. By applying this multidimensional framework, further studies aiming to evaluate the efficacy of MLLAs can systematically analyze how these apps align with established principles of effective language learning and technology integration in generalized perspectives. This approach offers insights into the potential of mobile technologies to enhance both teaching practices and student learning outcomes in language education.

The next section on reviews of previous studies introduces prior papers that focused on MALL and evaluation frameworks for MLLAs. The methodology section outlines the methods and major steps taken to identify and analyze existing evaluation frameworks from prior research, which served as references for developing the proposed framework. The results section presents the development process of the framework, detailing the steps for establishing the criteria across the four dimensions used to evaluate MLLAs. The conclusion section introduces how the framework was formed, emphasizing its potential applications in future research and practice. The limitations of the framework are addressed and areas for further refinement and improvement are recommended.

2. Reviews of Previous Studies

A growing number of researchers have conducted studies evaluating the effectiveness of mobile apps for language learning, reflecting the increasing prominence of MALL as a significant focus within academic research [14]. With the advancement of MALL, there has been a significant increase in the development and availability of MLLAs on the market [24,25]. This proliferation highlights the growing significance of MLLAs in the field of language education. Nonetheless, analysis and evaluation on the MLLAs is still at an early stage [14]. There is a growing need for a comprehensive domain-specific framework for educators and researchers to systematically evaluate and investigate their efficacy. Many researchers have developed their own evaluation frameworks or refined existing ones proposed by other scholars [22,26]. These frameworks were often applied to the examination of various aspects of MLLAs, such as their background, features, quality, and overall efficacy [21,27]. In some cases, scholars applied these frameworks to analyze specific apps as examples, demonstrating the effectiveness of these tools in supporting and contributing to language learning [2,22]. The following section reviews the relevant studies in this area. The prominent frameworks proposed to date are outlined below.

2.1. Evaluation Framework for MALL

By collecting data from user reviews and ratings on the Google Play Store, Singh and Suri [14] proposed a framework for evaluating the user experience of MLLAs. The framework was divided into two broad themes: App Quality, which included sub-themes such as technical quality, customer support quality, content quality, and teaching quality, and App Suitability, with usefulness as its sub-theme. Each sub-theme was accompanied by an operational definition and common markers, allowing for scoring on a scale from 1 to 3,

where 1 represented “low” and 3 represented “high”, with brief descriptions provided for each score. Singh and Suri [14] highlighted that technical quality, customer support quality, content quality, teaching quality, usefulness, compatibility, and the ability of learners to influence others collectively shape user experience, which subsequently affects the star ratings of MLLAs. Although this framework provided a structured approach to evaluation, it primarily focused on the app quality and app suitability, without including an evaluation of app design. In addition, the sub-themes within these categories were presented without detailed instructions, specific criteria, or a clearly organized checklist. As a result, the broad categorization of the framework may present challenges for researchers aiming to conduct a more comprehensive and detailed analysis of mobile apps.

Apart from using user experience as the basis for evaluation, Baloh et al. [26] and Paul et al. [27] proposed structured frameworks for assessing mobile learning apps. The framework of Baloh et al. [26] consisted of four levels: Technical, which included functionality, security, and performance; Educational, which covered pedagogical and usability aspects; Economic, which included support and service level; and Socio-cultural, which focused on communication and portability. Each category contained distinct properties for evaluation. Baloh et al. [26] identified overlaps between the properties of operability and learnability as well as the adequacy and just-in-time knowledge when evaluating the quality of mobile learning apps. This observation underscored the importance of addressing these overlaps in future research to prevent the inclusion of redundant or repetitive criteria in evaluation frameworks, thereby enhancing their precision and effectiveness. Similarly, the framework by Paul et al. [27] identified seven usability factors: Effectiveness, Understandability, Efficiency, Learnability, Operability, Serviceability, and Satisfaction, each with its own set of sub-factors. They emphasized effectiveness, productivity, safety, and satisfaction when evaluating mobile learning apps [27]. While both frameworks provide a clear and organized structure for analyzing mobile learning apps, they shared a common limitation: they provided neither detailed descriptions nor instructions for evaluating the properties or sub-factors. This lack of specificity made it challenging for researchers to effectively apply these frameworks when analyzing whether a mobile learning app meets the criteria.

Some scholars focus primarily on developing and proposing frameworks for evaluation, whereas others emphasize applying these frameworks to the assessment of specific apps. Sakalauskė and Leonavičiūtė [2] evaluated Duolingo, a popular language learning app that provides courses accessible through both its website and mobile app [2]. Their evaluation was structured around three aspects: first, they described the background and characteristics of Duolingo; they also applied the Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis; and lastly, they utilized the Value, Rarity, Inimitability, and Organization (VRIO) method. This framework was detailed and organized, as it went beyond usability to incorporate background factors such as app store ratings and the number of language courses offered, which could also influence the efficacy of the app. However, a limitation of this study was that it focused exclusively on Duolingo without applying the framework to other apps for a comparative analysis, which could have provided broader insights.

2.2. Key References for Framework Development

Three key studies, Essafi et al. [22], Almaiah et al. [21], and Rosell-Aguilar [23], were used as foundational references to develop the evaluation framework in this study, as they developed their own structured frameworks with detailed criteria.

Essafi et al. [22] conducted an evaluation of three MLLAs: Babbel, Memrise, and Duolingo. The analysis initially addressed Google Play ratings, language focus, learning

content, and progression pathways to provide contextual information about the apps. A structured framework was then developed to assess their efficacy, using key elements from the literature reviewed by Essafi et al. [22]. This ensured that the criteria were grounded in established research. The framework focused on three core aspects: design, content, and pedagogy. Each aspect was divided into four categories with two criteria per category to ensure a lucid and measurable evaluation process. Rather than relying on a checklist, the framework employed a numerical rating scale ranging from 0 (lowest) to 2 (highest), enhancing both the effectiveness and accuracy of the evaluation method and its results.

In contrast, Almaiah et al. [21] employed the Delphi method and collected data from 30 experts in Software Engineering, Information Systems, and the Mobile Learning field over the course of three rounds. In the first round, the experts compiled a list of quality dimensions accompanied by 21 technical requirements. In the second round, the experts evaluated and scored each quality dimension, resulting in a refined list of 19 technical requirements. Finally, in the third round, the experts revised their feedback and scores, settling on the final list of 19 technical requirements categorized under six quality dimensions: Interactivity, Functionality, Interface Design, Accessibility, Learning Content Quality, and Content Design Quality. By employing the Delphi method, this framework benefited from the integration of technical and professional expertise, providing a more rigorous and specialized approach to evaluating mobile learning apps which distinguished it from other frameworks.

Rosell-Aguilar [23] proposed a framework beginning with a taxonomy of apps associated with language learning and their role in language acquisition. This taxonomy classified apps into three types: those specifically designed for language learning, those not explicitly created for this purpose but still beneficial for language acquisition, and those that were irrelevant to this study, which are therefore excluded from discussion. Building on this taxonomy, Rosell-Aguilar [23] developed an evaluation framework divided into four categories: technology, user experience, pedagogy, and subject-specific content. To ensure its validity, the framework was tested and reviewed by educational professionals twice, thereby guaranteeing that it included sufficient and appropriate criteria for evaluating the pedagogy employed by the apps.

The methodologies used in each of these studies are introduced in this section, while their respective frameworks will be discussed in detail in the results section. Specifically, the framework proposed by Essafi et al. [22] serves as the core structure of the evaluation framework developed in this study. The studies by Almaiah et al. [21] and Rosell-Aguilar [23] address critical gaps and missing criteria in Essafi et al. [22], contributing to a more integrated and cohesive design.

The previous reviews described above examine the methods used by scholars to evaluate the design and content of MLLAs currently available on the market. While each study proposed its own evaluation framework or adapted those of others, there are notable similarities and differences among them. However, a common limitation was that they typically analyzed only one or two apps, which might not have offered a sufficiently generalized perspective. Therefore, this study aims to improve previous frameworks by designing a comprehensive evaluation model to assess the efficacy of MLLAs with high generalizability.

3. Materials and Methods

Rather than developing an entirely new evaluation framework, this study refines and integrates existing models to ensure a more comprehensive and objective assessment. As highlighted in the section of reviews of previous studies, most evaluation frameworks focus

on only one or two aspects, lacking a holistic perspective. Developing a completely new framework risks bias and limited applicability. However, previous studies have produced valuable and well-established insights, making them a solid foundation for reference. By scrutinizing existing frameworks, this study leveraged their strengths while addressing their limitations, ensuring that the proposed framework is both robust and applicable to a wide range of MALL evaluation contexts.

Therefore, the processes to develop an evaluation framework for MALL apps began with a systematic review of previous studies. It focused on identifying and analyzing existing frameworks that could serve as references for this research. By carefully selecting and refining these frameworks, the methodology aimed to ensure that the resulting evaluation framework not only builds on established principles but also addresses gaps identified in prior studies. The steps involved in this process are outlined in detail below.

The development of the evaluation framework began with a thorough review of previous studies related to MALL. The search was conducted on the Scopus database using the keywords “MALL”, “language learning”, “framework”, and “evaluation”. The review process involved three successive screening steps to identify the most suitable frameworks from previous studies, which would serve as key reference papers for developing the new evaluation framework proposed in this study.

The first screening aimed to identify papers that included the specified keywords in their titles or abstracts. While this step retrieved numerous studies, some papers included the keywords but did not propose frameworks or criteria for evaluating the effectiveness of MLLAs; therefore, they were excluded due to their lack of relevance. To refine the selection, a second screening was conducted to focus on papers that specifically proposed frameworks or criteria for analyzing MLLAs. After this second screening, six papers were identified as relevant and retained for further evaluation.

The third and final screening further assessed the selected papers to determine their suitability as references for this study. This step excluded papers that lacked well-structured or detailed frameworks with clear evaluation rubrics, criteria, or scoring methods. It was noted that while some papers offered broad dimensions or themes for evaluation, they failed to provide specific and actionable criteria, making their frameworks difficult to apply effectively. For example, the framework proposed by Singh and Suri [14] divided the evaluation into two broad categories—the quality and suitability of the app. However, the sub-themes within these categories lacked clear rubrics, leaving evaluators without detailed guidance and potentially causing frustration when applying the framework. Similarly, the frameworks by Baloh et al. [26] and Paul et al. [27] were relatively comprehensive, as they addressed multiple aspects of app evaluation. Nevertheless, their frameworks did not include clear and concise criteria or checklists for assessing each category, which could lead to confusion for evaluators attempting to use them. The content of these three papers has been introduced in detail in the “Reviews of Previous Studies” section.

After this third screening, three papers were identified as the primary reference sources for the development of the new evaluation framework. These key papers were authored by Essafi et al. [22], Almaiah et al. [21], and Rosell-Aguilar [23]. The framework proposed by Essafi et al. [22] was adopted as the central structure of the new framework, providing the primary basis for its design. Meanwhile, the frameworks developed by Almaiah et al. [21] and Rosell-Aguilar [23] offered supplementary criteria that addressed specific gaps and limitations within the framework of Essafi et al. [22]. The details regarding the development of this framework are presented in the results section.

4. Results and Discussion

The framework was derived, refined, and improved by integrating and adapting key elements from three existing frameworks identified as particularly relevant in the methodology section. Table A1 in the Appendix A presents a comparison of the three evaluation frameworks. These frameworks were analyzed to extract their strengths while addressing their limitations to create a more systematic tool. Through this process, the new framework was tailored to address the specific needs of evaluating MLLAs comprehensively. It avoids redundancy in criteria, incorporates commonly discussed factors from the scholarly literature [23], and ensures that it is concise, organized, and easy to use. The steps taken to retrieve, combine, and enhance these frameworks are introduced below.

Step 1. Main Structure from Essafi et al.

Essafi et al. [22] proposed a systematic framework for evaluating MLLAs which is organized into three main aspects: app design, app content, and app pedagogy. Each aspect was defined by four criteria and standards, assessed using a scoring system that ranges from low (0) to high (2) based on the extent to which the criteria are met in the apps. The framework by Essafi et al. [22] is shown in Table 1.

Table 1. Summarized evaluation framework of language learning apps [22] (pp. 207, 209–210).

Category		Criteria
App design	Multimedia	the app (i) utilizes different forms of multimedia; (ii) in a didactic and meaningful way.
	Offline mode	the app (i) functions online and offline; and (ii) offers learning materials for download.
	In-app advertising	the app (i) is generous with free content; and (ii) does not contain recurring disturbing ads.
	App support	the app provides (i) various channels for support; as well as (ii) instant and personalized responses.
App content	Learning objectives	these (i) are clear and aligned with the course items; (ii) are achievable and measurable.
	Learning content	the learning content (i) is rich and accurate; (ii) is logically built and structured.
	Learning activities	these (i) align with the Present-Practice-Test (PPT) model; (ii) are varied and interesting.
	Targeted skills	the app (i) effectively teaches their targeted skills; (ii) meaningfully integrates the other skills.
App pedagogy	Customization	the app offers (i) placement tests; and (ii) ease of access to different difficulty levels.
	Gamification	(i) the app utilizes gamified features; (ii) these have a pedagogical added value for the app users.
	Scaffolding	app users (i) can monitor their learning; (ii) are scaffolded and offered instant and detailed feedback.
	Interaction	the app (i) targets culture; and (ii) facilitates collaboration between the app users.

The app design aspect, which evaluated the structural and functional elements of MLLAs, was composed of the following four features:

- Multimedia, which requires that the app incorporates diverse forms of multimedia, such as audio, video, and images, and utilizes them effectively to enhance students' engagement, sustain their focus, and motivate active participation in learning activities;
- Offline Mode, which focuses on the app's ability to operate both online and offline and to provide downloadable materials for users to access without an internet connection;
- In-App Advertising, which ensures the free content the app provides and the absence of recurring or disruptive advertisements that could negatively impact the user experience;
- App Support, which requires the app to offer multiple support channels, such as FAQs or live chat, and delivers prompt and personalized assistance to users.

The app content aspect, which assesses the educational material and activities offered by MLLAs, includes four features:

- Learning Objectives, which requires that the app's learning goals are clearly defined and aligned with its course content and are achievable and measurable, ensuring that learners can track their progress effectively;
- Learning Content, which evaluates whether the content is rich in detail, accurate, and logically organized to provide a structured and cohesive learning experience;
- Learning Activities, which examines whether the app's activities follow the Present-Practice-Test (PPT) model and whether they are diverse and engaging to maintain learner motivation;
- Targeted Skills, which considers whether the app effectively teaches the specific skills it targets (e.g., vocabulary or grammar) and integrates other language skills in a meaningful way.

The final aspect, app pedagogy, evaluates the pedagogical strategies embedded within MLLAs to enhance the learning experience. It comprises four features:

- Customization, which emphasizes that the app provides placement tests to determine users' proficiency levels and offers easy access to content across various difficulty levels to enable tailored learning experiences;
- Gamification, which assesses whether the app incorporates game-like elements such as point systems, unlockable levels, virtual rewards, leaderboards, or competitions, and ensures that these features add educational value to enhance the learning process rather than serving as mere entertainment;
- Scaffolding, which requires that the app enables users to monitor their progress through tracking tools and provides support in the form of detailed, instant feedback to guide learners as they advance;
- Interaction, which evaluates whether the app enables learners to apply the knowledge in real-life contexts, such as travel or the workplace, while also fostering collaboration among users, encouraging social engagement, and promoting cooperative learning.

Essafi et al. [22] proposed a rounded and structured framework for evaluating MLLAs. They not only examined the content provided by the app, focusing on the nature of the learning material and the methods used to present it, but also evaluated the design of the app, including its interface and functionality. Furthermore, this framework incorporates a clear, three-level scoring method that minimizes complexity and avoids overwhelming evaluators with excessive grading scales. Therefore, this framework was adopted as the core structure of the evaluation model presented in this paper due to its suitability and effectiveness. Nonetheless, its assessment of interface design and functionality requires further elaboration for enhancement, as these two aspects were less emphasized in the framework.

Step 2. Additional Criteria from Almaiah et al.

To address the inadequacies in the evaluation of design and functionality, the framework by Almaiah et al. [21] is introduced, as it provides additional criteria for app design. Almaiah et al. [21] proposed this framework for evaluating mobile learning apps using the Delphi method, and Table 2 outlines the technical requirements under each quality dimension.

Table 2. List of technical quality dimensions and requirements [21] (pp. 11–12).

Quality Dimensions	Technical Requirements Items
Interactivity	1. The app enables learners to interact with instructors via online messages.
	2. The app enables learners to exchange and share the learning content.
	3. The app enables learners to discuss with learners and faculty by using discussion board.
Functionality	4. Learners can easily navigate between tasks.
	5. The app gives learners alerts for new notifications.
	6. Both students and instructors can access to the app.
	7. The app gives learners sufficient features.
Interface Design	8. The app provides a simple and flexible user-interface with a good icons design.
	9. Learners can easily identify the particular functions of the app.
	10. The app offers good organization of course content and activities.
Accessibility	11. Instructors can create courses and learning content items.
	12. Instructors and learners can access the documents of learning content in multiple formats.
	13. Instructors can upload and download attachments.
	14. Learners can submit assignments and homework.
Learning Content Quality	15. Learners can find the complete learning content when using the app.
	16. Learners can find the various activities of learning content when using the app.
Content Design Quality	17. The app provides learners different formats of learning content such as text, audio and video.
	18. The app provides learners up-to-date content.
	19. The app provides learners accurate content.

The evaluation framework in this study focuses on analyzing mobile apps designed to support self-directed learning. These apps operate without any guidance or teaching from instructors and tutors or with the inclusion of online or live lessons [28]. Therefore, any technical requirement items that depend on interaction with, or actions from, instructors were excluded from the framework. As a result, the following seven items were excluded from the framework. The remaining twelve technical requirement items are presented in Table 3.

- 1. The app enables learners to interact with instructors via online messages;
- 6. Both students and instructors can access the app;
- 11. Instructors can create courses and learning content items;

- 12. Instructors and learners can access the documents of learning content in multiple formats;
- 13. Instructors can upload and download attachments;
- 14. Learners can submit assignments and homework;
- 15. Learners can find the complete learning content when using the app.

Further exclusions were made to eliminate overlaps between the above framework and the framework structure outlined by Essafi et al. [22]. The framework proposed by Essafi et al. [22] clearly organized quality dimensions into three broader aspects: app design, app content, and app pedagogy, making it easier for readers to analyze apps in a structured manner. While the framework by Almaiah et al. [21] provided detailed rubrics specifying which quality dimensions included specific technical requirement items, these dimensions could be reorganized into the three aspects defined by Essafi et al. [22]. Thus, to avoid redundancy, the eight technical requirements, which included number 2, 3, 4, 7, 10, 16, 17, and 19 on the item list that aligned with criteria in the framework by Essafi et al. [22], were excluded.

After two rounds of exclusion, four items remained and were incorporated into Essafi et al.'s framework [22] as follows:

Under "App design":

- 5. The app gives learners alerts for new notifications;
- 8. The app provides a simple and flexible user-interface with a good icons design;
- 9. Learners can easily identify the particular functions of the app.

Under "App content":

- 18. The app provides learners up-to-date content.

The four technical requirement items above were identified as important criteria missing from Essafi et al.'s framework [22] and were added to it to address this gap and make the framework more extensive. Almaiah et al.'s framework [21] primarily focused on the technical aspects of evaluation, such as design and content presentation, but lacked a thorough analysis of pedagogical aspects. Therefore, integrating its strengths with Essafi et al.'s framework [22] ensures a more comprehensive evaluation model.

Table 3. Technical requirements items adopted after first round of exclusion.

Quality Dimensions	Technical Requirements Items
Interactivity	2. The app enables learners to exchange and share the learning content.
	3. The app enables learners to discuss with learners and faculty by using discussion board.
Functionality	4. Learners can easily navigate between tasks.
	5. The app gives learners alerts for new notifications.
	7. The app gives learners sufficient features.
Interface Design	8. The app provides a simple and flexible user-interface with a good icons design.
	9. Learners can easily identify the particular functions of the app.
	10. The app offers good organization of course content and activities.
Learning Content Quality	16. Learners can find the various activities of learning content when using the app.

Table 3. Cont.

Quality Dimensions	Technical Requirements Items
Content Design Quality	17. The app provides learners different formats of learning content such as text, audio and video.
	18. The app provides learners up-to-date content.
	19. The app provides learners accurate content.

Step 3. Additional Criteria from Rosell-Aguilar

The combined framework based on Essafi et al. [22] and Almaiah et al. [21] presents extensive criteria for evaluating the design and content of language learning apps. While it incorporates multiple criteria related to pedagogy, there remains room for further enhancement in this area. Rosell-Aguilar [23] introduced a framework that offered more valid and reliable criteria for assessing pedagogy, having been rigorously tested and reviewed by teaching professionals. Rosell-Aguilar's framework [23] presented questions rather than rubrics with detailed descriptions, which might make it less systematic and structured for use as a core reference framework. However, integrating this framework strengthens the overall evaluation process by complementing the criteria established in the final framework. The framework by Rosell-Aguilar [23] was organized into four main categories: language learning, pedagogy, user experience, and technology. Each category contains a set of criteria formulated as questions, with a total of 31 designed to assess various aspects of an app's functionality and effectiveness.

The language learning category evaluates whether the app supports essential language skills such as reading, listening, writing, speaking, vocabulary, grammar, and pronunciation. It also considers whether the app includes cultural information, diverse visual content, and language varieties. The pedagogy category examines the app's educational value, focusing on alignment between its description and functionality, its approach to teaching (e.g., modeling versus testing), progress tracking, scaffolding of activities, feedback quality, content accuracy, use of media, differentiation for various skill levels, and user engagement. The user experience category assesses features like user interaction, the level of interactivity with content, options for sharing, badging for recognition, pricing structure, registration requirements, and the presence of distracting advertisements. Finally, the technology category evaluates the app's interface clarity, navigational ease, availability of instructions, stability, gamification features, support options, and whether the app works offline. Together, this comprehensive framework provides a detailed tool for analyzing the effectiveness, usability, and educational potential of language learning apps.

Drawing on the structured frameworks established by the previous two framework, only four criteria from the thirty-three identified by Rosell-Aguilar [23] were incorporated into the final framework. These criteria were selected because they were not included in the combined framework derived from Essafi et al. [22] and Almaiah et al. [21], and their inclusion helps to address gaps in the framework by improving coverage of the pedagogical aspects. They are detailed as follows:

Under "App design":

- "Instructions" from the "Technology" aspect evaluated whether the app provides clear guidance to users, ensuring usability and accommodating first-time users who might not understand how to navigate or operate the app.

Under "App content":

- “Teaching” from the “Pedagogy” aspect examined whether the app offers step-by-step teaching and sufficient explanations of the language and the material it covers before presenting tests or assessments;
- “Quality of content” from the “Pedagogy” aspect assessed the accuracy of the app’s content, ensuring that it is free from errors and suitable for effective learning;
- “Cultural information” from the “Language Learning” aspect considered whether the app provides cultural and contextual information related to the target language. Understanding cultural aspects enhances learners’ comprehension of the language’s structure and increases engagement and interest in the learning process.

The primary structure of the framework, including its core aspects and criteria, was adapted from the systematic model proposed by Essafi et al. [22], while the frameworks presented by Almaiah et al. [21] and Rosell-Aguilar [23] provided additional details to finalize its design. By integrating and refining these elements, the revised framework aims to resolve overlaps, address gaps, and offer a cohesive and systematic structure for evaluating the efficacy of MLLAs. The finalized evaluation framework is presented in Table 4.

Before evaluating the design, content, and pedagogy of MLLAs, it is essential to examine their background and characteristics, as this foundational understanding may provide critical context for interpreting their performance, reception, and potential impact. The background and characteristics section included six criteria: Google Play downloads, Google Play ratings and Apple App Store ratings, year founded, headquarters location, number of language courses offer, and learning structure.

Table 4. Evaluation framework of mobile language learning apps.

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) utilizes different forms of multimedia.
		(2) utilizes 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 personalized responses.
		(3) provides instructions on using.
		(4) gives learners alerts for new notifications.
	Interface design	(1) provides good icon design in user interface.
		(2) identifies the particular functions easily.

Table 4. Cont.

Category		Criteria
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, not only test it.
		(3) includes background information about the customs and traditions of the language (or the area where the language is spoken).
		(4) error-free.
	Learning activities	(1) aligns with the Present-Practice-Test (PPT) model.
		(2) varied and interesting.
(3) offers structured revision test.		
Targeted skills	(1) effectively teaches their targeted skills.	
	(2) meaningfully integrates the other skills.	
App updates	(1) features and contents update regularly.	
App pedagogy	Personalization	(1) offers placement tests.
		(2) offers ease of access to different difficulty levels.
	Gamification	(1) utilizes gamified features (e.g., unlock levels, earn points, rewarded with virtual tokens, win virtual currency, compete in leaderboards).
		(2) gamified features have a 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 the app users (e.g., content exchange and sharing, discussion board).		

The above background information criteria provided objective insights into the performance and reception of MLLAs. Google Play downloads highlighted the app's achievement and widespread usage, as the number of downloads reflects its popularity and adoption among users [29]. The founder of Duolingo celebrated the app as the most downloaded educational app in the history of the Apple App Store [2], emphasizing the strong correlation between high download numbers and an app's impact. Although the Apple App Store does not display download statistics, incorporating Google Play downloads has offered valuable context by showcasing the scale of user engagement with the app. Google Play ratings and Apple App Store ratings were included as straightforward indicators of user satisfaction and app usability as well. These metrics, referenced by Essafi et al. [22] when introducing example apps, provided accessible and objective measures of an app's effectiveness and user experience.

The year that the app was founded was added to highlight the app's history, which might reflect the maturity and development of its features over time. The headquarters location was included as a criterion because teaching methods can vary across countries and regions due to cultural influences. Joy and Kolb [30] demonstrated that culture, defined in their study as countries, significantly impacts learning styles. Therefore, identifying

the headquarters location offers general implications regarding the potential pedagogical approaches and design principles of each app.

The number of languages on offer was also included, as language variety is a key criterion in evaluating a language learning app. This criterion, as noted by Rosell-Aguilar [23], emphasized the regional or national varieties of a language that an app provides. Finally, learning structure was incorporated to examine how each app designed and organized its courses, as every app offers a unique approach to structuring its learning content. Since the categories in the background and characteristics section consist solely of objective data, they were not scored like the other parts of the framework; instead, they were recorded as factual data to provide context for the evaluation.

Under the app content section, the criterion “offers structured revision test” has been added to the category of “Learning activities”. Kuimova et al. [3] highlighted that mobile language learning should be consistent and systematic to enhance retention and consolidate memory. Structured revision tests align with this principle by helping users to review and assess their knowledge after completing a unit or chapter, ensuring their learning progresses in an organized and effective manner. Thus, the criterion is added to assess whether the app facilitates structured and consistent learning through revision.

For app design, app content, and app pedagogy aspects, the evaluation followed a standardized rating scale, ranging from 0 to 2, as adopted from the framework proposed by Essafi et al. [22]. The implementation of this numerical rating system ensured the validity and reliability of the evaluation rubric and the results derived from it [22]. In this scale, a score of 0 represents a low level of fulfillment, 1 represents a medium level, and 2 represents a high level of fulfillment for each criterion. Consequently, the “Score” column in each of these three parts must be completed using one of these values (0,1, or 2), ensuring a consistent and systematic evaluation process across all criteria. To ensure objectivity and minimize bias in the evaluation process, users are encouraged to document their assessment procedures and provide a rationale for the scores assigned to each criterion. To demonstrate the feasibility and practicality of the developed framework, this study has implemented this using a MLLA available on the market. The results highlight the functionality of the scoring system and provide examples of criteria evaluation based on assigned scores. Table A2 in the Appendix A presents the demonstration for user reference.

In brief, the proposed evaluation framework integrates the strengths of the three existing frameworks, offering a comprehensive model for assessing MLLAs. This framework ensures a balanced evaluation by addressing key aspects of app design, content, and pedagogy, thereby accommodating both its functionality as a mobile app and its effectiveness as a learning tool. By incorporating a structured scoring system, it eliminates the complexities of ambiguous rating methods, providing a clear and systematic approach to evaluation. Additionally, the inclusion of detailed rubrics with explanations enhances clarity and usability, reducing the risk of misinterpretation that may arise from frameworks relying solely on question-based assessments. This contribution facilitates a more objective and user-friendly evaluation process, supporting future research and practical applications in MLLA assessment.

5. Conclusions

This study proposes a framework for evaluating MALL apps, developed through the integration and adaptation of elements from three established models. The fundamental structure of the framework was derived from the model proposed by Essafi et al., and additional perspectives from Almaiah et al. and Rosell-Aguilar were incorporated to refine and expand its scope. It is organized into four dimensions: background and characteristics, app design, app content, and app pedagogy. To facilitate consistent appraisals, a standardized

rating scale ranging from 0 to 2 has been included, providing a clear basis for assessing the quality and effectiveness of apps. By integrating these contributions, the framework addresses previous challenges, such as overlapping criteria and gaps, offering a more cohesive approach to app evaluation.

The resulting framework is designed to support the evaluation of MALL apps in research and educational practice. First, it offers educators a framework to evaluate the efficacy of MLLAs. The framework enables informed decisions in selecting apps that align with pedagogical objectives and support students' learning needs, thereby enhancing the effectiveness of language instruction. While further testing and validation are necessary, this framework provides a foundation for analyzing and comparing MLLAs, addressing gaps in existing evaluation models and potentially improving app development and evaluation methodologies. It comprehensively integrates three key evaluation aspects from previous studies and provides a clear scoring system. This framework offers a reference from more generalized perspectives for future researchers and educators seeking to assess the effectiveness of mobile language learning apps and select the most suitable one for their needs.

Despite its contributions, this study has certain limitations that should be acknowledged. First, the framework was developed by synthesizing elements from three existing models, which, while informed by the relevant literature, may not fully address all aspects of evaluating MALL apps across diverse educational contexts. Furthermore, the framework emphasizes pedagogical and usability dimensions, potentially overlooking other critical factors such as cultural adaptability, accessibility for users with disabilities, and long-term user engagement. Future research could address these limitations by applying the framework in varied contexts, such as evaluating the efficacy of MLLAs for different age groups or proficiency levels. Additionally, researchers could refine its components and incorporate feedback from diverse stakeholders, including educators, learners, and app developers. The process of selecting and integrating criteria, although based on established research retrieved from the Scopus database, may have excluded relevant studies indexed in other databases. This limitation could affect the comprehensiveness of the final framework. Future research should consider including additional databases, such as Web of Science or ERIC, to ensure broader literature coverage and reduce the potential for bias. The predefined scoring system (0–2) of this framework may not be suitable for all users or evaluation contexts; future studies may adapt the scoring system based on their specific application needs, as the system presented in this study serves as a reference rather than a fixed standard. Moreover, the framework has not yet undergone extensive empirical testing or validation with a wide range of MALL apps, which limits its applicability and generalizability. Further empirical studies could also be conducted to validate and enhance the framework's effectiveness.

Author Contributions: Conceptualization, K.-C.L., K.-P.S. and B.T.M.W.; methodology, K.-C.L., K.-P.S. and B.T.M.W.; formal analysis, K.-C.L. and K.-P.S.; writing—original draft preparation, K.-C.L. and K.-P.S.; writing—review and editing, K.-C.L., K.-P.S. and M.M.F.W.; supervision, K.-C.L., B.T.M.W. and M.M.F.W.; project administration, K.-C.L., B.T.M.W. and M.M.F.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

- MALL Mobile-assisted language learning;
- MLLA Mobile language learning app;
- SWOT Strengths, Weaknesses, Opportunities, and Threats;
- VRIO Value, Rarity, Inimitability, and Organization;
- PPT Present-Practice-Test.

Appendix A

Table A1. Comparison of the Three Evaluation Frameworks for App Assessment.

Criteria	Framework 1 Essafi et al. (2024)	Framework 2 Almaiah et al. (2022)	Framework 3 Rosell-Aguilar (2017)
Purpose	Framework development and evaluation process	Framework development	Framework development
Evaluation Aspects	App design, app content, app pedagogy	App design	App design, app content, app pedagogy
Evaluation Criteria	Rubrics with explanations	Rubrics with explanations	List of questions
Scoring System	0 (low), 1 (medium), 2 (high)	Not specified	Not specified
Empirical Validation	Tested with 3 apps	Not yet tested	Not yet tested

Table A2. Demonstration of the Evaluation Framework Using a MLLA.

Aspects	Category	Criteria	Score	Comments
App design	Multimedia integration	(1) utilizes different forms of multimedia	1	Includes pictures and audios
		(2) utilizes multimedia in a didactic and meaningful way	1	Fewer visual elements may lower user engagement
	Offline functionality	(1) functions online and offline	1	Can only access offline the lesson you are currently working on
		(2) offers learning materials for download	0	No materials are provided for download
	In-app advertising	(1) offers free content	1	Users can access free content for up to 5 min per day but must take a 10-min break after each lesson
		(2) does not contain recurring disturbing ads	2	No ads during, before or after lesson
	App support	(1) provides various channels for support	2	Provides FAQs, instant chatbot assist, and email assistance
		(2) provides instant and personalized responses	2	Instant chatbot assistance to answer your inquiries
		(3) provides instructions on using	1	Guides users on how to answer questions during the lesson, while providing no instructions on the functions
		(4) gives learners alerts for new notifications	2	Sends user an email every day at a fixed time to remind you to take a lesson
	Interface design	(1) provides good icon design in user interface	2	Concise
		(2) identify the particular functions easily	2	Easy to recognize and understand which function each icon represents
	Sub-total scores		17	
App content	Learning objectives	(1) clear and aligned with the course items	2	“Make learning awesome!” by “Captivating game-based learning for all ages”, “Get going in just 5 min a day”, “Choose from 50 languages”, “Turn language learning into a daily habit”, “Personalized learning journey”, “Learn and compete in multiplayer mode”
		(2) achievable and measurable	2	Its corresponding functions ensure that its mission is achievable

Table A2. Cont.

Aspects	Category	Criteria	Score	Comments
App content	Learning content	(1) logically built and structured	1	Focuses on daily life topics but lacking a strong and structured foundational framework
		(2) presents, explains, or models the languages, not only test it	0	Does not explain or provide a translation of the vocabulary before testing the user
		(3) includes background information about the customs and traditions of the language (or the area where the language is spoken)	0	Does not include any background information
		(4) error-free	2	No errors found
	Learning activities	(1) align with the Present-Practice-Test (PPT) model	1	Combines learning materials with quick follow-up quizzes in a review-based format
		(2) varied and interesting	1	Limited to matching and spelling exercises
		(3) offers structured revision test	0	No revision test provided
	Targeted skills	(1) effectively teaches their targeted skills	2	Focuses on vocabulary skill
		(2) meaningfully integrates the other skills	0	Does not offer any exercises that target other skills
	App updates	(1) features and contents update regularly	2	Updates are provided at least once a month
Sub-total scores			13	
App pedagogy	Personalization	(1) offers placement tests	0	No placement test provided
		(2) offers ease of access to different difficulty levels	0	The learning materials are the same regardless of the level you choose
	Gamification	(1) utilizes gamified features	1	Fewer gamified features, only retaining streaks, and unlocking new functions and challenges
		(2) gamified features have a pedagogical added value	1	Unlocking levels and achievements may not be sufficiently encouraging or motivating
	Scaffolding	(1) users can monitor their learning	1	Users can see their current level, but there is no progress bar to indicate overall learning progress
		(2) users are scaffolded and offered instant and detailed feedback	1	Users can see if their answer is correct, but it only practices vocabulary and does not explain incorrect answers

Table A2. Cont.

Aspects	Category	Criteria	Score	Comments
App pedagogy	Interaction	(1) targets culture	1	Enhances your vocabulary on daily topics but is hard to apply in real-life situations due to a lack of conversation practice
		(2) facilitates collaboration between the app users	1	Users can add friends and message them directly within the app, but there is no discussion forum for exchanging ideas or sharing content
	Sub-total scores		6	
	Total scores		36	

References

1. Al Mulhem, A.; Almaiah, M.A. A Conceptual Model to Investigate the Role of Mobile Game Applications in Education during the COVID-19 Pandemic. *Electronics* **2021**, *10*, 2106. [\[CrossRef\]](#)
2. Sakalauskė, A.; Leonavičiūtė, V. Strategic Analysis of Duolingo Language Learning Platform. *Sci.–Future Lith.* **2022**, *14*, 1–9. [\[CrossRef\]](#)
3. Kuimova, M.; Burleigh, D.; Uzunboyly, H.; Bazhenov, R. Positive Effects of Mobile Learning on Foreign Language Learning. *TEM J.* **2018**, *7*, 837–841. [\[CrossRef\]](#)
4. Loewen, S.; Crowther, D.; Isbell, D.R.; Kim, K.M.; Maloney, J.; Miller, Z.F.; Rawal, H. Mobile-Assisted Language Learning: A Duolingo Case Study. *ReCALL* **2019**, *31*, 293–311. [\[CrossRef\]](#)
5. Kravchenko, O.; Dokuchaieva, V.; Sbitnieva, L.; Sakhatska, V.; Akinshyna, I. Effectiveness of Generative Learning Strategies Based on Mobile Learning Technologies in Higher Education. *Int. J. Eval. Res. Educ.* **2024**, *13*, 2279–2287. [\[CrossRef\]](#)
6. Liao, L. Artificial Intelligence-Based English Vocabulary Test Research Using Log Analysis with Virtual Reality Assistance. *Comput.-Aided Des. Appl.* **2023**, *20*, 23–39. [\[CrossRef\]](#)
7. Raj, K.A.; Baisel, A. Empirical Study on the Influence of Mobile Apps on Improving English Speaking Skills in School Students. *World J. Engl. Lang.* **2024**, *14*, 339–348. [\[CrossRef\]](#)
8. Bacca-Acosta, J.; Fabregat, R.; Baldiris, S.; Kinshuk; Guevara, J. Determinants of Student Performance with Mobile-Based Assessment Systems for English as a Foreign Language Courses. *J. Comput. Assist. Learn.* **2023**, *39*, 797–810. [\[CrossRef\]](#)
9. Khlaisang, J.; Sukavatee, P. Mobile-Assisted Language Learning to Support English Language Communication among Higher Education Learners in Thailand. *Electron. J. e-Learn.* **2023**, *21*, 234–247. [\[CrossRef\]](#)
10. Zou, B.; Lyu, Q.; Han, Y.; Li, Z.; Zhang, W. Exploring Students' Acceptance of an Artificial Intelligence Speech Evaluation Program for EFL Speaking Practice: An Application of the Integrated Model of Technology Acceptance. *Comput. Assist. Lang. Learn.* **2023**, 1–26. [\[CrossRef\]](#)
11. Belda-Medina, J.; Marrahi-Gomez, V. The Impact of Augmented Reality (AR) on Vocabulary Acquisition and Student Motivation. *Electronics* **2023**, *12*, 749. [\[CrossRef\]](#)
12. Rachman, D.; Margana, M.; Priyanto, P.; Mahayanti, N.W.S. Designing Model for Oral Presentation Instruction in Indonesian Tertiary Context. *World J. Engl. Lang.* **2023**, *13*, 559–568. [\[CrossRef\]](#)
13. Xodabande, I.; Hashemi, M.R. Learning English with Electronic Textbooks on Mobile Devices: Impacts on University Students' Vocabulary Development. *Educ. Inf. Technol.* **2023**, *28*, 1587–1611. [\[CrossRef\]](#)
14. Singh, Y.; Suri, P.K. An Empirical Analysis of Mobile Learning App Usage Experience. *Technol. Soc.* **2022**, *68*, 101929. [\[CrossRef\]](#)
15. Kessler, M.; Ferronato, T.; Centurion, M.J.T.; Akay, M.; Kim, J. Mobile-assisted language learning with commercial apps: A focused methodological review of quantitative/mixed methods research and ethics. *Res. Methods Appl. Linguist.* **2025**, *4*, 100186. [\[CrossRef\]](#)
16. Fang, J.; Chew, F.P.; Shaharom, M.S.N. Mobile-assisted learning of Chinese as a second/foreign language abroad: A systematic literature review of studies between 2010 and 2022. *Knowl. Manag. e-Learn.* **2024**, *16*, 501–520. [\[CrossRef\]](#)
17. Moghaddam, M.M.; Esmailpour, F.; Ranjbaran, F. Insights into Mobile Assisted Language Learning Research in Iran: A Decade Review (2010–2023). *Educ. Inf. Technol.* **2025**, *30*, 2155–2181. [\[CrossRef\]](#)
18. Almaiah, M.A.; Hajjej, F.; Shishakly, R.; Lutfi, A.; Amin, A.; Awad, A.B. The Role of Quality Measurements in Enhancing the Usability of Mobile Learning Applications during COVID-19. *Electronics* **2022**, *11*, 1951. [\[CrossRef\]](#)
19. Lin, E.Y.-C.; Hsu, H.-T.; Chen, K.T.-C. Factors That Influence Students' Acceptance of Mobile Learning for EFL in Higher Education. *Eurasia J. Math. Sci. Technol. Educ.* **2023**, *19*, em2279. [\[CrossRef\]](#)
20. Söğüt, S.; Belli, S.A. QR Code Enriched Writing and Speaking Practices: Insights from EFL Learners at Tertiary Level. *Iran. J. Lang. Teach. Res.* **2024**, *12*, 1–18. [\[CrossRef\]](#)
21. Almaiah, M.A.; Hajjej, F.; Lutfi, A.; Al-Khasawneh, A.; Alkhdour, T.; Almomani, O.; Shehab, R. A Conceptual Framework for Determining Quality Requirements for Mobile Learning Applications Using Delphi Method. *Electronics* **2022**, *11*, 788. [\[CrossRef\]](#)
22. Essafi, M.; Belfakir, L.; Moubtassime, M. Investigating Mobile-Assisted Language Learning Apps: Babel, Memrise, and Duolingo as a Case Study. *J. Curric. Teach.* **2024**, *13*, 197–215. [\[CrossRef\]](#)
23. Rosell-Aguilar, F. State of the App: A Taxonomy and Framework for Evaluating Language Learning Mobile Applications. *CALICO J.* **2017**, *34*, 243–258. [\[CrossRef\]](#)
24. Polakova, P.; Klimova, B. Vocabulary Mobile Learning Application in Blended English Language Learning. *Front. Psychol.* **2022**, *13*, 869055. [\[CrossRef\]](#)
25. Lindaman, D.; Nolan, D. Mobile-Assisted Language Learning: Application Development Projects Within Reach for Language Teachers. *IALLT J. Lang. Learn. Technol.* **2015**, *45*, 1–22. [\[CrossRef\]](#)
26. Baloh, M.; Zupanc, K.; Košir, D.; Bosnić, Z.; Scepjanovic, S. A Quality Evaluation Framework for Mobile Learning Applications. In Proceedings of the 2015 4th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 14–18 June 2015; pp. 280–283. [\[CrossRef\]](#)

27. Paul, A.; Aaron, A.; Victor, T.; Muheise, H.; Brian, M.; Joe, M. A Framework for Evaluating the Usability of Mobile Learning Applications in Universities. *J. Sci. Technol.* **2023**, *7*, 42–59. [[CrossRef](#)]
28. Voskamp, A.; Kuiper, E.; Volman, M. Teaching Practices for Self-Directed and Self-Regulated Learning: Case Studies in Dutch Innovative Secondary Schools. *Educ. Stud.* **2020**, *48*, 772–789. [[CrossRef](#)]
29. Yahya, A.E.; Gharbi, A.; Yafooz, W.M.S.; Al-Dhaqm, A. A Novel Hybrid Deep Learning Model for Detecting and Classifying Non-Functional Requirements of Mobile Apps Issues. *Electronics* **2023**, *12*, 1258. [[CrossRef](#)]
30. Joy, S.; Kolb, D.A. Are There Cultural Differences in Learning Style? *Int. J. Intercult. Relat.* **2009**, *33*, 69–85. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.