**List of Abbreviations**

ANOVA – Analysis of Variance

ANCOVA – Analysis of Covariance

VLT – Vocabulary Levels Test

XP – Experience Points

M – Mean

SD – Standard Deviation

SE – Standard Error

d – Cohen’s d

η² – Eta squared

**List of Formulas**

* VLT\_i = PostVLT\_i – PreVLT\_i
* Mean Difference = M\_post – M\_pre
* Cohen’s d = (M1 – M2) / SD\_pooled,

where SD\_pooled = √[(SD1² + SD2²) / 2]

* Adjusted Mean\_{Group} = Raw Mean\_{Group} – b(PreCovariate – Overall Mean\_{Pre}),

where b = regression coefficient

**Chapter 4: Findings**

This chapter presents the findings of the mixed‑methods study that assessed mobile‑app teaching and traditional classroom instruction for Spanish language learning (Creswell & Plano Clark, 2018). The procedures mentioned in Chapter 3 are carried out for the analyses. Section 4.1 contains numerical information about vocabulary, proficiency in speaking and student involvement. Section 4.2 provides a list of main themes and interesting quotes shared in interviews transcripts and survey answers. The section shows how qualitative and quantitative findings can be connected along with a list of key findings.

**4.1 Quantitative Findings**

As described in Chapter 3, 100 participants, equally divided into two groups according to instruction method as described in chapter 3. The data showed no missing values, both skewness and kurtosis were within acceptable bounds and there was no problem with the variance of the data (Levene’s test results were p >0.05). All analysis processes were carried out using SPSS v28 with the alpha level set at 0.05 (IBM Corp, 2021).

**4.1.1 Vocabulary Growth (VLT)**

* The research used a two-way ANOVA to assess whether the method used, students’ age and levels of experience affected results in the VLT. The instructional method had the strongest influence on students’ performance, F(1,196) = 74.53, p < 0.001, η² = 0.28.
* For the classroom students, the Mean Vocabulary Level Test (M = 18.74, SD = 6.23) was higher than that of the App group (M = 12.82, SD = 4.73).
* None of the interactions between age groups and learning methods or between learning experience levels and interaction methods, reached statistical significance, as described in Chapter 3.
* At Week 6, the mid‑test results showed treatment was continuing to work as before.

**4.1.2 Oral Proficiency (PostOral)**

* The ANCOVA examined how well students spoke after the lessons, as this was compared with their scores at the beginning of the course.
* The overall assumptions required for regression are fulfilled.
* The method through which lessons were taught had an important impact, F(1,197) = 36.27, p < 0.001, η² = 0.16.
* The adjusted means were generally in favour of Classroom learners (M = 11.12, SE = 0.23) compared to those who learned through the app (M = 9.80, SE = 0.24).
* The medium‑to‑large effect was associated with a Cohen’s d of 0.68.

**4.1.3 Engagement and Motivation**

Scores on the surveys of engagement and motivation were analysed with an independent‑samples t‑test.

* The students in the classroom group reported greater engagement (M = 4.51, SD =0 .50) than those in the App group (M = 3.97, SD = 0.73), t(198) = 5.34, p < 0.001, d = 0.76.
* Classroom score for motivation was higher than App score (M = 4.27, SD = 0.62 v.s. M = 3.88, SD = 0.84), t(198) = 3.61, p < .001, d = 0.52.
* Average app use data was 344.22 XP per day (SD = 100.09) and record a streak over 42.38 days (SD = 14.60), classroom users did not have any interaction with the app, meeting the necessary requirements

***Table 4.1 Descriptive Statistics for Quantitative Measures by Instructional Method***

|  |  |  |  |
| --- | --- | --- | --- |
| Measure | Group | M | SD |
| PreVLT | App | 39.44 | 6.41 |
|  | Classroom | 40.84 | 7.55 |
| PostVLT | App | 52.25 | 7.89 |
|  | Classroom | 59.57 | 8.81 |
| ΔVLT | App | 12.82 | 4.73 |
|  | Classroom | 18.74 | 6.23 |
| PreOral | App | 7.25 | 2.03 |
|  | Classroom | 7.23 | 1.96 |
| PostOral | App | 9.83 | 2.03 |
|  | Classroom | 11.18 | 1.96 |
| AvgXPperDay | App | 344.22 | 100.09 |
| StreakLength | App | 42.38 | 14.60 |
| Likert Engagement | App | 3.97 | 0.73 |
|  | Classroom | 4.51 | 0.50 |
| Likert Motivation | App | 3.88 | 0.84 |
|  | Classroom | 4.27 | 0.62 |

**4.2 Qualitative Findings**

Open-ended responses were obtained through the Week 12 online questionnaire (100 responses) and through semi-structured interviews conducted on Zoom and in person (collection of 50 interview excerpts). The answers were transcribed and moved into NVivo 14, where Braun and Clarke’s (2006) method of inductive thematic analysis was implemented. Independent coders were given the entire dataset, set up initial categories and later organised them into 5 main redundant themes. The coding applied by each person was found to be very similar (Cohen’s κ = 0.82). The researchers compared the main themes with both the scores related to involvement and participants’ increase in skills to ensure agreement among the different approaches.

**4.2.1 Theme Frequency**

Five themes emerged, summarized in the table below.

***Table 4.2 Frequency of Emergent Themes in Learner Feedback***

|  |  |  |
| --- | --- | --- |
| Theme | Count | Percentage |
| Gamification Motivation | 13 | 25.7% |
| Isolation Frustration | 11 | 22.9% |
| Peer Interaction Value | 10 | 20.0% |
| Technical Usability | 9 | 17.1% |
| Feedback Appreciation | 7 | 14.3% |

**4.2.2 Exemplar Quotations**

**Gamification Motivation**

“Seeing my streak climb every morning pushes me to open the app daily.” (ID 1)

“I felt proud when I hit 30 XP in a day—it kept me engaged.” (ID 2)

**Isolation Frustration**

“No real conversations; Duolingo can’t replace talking.” (ID 3)

“I miss feedback from real people—I’ll never know if I’m pronouncing correctly.” (ID 4)

**Peer Interaction Value**

“Group role‑plays were the highlight of our lessons.” (ID 79)

“I learned from my peers’ mistakes more than from the teacher.” (ID 80)

**Technical Usability**

“The app sometimes froze, and I lost my progress mid‑lesson.” (ID 78)

“Navigating the grammar section was confusing on mobile.” (ID 5)

**Feedback Appreciation**

“I appreciated the structured feedback on pronunciation in class.” (ID 76)

“The teacher’s immediate corrections were invaluable.” (ID 77)

**4.3 Joint Display of Findings**

***Table 4.3 Joint Display Integrating Quantitative Results and Qualitative Themes***

|  |  |
| --- | --- |
| Quantitative Result | Qualitative Insight (Theme + Excerpt) |
| ΔVLT: Classroom > App, F(1,196)=74.53, p<0.001 | Gamification Motivation: “Seeing my streak climb…” (ID 1) |
| PostOral: Classroom > App, F(1,197)=36.27, p<0.001 | Isolation Frustration: “I miss feedback from real people…” (ID 4) |
| Engagement: Classroom > App, t(198)=5.34, p<0.001 | Peer Interaction Value: “Group role‑plays were the highlight…” (ID 79) |
| Motivation: Classroom > App, t(198)=3.61, p<0.001 | Feedback Appreciation: “I appreciated the structured feedback…” (ID 76) |

**4.4 Summary of Findings**

* Learners in a classroom made more progress in expanding their vocabularies than those learning through the App (η² = 0.28).
* Participants in the classroom did better at speaking than those using the App (η² = 0.16).
* Both the incoming motivation and current engagement in the classroom were rated higher than those in the online group (d = 0.76 and d = 0.52).
* A total of five categories came out on top: Gamification Motivation (25.7%), Isolation Frustration (22.9%), Peer Interaction Value (20.0%), Technical Usability (17.1%) and Appreciation for Feedback (14.3%).
* Quotations from participants show that the gamified experience promoted daily interaction, lack of interaction with others hindered oral improvement and the advice in classrooms helped them improve their skills in the language.

**References**

* Braun, V., & Clarke, V. (2006). "Using thematic analysis in psychology." Qualitative Research in Psychology, 3(2), 77–101.
* Creswell, J. W., & Plano Clark, V. L. (2018). "Designing and Conducting Mixed Methods Research" (3rd ed.). SAGE Publications.
* IBM Corp. (2021). IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp.
* Cohen, J. (1988). Statistical power analysis for the behavioural sciences (2nd ed.). Routledge.