

# User Experience in Mobile Language Learning Applications (MLLA): Insight from Mixed-Methods Triangulation Study

Yudhy Setyo Purwanto<sup>1,3\*</sup>, Rahmat Gernowo<sup>2</sup> and  
Dinar Mutiara Kusumo Nugraheni<sup>2</sup>

<sup>1</sup>Doctoral Program of Information Systems, School of Postgraduate Studies, Universitas Diponegoro, Semarang, Indonesia

<sup>2</sup>Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia

<sup>3</sup>Faculty of Energy Telematics, Institut Teknologi PLN, Jakarta, Indonesia

\*Correspondence: [yudhysp@students.undip.ac.id](mailto:yudhysp@students.undip.ac.id)

## PAPER INFO

### Paper history:

Received 11 September 2025

Accepted 19 February 2026

### Citation:

Purwanto, Y. S., Gernowo, R. & Nugraheni, D. M. K. (2026). User Experience in Mobile Language Learning Applications (MLLA): Insight from Mixed-Methods Triangulation Study. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 105-127

### Copyright:

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## ABSTRACT

Mobile Language Learning Applications (MLLAs) are gaining widespread use in higher education because of the flexible nature of practice opportunities with languages. The majority have easy-to-use interfaces, but few enable long-term engagement, retention, and productive learning. There is much empirical work that isolates usability or outcomes singularly without noting how user experience, motivation, and pedagogy intersect in the mobile environment. This study investigates MLLA user experience among university students in Indonesia using a mixed-methods design. Structural Equation Modelling (SEM) was used to investigate the inter-plays of usability, acceptance, engagement, and retention. Qualitative data provided information about user views and context-based limitations. Findings suggest that while overall MLLAs are usable and widely accepted, they tend to lack instructional intensity and intrinsic motivation support. Motivation and perceived usefulness significantly impact engagement, but redundant repetition, superficial individualization, and minimal interaction reduce retention. With reference to TPACK, FRAME, and SDT, the study highlights learner-centered design featuring effective pedagogy, social interaction, and adaptive functionality as being vital. To transcend surface gamification, MLLAs must support deep, long-term language learning. These results provide actionable recommendations for developers and instructors seeking to optimize mobile language learning by synergizing technology, pedagogy, and learner psychology.

**Keywords:** mobile language learning applications, user experience, mixed methods, TPACK- FRAME-SDT integration, SEM

## 1. Introduction

Mobile technology has changed access to and engagement with learning materials. In language acquisition, mobile-assisted language learning (MALL) applications have gained popularity since they are convenient, handy, and increasingly accepted by users. Among these, Mobile Language learning Applications (MLLA) is a subcategory that is enabled by the high smartphone penetration rate and the growing need for anytime-anywhere learning. These mobile applications seek to cater to learners with interactive tasks, multimedia learning, and on-demand language training that suit the habits and expectations of modern learners (Solihin, 2021). However, while such apps are on the rise in terms of availability, empirical studies are still limited

concerning the way users perceive MLLA as it relates to participation, ease of use, retention, and acceptance (Febriyanti et al., 2021).

Previous studies in MLLA have predominantly been focused on the general mobile learning benefits, such as convenience, accessibility, and instant feedback, and most of the existing literature has been focused on functionality, learning outcomes, or app performance, as opposed to how the users interact with the apps over time and what motivates their continued usage (Ekoç, 2021; Madhura Govind Date, 2020; Phuc & Nghi, 2023). This creates a gap in understanding the experiential aspects of mobile learning, especially how factors such as engagement, retention, usability, and acceptance work together to shape learners' long-term relationship with MLLAs. User experience (UX) has become increasingly important in educational settings because long-term usage, persistence, and perceived usefulness influence whether an app becomes part of a meaningful learning routine.

This study positions MLLAs as a set of tools with unique affordances such as personalization, contextual relevance, and continuous access, features that often differentiate them from traditional digital platforms. These affordances may play a significant role in shaping user satisfaction, frequent use, and perceived learning effectiveness. UX here, however, is complex and susceptible to the effects of numerous variables like individual user characteristics, environmental context, and technology design. To address this complexity, this study employs a mixed-method triangulation design that combines quantitative analysis of user survey data with qualitative data from student interviews.

The research is founded on three theory frameworks: Technological Pedagogical Content Knowledge, i.e., TPACK (Mishra & Koehler, 2006), which supports the confluence of pedagogy and technology; the Framework for the Rational Analysis of Mobile Education or FRAME (M. Koole, 2022; M. L. Koole, 2009), which emphasizes device, learner, and context interaction; and Self-Determination Theory or SDT (R. Ryan & Deci, 2017; R. M. Ryan & Deci, 2020), which emphasizes autonomy, competence, and relatedness motivation. These frameworks are used not only as descriptive models but as analytical lenses in interpreting the relationships among UX dimensions in MLLA use.

This study is guided by two refined research questions addressing (1) how users experience engagement, retention, usability, and acceptance when using MLLAs, and (2) how mixed-method triangulation can deepen the understanding of user interaction and satisfaction.

Through these questions, the study aims to expand the broader literature on user-centered design in educational technology, especially for location-aware systems. It is also offering worthwhile feedback for practice, pedagogy, and policy makers interested in enhancing language learning experience through the deployment of mobile technology in a context-aware manner.

## **2. Literature Review**

### **2.1. MALL and MLLA**

Mobile-Assisted Language Learning (MALL) entails the use of mobile technology to support and facilitate language learning. Development has been made possible by expanding smartphone penetration, making learning location-independent, self-paced, and integrated into daily routines. Early research into MALL highlighted accessibility and learner control advantages (Kukulska-Hulme & Shield, 2008; Loewen et al., 2019). Current studies have explored the ways in which mobile applications offer limited learning paths that aid in vocabulary building (Janebi Enayat et al., 2025), teaching grammar (Guerrero et al., 2010; Lin et al., 2020), pronunciation training (Dai & Wu, 2022), and communicative practice exercises (Hwang et al., 2024).

In this study, the term Mobile Language Learning Applications (MLLAs) is used to refer specifically to general-purpose language learning apps (e.g., Duolingo, ELSA Speak, Babbel, HelloTalk). However, critical examination of such platforms tends to identify a gap between the way the app was designed and the way it supports learning. For instance, (Kessler et al., 2023; Loewen et al., 2019) reported that while Duolingo strengthened vocabulary gains, it fell short in fostering communicative competence. Similarly, recent analysis (Al-Jamili et al., 2024; Sakkir & Syamsuddin, 2023; Sevara, 2024) note that many MLLAs rely heavily on repetitive gamification, limited contextual depth, and content that remains at a beginner level regardless of learner progress.

Some studies have looked to outcome-based metrics (like improved test results or task accomplishment rates) to evaluate such apps, but that only tells half the tale. Increasingly, researchers are attributing significance to user experience (UX) as a key factor in determining mobile learning effectiveness (Lubis et al., 2019; Okonkwo, 2024). Poor UX, especially engagement or app usability, has been found to lead to low motivation and high drop-out rates even when the instructional content of the app is of high quality. While

UX is essential, relatively little systematic, integrated research of MLLAs has examined how UX components interact and combine to affect lasting learning.

This research addresses that gap by examining how users interact with MLLA in authentic contexts through four key UX dimensions: engagement, retention, usability, and acceptance. Rather than focusing solely on learning outcomes, the study emphasizes the patterns of use, motivation, and adoption that determine whether MLLAs become sustainable learning tools.

## 2.2. User Experience in Educational Applications

User Experience (UX) in educational technology extends beyond interface design or usability. UX, to (Okonkwo, 2024), encompasses all the user's attitudes and emotions towards digital experience shaped by expectations, motivation, and context. UX in learning environments directly affects the ways learners experience, navigate, and engage with learning tools (Lubis et al., 2019; Royo & Laborda, 2018).

There are a variety of UX models that have been created in the past few years. In one such study by (Alshammare et al., 2025), UX is categorized into three dimensions, i.e., application, user, and context dimensions. Among its conclusions were efficiency, satisfaction, effectiveness, learnability, memorability, attractiveness, errors, and ease of use were the attributes that were most utilized. These are especially relevant to language learning applications, which must be good to look at, easy to use, and meaningful in the context of learners' personal or professional goals.

ISO 9241-210 (ISO 9241-210, 2010) also defines UX as “*a person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service,*” with emphasis on an overall framework that includes emotional, psychological, and situational factors. In the context of learning, UX is not merely a question of how students utilize the app, but how their utilization of the app affects their confidence, motivation, and persistence.

Earlier studies have examined individual determinants of UX in mobile learning. For example, (Kelly et al., 2023) studied usability and acceptance while learning through mobile devices and concluded that intuitive interaction and perfect error management enhanced learners' satisfaction. On the other hand, (M. Koole, 2022) examined engagement and established that adaptive tasks and instant feedback encouraged more cognitive engagement. However, these studies often examine UX dimensions in isolation. There remains limited research integrating these dimensions to explain how they collectively influence adoption and long-term learning behavior in MLLAs.

This work addresses this gap by focusing on four key aspects of user experience. The first is engagement, emotional and cognitive attachment to learning activities. The second is retention, users repeated use of the app and long-term adherence. Third is usability, which focuses on how easy the app is to use, how clear it is, and the quality of the app's interface overall. Lastly, the study considers acceptance, which reflects users' willingness to adopt the app and recommend it to others.

By analyzing these four dimensions together, this research presents UX as both an outcome of design and a predictor of educational persistence. It situates UX not merely as a design issue, but as a structural factor influencing learning success in mobile language environments.

## 2.3. Theoretical Foundations

In order to theorize the intersection of mobile learning, user motivation, and user experience, this study invokes three other frameworks: TPACK, FRAME, and Self-Determination Theory (SDT). However, in order to more fully understand the importance of UX in MLLA, this section also considers the broader theory of MALL and mobile app engagement literature.

This study invokes three theoretical frameworks for investigating user experience in mobile language learning applications. The first is Technological Pedagogical Content Knowledge (TPACK), designed by (Mishra & Koehler, 2006), which emphasizes that successful digital learning requires balanced integration of technology, pedagogy, and content. In MLLAs, misalignment between these components, such as content presented without pedagogical grounding or gamified interfaces that do not connect to learning goals, can undermine UX and retention. The second is the FRAME model of (M. L. Koole, 2009), which theorizes mobile learning as the intersection of device ease of use, social interaction, and learner profile. This is a very practical view since learners employ personal devices to pursue MLLAs in different social settings with different learning goals, factors largely determining user experience. The third is Self-Determination Theory which explains how autonomy, competence, and relatedness drive sustained motivation (Ryan & Deci, 2017). MLLAs that provide meaningful feedback, personalized challenges, or social connection opportunities are more likely to support intrinsic motivation and long-term use. In this study, TPACK informs the alignment between

content quality and usability, FRAME contextualizes engagement and retention through device and social interaction, and Self-Determination Theory explains acceptance and continued use through autonomy, competence, and relatedness.

In addition, the MALL theory places more attention on learner-centered design. Mobile learning is not merely about delivering content while on the move; it involves context-aware, affective, and behaviorally facilitating interaction (Kukulka-Hulme, 2009, 2021; Kukulka-Hulme & Viberg, 2018). UX becomes the channel through which pedagogical intentions are either reinforced or weakened.

Figure 1 summarize the theoretical framework. The model integrates TPACK (alignment of technology, pedagogy, and content), the FRAME model (device, learner, and social context), and Self-Determination Theory (autonomy, competence, and relatedness) to explain how four user experience dimensions (engagement, retention, usability, and acceptance) influence learners’ long-term use of mobile language learning applications.

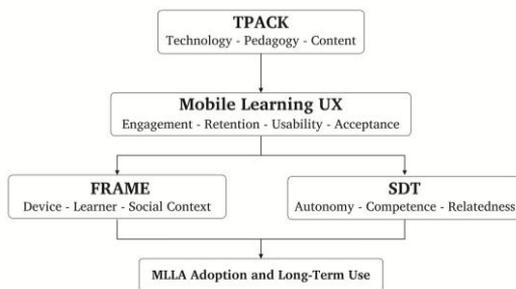


Figure 1. Theoretical Framework Overview

By integrating TPACK, FRAME, SDT, and MALL perspectives, this study positions UX as a central mechanism that connects design, motivation, and educational outcomes. This theoretical integration supports the study’s focus on examining how engagement, retention, usability, and acceptance interact to shape learners’ adoption of MLLAs.

### 3. Methodology

This study adopted a sequential explanatory mixed-methods design, in which quantitative data collection and analysis preceded the qualitative phase. The initial survey phase examined relationships among four user experience (UX) constructs: Engagement, Retention, Usability, and Acceptance, using a structural path model based on composite scores. The subsequent qualitative phase involved purposive interviews with selected participants to further interpret and contextualize the quantitative results. Rather than collecting both strands concurrently, the qualitative data were used to explain patterns observed in the survey findings, particularly where statistical relationships required deeper insight. Integration occurred at the interpretation stage, where qualitative themes were connected to the quantitative associations to provide a more comprehensive understanding of user experience in mobile language learning applications. This study follows Creswell’s convergent mixed-methods triangulation design (Creswell & Clark, 2018), in which quantitative and qualitative data are collected in parallel and integrated during interpretation.

Participants were adult learners (18 years and above) who had used one or more MLLAs (such as Duolingo, Cake, or ELSA Speak) within the past six months. Convenience-purposive sampling (Andrade, 2021; Etikan, 2016) was used due to the researcher’s limited access to a comprehensive student list. The survey link was distributed through WhatsApp groups (including class and student organizations) and university email networks.

A total of 266 responses were collected: 198 were active MLLA users and 68 were non-users. From the pool of MLLA users, 15 Interview participants were selected purposively based on survey responses to ensure that the qualitative phase captured diverse engagement and usability experiences.

Although the study was conducted within a single private university in Jakarta, the campus hosts students originating from various cities across Indonesia, representing diverse linguistic and educational backgrounds. This composition makes the sample contextually appropriate for the study’s focus on Indonesian higher education learners. Conducting the study within one institution also ensures tighter control over demographic consistency and reduces institutional variability, which supports internal validity while remaining representative of typical urban Indonesian university environments.

Participation was voluntary, and informed consent was obtained from all respondents. Anonymity and confidentiality were maintained during data handling and reporting. The research protocol was reviewed and approved by the ethics board at the researcher's affiliated institution.

The survey instrument consisted of 28 questions: 4 demographic questions (age, gender, study program, and MLLA usage status), 4 Application-usage questions, and 20 items measuring UX aspects; five items each for engagement, retention, usability, and acceptance (Appendix A). These items were designed based on literature in user experience and mobile language learning and were informed by TPACK, FRAME, and SDT frameworks.

The questionnaire was pilot tested with 20 participants (18 MLLA users, 2 non-users). Validity and reliability testing (Appendix F2) revealed that two items (Q11-Usability; Q16-Acceptance) scored below the acceptable threshold ( $\alpha < 0.7$ ). These items were removed before full deployment to avoid reducing internal consistency. Removing low-performing items is a standard refinement step in survey development and improves the reliability and precision of subsequent analyses. The qualitative phase involved semi-structured interviews with the 15 selected MLLA users. The interviews were conducted via Microsoft Teams, using built-in recording and automatic transcription features. Transcripts were manually cleaned and checked before coding and analysis.

The quantitative structural analysis in this study was conducted using a single structural path model based on composite user experience (UX) scores. Specifically, Engagement, Retention, Usability, and Acceptance were computed as averaged scale scores and entered the model to examine their associative relationships. No additional structural equation models or extended UTAUT-type models were estimated beyond this four-construct configuration. Accordingly, the model should be understood as a structural path analysis on composite UX dimensions rather than as a multi-model or latent-variable SEM framework. Although the model comprises four composite UX dimensions, a structural modeling approach was selected instead of separate regression analyses to examine the system of relationships simultaneously within a single theoretically specified framework. This approach allows the estimation of direct and indirect associations among engagement, retention, usability, and acceptance in one coherent model, rather than testing isolated pairwise regressions. The structural path model therefore reflects the theoretically driven structure of UX dynamics while maintaining parsimony in model specification. Quantitative data were analyzed using Python in Jupyter Notebook (Fangohr et al., 2019; Mckinney, 2017). Cronbach's alpha (Bonett & Wright, 2015; Tavakol & Dennick, 2011) was used to test internal consistency for each UX dimension. Descriptive statistics (mean, standard deviation, frequency) were used to summarize the data (Alabi & Bukola, 2023; Fisher, 2009; Marshall & Jonker, 2010), followed by regression testing to identify relationships among UX factors (Blatna, 2013; Freedman, 2007; Xin & Ren, 2016), correlation analysis using Structural Equation Modelling (SEM) (Blanchin et al., 2019; Elliott & Timulak, 2015; Kunnan, 1998). Because the data are cross-sectional, the SEM results indicate associative rather than causal relationships among engagement, retention, usability, and acceptance. Model fit indices including CFI, TLI, GFI, AGFI, and RMSEA were computed to evaluate overall SEM adequacy. K-Means clustering was used to segment users into groups based on UX patterns (Bock, 2007; Chen & Hoe, 2025; Oti et al., 2021; Priyatma et al., 2024). To ensure conceptual clarity, it should be noted that the structural analysis in this study involved only four empirically measured user experience constructs: Engagement, Retention, Usability, and Acceptance. These constructs were operationalized through validated questionnaire items and aggregated into composite scores prior to model estimation. No additional UTAUT2 constructs (e.g., perceived usefulness, effort expectancy, hedonic motivation, price value) were directly measured or statistically modeled in the structural analysis. References to UTAUT2, TPACK, FRAME, and Self-Determination Theory in later sections serve as interpretive frameworks to contextualize and explain the observed relationships among the four UX constructs, rather than as empirically estimated latent variables within the structural model. The purpose of applying a structural modeling approach was to examine the simultaneous relationships among multiple UX dimensions within a single coherent framework, rather than to compare alternative model specifications.

Qualitative data were analyzed using thematic analysis, following an inductive coding approach (Cooper, 2010; DiCicco-Bloom & Crabtree, 2006; Elhami, 2022; Robinson, 2023). Codes captured recurring patterns: motivation, satisfaction, challenges, and design expectations. The resulting themes were triangulated with quantitative findings to identify convergence, divergence, and complementarity across data sources, thereby fulfilling the mixed-method triangulation requirement.

Integration of quantitative and qualitative findings was conducted at the interpretation stage, consistent with a sequential explanatory mixed-methods design. Rather than merging datasets statistically, integration was achieved through side-by-side comparison and explanatory linking between quantitative patterns and qualitative themes.

Specifically, results from the structural path model and descriptive analyses were first examined to identify key relationships among engagement, retention, usability, and acceptance. The subsequent qualitative interview data were then used to explain, contextualize, and elaborate on these quantitative patterns, particularly where statistical results indicated moderate or ambiguous user experience outcomes.

To make this integration explicit, selected summary tables were structured as joint displays. These tables indicate whether findings derive from survey results, structural modeling, qualitative themes, or their combined interpretation. This approach enables transparent tracing of how qualitative insights inform the interpretation of quantitative relationships without implying that qualitative themes were statistically modeled.

## 4. Result

### 4.1. Quantitative Results

The quantitative results of 198 out of 266 Mobile Language Learning Application (MLLA) users are included in the section. Emphasis is on four key variables: engagement, retention, usability, and acceptance. These were measured with Likert-scale ratings and contrasted to discern general user trends, degrees of satisfaction, and points of improvement. The results provide a broad picture of the way in which students interact with MLLAs and serve as the basis for comparison with qualitative remarks below.

#### 4.1.1. General and Demographic Questions

This section reports the findings from the general and demographic questions completed by all participants ( $n = 266$ ). The information collected involves age, gender, study program, and whether respondents have experience with mobile-based language learning apps (MLLAs). These variables set context for knowing the background of the participants and for comparing opinions between users and non-users, which is necessary for the following analysis of user experience and engagement behaviors.

In terms of experience, most of the participants (74.4%) or 198 respondents reported ever having used mobile-based language learning apps (MLLA). Conversely, 68 respondents (25.6%) reported never having tried using such apps. This difference provides a good comparison basis for evaluating user and non-user attitudes toward MLLAs and vindicates the UX (user experience) comparison analysis used in this study.

The age distribution among the respondents was diverse, with the ages ranging from 18 to over 23 years. The predominant ages were 20 years old (21.8%) and 22 years old (24.4%), followed by 19 years old (19.2%) and 21 years old (15.0%). Among the ages 18 and older than 23, the percentages were lower at 6.8% and 12.8%, respectively. This trend shows that most of the respondents were current university students at the 2<sup>nd</sup> to 6<sup>th</sup> semester, who are likely to start using mobile learning applications. And gender-wise, 165 respondents (62.0%) were male and 101 respondents (38.0%) were female. This proportion is a reasonable representative of the overall population of students in technology and engineering-related programs, which are the subjects of this study.

Respondents were from various academic programs, though the largest proportion was from the bachelor's in informatics engineering (39.5%) and bachelor's in electrical engineering (15.0%). Other academic programs with a significant number of respondents were bachelor's in information systems (11.3%) and bachelor's in mechanical engineering (8.6%). A few respondents were from vocational programs like the Diploma in Electrical Technology and Diploma in Mechanical Engineering, though in smaller numbers. This heterogeneity enriches the analysis by allowing comparisons across majors with different levels of exposure to technology and English-learning needs.

Meanwhile, among the 198 respondents who said they had utilized mobile language learning apps, most employed Duolingo (140 respondents), followed by ELSA Speak (46 respondents), and the remaining few employed other apps such as Babbel, Busuu, or domestic apps (12 respondents). This shows the dominance of two mainstream apps that rely on contrasting instructional designs: gamification vs. AI-driven pronunciation.

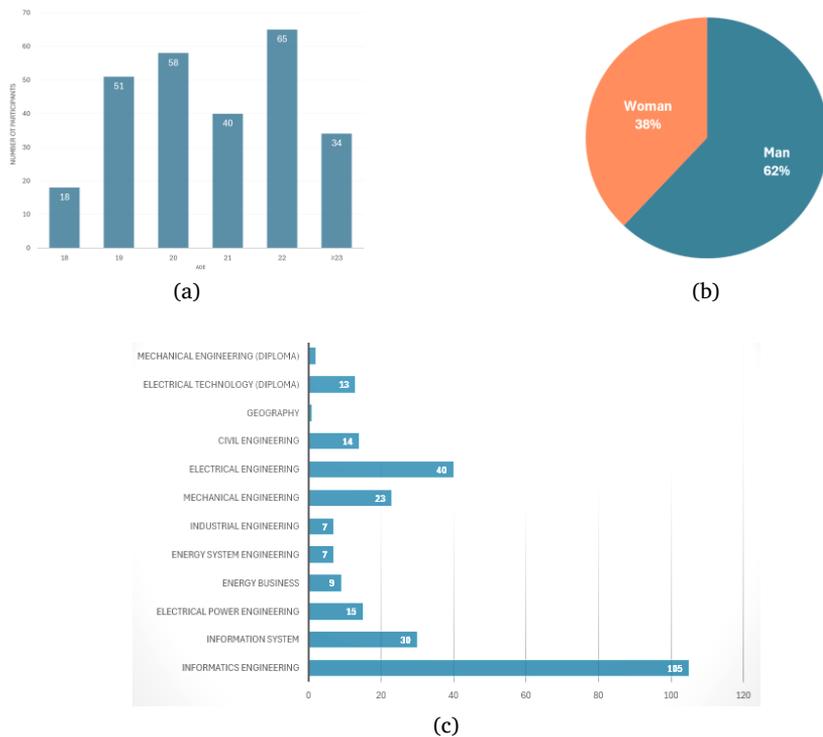


Figure 2. Demographic Questions Results, (a) age; (b) gender; (c) study programs

With respect to usage duration, the majority of the respondents had used the apps for over 6 months (48 respondents), while 45 respondents had used them between 1–2 months; the remaining respondents were evenly distributed across less than 1 month and 5–6 months. This indicates that some of the respondents had had long-term exposure to MLLAs, which can provide richer data on areas of user experience (UX).

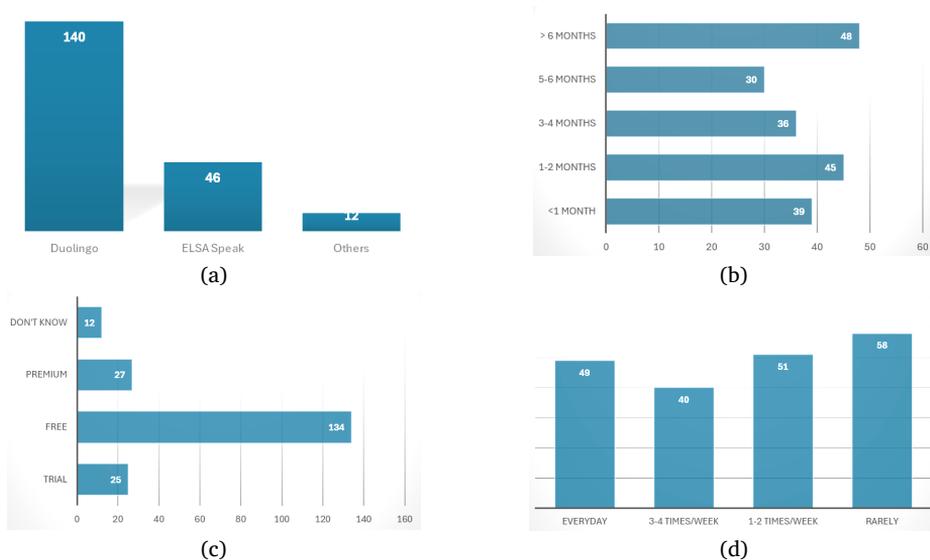


Figure 3. MLLAs General Questions Results: (a) App(s) used; (b) Duration; (c) Plan; (d) Intensity

With this data composition, the subsequent analysis will be able to better determine variations in patterns of use, retention, usability, and MLLA acceptance based on users' demographic and academic backgrounds.

**4.1.2. Reliability and Validity**

During pilot testing, two items (Q11 and Q16) were identified as underperforming items through overall reliability contribution and item-total correlation. These two items were excluded from the final analysis to improve the scale's validity and to offer more reliable results.

The result of full questionnaire's reliability ratings for all UX dimensions (Engagement, Retention, Usability, and Acceptance) was strong. Cronbach's Alpha varies between 0.858 and 0.868, implying that the measurement tools employed had high internal consistency on the whole.

Section 1 (Engagement)			Section 2 (Retention)			Section 3 (Usability)			Section 4 (Acceptance)		
CA: 0.868			CA: 0.859			CA: 0.858			CA: 0.868		
Q	CI	CA	Q	CI	CA	Q	CI	CA	Q	CI	CA
Q1	0.55	0.87	Q6	0.68	0.82	Q11	0.69	0.82	Q15	0.69	0.84
Q2	0.68	0.84	Q7	0.77	0.80	Q12	0.78	0.78	Q16	0.74	0.82
Q3	0.72	0.83	Q8	0.72	0.82	Q13	0.60	0.86	Q17	0.78	0.80
Q4	0.73	0.82	Q9	0.56	0.86	Q14	0.72	0.80	Q18	0.66	0.85
Q5	0.76	0.82	Q10	0.65	0.83						

**Table 1.** Full Questions Reliability and Validity Tests Results

**4.1.3. User Experience Patterns and Statistical Insights**

*Descriptive Analysis*

The descriptive statistics among 198 MLLA users show that perceptions of engagement, retention, usability, and acceptance are on the moderate level as shown by the table below.

	Engagement	Retention	Usability	Acceptance
count	198.000000	198.000000	198.000000	198.000000
mean	2.782828	2.779798	2.861111	2.925505
std	1.000867	0.948842	0.937598	0.934977
min	1.000000	1.000000	1.000000	1.000000
25%	2.000000	2.200000	2.250000	2.250000
50%	2.800000	2.800000	2.750000	3.000000
75%	3.400000	3.400000	3.250000	3.500000
max	5.000000	5.000000	5.000000	5.000000

**Table 2.** Descriptive Analysis' Results

The mean scores for engagement, retention, usability, and acceptance ranged from 2.78 to 2.93 on a five-point Likert scale (1 = strongly disagree, 3 = neutral, 5 = strongly agree). These values are slightly below the neutral midpoint, indicating neither clear dissatisfaction nor strong endorsement of the applications. Accordingly, the results suggest a generally neutral to mildly positive user experience rather than a distinctly high level of engagement, retention, usability, or acceptance.

For data distribution, the measures of standard deviation range from 0.93 to 1.00, which depicts considerable variability in user ratings. This means that some users were satisfied, but others gave lower ratings. This is also evidenced in the lower quartile (25%) values that describe the 25% of the respondents who gave scores for engagement and retention below 2.0–2.2. On the contrary, the upper quartile (75%) signifies that 25% of the participants rated usability and acceptance over 3.25–3.5, meaning that more than half of the users rated the dimensions positively.

Although usability obtained comparatively higher central tendency values than the other UX dimensions, both the mean and median scores remained close to the neutral midpoint of the scale. Interquartile ranges further indicate moderate dispersion, suggesting variability in user evaluations. These statistics do not reflect strong endorsement of ease of use or readiness to adopt MLLAs, but rather indicate functional adequacy with

considerable room for improvement. Overall, these findings suggest a gap between “functional usability” and “motivational experience,” with users finding the apps easy to operate but less capable of sustaining long-term interest.

**UX Factor Prediction**

To build upon these correlations, a bootstrapped regression was run. The results reiterated the trend discovered in the correlations. Acceptance proved to be a strong predictor of usability ( $\beta = 0.755, p < 0.001$ ), and usability strongly impacted retention ( $\beta = 0.379, p < 0.001$ ).

The predictive models show strong relationships between engagement, retention, usability, and acceptance. All these elements have a strong positive effect on the others, resulting in a reinforcing loop. For example, users who can easily employ an app (usability) and comprehend its value (acceptance) will be in favor of using it for a long period (retention). This is indeed evident through the engagement model, where retention, usability, and acceptance have a positive effect on user engagement. Unexpectedly, users of ELSA were more engaged than for the majority of other apps, although in this instance, it did not necessarily transfer to long-term retention.

When comparing all models separately, retention was the best predictor overall, with particular influence on both engagement and acceptance. However, while ELSA had success in boosting early engagement, its negative correlation with retention suggests potential issues with sustained interest. Usability, though, appears to be less brand-specific and more broadly understood. Overall, the findings suggest that to build a successful MLLA, developers need to be worried not just about first-time appeal or technical usability, but also about sustaining user motivation and delivering perceived value throughout the learning process.

**Cluster Analysis**

The K-Means segmentation validated three UX pattern and engagement-intensive MLLA user groups with varying levels of retention, usability, acceptance, and UX patterns. Cluster 1 (n = 89), the low-performing group, are users who are likely to drop off, and they would need ease of design, easier value propositions, and support mechanisms that are improved (all of which had low scores of ~2.0). The medium cluster (Cluster 0, n = 24) has moderate satisfaction and room to grow with focused educational and interactive capabilities (rated all aspects around 3.0). In contrast, the high-performance cluster (Cluster 2, n = 85) is comprised of active, devoted users who perceive the app as valuable, intuitive, and suitable for their learning needs (rated all UX dimensions highly in ~4.5). These results are portrayed in Figure 4 and Table 3 below.

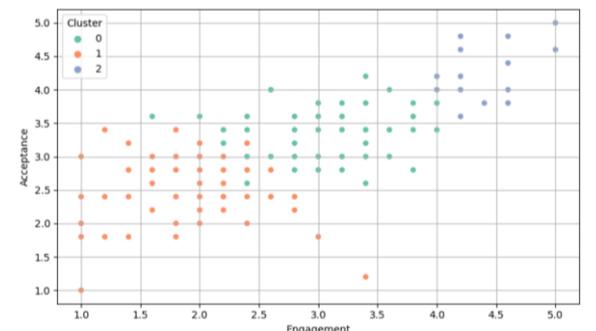


Figure 4. Users’ Cluster based on UX Profile

Cluster	Engagement	Retention	Usability	Acceptance
0	3.01	3.00	3.2	3.25
1	1.90	1.96	2.5	2.52
2	4.56	4.38	4.5	4.43

Table 4. UX mean score per Cluster

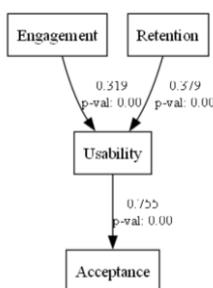
These clusters offer actionable guidance for tailoring UX strategies, showing that different groups require different design emphases, from simplifying workflows for low-performing groups to sustaining challenge and value for advanced users.

**Structure Equation Modelling (SEM)**

The relationships among Engagement, Retention, Usability, and Acceptance were examined using a structural path model based on composite scale scores. Each construct was operationalized as the mean of its validated questionnaire items and treated as an observed variable in the model. Accordingly, the analysis should be interpreted as a path analysis of aggregated UX dimensions rather than as a latent-variable structural equation model. Given the cross-sectional nature of the data and the use of composite scores, the estimated paths indicate statistical associations among UX dimensions rather than causal effects.

	lval	op	rval	Estimate	Std. Err	z-value	p-value
0	Usability	~	Engagement	0.318846	0.055907	5.703165	1.176029e-08
1	Usability	~	Retention	0.379415	0.058972	6.433787	1.244636e-10
2	Acceptance	~	Usability	0.754780	0.043290	17.435509	0.000000e+00
3	Usability	~~	Usability	0.196191	0.019718	9.949874	0.000000e+00
4	Acceptance	~~	Acceptance	0.228290	0.022944	9.949874	0.000000e+00

**Table 4.** Causal Relationships Among Key UX Factors in MLLA



**Figure 5.** Key UX Factors in MLLA Causal Relationships

The results show that Engagement and Retention play a strong role in influencing perceptions of Usability. More specifically, Engagement showed a significant positive association with usability (estimate = 0.3188,  $p < 0.000001$ ) to show that more motivated and engaged users find the app easier to use. Similarly, Retention was positively associated with usability (estimate = 0.3794,  $p < 0.000001$ ), in that repeated usage of the app makes the user progressively more comfortable and familiar with its structure, further strengthening their sense of ease in using it. The strongest effect in the model was from Usability to Acceptance, highlighting that usability is the most basic element in determining whether users accept the MLLA as a learning tool. This finding supports the hypothesis that an effortless experience with simple navigation, understandable content arrangement, and easily accessible features can have a strong influence on user acceptance and future use.

In summary, the SEM analysis confirms that Usability actually serves as the central mediating variable in the relationship between users' engagement/retention and application acceptance. The feeling of the experience being intuitive and familiar not only provides support towards long-term use but also has a straight effect on increasing users' adoption willingness. Although there is a little bit of unexplained variance in the model, the overall low residual values are indicative of the fact that the model is well-suited to fitting the overall dynamics between these UX factors.

This result points towards designing MLLAs with a strong focus on usability in an attempt to maximize adoption rates. In order to validate the structural relationships, we correlated Spearman's Rho over the four pillars. All the cross-variable correlations were strong (0.67-0.78), validating the SEM-based relationships.

The model demonstrated acceptable fit based on incremental indices (CFI  $\approx 0.93$ ; TLI  $\approx 0.90$ ), while the RMSEA value was relatively high ( $\approx 0.21$ ). Because RMSEA is known to be unstable and prone to inflation in models with very low degrees of freedom, particularly when composite variables are used, overall model evaluation relied on multiple fit indices rather than RMSEA alone. A more detailed discussion of fit considerations is provided in Appendix E. Semopy does not compute SRMR for composite-based structural models, so SRMR is not reported. Overall, the CFI and TLI values support the acceptability of the model.

## 4.2. Qualitative Findings

Thematic analysis of interview responses yielded nine emergent themes, five of which were strongly dominant. The most frequently mentioned problem was repetitive and monotonous content emerged as a dominant concern among participants. Users described a learning experience that quickly became predictable, reducing cognitive challenge and diminishing motivation to continue. This perceived lack of progression helps explain why engagement and retention scores remained only moderate in the quantitative results, despite comparatively higher usability relative to other dimensions ratings. Participants did not disengage because the apps were difficult to use, but because repeated exposure failed to generate a sense of development or mastery. Despite progressing through various levels, users felt the apps failed to introduce advanced material or content relevant to real-life contexts, leaving their learning plateaued.

Another extremely close issue was the user interface design that was artful but superficial, especially for apps like Duolingo. Though 80% of the users loved the gamified interface and encouraging graphics, they didn't believe the approach dug deep enough and didn't support long-term or substantial learning. In addition, low retention and non-practical application were huge problems. Some 73% of those polled admitted to forgetting most of what they had learned and having trouble applying new grammar or vocabulary to real-life situations. Lastly, paid walls and advertisements were cited by 40% of users as being frustrated with the interruption caused by ads or limitations on free content.

Theme	Description	%
Repetitive Content	Exercises felt predictable and lacked variety	87%
Stagnation at Basic Levels	Lack of advanced or real-world content even after progress	73%
Fun UI, Superficial Depth	Visually engaging but did not support deep learning	80%
Poor Retention & Practical Use	Knowledge quickly forgotten and hard to apply	73%
Premium Barriers & Ads	Ads and paywalls disrupted learning	40%

**Table 6.** Qualitative Results

Only 3 of 15 users perceived the apps as supporting long-term learning, indicating that while MLLAs may engage learners initially, they struggle to translate early engagement into sustainable progress. Users recommended improvements such as contextualized tasks, adaptive difficulty, and deeper speaking evaluation, reinforcing the quantitative findings that MLLAs need stronger engagement and retention mechanisms (See Appendix D for the full participant-by-theme matrix).

## 5. Discussions

This section synthesizes the study's findings by moving beyond statistical reporting to interpret the results in relation to theory and prior research. To avoid redundancy, detailed statistical results are presented in Section 4, while Section 5 focuses on interpretation and theoretical implications. The discussion is grounded in the four UX constructs empirically examined in the structural path model: Engagement, Retention, Usability, and Acceptance. The model reveals patterned associations among these dimensions rather than directional causal mechanisms: higher levels of engagement and retention were associated with more favorable usability perceptions, which in turn were strongly associated with acceptance. These relationships should be understood as co-occurring tendencies within users' reported experiences, not as evidence of causal progression. Concepts from Self-Determination Theory (SDT), FRAME, TPACK, and UTAUT-related literature are used solely as theoretical lenses to contextualize these associations and were not included as measured predictors in the quantitative model.

### 5.1. Quantitative Results

Quantitative findings showed moderate UX scores on four dimensions assessed (engagement, retention, usability, and acceptance). These values represent a middle ground: users neither strongly accepted nor rejected their experience with MLLAs. Still, the strength of the correlations among these variables, and particularly between engagement and retention ( $\rho = 0.78$ ) and usability and acceptance ( $\rho = 0.67$ ), implies a highly integrated UX system. Improving one facet, such as making an app more enjoyable, can indirectly heighten users' overall satisfaction and desire to continue using it.

Structural path modeling was used to examine patterned associations among the four UX dimensions. The results indicate that higher levels of engagement and retention were associated with more favorable

usability perceptions ( $\beta = 0.3794$  and  $\beta = 0.3188$  respectively,  $p < 0.001$ ). Usability, in turn, showed a strong association with acceptance ( $\beta = 0.7548$ ,  $p < 0.001$ ). Rather than implying a directional causal mechanism, these findings suggest that usability functions as a central relational link connecting users' affective experiences (engagement and retention) with their reported willingness to adopt or continue using an application. From this perspective, interface and experience design should be understood as embedded within broader patterns of user interaction, not evaluated in isolation. Notably, although usability scores were relatively higher, engagement and retention remained closer to neutral levels, pointing to a disconnect between ease of use and sustained learning impact. Consistent with prior research on MLLA user experience, this pattern indicates that high usability alone does not necessarily coincide with enduring engagement or deeper learning outcomes.

The qualitative interviews, however, introduced a more critical interpretive layer. Participants would praise the user-friendly interface and gamified design of popular applications like Duolingo but would commonly complain of a lack of meaningful progress in the material. While the SEM results suggest usability as a stimulant to acceptance, the qualitative data show that perceived usability does not always translate to actual learning value. For instance, most users reported content being repetitive or too easy, and that they felt stuck at a beginner level. Even the users who were clustered into the "Satisfied" cluster through K-Means analysis complained about the inability of the app to facilitate long-term retention or real-world application of language skills.

This discrepancy reveals a shortcoming in how users experience "usability" in learning environments. SEM posits that users who have fun using an app and find it easy to return to will find it more usable, but that this perceived usability can be an indicator of interface design instead of instructional effectiveness. That is, users may enjoy the app because it is comfortable and relatively accessible, not because it is necessarily making them more proficient in the language. Here is the key insight: acceptance driven by surface-level usability can lead to high adoption without corresponding learning outcomes.

Briefly, while the quantitative and SEM results affirm a well-functioning UX model wherein engagement, retention, usability, and acceptance predict each other in anticipated ways, the qualitative feedback suggests an underlying issue: a gap between surface-level fun and deep learning. This gap signals that UX satisfaction alone does not guarantee meaningful progress, especially when learning activities lack structure and cognitive challenge.

Furthermore, assessing MLLAs against theoretical frameworks like TPACK, FRAME, and Self-Determination Theory (SDT) helps to identify significant gaps that can account for the mismatch between usability ratings and more profound learning issues. Taking TPACK as a starting point, the apps show good technological knowledge (TK), something reflected in their smooth interfaces, responsive navigation, and gamification. Yet the same apps lack pedagogical knowledge (PK) and content knowledge (CK). The majority fail to scaffold learning or provide organized feedback, especially on speaking and productive language use. This imbalance creates a heavy T layout design, where interface elegance masks superficial instructional depth.

These patterns are reflected in the structural path model results. Usability showed a strong association with acceptance, indicating that users tended to value applications that were easy to navigate and comfortable to use. However, engagement and retention remained closer to neutral, suggesting that ease of use alone does not compensate for limitations in pedagogical depth or learning progression. This gap helps explain why technically polished applications may be readily accepted but still struggle to sustain long-term learning involvement.

The observed relationships among engagement, usability, and acceptance can be interpreted through the perspectives of TPACK and the FRAME framework. From a FRAME standpoint, the alignment between device usability and learner interaction appears central to positive acceptance outcomes. At the same time, FRAME also emphasizes social and contextual dimensions of mobile learning, which were largely absent in participants' accounts. Interview data revealed that learning activities were often decontextualized and repetitive, with limited opportunities for real-world use or peer interaction. This lack of contextual grounding may help explain why engagement and retention scores remained moderate despite relatively positive usability perceptions.

Similarly, TPACK highlights the importance of balancing technological efficiency with pedagogical and content integration. While participants generally reported low technical barriers and intuitive interfaces, qualitative responses suggested that instructional design and content sequencing were insufficient to support sustained learning development. These perspectives help contextualize why usability was strongly associated with acceptance in the present model without implying that pedagogical or content constructs were directly measured.

The association between engagement and retention can also be interpreted through concepts from Self-Determination Theory (SDT). Although psychological needs were not directly measured in the quantitative

model, qualitative findings suggest that when learners experienced a sense of progression or mastery, engagement was more likely to co-occur with continued use. Conversely, reliance on surface-level incentives attracted initial attention but did not sustain motivation in the absence of perceived competence, autonomy, or meaningful social connection. In this sense, SDT provides an explanatory framework for understanding why engagement and retention remained moderate despite favorable usability evaluations.

Framework	Focus Area	Findings	Implication
TPACK	Pedagogical & Content Fit	Weak feedback; little advanced/contextual content.	Tech-driven, limited pedagogical depth.
FRAME	Social & Contextual Learning	Few real-world or collaborative tasks.	Learning is isolated and less authentic, reducing engagement.
	Device & Personal Usability	Good interface design; intuitive navigation.	A strong interface, but limited instructional depth.
SDT	Motivation (Autonomy, Competence, Relatedness)	Motivation from streaks/peers; users feel isolated.	Low intrinsic motivation leads to poor retention and engagement.

**Table 7.** Joint display of framework-based interpretation of quantitative and qualitative findings

These findings can be mapped across the three frameworks to show where current designs fall short (Table 7). MLLAs excel in interface quality and surface-level engagement but fail to offer the pedagogical integration, contextual richness, and psychological scaffolding that is required for deep, lasting learning. Quantitative associations from the structural path model are interpreted alongside qualitative themes to illustrate how theoretical frameworks contextualize user experience patterns. To advance beyond superficial engagement, future apps must integrate instructional design with contextual and motivational affordances from the outset.

In conclusion, SEM data reinforces what these models suggest: usability enhancement alone will not yield long-term learning or engagement results. The next wave of MLLAs must be instructionally smart, socially and contextually appropriate, and psychologically supportive if they are to realize their full educational potential.

## 5.2. Qualitative Results Discussion

Semi-structured interviews with 15 active MLLA users resulted in nine themes, five of which were strongly represented across participants. The themes describe perceptions of learning depth, task variety, feedback quality, and motivational experience with the current mobile language learning applications. On the whole, qualitative results point to the trend of initial enthusiasm followed by declining engagement due to repetitive content and lack of meaningful language improvement. Table 6 summarises the most salient themes.

**Repetitive and Predictable Content.** The issue most reported was related to task repetitiveness. Thirteen of the fifteen participants stated that activities became predictable after several days of use. Users described a feeling that the lessons recycle the same sentence structures or vocabulary; this makes progress feel "stagnant" and "no longer challenging." This was considered a central cause of disengagement and early drop-off.

**Stagnation at Basic Levels.** Eleven participants, or 73%, reported that even after progressing to higher levels within apps, advanced or real-world content was not introduced. Participants reported being "stuck in beginner mode" and having few opportunities to put learned vocabulary or grammar into useful applications. This perceived lack of progression contributed to frustration and ultimately reduced long-term retention.

**Visually Appealing yet Superficial UI.** Twelve of the participants, or 80%, addressed the aesthetically appealing, gamified aspects, and interface design of mainstream apps like Duolingo, but they followed up this positive framing by indicating the design "looked fun but didn't go deep". Gamified elements were motivating at the outset, but participants uniformly reported that interface attractiveness did not translate to deeper learning or improved competence.

**Weak retention and limited real-world application.** Eleven participants (73%) reported difficulty remembering what they had learned, especially when trying to apply it in conversations or academic tasks. Users noted that vocabulary learned through repeated drills was not reinforced within diverse contexts and thus could hardly be transferred outside the app. This problem has been stated to be most serious concerning speaking and productive language skills.

**Premium Paywalls and Advertisements.** Six participants (40%) reported frustration either with ads interrupting or limitations on free content. For example, some noted that such interruptions "break the learning flow" and "force a stop before I'm ready." Several participants said they stopped using the apps because essential features were locked behind paid subscriptions.

Only three of the fifteen interviewees felt that current MLLAs supported long-term learning or would be worth recommending to others. Most participants described a pattern in which initial motivation was high, especially because of visual appeal, streak systems, or ease of navigation, but this motivation declined sharply after one to two weeks of use. Several users mentioned that they continued using the apps only to maintain the streaks rather than actually feeling they were learning something new. Others reported that despite daily use, they were "unsure what progress actually meant" since the apps did not provide clear indicators of improvement beyond points or badges.

Participants also commented that the development of skills felt fragmented. For instance, while learners felt comfortable with recognition tasks, they were still unable to speak or construct sentences on their own. This mismatch between perceived progress in the app and actual language ability led users to question the value of continued use. In fact, several participants stated that they used additional materials such as YouTube lessons, classroom exercises, or conversation practice because the MLLAs alone did not help them make progress beyond beginning levels of proficiency.

Taken together, the interviews paint a picture of users who appreciate the convenience and accessibility of MLLAs yet believe that such tools cannot maintain meaningful development without supplementary support, richer content, or stronger feedback mechanisms. (See Appendix D for the full participant-by-theme matrix).

### 5.3. User Segmentation and Behavioural Patterns

To better understand the range of user experiences, a cluster analysis of interview and survey data was conducted. This identified patterns within how different users interact with mobile language learning apps (MLLAs), and more importantly, why some continue, and others drop off. The segmentation identifies not just preferences, but deeper behavioral trends that inform how MLLAs can better serve learners of variable motivation, autonomy, and digital readiness.

**Cluster 2: Highly Satisfied Users.** These learners had consistently good experiences. They found the apps to be relatively accessible and helpful, naming ease of use and perception of progress as reasons for continued use. Many were apparently self-regulated learners with high digital fluency and clear goals. Their usage suggests that current features in apps—like progress tracking, instant feedback, and conventional interface design—are effective for learners already proficient at controlling their own learning.

**Cluster 1: Frustrated or Passive Users.** This was the largest and most problematic cluster. While they generally gave neutral-to-positive first impressions, they also tended to lose interest early on. Interviews suggest this was due to shallow content, lack of feedback (especially for speaking exercises), and repetitious activities. These users describe a common experience: the application is usable at first, but the learning process becomes mechanical and unrewarding after a while. They expected more depth and real-world relevance in their practice but did not find it. This makes Cluster 1 a critical target group for redesign, as they mirror the majority of disengaged learners in real usage settings.

**Cluster 0: Neutral Users.** This group of users was in the middle. They did not despise the app, nor were they fully engaged. They used the app occasionally, questioned its long-term usefulness, or indicated that they were waiting for more personalized or meatier content. For this group, subtle design improvements, such as onboarding improvements, clearer learning goals, or smarter feedback, would go a long way in keeping them engaged. Their moderate stance suggests they are the easiest group to convert into active, long-term users if given the right nudges.

Those three clusters subsequently were labeled, key characteristics challenges and areas of improvement opportunities. The summary of those segmentations can be viewed on Table 8 below:

Cluster	Label	Key Characteristics	Challenges Identified	Opportunities for Improvement
Cluster 2	Highly Satisfied Users	Consistent engagement, confident use, benefit from current features	May represent users with high self-regulation or extrinsic motivation	Maintain strong UX; consider adding advanced modules to retain long-term engagement

Cluster	Label	Key Characteristics	Challenges Identified	Opportunities for Improvement
Cluster 1	Frustrated or Passive Users	Initial interest fades; high early use but low retention	Content lacks depth; poor feedback on skills; repetitive tasks	Improve feedback quality, diversify content, connect learning to real-life use
Cluster 0	Neutral Users	Mixed engagement, unsure about app's long-term value	Unclear purpose, weak onboarding, low personalization	Add adaptive learning paths, set clearer goals, provide targeted nudges to increase involvement

**Table 8.** User Segmentation and Behavioral Patterns

To sum up, usability alone is not enough to drive retention or real learning. While Cluster 2 represents a group that performs well within the current app structure, Clusters 1 and 0 represent clear design gaps. Users want more than just navigation, they want content that is meaty, feedback that is constructive, and learning that is compelling. Interview data supports this: most disengaged users drop off within a two-week window, not because of a buggy app, but one that does not develop with learners' demands. This early drop-off window highlights the urgency of improving depth, relevance, and feedback in the initial weeks of use.

#### 5.4. Towards Better MLLA Design: Analysis-Derived Recommendations

Results of both statistical modeling and user interviews reveal several main gaps that current mobile language learning apps (MLLAs) should address. Firstly, content depth and thematic organization was a recurring problem. Repetitive drills and insufficient progression annoyed numerous users. Despite showing moderate to high usability scores, the quantitative data also revealed long-term stagnation of engagement. Addressing these issues should help make MLLAs more engaging and effective for users. That would mean that usability would not be synonymous with meaningful learning. More content that is structured and scenario-based, such as modules based on simulating real-life tasks like job interviews, campus life, or traveling, would be required to more actively support real language use and intellectual stimulation. This shift from drill-based content to scenario-driven learning is likely to enhance perceived usefulness and competence simultaneously.

The other central issue is the lack of substantive speaking feedback. While some apps offer pronunciation scoring, users consistently reported that such feedback appeared superficial and often inaccurate. What learners need is feedback that reflects genuine communicative competence, e.g., grammar integration, intonation cues, and markers of fluency. This would align with the SDT principle of competence, in which learners are more motivated as their progress is evaluated in terms that reflect actual real-world accomplishment, rather than merely technical performance.

From a motivational design standpoint, features such as streaks and badges may capture users' initial interest, but they are insufficient for sustaining motivation over time. Both interview themes and UTAUT2-related metrics suggest users need more than just "fun" to stay engaged. Features that allow users to set personal goals, reflect on their progress, and build mastery over time can help promote intrinsic motivation. Providing clear pathways with adaptive difficulty and personalized milestones can help support autonomy and relevance, two vital pillars of Self-Determination Theory. Integrating these features early in the learner journey may significantly reduce the two-week drop-off observed in Clusters 1 and 0.

Lastly, two practical but overlooked issues emerged: premium paywalls and the absence of social features. Many learners reported that their learning momentum was abruptly halted by ads or paywalls, leading to disengagement. While monetization is understandable, pushing premium products too aggressively undermines retention and user trust. On the other hand, none of the apps reviewed provided intrinsic possibilities for peer interaction or collaborative learning. Not only is this at odds with FRAME's emphasis on mobile learning's social-contextual dimension, but it also denies users the chance to use language in real communicative situations. This omission could be remedied while promoting learner relatedness by including peer review, discussion forums, or collaborative challenges. Adding meaningful social layers would align mobile learning with real communicative practice, a core part of language acquisition missing in most existing apps.

Design Concern	Qualitative Evidence	Supported by SEM Results
Content Depth & Real-Life Scenarios	Repetitive drills, weak progression.	<i>Perceived Usefulness</i> and <i>Content Quality</i> had significant positive effects on Engagement.
Speaking Feedback Quality	Inaccuracy, lack of real communicative value.	<i>Perceived Competence</i> influences <i>Continuance Intention</i> (SDT path).
Shallow Gamification	Initial attraction but short-term impact.	<i>Hedonic Motivation</i> had weaker influence; <i>Intrinsic Motivation</i> was stronger in sustaining use.
Autonomy & Goal Setting	Need for progress tracking and goal orientation.	<i>Autonomy</i> positively affected Engagement and <i>Continuance Intention</i> .
Paywalls & Ads	Disruptive to learning flow.	<i>Effort Expectancy</i> and <i>Price Value</i> paths suggest negative impacts on intention.
Lack of Social Interaction	Absence of peer feedback or collaboration.	FRAME's <i>Social</i> component missing; <i>Relatedness</i> in SDT unfulfilled.

**Note:** Constructs listed under UTAUT2 and related frameworks are presented for theoretical interpretation and were not directly measured or included in the structural path model.

**Table 9.** Joint display aligning structural path model results with qualitative interpretations and theoretical lenses

Following user segmentation and statistical modeling (SEM) results, this section offers design guideline recommendations based on both quantitative results and qualitative observation. Usability and ease of access are crucial but are not sufficient to encourage ongoing use or compelling learning outcomes. SEM results indicated the central role of perceived usefulness, content quality, and motivational factors, all of which were confirmed through user interviews. The following recommendations address key areas of tension observed across user groups and modeled paths, with the goal to inform MLLA design with user-centered principles. Any references to additional theoretical constructs in this section are intended to support conceptual interpretation and do not reflect separate empirical models or estimated structural paths.

### 5.5. Final Reflection

This study confirms that most MLLAs are well-designed in the usability dimension, which aligns with both the technological aspect of TPACK and the device-focused aspects of the FRAME model. Users find such apps intuitive and visually appealing. However, both user interviews and SEM results reveal deeper concerns with pedagogy, context, and motivation, areas where many MLLAs are deficient.

These concerns are confirmed by the structural equation modeling results. While usability positively affects initial adoption, it has little effect on continued use. The observed relationships among engagement, retention, usability, and acceptance can be interpreted through constructs commonly discussed in UTAUT2, such as perceived usefulness and effort expectancy. However, these concepts are invoked here as theoretical lenses to contextualize the findings, rather than as empirically estimated predictors within the model.

Collectively, these findings point to the need for MLLAs to go beyond neat interfaces or extrinsic rewards. Effective mobile-supported language learning apps must include sound instructional design, facilitate authentic language use in daily life, and foster internal motivation. Adding features that allow learners to feel competent, make meaningful choices, and engage with others has the potential to transform MLLAs from tools of convenience to spaces of sustained learning and personal growth. Such integration would create a more durable motivational ecosystem, reducing dropout and enhancing long-term language development. Future research employing longitudinal or experimental designs would be required to examine causal sequencing among engagement, retention, usability, and acceptance.

Table 10 presents the mapping of SEM findings onto the theoretical frameworks, demonstrating how empirical evidence supports the conceptual underpinnings of good MLLA design.

SEM Construct	Key Finding	Related Framework	Theoretical Link
Usability	Positively affects initial acceptance but not long-term engagement	TPACK, FRAME	Matches <i>Technological Knowledge</i> in TPACK and <i>Device Usability</i> in FRAME, showing ease of use alone is not enough for sustained learning.
Perceived Usefulness	Strong predictor of continuance intention	TPACK	Reflects <i>Pedagogical-Technological</i> integration where users stay engaged when apps contribute meaningfully to their language development.

SEM Construct	Key Finding	Related Framework	Theoretical Link
Content Quality	High-quality, structured content improves satisfaction and motivation	TPACK, SDT	Tied to <i>Content Knowledge</i> in TPACK and <i>Competence</i> in SDT—users want rich, contextualized materials that mirror real-world usage.
Intrinsic Motivation	Strongest predictor of sustained use and satisfaction	SDT	Confirms the need for autonomy, competence, and relatedness to keep learners engaged beyond extrinsic incentives.
Social Interaction Absence	Weak or missing in most apps; limits user engagement	FRAME, SDT	Aligns with <i>Social Context</i> in FRAME and <i>Relatedness</i> in SDT—users benefit from peer interaction, collaboration, and feedback.
External Barriers (e.g., Ads)	Negative influence on user retention and trust	FRAME	Highlights how <i>Device + Context</i> interaction can disrupt flow—ads and paywalls reduce usability and perceived learning value.

Table 10. Linking SEM Results with Theoretical Frameworks

## 6. Conclusion and Recommendations

The study explored the user experience of Mobile Language Learning Applications (MLLAs) for engagement, retention, usability, and acceptance. While people accept MLLAs in general and find them relatively accessible, the majority of applications still fall short in maintaining long-term engagement and learning efficiently. Analysis via user feedback and statistics shows that the challenge lies mainly due to ineffective pedagogical approaches, too little personalization, and a lack of interactive or real-world aspects. These shortcomings suggest that current MLLAs prioritize interface design over deeper learning mechanisms needed for sustained progress.

The theoretical analysis using TPACK, FRAME, and Self-Determination Theory (SDT) highlights several key gaps. The majority of apps reflect technological design capability but are poor in pedagogical depth and learner-centered functionality. SDT, for instance, reports that the apps tend to neglect basic motivational drivers—autonomy, competence, and relatedness—necessary for long-term learning. When these needs are unmet, user engagement drops quickly despite initially positive impressions.

These theoretical results were confirmed by Structural Equation Modeling (SEM), which determined that usability and perceived ease of use are not enough to guarantee engagement and retention. SEM segregated the interaction among the four dimensions assessed, signaling that motivation and pedagogical alignment have more influential effects on retention than interface quality per se. This reinforces the idea that well-designed interfaces must be paired with meaningful learning pathways to achieve lasting impact.

Better, though, app developers require starting with better instructional design. In addition to the endless repetition of drills, features need to support scenario-based activities and simulations that reflect real communication. And students need more control—adaptive content, goal-setting tools, and interest-based modules can personalize the experience and create motivation. Embedding these elements early in the learner journey can help reduce the common two-week disengagement window identified in this study.

A good interface matters too. Clean navigation, timely feedback, and localization improve usability and user trust. For social interaction, functionality like discussion forums, peer activities, or live chat can build community. To address retention, developers need to look past badges and leaderboards, offering varied content progression and intrinsic rewards. Such enhancements would align MLLAs more closely with the social and contextual dimensions emphasized in FRAME and SDT.

Later MLLAs can also derive further benefits from context-aware or location-based features. Taking advantage of the use of GPS or geotagging, applications can link activities to geographical locations, compelling learners to exercise the language in natural settings. Such experiential learning mediates between computer-based learning and everyday communication. These location-aware interactions can transform passive practice into practical, meaningful engagement.

In brief, effective MLLAs must evolve beyond mere useful apps to seamless learning platforms. Supported by good pedagogy, for real-world application, and backed by motivational psychology, they have the capacity to foster deeper engagement and enhanced learning yield. By integrating insights from TPACK, FRAME, SDT, and SEM, future MLLAs can be designed in ways that meaningfully strengthen user motivation, contextual relevance, and long-term language development.

## References

- Al-Jamili, O., Aziz, M., Mohammed, F., Almogahed, A., & Alawadhi, A. (2024). Evaluating the efficacy of computer games-based learning intervention in enhancing English speaking proficiency. *Heliyon*, *10*(16), e36440. <https://doi.org/10.1016/j.heliyon.2024.e36440>
- Alabi, O., & Bukola, T. (2023). Introduction to Descriptive statistics. *Recent Advances in Biostatistics*. <https://doi.org/10.5772/intechopen.1002475>
- Alshammare, H., Alshayeb, M., & Baslyman, M. (2025). Revealing the mobile UX horizon: Exploring user experience aspects, attributes, and measurement methods - A systematic mapping study. *Computer Standards and Interfaces*, *94*(March), 103999. <https://doi.org/10.1016/j.csi.2025.103999>
- Andrade, C. (2021). The Inconvenient Truth About Convenience and Purposive Samples. *Indian Journal of Psychological Medicine*, *43*(1), 86–88. <https://doi.org/10.1177/0253717620977000>
- Blanchin, M., Guilleux, A., Hardouin, J.-B., & Sébille, V. (2019). Comparison of structural equation modelling, item response theory and Rasch measurement theory-based methods for response shift detection at item level: A simulation study. In *Statistical Methods in Medical Research* (Vol. 29, Issue 4, pp. 1015–1029). SAGE Publications. <https://doi.org/10.1177/0962280219884574>
- Blatna, D. (2013). Bootstrapping In Regression Analysis Of Tertiary Education Attainment In European Countries. *The 7th International Days of Statistics and Economics*, 154–163.
- Bock, H.-H. (2007). *Clustering Methods: A History of k-Means Algorithms*. 1957, 161–172. [https://doi.org/10.1007/978-3-540-73560-1\\_15](https://doi.org/10.1007/978-3-540-73560-1_15)
- Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior*, *36*(1), 3–15. <https://doi.org/10.1002/job.1960>
- Chen, M., & Hoe, T. W. (2025). K-Means Clustering: A Tool for English Language Teaching Innovations. *Forum for Linguistic Studies*, *7*(2), 988–998. <https://doi.org/10.30564/fls.v7i2.8379>
- Cooper, R. (2010). Theoretical considerations in qualitative interviewing. *Qualitative Report*, *15*(4), 1002–1005. <https://doi.org/10.46743/2160-3715/2010.1195>
- Creswell, J. W., & Clark, V. L. P. (2018). Designing and Conducting Mixed Methods Research - Third Edition. In *Encyclopedia of Research Design* (Third Edit). <https://doi.org/10.4135/9781412961288.n245>
- Dai, Y., & Wu, Z. (2022). Mobile-assisted peer feedback on EFL pronunciation: Outcome effects, interactional processes, and shaping factors. *System*, *111*(July 2021), 102953. <https://doi.org/10.1016/j.system.2022.102953>
- DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, *40*(4), 314–321. <https://doi.org/10.1111/j.1365-2929.2006.02418.x>
- Elhami, A. (2022). Conducting an Interview in Qualitative Research : *MEXTESOL Journal*, *46*(1), 1–7.
- Elliott, R., & Timulak, L. (2015). A Handbook of Research Methods for Clinical and Health Psychology (DRAFT). *A Handbook of Research Methods for Clinical and Health Psychology (DRAFT)*, 147–160. <https://doi.org/10.1093/med:psych/9780198527565.001.0001>
- Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fangohr, H., Beg, M., Bergemann, M., Bondar, V., Brockhauser, S., Carinan, C., Costa, R., Antonia, F. D., Danilevski, C., Ehsan, W., Esenov, S. G., Fabbri, R., Fangohr, S., Flucke, G., Fortmann, C., Marsa, D. F., Giovanetti, G., Goeries, D., Hauf, S., ... Perrin, J. F. (2019). Data Exploration and analysis with Jupyter notebooks. *Proceedings of 17th International Conference on Accelerator and Large Experimental Physics Control Systems*, January. <https://doi.org/10.18429/JACoW-ICALEPS2019-TUCPR02>
- Fisher, R. A. (2009). Descriptive Statistics Objective: Review the basic concepts of elementary statistics. *Mathematical Statistics with Applications*, 1–51. [www.stetson.edu/~efriedma/periodictable/jpg/Fisher.jpg](http://www.stetson.edu/~efriedma/periodictable/jpg/Fisher.jpg)
- Freedman, D. A. (2007). Bootstrapping Regression Models. *The Annals of Statistics*, *9*(6), 1–14. <https://doi.org/10.1214/aos/1176345638>
- Guerrero, L. A., Ochoa, S., & Collazos, C. (2010). A mobile learning tool for improving grammar skills. *Procedia - Social and Behavioral Sciences*, *2*(2), 1735–1739. <https://doi.org/10.1016/j.sbspro.2010.03.975>
- Hwang, G. J., Rahimi, M., & Fathi, J. (2024). Enhancing EFL learners' speaking skills, foreign language enjoyment, and language-specific grit utilising the affordances of a MALL app: A microgenetic perspective. *Computers and Education*, *214*(February), 105015. <https://doi.org/10.1016/j.compedu.2024.105015>

- ISO 9241-210. (2010). ISO 9241-210: Ergonomics of human–system interaction - Human-centred design for interactive systems. *International Organization for Standardization*, 2010, 32.
- Janebi Enayat, M., Asadi Ghadim, N., & Arabmofrad, A. (2025). Effects of two mobile-assisted language learning apps on L2 receptive and productive vocabulary knowledge: A mixed-methods study. *System*, 133(April), 103763. <https://doi.org/10.1016/j.system.2025.103763>
- Kelly, S., Kaye, S. A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, 77(December 2022), 101925. <https://doi.org/10.1016/j.tele.2022.101925>
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods and Research*, 44(3), 486–507. <https://doi.org/10.1177/0049124114543236>
- Kessler, M., Loewen, S., & Gönülal, T. (2023). Mobile-assisted language learning with Babbel and Duolingo: comparing L2 learning gains and user experience. *Computer Assisted Language Learning*, May. <https://doi.org/10.1080/09588221.2023.2215294>
- Koole, M. (2022). *The Framework for the Rational Analysis of Mobile Education (Frame) Model: An Evaluation of Mobile Devices for Distance Education. February 2006*.
- Koole, M. L. (2009). Mobile Learning - A Model for Framing Mobile Learning. *Mobile Learning: Transforming the Delivery of Education and Training*, 1(2), 25–47.
- Kukulka-Hulme, A. (2009). Will mobile learning change language learning? *ReCALL*, 21(2), 157–165. <https://doi.org/10.1017/S0958344009000202>
- Kukulka-Hulme, A. (2021). Reflections on research questions in mobile assisted language learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 28–46. <https://doi.org/10.1515/jccall-2021-2002>
- Kukulka-Hulme, A., & Viberg, O. (2018). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology*, 49(2), 207–218. <https://doi.org/10.1111/bjet.12580>
- Kunnan, A. J. (1998). An introduction to structural equation modelling for language assessment research. In *Language Testing* (Vol. 15, Issue 3, pp. 295–332). SAGE Publications. <https://doi.org/10.1177/026553229801500302>
- Lin, C. J., Hwang, G. J., Fu, Q. K., & Cao, Y. H. (2020). Facilitating EFL students' English grammar learning performance and behaviors: A contextual gaming approach. *Computers and Education*, 152(November 2019), 103876. <https://doi.org/10.1016/j.compedu.2020.103876>
- Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, 31(3), 293–311. <https://doi.org/10.1017/S0958344019000065>
- Lubis, M., Sutoyo, E., Azuddin, M., & Handayani, D. (2019). User Experience in Mobile Application Design: Utility Defined Context of Use. *Journal of Physics: Conference Series*, 1361(1). <https://doi.org/10.1088/1742-6596/1361/1/012043>
- Marshall, G., & Jonker, L. (2010). An introduction to descriptive statistics: A review and practical guide. *Radiography*, 16(4), e1–e7. <https://doi.org/10.1016/j.radi.2010.01.001>
- Mckinney, W. (2017). *Python for Data Analysis - 2nd edition*.
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education*, 108(6), 1017–1054. <https://doi.org/10.1177/016146810610800610>
- Okonkwo, C. (2024). Assessment of User Experience (UX) Design Trends in Mobile Applications. *Journal of Technology and Systems*, 6(5), 29–41. <https://doi.org/10.47941/jts.2147>
- Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*, 07(08), 64–69. <https://doi.org/10.31695/ijasre.2021.34050>
- Priyatna, J. E., Sriwindono, H., Prima, P. H., & Polina, A. M. (2024). *The Application of K-Means Clustering Algorithm for Initial Analysis of Students Online Learning*. 6(07), 67–75. <https://doi.org/10.35629/5252-06076775>
- Robinson, O. C. (2023). Probing in qualitative research interviews: Theory and practice. *Qualitative Research in Psychology*, 20(3), 382–397. <https://doi.org/10.1080/14780887.2023.2238625>

Royo, T., & Laborda, J. G. (2018). *Standardization of Design Interfaces Applied to Language Test on-line through Ubiquitous Devices*. <https://doi.org/10.3991/ijim.v12i4.9197>

Ryan, R. M., & Deci, E. L. (2017). Self-determination theory. Basic psychological needs in motivation, development and wellness. *Revue Québécoise de Psychologie*, 38(3), 231. <https://doi.org/10.7202/1041847ar>

Sakkir, G., & Syamsuddin, N. A. (2023). Students' Perceptions of Duolingo Mobile Assisted Language Learning (MALL) in Learning English Vocabulary. *EduLine: Journal of Education and Learning Innovation*, 3(3), 381–388. <https://doi.org/10.35877/454ri.eduline1970>

Sevara, M. (2024). The Impact of Language Learning Apps on English Proficiency. *European Journal of Emerging Technology and Discoveries*, 2(9), 45–58.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>

Xin, S., & Ren, B. (2016). Application of Regression Analysis for Small Samples Based on Bootstrap Method. *Proceedings of the 2016 International Conference on Education, Management, Computer and Society*, 37(Emcs), 816–819. <https://doi.org/10.2991/emcs-16.2016.200>

**APPENDICES**

**Appendix A. Demographic and Application-Usage Questionnaire**

This appendix presents the demographic items and preliminary usage questions administered prior to the main UX instrument.

**A1. Demographic Questions**

No.	Question	Response Type
1	Age	Dropdown: 18, 19, 20, 21, 22, ≥23
2	Gender	Multiple choice: Male, Female
3	Study Program	Short answer (program name)
4	Have you ever used a mobile language learning application?	Multiple choice: Yes, No

**A2. Application-Usage Questions**

No.	Question	Response Type
1	Which application do you use?	Multiple choice: Duolingo, ELSA Speak, Other
2	How long have you used the application?	Dropdown: <1 month, 1–2 months, 3–4 months, 5–6 months, >6 months
3	What type of plan do you use?	Multiple choice: Trial, Free, Premium, Not sure
4	On average, how often do you use the application?	Multiple choice: Every day, 3–4 times/week, 1–2 times/week, Rarely

**Appendix B. User Experience Questionnaire**

This appendix contains the full set of UX items used in the study. The final instrument includes **18 validated items** measuring Engagement, Retention, Usability, and Acceptance on a **5-point Likert scale** (1 = *Strongly Disagree*, 5 = *Strongly Agree*). Items Q11 and Q16 were removed after the pilot due to low reliability.

**B1. Engagement (5 items)**

1. I feel motivated to keep using this application even when I am not required to.
2. The materials in this application feel boring because they repeat too much and rarely offer new challenges.
3. I rarely find opportunities to use the language in real-life contexts through this application.
4. I often feel like I am just clicking through the lessons without truly understanding the content.
5. The content in this application does not feel relevant to my daily life or local culture.

**B2. Retention (5 items)**

6. I still remember a good portion of the material I learned from this application over time.
7. I struggle to remember the material because the application emphasizes memorization rather than real usage.
8. The examples in this application are not based on situations I encounter in daily life.
9. Excessive repetition in the application makes the learning experience less memorable.

10. I still cannot apply what I learned in real conversations or practical tasks.

### **B3. Usability (4 retained items)**

11. — *Removed after pilot testing (low item-total correlation)*

12. I find it difficult to locate features that help me practice the language in a practical and meaningful way.
13. The application does not support activities that resemble natural interaction in real-life situations.
14. Some lessons feel too repetitive, even when I have already mastered the material.
15. The content feels unfamiliar or disconnected from my everyday environment.

### **B4. Acceptance (4 retained items)**

16. — *Removed after pilot testing (low reliability)*

17. I do not feel that I am actually learning how to use the language in real tasks or real-life situations.
18. The lessons focus too much on repetitive exercises and do not encourage deeper understanding.
19. I do not feel represented in the examples or cultural contexts shown in the application.
20. The repetitive format of the exercises reduces my motivation over time.

## **Appendix C. Interview Protocol**

This appendix provides the complete interview guide used during qualitative data collection.

### **C1. Warm-up and Demographics**

1. Could you tell me a bit about your major and year of study?
2. Have you used language-learning applications such as Duolingo or ELSA Speak? For how long?

### **C2. Engagement (4 questions)**

*Linked with SDT and FRAME*

3. What usually makes you feel interested or excited to use this language-learning app?
4. Have there been moments when you felt bored or lost interest? What caused that?
5. Does the way the app presents lessons feel lively and enjoyable or more routine?
6. Are there features that make you feel more motivated to keep learning?

### **C3. Retention (4 questions)**

*Linked with TPACK and SDT*

7. Do you still remember and use the material you learned from the app?
8. How often are you able to apply what you learned in real life?
9. Does the app focus more on memorization or deeper understanding?
10. Are there features that help strengthen your memory or language skills?

### **C4. Usability (3 questions)**

*Linked with FRAME*

11. What was your first impression when using the app—easy or confusing?
12. Are there features that feel unintuitive or frustrating?
13. Have you experienced technical issues such as bugs, errors, or slow loading?

### **C5. Acceptance (4 questions)**

*Linked with SDT, TPACK, FRAME*

14. Do you think this app is suitable for long-term use in language learning? Why?
15. Have you ever recommended the app to others? Why or why not?
16. What are your expectations for future improvements?
17. If you could change one thing to make the app better suited for your needs, what would it be?

### **C6. Closing**

18. Is there anything else you would like to add?
19. Thank you for participating.

## **Appendix D. Thematic Coding Matrix (15 interviewees)**

**Description:** This table lists each interview participant (anonymized ID), the app they used, and the themes/codes that emerged in their interview. Checkboxes indicate which themes were present for that participant.

Theme/Code	Interviewee														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Repetitive & monotonous content	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gamified UI Fun (superficial)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Real-world Use/Retention(No)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Speaking Feature Limitations			✓		✓			✓		✓		✓		✓	✓
Premium Walls/Ads							✓	✓	✓				✓		✓
Rarely Recommend	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Suggested Improvements			✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓

**Appendix E. SEM Model Results and Fit Indices**

**E1. Standardized Path Estimates**

Path	Estimate	p-value	Interpretation
Engagement → Usability	0.3188	1.18e-08	Higher engagement is associated with higher perceived usability
Retention → Usability	0.3794	1.24e-10	More consistent use predicts higher usability
Usability → Acceptance	0.7548	< 0.000001	Strongest effect; usability strongly predicts acceptance

Residual variances: Usability = 0.196; Acceptance = 0.228

**E2. Model Fit Indices**

Fit Index	Value	Interpretation
CFI	0.9348	Acceptable model fit
TLI	0.8957	Marginal but acceptable for low-df structural models
RMSEA	0.2119	Inflated due to df = 5; RMSEA is unreliable with low degrees of freedom
SRMR	<i>Not computed in semopy for this model</i>	

The relatively high RMSEA value should be interpreted with caution. RMSEA has been shown to be sensitive to models with very low degrees of freedom, where it may overestimate lack of fit even when other indices indicate acceptable model performance (Kenny et al., 2015). Given that the present structural path model uses composite variables and a parsimonious specification, the RMSEA value is likely influenced by these characteristics. Therefore, model adequacy was evaluated using a combination of absolute and incremental fit indices rather than relying solely on RMSEA. In low-df path models with observed composite variables, CFI and TLI are generally considered more informative indicators of comparative model fit.

**Appendix F. Pilot Test: Validity & Reliability (Item and Scale Statistics)**

A pilot test was conducted with 20 participants to assess item clarity and internal consistency. Item-total correlations and factor diagnostics indicated that two items (Q11 and Q16) did not meet the pre-specified thresholds for retention (CI-TC Q11 = 0.661; Q16 = 0.415).

**F1. Summary reliability (pilot → final)**

Construct	Initial items (before pilot)	Final items (after removal)	Cronbach's α (final)
Engagement	5	5	0.868
Retention	5	5	0.859
Usability	5 → 4 (Q11 removed)	4	0.858
Acceptance	5 → 4 (Q16 removed)	4	0.868
<b>Overall UX scales</b>	20 → 18	18	α range: 0.858 – 0.868

**F2. Item diagnostics table (Pilot)**

Section 1 (Engagement) Analysis			Section 2 (Retention) Analysis			Section 3 (Usability) Analysis			Section 4 (Acceptance) Analysis		
Cronbach's Alpha: <b>0.924</b>			Cronbach's Alpha: <b>0.946</b>			Cronbach's Alpha: <b>0.903</b>			Cronbach's Alpha: <b>0.879</b>		
Q Code	CI-TC	CA-If Del	Q Code	CI-TC	CA-If Del	Q Code	CI-TC	CA-If Del	Q Code	CI-TC	CA-If Del
Q1	0.723	0.922	Q6	0.832	0.938	Q11	0.661	0.902	Q16	0.415	0.925
Q2	0.812	0.905	Q7	0.853	0.934	Q12	0.794	0.875	Q17	0.765	0.842
Q3	0.866	0.894	Q8	0.884	0.929	Q13	0.700	0.896	Q18	0.886	0.814
Q4	0.746	0.920	Q9	0.818	0.940	Q14	0.783	0.876	Q19	0.777	0.838
Q5	0.877	0.892	Q10	0.882	0.929	Q15	0.864	0.859	Q20	0.781	0.837

