Master Degree Project

UNIVERSITY
OF SKÖVDE

GAMIFICATION AND PERCEIVED LONG-TERM ENGAGEMENT IN LANGUAGE LEARNING

Master Degree Project in Informatics
One year Level 22.5 ECTS
Spring term 2026

Jiaxin Wang
Supervisor: Per Backlund
Examiner: Tarja Susi

Abstract

Although gamification is widely used to support motivation and engagement in language learning, less is known about how learners perceive its value for sustained participation over time. This study examines how individual gamification mechanics in Duolingo are associated with perceived long-term engagement and perceived learning value. A sequential mixed-methods case study was conducted, combining a design analysis of Duolingo's visible gamification mechanics, an online survey of 155 users, and structured written follow-up responses from seven participants. Descriptively, the streak received the highest mean helpfulness rating for engagement (M = 3.03) and was ranked as the most helpful mechanic by 58.06% of respondents. At the same time, 40.00% associated the streak with pressure or discomfort. Leaderboards were the most frequently reported source of pressure or discomfort (45.16%), while the Hearts system received the lowest mean helpfulness rating (M = 2.76). The qualitative findings further indicated that gamification could support habit formation, regular language exposure, and continued app use, but could also encourage minimum-effort completion, competitive fatigue, and attention to game-related goals rather than learning goals. Overall, the findings suggest that gamification is perceived as most useful when it supports continuity and meaningful progress without becoming overly controlling or punitive. Because most respondents had used Duolingo for six months or less, the findings concern perceived sustained engagement rather than objectively measured long-term retention or language proficiency gains.

Keywords: gamification, language learning, perceived long-term engagement, motivation, learner engagement, Duolingo

Table of Contents

1 Introduction 1

2 Background 3

2.1 Gamification and Language Learning 3

2.2 Duolingo's Gamification Mechanisms 5

2.3 Self-Determination Theory, Reinforcement Theory and Feedback Mechanisms 8

3 Research Design 11

3.1 Research Questions 11

3.2 Research Methodology 12

3.2.1 Overall Research Design 12

3.2.2 Design Analysis 13

3.2.3 Questionnaire 14

3.2.4 Structured Written Follow-up Interviews 15

3.2.5 Pilot Study 16

3.3 Ethical Considerations 17

3.4 Research Participants and Procedure 19

3.4.1 Participants and Recruitment 19

3.4.2 Research Procedure 20

3.5 Data Analysis 20

3.5.1 Design Analysis 21

3.5.2 Questionnaire Analysis 21

3.5.3 Analysis of the Structured Written Follow-up Interviews 22

3.5.4 Integration and Theoretical Interpretation 22

4 Results and Analysis 24

4.1 Design Analysis of Duolingo Gamification Mechanics 24

4.1.1 Streak 25

4.1.2 XP, Daily Goals, and Progress System 27

4.1.3 Leaderboard 28

4.1.4 Hearts, Streak Freeze, Reminders, and Badges 29

4.2 Survey Findings 30

4.2.1 Sample Overview 30

4.2.2 Streak-Driven Behaviour 32

4.2.3 Feature Helpfulness Ratings 33

4.2.4 User Preference Rankings 34

4.2.5 Pressure and Discomfort 35

4.2.6 Long-Term User Segment 35

4.3 Findings from the Structured Written Follow-up Interviews 37

4.3.1 Streak as a Habit-Supporting Mechanism and a Source of Pressure 38

4.3.2 Minimum-Effort Completion 38

4.3.3 Leaderboard Motivation and Fatigue 39

4.3.4 Hearts and Interrupted Learning Flow 40

4.3.5 Perceived Learning Value and Daily Language Exposure 41

4.4 Integrated Analysis 41

4.5 Theoretical Interpretation 44

4.5.1 Self-Determination Theory 44

4.5.2 Reinforcement Theory 45

4.5.3 Feedback Mechanisms 46

4.6 Summary of Analysis 47
5 Conclusion 49
5.1 Overview 49
5.2 Limitations 49
5.3 Future Research 51
6 Acknowledgements 53
References 54
Appendix A - Survey 58
Appendix B - Structured Written Follow-up Interview Questions 60

1 Introduction

In recent years, gamification has become a common design strategy across a wide range of application contexts, including fitness applications, professional skills training, team management, and language learning. Gamification is generally defined as the use of game-design elements and incentive structures in non-game contexts to motivate users and increase their engagement (Deterding et al., 2011; Hamari et al., 2014). It commonly employs mechanisms such as points, streaks, badges, progress indicators, rewards, and leaderboards to provide users with goals, feedback, and incentives, thereby encouraging more frequent engagement and supporting sustained motivation (Sailer and Homner, 2020). The underlying definition and the broader claim that gamification can influence motivational and behavioural learning outcomes are consistent with the cited literature.

Language learning is one of the prominent contexts in which gamification is applied. Because improvements in language proficiency typically depend on sustained, regular, and long-term practice, learners need to maintain a relatively stable frequency of study and a willingness to remain engaged over an extended period. However, language learning often involves repetitive practice, substantial memory demands, and periods in which progress is difficult to perceive, all of which may contribute to learner fatigue or discontinuation. Gamification mechanisms are therefore frequently incorporated into language-learning applications to help learners establish study habits, maintain learning motivation, and reduce the sense of monotony associated with long-term learning (Zhang and Hasim, 2023).

Duolingo can be regarded as a representative example of the use of gamification in language learning. As a widely used mobile language-learning application, Duolingo employs a range of gamification mechanisms to encourage regular practice, including streaks, experience points (XP), daily goals, leaderboards, achievement badges, reminders, and the Hearts system (Duolingo, 2023). These mechanisms are designed to make the language-learning process more structured, reward-oriented, and interactive, while supporting learner engagement through extrinsic rewards, progress feedback, and social competition. More broadly, such mechanisms may play an important role in attracting users, strengthening learning motivation, and encouraging continued use. The individual feature names are supported by Duolingo's own documentation, although the Hearts terminology must be aligned with the period in which the empirical data were collected because Duolingo began introducing its replacement Energy system in 2025.

However, despite the widespread adoption of gamification in language-learning applications such as Duolingo, existing research often treats gamification as a holistic intervention and pays comparatively less attention to differences in the effects of individual gamification mechanisms. Further investigation is therefore needed to determine which specific mechanisms learners perceive as most supportive of sustained engagement, which mechanisms may generate pressure, fatigue, or superficial engagement, and which mechanisms may increase frequency of use without necessarily contributing to genuine improvements in language proficiency. Recent research has suggested that some gamification designs may lead learners to become excessively focused on gamified rewards and goals, thereby distracting them from the underlying learning activity (Mogavi et al., 2022).

Research on the negative effects of gamification in educational software has likewise found that mechanisms such as badges, leaderboards, competition, and points may, in some contexts, be associated with motivational problems, poorer learning performance, lack of understanding, and other forms of ineffective engagement (Almeida et al., 2023).

It is therefore necessary to examine more closely the specific roles and potential limitations of different gamification mechanisms in language-learning applications.

Against this background, this study aims to examine the perceptions of different gamification mechanisms in language-learning contexts on user perceptions, sustained engagement, and perceived learning outcomes. Drawing on Self-Determination Theory, reinforcement theory, and research on feedback mechanisms as analytical perspectives, the study uses Duolingo as a representative case. It first analyses the principal gamification mechanisms employed in the language-learning context and the design logic underlying them, including streaks, experience points, leaderboards, achievement badges, daily goals, reminders, and the Hearts system. It then draws on data collected through an online questionnaire and supplementary structured written qualitative responses (Gill et al., 2008) to investigate how users perceive these mechanisms, how they influence users' continued-use intention and engagement behaviour, and the extent to which users perceive them as contributing to their language learning. By integrating questionnaire data with structured written follow-up interview data, this study seeks to identify differences among gamification mechanisms in their capacity to support engagement in language learning and to further examine their potential limitations.

2 Background

2.1 Gamification and Language Learning

Research on gamification in education commonly conceptualises gamification as a strategy that uses design mechanisms to influence learners' behaviour and learning experiences. Dicheva et al. (2015), in their systematic mapping study of gamification in education, identified points, badges, leaderboards, levels, challenges, progress indicators and immediate feedback as some of the most frequently used gamification mechanisms in educational contexts. By establishing explicit goals, visualising progress and providing rewards, these mechanisms can influence learners' engagement and make learning activities more structured and interactive. However, Dicheva et al. (2015) also emphasised that the empirical evidence available in early educational gamification research remained limited, particularly with regard to controlled evaluations of individual gamification mechanisms in specific educational contexts and among particular learner groups. The potential value of gamification therefore does not imply that all gamification mechanisms will produce equally positive effects across different learning contexts.

With the development of mobile learning, language learning has increasingly extended beyond the traditional classroom into more fragmented and everyday digital contexts. Godwin-Jones (2011) argues that mobile applications provide learners with more flexible opportunities for language learning, allowing them to make use of short periods of time in their daily lives for language practice. Kukulska-Hulme and Viberg (2018) similarly highlights the portability and situated nature of mobile-assisted language learning, through which mobile technologies can connect learning activities across formal and informal contexts. Such flexibility means that language-learning applications need not only to provide learning content, but also to encourage users to return regularly through mechanisms such as reminders, goals, feedback and rewards.

At the same time, the integration of mobile learning into everyday life may make the distinction between private time and study time less clear. This issue is particularly relevant when learning technologies are constantly accessible through personal mobile devices. Earlier research on mobile-assisted language learning has explicitly identified the distinction between private time and study time as a potential psychological barrier associated with mobile learning. It is therefore important to recognise that mechanisms designed to encourage repeated engagement may not always be experienced positively by learners.

In studies focusing specifically on language-learning applications, Duolingo has become one of the most frequently discussed cases. Munday (2016) examined the potential use of Duolingo as a supplementary tool in classroom language learning and suggested that its accessibility, relatively low barrier to use and task-based

structure could support additional practice, particularly for the reinforcement of vocabulary and grammar. At the same time, the study suggests that Duolingo is more appropriately understood as a supplement to classroom instruction than as a complete replacement for richer and more contextualised language-learning activities.

Loewen et al. (2019) conducted a semester-long study involving nine participants who used Duolingo to learn Turkish. The participants demonstrated overall improvement on the study's second-language measures, and their perceptions of Duolingo's flexibility and gamification were generally positive. However, substantial differences were observed in participants' motivation and persistence, while some participants also reported frustration with the instructional materials. These findings indicate that although the application may provide a flexible and engaging environment for additional language practice, users do not necessarily respond to its learning design in the same way.

From another perspective, Shortt et al. (2023), in a systematic review of research on Duolingo, argue that existing studies have often focused more heavily on the application's design features than on critically examining its underlying models of language learning and gamification. Their review identifies a strong emphasis on design-focused research and notes that Duolingo has traditionally relied extensively on drill-based activities, repetition, competition and extrinsic rewards. Although such mechanisms may support regular engagement, they do not necessarily address language learning as a fundamentally social and culturally situated process. Shortt et al. (2023) therefore raise an important distinction between encouraging learners to study regularly and encouraging deeper information processing, authentic language use and communication. From this perspective, repeatedly opening the application, completing tasks or accumulating XP should not automatically be interpreted as evidence of deeper language processing or the development of communicative competence.

Other research has more directly identified potentially detrimental consequences of gamification. Mogavi et al. (2022) describe gamification misuse as a situation in which users become excessively focused on gamification and are consequently distracted from the underlying learning activity. Their qualitative study of Duolingo found that competitiveness and excessive engagement with game-like aspects could redirect users' attention away from language learning itself, with implications for learning performance and well-being. Such behaviour may include repeatedly completing relatively easy activities, attempting to maximise XP efficiently, or prioritising competition with other users over engagement with more demanding learning content. More broadly, Almeida et al. (2023) found that badges, leaderboards, competition and points were among the gamification mechanisms most frequently associated with reported negative effects in educational software. These effects included a lack of the intended effect, poorer performance,

motivational problems, insufficient understanding and irrelevant engagement, as well as behaviours such as gaming the system or cheating.

Overall, gamification offers clear potential within educational contexts, but its effects are neither uniformly positive nor independent of the particular mechanisms through which it is implemented. Using Duolingo as a case study, this study therefore examines how different gamification mechanisms affect user perceptions, long-term engagement and perceived learning outcomes in a specific language-learning context. The distinction between individual mechanisms is particularly important because existing research indicates that design features capable of increasing engagement may not necessarily produce corresponding improvements in language learning.

2.2 Duolingo's Gamification Mechanisms

Duolingo represents a prominent example of gamification within contemporary language-learning applications. In comparison with more traditional language-learning tools, Duolingo does not simply provide vocabulary, grammar, listening, speaking, reading and writing exercises; instead, learning tasks are embedded within an ongoing system of feedback and incentives. As users complete lessons, they may simultaneously receive XP, progress information, streak updates, leaderboard positions and other forms of reward-related feedback. In Duolingo′s own description of the application in 2024, XP, leaderboards, streaks, in-app rewards and reminders were explicitly presented as features intended to support motivation and continued learning. This design transforms language learning from a process based solely on exposure to learning content into an interactive experience continuously organised through goals, rewards and feedback. Figure 1 provides an example of the main gamification mechanics discussed in this thesis, including hearts and streak count.

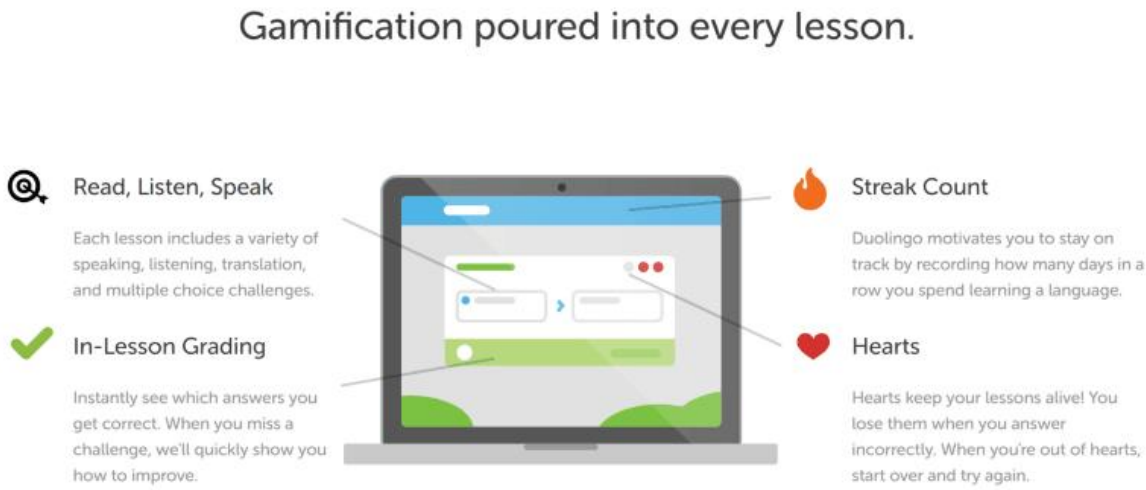

Figure 1 Gamification Elements in Duolingo

Source: Duolingo

For the purposes of this study, Duolingo's gamification mechanics can be broadly divided into four functional categories. The first category consists of habit-formation mechanisms, primarily including streaks, daily goals and reminders. These mechanisms are designed to encourage users to establish a regular learning routine. Streaks record the number of consecutive days on which a user has studied, thereby making continuity in learning visible. Daily goals divide a larger and potentially long-term learning process into smaller and more manageable daily objectives, while reminders provide external prompts encouraging users to return to the application. Duolingo itself explicitly describes streaks as a way of keeping learners on track and reminders as a means of helping users develop a study habit. Such mechanisms are particularly relevant to long-term language learning because the development of language ability generally depends on sustained and repeated practice rather than short periods of intensive activity alone.

The second category consists of reward and achievement mechanisms, primarily including XP, badges and other rewards. XP converts users' completion of learning activities into quantifiable scores, allowing learning activity to be immediately recorded and, in some contexts, compared with that of other users. Badges are used to recognise the achievement of particular milestones or accomplishments. These mechanisms can enhance users' sense of achievement by making learning effort and intermediate progress more visible. Duolingo's 2024 product description confirms that users earn XP through lessons and other activities and that XP can also be used within leaderboard competition. However, when users become excessively focused on XP and rewards, they may prioritise activities that generate points efficiently rather than those that require deeper engagement with language content. Research

on gamification misuse in Duolingo provides evidence that the game-like reward structure can, for some learners, become an objective in itself rather than merely a means of supporting learning.

The third category consists of feedback and progress mechanisms, primarily including progress indicators, immediate corrective feedback and the Hearts system. Progress indicators help users identify their current position within the course and make the overall learning path more visible. Immediate corrective feedback is provided after users answer questions, enabling them to identify whether a response is correct and, to some extent, adjust their subsequent performance. The Hearts system, which was in use during the main 2024–2025 period relevant to this study, connected errors with users' ability to continue completing lessons: mistakes could reduce the number of Hearts available to free users, whereas premium users could access unlimited Hearts. Duolingo's own 2024 materials still explicitly referred to unlimited Hearts as a subscription feature, while the company did not announce the experimental Energy replacement until July 2025. The Hearts system may have a dual effect. On the one hand, linking mistakes to a limited resource may encourage some users to approach exercises more carefully. On the other hand, it may cause some learners to experience mistakes as a form of penalty and consequently create pressure or frustration, particularly when errors limit their ability to continue practising. Significantly, when Duolingo announced its 2025 experiment with the Energy system, the company itself acknowledged that its previous Hearts system could be frustrating, especially for beginners, and that running out of Hearts during a lesson could be discouraging. For this reason, the Hearts system is especially relevant to the present study's interest in both the motivational benefits and potential limitations of individual gamification mechanisms.

The fourth category consists of social and competitive mechanisms, with leaderboards being the most prominent example. Leaderboards introduce competition and social comparison by ranking users according to XP earned over a given period. Duolingo's official description explains that users accumulate XP in weekly leaderboards and that higher-performing users can progress to subsequent leagues, while users who do not wish to participate can opt out of the competitive feature. For some users, leaderboards may increase the sense of challenge and motivate them to devote more time to learning tasks. For others, however, rankings may generate pressure and cause competition to become disproportionately important. Research on both Duolingo and gamification more broadly indicates that competition and leaderboards can sometimes redirect attention towards measurable performance and away from the substantive objectives of the learning activity. In the context of language learning, an excessive emphasis on competitive goals may therefore encourage users to maximise short-term, visible task completion rather than concentrate on longer-term language development.

Taken together, Duolingo’ s gamification design should not be understood as a simple collection of independent elements, but rather as an interconnected system of behavioural guidance. Different gamification mechanisms operate through habit formation, rewards, perceptions of progress, responses to errors and social comparison, and collectively shape users’ learning experiences and patterns of engagement. At the same time, an increase in frequency of use should not automatically be interpreted as an equivalent improvement in language learning. Existing research on Duolingo has explicitly raised the question of whether extrinsic motivation to study regularly necessarily supports deeper processing, authentic language use or communication, while research on gamification misuse demonstrates that engagement with game-related objectives can in some cases become detached from learning itself. It is therefore necessary to examine users’ evaluations of individual gamification mechanisms rather than treating gamification as a single, undifferentiated intervention.

2.3 Self-Determination Theory, Reinforcement Theory and Feedback Mechanisms

To explain why different gamification mechanisms may affect users in different ways, this study draws on Self-Determination Theory, reinforcement theory and research on feedback mechanisms as its theoretical foundations. These theories are not treated as independent objects of investigation. Rather, they provide analytical perspectives through which users′ positive experiences, perceptions of pressure and evaluations of learning effectiveness in relation to Duolingo′s different gamification mechanisms can be interpreted.

Self-Determination Theory proposes that sustained motivation is closely associated with the satisfaction of three basic psychological needs: autonomy, competence and relatedness (Deci and Ryan, 1985; Ryan and Deci, 2000). The theory distinguishes between conditions that support more self-determined forms of motivation and those that may undermine them, making it particularly relevant to the analysis of gamified learning environments. In language-learning applications, different gamification mechanisms may influence these psychological needs in different ways. For example, progress indicators, intermediate goals and immediate feedback can enable users to perceive that they are making progress and may therefore support a sense of competence. Leaderboards and social comparison can make other learners more visible and may influence users′ sense of relatedness through either connection or competition. Daily goals and streaks may also help users organise their learning behaviour by providing regular and visible objectives.

At the same time, these same mechanisms may be experienced as controlling rather than supportive. When users feel pressured to maintain a streak, achieve a particular

ranking or respond to repeated prompts from the application, their perceived autonomy may be reduced. This provides one possible theoretical explanation for why the same gamification mechanism may simultaneously motivate some users while creating pressure for others. The qualitative findings of Mogavi et al. (2022), which identify stress and excessive fixation on competitive aspects of Duolingo among some users, are consistent with the need to consider this distinction rather than assuming that more engagement necessarily represents more autonomous motivation. Self-Determination Theory therefore provides an appropriate framework for examining the coexistence of motivational and pressuring effects within the same gamification mechanism.

Reinforcement theory focuses on the ways in which behaviour can be shaped by its consequences, including rewards and reinforcement. In Skinner's behaviourist account, behaviour followed by reinforcing consequences becomes more likely to recur under relevant conditions (Skinner, 1953). Within Duolingo, XP, badges, streak continuation and other rewards can be interpreted as forms of positive reinforcement associated with learning-related behaviour. When users receive an immediate score, reward or continuation of a valued streak after completing a lesson, these consequences may increase the likelihood that they will return to the application and repeat the behaviour.

The Hearts system, by contrast, incorporates a restrictive consequence associated with errors. During the period relevant to this research, free users could lose access to Hearts through mistakes, while premium plans offered unlimited Hearts; Duolingo later acknowledged that the older system could frustrate learners when mistakes prevented them from continuing. From a reinforcement perspective, this mechanism may encourage users to place greater emphasis on accuracy, but it may also result in negative emotional responses to errors. Reinforcement theory therefore provides a useful analytical perspective for examining how rewards and restrictions shape repeated patterns of application use without assuming that repeated use necessarily corresponds to deeper learning.

Research on feedback mechanisms, meanwhile, emphasises the regulatory role of feedback within the learning process. Hattie and Timperley (2007) argue that effective feedback can help learners understand their current performance in relation to learning goals and identify how the gap between current and desired performance can be reduced. Their review also emphasises that feedback is not uniformly beneficial; its effects depend on the form of feedback and the level at which it operates. In gamified language-learning applications, feedback extends beyond conventional indications of whether an answer is correct or incorrect. It can also include changes in XP, movement along a progress path, the acquisition of badges,

changes in leaderboard positions and updates to streak records. These forms of visualised feedback make aspects of learning behaviour immediately visible to users. However, the direction of attention created by feedback is also important. When users become highly dependent on quantifiable indicators, they may begin to interpret scores, progress, streak length or leaderboard position as the primary evidence of successful learning, even though these indicators principally measure particular forms of activity within the application. Shortt et al. (2023) explicitly question whether gamification that encourages daily study necessarily promotes deeper information processing or authentic language use, while Mogavi et al. (2022) demonstrate how fixation on gamification can distract users from learning itself. Feedback mechanisms therefore provide not only information about progress but also a potential means through which users ' attention can be directed towards particular definitions of progress.

Taken together, Self-Determination Theory, reinforcement theory and research on feedback mechanisms provide complementary perspectives on the influence of gamification. Self-Determination Theory helps explain differences in users ' motivation and experiences of autonomy, competence and relatedness; reinforcement theory helps explain how rewards and restrictions may shape repeated behaviour; and feedback research helps explain how different forms of information about performance and progress can guide learners ' attention and subsequent actions. These perspectives therefore provide the theoretical foundation for analysing how individual gamification mechanisms may influence user perceptions, long-term engagement and perceived language-learning outcomes within Duolingo.

3 Research Design

This chapter presents the research design, data collection methods, participants and research procedure, and data analysis methods used in this study. Using Duolingo as the focal case, the study examines how users perceive different gamification mechanics and how these perceptions relate to sustained engagement, negative experiences such as pressure or fatigue, and perceived language-learning value.

To address the research questions, the study adopted a case-study design incorporating an online questionnaire, structured written follow-up interviews, and a design analysis of Duolingo's visible gamification mechanics. The user-data component followed a sequential structure in which quantitative survey data were collected first and qualitative written responses were subsequently used to provide further explanation. The design analysis constituted a separate source of contextual evidence and was used to support the interpretation of the user data.

3.1 Research Questions

The previous two chapters reviewed the application of gamification in language learning, the principal gamification mechanics used by Duolingo, and the theoretical foundations relevant to the study, including Self-Determination Theory, Reinforcement Theory, and research on feedback mechanisms. Existing research suggests that gamification can support continued engagement in language learning, but individual gamification mechanics do not necessarily influence user experience, sustained engagement, or perceived learning value in the same way.

Against this background, the present study uses Duolingo as a focal case to examine how users evaluate specific gamification mechanics and how different mechanics may be associated with both positive and negative experiences.

Based on these research objectives, the study addresses the following four research questions:

RQ1: How do users' overall perceptions and evaluations vary across different gamification mechanisms in Duolingo?

RQ2: Which gamification mechanisms do users perceive as most supportive of sustained engagement?

RQ3: Which gamification mechanisms are perceived as contributing to pressure, fatigue, or superficial engagement?

RQ4: How do users evaluate the extent to which different gamification mechanisms support their perceived language learning?

These four research questions correspond respectively to four analytical dimensions: overall user perceptions, sustained engagement, potential negative experiences, and perceived learning effectiveness. Together, they provide the main analytical framework for the subsequent presentation and discussion of the empirical findings.

3.2 Research Methodology

This section presents the overall methodological framework of the study and describes the design of its three main sources of evidence: the Duolingo design analysis, the online questionnaire, and the structured written follow-up interviews. A separate subsection subsequently addresses the pilot study in order to distinguish the development of the research instruments from the formal data collection procedure.

3.2.1 Overall Research Design

The study adopted a case-study design centred on Duolingo and incorporated a sequential explanatory component for the collection of user data. Specifically, an online questionnaire was first used to identify broad descriptive patterns in participants' evaluations of different gamification mechanics. Structured written follow-up interviews were then used to explore the reasons, experiences, and interpretations underlying these patterns.

The quantitative and qualitative components therefore served complementary purposes. The questionnaire provided standardised evidence about overall distributions and relative tendencies within the sample, while the open-ended written responses provided greater explanatory depth regarding why particular patterns occurred and why the same gamification mechanic could be experienced positively by some users and negatively by others. Combining quantitative and qualitative evidence in this way can support a more comprehensive interpretation of the research problem (Creswell and Creswell, 2018).

In addition to the user-data component, the study included a separate design analysis of Duolingo. The design analysis was not treated as part of the sequential quantitative-to-qualitative user-data process. Instead, it provided contextual evidence about the observable design of the application, including how visible gamification mechanics presented feedback, represented progress, and appeared to encourage particular forms of behaviour. Comparing this design-oriented evidence with the questionnaire and interview data made it possible to examine both the observable design logic of the application and the experiences reported by participants.

Duolingo was selected as the focal case because its language-learning environment incorporates a wide range of visible gamification mechanics, including streaks, XP, daily goals, leaderboards, progress feedback, error-management mechanisms, reminders, and other reward- and behaviour-oriented features. Chapter 2 discussed these mechanics in relation to habit formation, rewards and achievement, feedback

and progress, and social comparison and competition. The methodological design presented in this chapter explains how these mechanics were subsequently observed, evaluated, and interpreted within the empirical study.

It is important to emphasise that the study primarily examines users' reported perceptions and experiences. It did not independently measure language proficiency through standardised tests, pre- and post-tests, or other objective measures of learning gain. Similarly, it did not track actual long-term retention or application use through longitudinal behavioural data. Consequently, findings concerning sustained engagement and learning value are interpreted as participants' self-reported perceptions and experiences rather than as evidence of objective or causal effects produced by particular gamification mechanics.

3.2.2 Design Analysis

A design analysis was conducted to examine how Duolingo's gamification mechanics are organised and presented within the application. The analysis was informed by qualitative approaches to game analysis that emphasise the systematic examination of interfaces, interaction structures, and gameplay processes (Consalvo and Dutton, 2006). Whereas the questionnaire and interviews focused on users' subjective experiences, the design analysis focused on the observable structure of the application itself.

During the research period, Duolingo was used and observed regularly over several weeks. Particular attention was paid to visible gamification mechanics, including streaks, XP, daily goals, leaderboards, Hearts, badges, notifications, progress indicators, widgets, and streak freezes. Screenshots and research notes were collected to document how these mechanics appeared during routine use, including after completing a lesson, making an error, reaching a daily goal, or returning to the application after a period of inactivity.

Each mechanic was examined according to five analytical dimensions: its name and visible form, its observable design function, the type of feedback it provided, the behaviour its design appeared to encourage, and its potential side effects or negative user experiences. Using the same analytical dimensions across mechanics made it possible to compare different forms of gamification within a common framework.

For example, streaks were considered in relation to continuity, habit formation, and repeated return behaviour; XP was examined in relation to rewards and the visualisation of activity; leaderboards were considered in relation to social comparison and competition; and Hearts were examined in relation to error management, restriction, and penalty-like feedback. These dimensions were also used to connect the observable design features of Duolingo with the theoretical perspectives introduced in Chapter 2.

The design analysis does not claim to identify Duolingo's internal design intentions. It is based on user-facing interfaces and observable interactions rather than access to the company's internal design process. The analysis therefore concerns the application's observable design logic, rather than unverified assumptions about internal product decisions. The resulting design analysis provides contextual evidence that can later be compared with participants' evaluations and experiences in the questionnaire and structured written follow-up interviews.

3.2.3 Questionnaire

The online questionnaire was the primary quantitative user-data collection instrument in the study. Its purpose was to collect standardised information about Duolingo users' experiences and evaluations of different gamification mechanics. Online surveys provide an appropriate means of collecting comparable information concerning participants' attitudes, behavioural tendencies, and experiences across a relatively broad group of respondents (Dillman, Smyth and Christian, 2014).

The complete questionnaire is provided in Appendix A. It first collected background information about participants' use of Duolingo, including duration of use, frequency of use, and subscription status.

The main rating section asked participants to evaluate how helpful different gamification mechanics were for keeping them engaged with Duolingo. The mechanics included streaks, XP rewards, daily goals, leaderboards, skill tree or progress indicators, the Heart system, notifications, widgets or lock-screen displays, and streak freeze. Participants evaluated these mechanics using a five-point Likert-type response scale ranging from "not helpful at all" to "extremely helpful." Likert-type items allowed participants' evaluations of different mechanics to be compared descriptively (Sullivan and Artino, 2013).

In addition to the helpfulness ratings, the questionnaire included an item concerning streak-maintenance behaviour. Participants were asked whether they had ever opened Duolingo mainly to avoid losing their streak, with the response options Yes, No, and Not sure. This item was intended to examine whether maintaining an accumulated streak was a self-reported reason for returning to the application.

The questionnaire also included a ranking task in which participants were required to compare selected gamification mechanics directly and rank them according to how helpful they considered them for maintaining engagement. The ranking options included streak, leaderboard, daily goal, XP and progress, notifications, and widgets or lock-screen displays. Unlike the individual rating items, this task required participants to make relative comparisons across several mechanics and therefore provided a complementary indication of perceived importance.

Finally, the questionnaire included a multiple-response item concerning negative experiences. Participants were asked whether particular gamification mechanics made them feel pressured or uncomfortable. This part of the questionnaire reflected the study's intention not to assume that gamification is uniformly beneficial, but instead to examine both its perceived engagement-supporting functions and its possible negative effects. This item was particularly relevant to RQ3.

The questionnaire primarily measured participants' engagement-related evaluations, behavioural tendencies, and experiences of pressure or discomfort. It did not objectively measure language-learning outcomes. Perceived learning value, particularly the distinction between supporting learning and merely encouraging continued application use, was explored more directly in the structured written follow-up interviews.

3.2.4 Structured Written Follow-up Interviews

Following the quantitative questionnaire, structured written follow-up interviews were used to explore participants' experiences of Duolingo's gamification mechanics in greater depth. Qualitative interviews are particularly useful for examining participants' views, experiences, beliefs, and motivations and can provide forms of explanation that cannot always be captured through closed-form questionnaire items (Gill et al., 2008).

The formal written follow-up interview consisted of eight fixed open-ended questions. All participants received the same questions in the same order, and no participant-specific probing questions were introduced. The method is therefore described as a structured written follow-up interview, rather than a semi-structured interview.

The interviews were completed online in an asynchronous written format. Written asynchronous interviewing can provide participants with greater opportunity to reflect on and formulate their responses. At the same time, it does not provide the same opportunities for immediate probing or for interpreting non-verbal cues as face-to-face or synchronous interviewing (Meho, 2006). The responses were therefore treated as qualitative written interview data.

The eight questions addressed participants' duration and frequency of Duolingo use; their reasons for initially using the application; the mechanics that most encouraged continued use; their experiences of the streak feature; mechanics associated with pressure, discomfort, or reduced interest; their perceptions of leaderboards; whether gamification mechanics supported language learning or primarily encouraged users to log in and continue using the application; and their overall evaluation of Duolingo's ability to support long-term engagement in language learning.

This design maintained a basic level of comparability across participants while still allowing them to explain their experiences in their own words. The qualitative

responses were particularly useful for interpreting apparently contradictory questionnaire patterns. For example, a mechanic could be evaluated relatively positively in relation to engagement while also being associated with pressure or discomfort. The written responses provided additional context for understanding how these apparently conflicting experiences could coexist.

The purpose of the follow-up interviews was not to generate a second statistically representative dataset. Instead, they were intended to provide explanatory depth and contextual interpretation for the descriptive patterns identified in the questionnaire. The complete set of eight questions is provided in Appendix B – Structured Written Follow-up Interview Questions.

3.2.5 Pilot Study

Before the formal questionnaire was distributed, a small pilot study was conducted to assess the clarity, structure, and suitability of the survey instrument. The pilot involved two long-term Duolingo users, both of whom had used the application for more than six months and had maintained streaks of more than 100 days. Their experience with Duolingo made them useful pilot participants because they were familiar with a range of gamification mechanics and could comment on whether the questionnaire adequately reflected features relevant to continued use.

The pilot indicated that the questionnaire items were generally understandable and capable of eliciting differentiated responses. The two participants did not evaluate all gamification mechanics in the same way. Both regarded the streak as highly important, whereas XP and badges were evaluated much less positively. Their views of the leaderboard also differed: one participant regarded competition as motivating, while the other experienced it as a source of pressure. These contrasting responses suggested that the questionnaire was capable of capturing differences in how individual users experienced the same gamification mechanic.

More importantly, the pilot identified several aspects of the questionnaire that required revision before formal data collection. First, the initial version did not include all of the mechanics that participants considered relevant to their continued engagement. One participant, for example, referred to reminders as an important cue for returning to the application. The pilot therefore prompted a broader consideration of externally visible engagement mechanisms, including notifications, widgets or lock-screen displays, progress-related feedback, the Heart system, and streak freeze. These mechanics were incorporated into the revised questionnaire where relevant to the study's research questions.

Second, the pilot suggested that asking participants only to rate individual mechanics in terms of helpfulness might produce insufficient differentiation between features. Because many gamification mechanics are explicitly designed to encourage engagement, participants could potentially rate several of them positively without

clearly indicating which mechanics they regarded as relatively more important. To address this limitation, the revised questionnaire included a ranking task in addition to the individual helpfulness ratings. This required participants to compare selected mechanics directly and indicate their relative importance for maintaining engagement.

The pilot also highlighted the importance of collecting information about negative as well as positive experiences. Both pilot participants described some gamification features as simultaneously motivating and pressuring. The revised questionnaire therefore included a multiple-response item asking whether particular mechanics were associated with pressure or discomfort. This allowed the formal questionnaire to capture the possibility that a mechanic could support continued use while also creating negative emotional experiences.

Finally, the pilot prompted further refinement of the questions concerning participants' duration of Duolingo use. Because the study focuses on sustained engagement, the final questionnaire used separate duration categories to distinguish participants with different levels of experience with the application. This made it possible to examine descriptive differences between relatively new and longer-term users in the subsequent analysis.

Overall, the pilot study was used primarily as an instrument-development stage rather than as part of the main empirical analysis. Its purpose was to assess whether the questionnaire was sufficiently clear, comprehensive, and capable of distinguishing between users' perceptions of different gamification mechanics. The revisions made after the pilot broadened the range of mechanics considered, added direct relative comparison through ranking, incorporated negative-experience items, and refined the categorisation of user experience duration.

3.3 Ethical Considerations

This study involved human participants through an online questionnaire and structured written follow-up interviews. Participation was voluntary. Before taking part, participants were provided with information about the general purpose of the study, what participation would involve, how the information they provided would be used for research purposes, and their right to discontinue participation without having to provide a reason or experience negative consequences. These procedures were intended to support informed and voluntary consent and to respect participants' autonomy throughout the research process (British Educational Research Association, 2024; British Psychological Society, 2021).

Because the user data were collected through online methods, particular attention was given to privacy, confidentiality, and participant identifiability. The

questionnaire did not require participants to provide their names or contact details as part of the main survey responses. However, at the end of the questionnaire, respondents who were willing to participate in the follow-up stage could voluntarily provide an email address for further contact. These contact details were used only for follow-up recruitment and were not treated as variables in the analysis or reported in the thesis. No highly sensitive personal information was intentionally collected. As online research may involve additional issues relating to data transmission, confidentiality, and platform-level data processing, the study does not make stronger claims about technical anonymity than can be supported by the data collection procedure itself (Barchard and Williams, 2008; Buchanan and Hvizdak, 2009).

For the structured written follow-up interviews, identifying information was removed or excluded from the material used for analysis and reporting. Personal details that could reasonably permit the identification of individual participants were not included in the thesis. This approach was intended to protect participants' privacy and confidentiality and to reduce potential risks of re-identification in internet-mediated research (British Psychological Society, 2021; Franzke et al., 2020).

Participants were informed that they could stop participating before completing the questionnaire or interview without providing a reason. In relation to the questionnaire, this principle refers to the ability to discontinue participation before submission. The study does not claim that already-submitted responses could always be withdrawn after submission, because this would only be possible where a particular response could subsequently be identified without undermining the privacy protections applied to the dataset. This distinction reflects the need to balance withdrawal procedures with confidentiality and anonymity considerations in online research (Barchard and Williams, 2008; British Educational Research Association, 2024).

The study focused on participants' ordinary experiences and perceptions of using Duolingo. It did not intentionally target vulnerable populations or ask participants to disclose highly sensitive personal experiences. The research was therefore designed as a low-risk study. Nevertheless, low expected risk does not remove the responsibility to minimise potential harm and to respect participants' autonomy, dignity, privacy, and confidentiality throughout the research process (British Educational Research Association, 2024).

The collected data were accessed and used solely for the purposes of completing this thesis. Questionnaire findings are reported primarily in aggregate form, while qualitative material from the structured written follow-up interviews is presented only after de-identification. Individual responses are not intentionally reported in a manner that would make participants identifiable. These procedures are consistent with broader principles of data minimisation, confidentiality, and privacy protection

in educational and internet-mediated research (British Educational Research Association, 2024; British Psychological Society, 2021; Franzke et al., 2020).

3.4 Research Participants and Procedure

This section describes the recruitment and composition of the study participants and the procedure followed during the formal data collection. The design of the individual research instruments has already been described in Section 3.2 and is therefore not repeated here.

3.4.1 Participants and Recruitment

The study used a purposive, non-probability sampling strategy to recruit participants with experience of using Duolingo. Recruitment focused on individuals whose previous use of the application made their experiences relevant to the research questions, including university students and members of the wider public from different occupational backgrounds. Participants were therefore recruited on the basis of characteristics relevant to the purpose of the study rather than through random probability sampling.

Purposive sampling provides a practical means of recruiting participants with experience relevant to a particular research question. However, findings obtained through non-probability sampling should not automatically be interpreted as statistically representative of the wider population (Andrade, 2021). Accordingly, the quantitative findings in this study are interpreted as descriptive patterns within the participating sample rather than as population-level estimates for all Duolingo users.

A total of 155 questionnaire responses were retained for the quantitative analysis. The respondents had varying levels of experience with Duolingo, allowing the study to examine perceptions across users with different durations and frequencies of application use. Detailed characteristics of the questionnaire sample are presented in Section 4.2.1 rather than repeated in the methodology chapter.

At the end of the questionnaire, respondents were asked whether they would be willing to participate in a follow-up stage. Those who expressed interest could voluntarily provide an email address for further contact. From this pool of volunteers, seven participants subsequently completed the structured written follow-up interviews. The contact information was used only for arranging and distributing the follow-up questions and was not used as part of the analytical dataset.

The purpose of the smaller qualitative sample was not to provide statistical generalisation, but to offer more detailed explanations of individual experiences and to help interpret patterns identified in the questionnaire.

3.4.2 Research Procedure

Before the formal questionnaire was distributed, the questionnaire was pilot-tested and revised as described in Section 3.2.5. The pilot stage was used to assess the clarity and structure of the questionnaire and to identify areas requiring modification before the main data collection.

The formal questionnaire was subsequently administered online through Wenjuanxing. Participants were provided with information about the research and the voluntary nature of participation before completing the questionnaire. The questionnaire collected background information concerning participants' use of Duolingo and their evaluations of the gamification mechanics described in Section 3.2.3. Following the completion of data collection, responses were retained as the formal quantitative dataset.

During the research period, a separate design-analysis strand was also conducted. Duolingo was used and observed regularly over several weeks, and screenshots and research notes were collected to document visible gamification mechanics and relevant interaction processes. As described in Section 3.2.2, this design analysis did not involve human-participant data and served as an independent source of contextual evidence about the observable design of the application.

Following the questionnaire stage, respondents who had voluntarily indicated their willingness to participate in the follow-up stage were contacted by email. Seven participants subsequently completed the structured written follow-up interviews online. All participants received the same eight open-ended questions in the same order, without participant-specific probing. The interviews were conducted asynchronously in written form and were intended to provide more detailed explanations of participants' experiences with the gamification mechanics and of patterns identified in the questionnaire.

The formal research procedure therefore generated three complementary sources of evidence: questionnaire data concerning participants' evaluations of Duolingo's gamification mechanics, qualitative written responses providing explanatory detail about these evaluations, and design-analysis material documenting the application's observable gamification structure. The methods used to analyse and integrate these three sources of evidence are described in Section 3.5.

3.5 Data Analysis

The analysis consisted of three complementary strands: analysis of the Duolingo design material, descriptive analysis of the questionnaire data, and qualitative analysis of the structured written follow-up interviews. These three sources of evidence were subsequently

integrated and interpreted in relation to the research questions and the theoretical framework presented in Chapter 2.

3.5.1 Design Analysis

The screenshots and research notes collected during the observation of Duolingo were analysed using the analytical framework described in Section 3.2.2. Each gamification mechanic was examined in relation to its observable design function, the type of feedback it provided, the behaviour its visible design appeared to encourage, and its potential side effects or negative user experiences.

The purpose of this analysis was to identify recurring design patterns across mechanics rather than to infer Duolingo's internal product intentions. For example, streaks were examined in relation to continuity and repeated return behaviour, XP in relation to rewards and quantification of activity, leaderboards in relation to competition and social comparison, and Hearts in relation to error management and restrictive feedback.

The resulting design analysis provided contextual evidence for interpreting the user-data findings. It was therefore used primarily to compare the observable design characteristics of individual mechanics with participants' reported evaluations and experiences.

3.5.2 Questionnaire Analysis

The questionnaire data were analysed using descriptive statistics. Frequencies and percentages were used to summarise categorical responses and the distributions of participants' evaluations.

For the Likert-type helpfulness items, response distributions were treated as the primary descriptive evidence. Mean scores were also calculated as concise comparative summaries of how the different gamification mechanics were evaluated. These means were interpreted cautiously because individual Likert-type responses are ordinal, and differences between mean values should not automatically be treated as precise interval-level differences (Sullivan and Artino, 2013).

The ranking responses were analysed to identify which selected gamification mechanics were most frequently prioritised by participants as helpful for maintaining engagement. The ranking results were treated as complementary to the individual helpfulness ratings because the two question formats required different types of judgement: the rating items evaluated mechanics individually, whereas the ranking task required participants to compare selected mechanics directly.

Responses to the item concerning streak-maintenance behaviour were summarised using frequencies and percentages. Similarly, the multiple-response item concerning pressure or discomfort was analysed using response frequencies and percentages in

order to identify mechanics that were more frequently associated with negative experiences.

Where comparisons were made between participants with different durations of Duolingo use, these comparisons remained descriptive and exploratory. The duration categories used in the questionnaire were used to examine patterns within the sample, but no inferential statistical tests were conducted to determine whether observed differences between user groups were statistically significant. The findings were therefore not interpreted as evidence of causal or population-level differences.

3.5.3 Analysis of the Structured Written Follow-up Interviews

The structured written follow-up interviews were analysed qualitatively using a question-led thematic approach. The written responses were read repeatedly and coded for recurring patterns relevant to the research questions.

Initial coding was informed by the topics addressed in the structured interview questions, including habit formation, streak-related pressure, minimum-effort participation, competition, leaderboard fatigue, frustration associated with the Heart system, notifications, and perceived learning value. At the same time, responses that introduced experiences or explanations not anticipated by the questionnaire or interview structure were retained where they were relevant to the research questions.

The coded material was subsequently grouped into broader recurring patterns that could be used to compare similarities and differences across participants' experiences. This approach drew on thematic analysis as a flexible method for identifying and interpreting patterns within qualitative data (Braun and Clarke, 2006).

Because the participants provided their written responses in Chinese, coding was conducted on the original Chinese responses in order to minimise unnecessary loss of meaning during the analytical process. Themes were developed from the original-language material, after which selected quotations used in the thesis were translated into English. The translations were kept as close as possible to the participants' original meaning.

3.5.4 Integration and Theoretical Interpretation

The final stage of the analysis integrated the three sources of evidence. Patterns identified in the questionnaire were compared with participants' explanations in the structured written follow-up interviews and with observations from the design analysis.

For example, where a gamification mechanic received relatively positive ratings for supporting engagement but was also associated with pressure or discomfort, the

qualitative responses were used to examine how participants understood and experienced this apparent tension. The design analysis was then used to provide contextual evidence about how the mechanic operated within the visible interface. Integrating quantitative patterns, qualitative explanations, and contextual design evidence in this way enabled the analysis to move beyond identifying whether individual mechanics were evaluated positively or negatively and towards explaining why these evaluations occurred (Fetters, Curry and Creswell, 2013).

The integrated findings were subsequently interpreted using the theoretical perspectives established in Chapter 2: Self-Determination Theory, Reinforcement Theory, and research on feedback mechanisms. These perspectives were used as analytical lenses rather than as variables directly tested in the study.

Self-Determination Theory was used to interpret experiences relating to autonomy, competence, and social comparison; Reinforcement Theory was used to consider repeated behaviour in relation to rewards and restrictions; and research on feedback mechanisms was used to interpret how visible progress, corrective information, and performance-related feedback influenced users' perceptions of engagement and learning. The theoretical interpretation was applied across the four research questions rather than treated as a separate research question.

Table 1 summarises the relationship between the four research questions, the principal data sources, and their analytical purposes. The design analysis provides contextual evidence across all four research questions, while the theoretical perspectives are applied during the integrated interpretation of the findings.

Table 1. Alignment between the research questions, data sources and analytical purposes

Research question	Primary data source(s)	Analytical purpose
RQ1: How do users' overall perceptions and evaluations vary across different gamification mechanisms in Duolingo?	Online questionnaire; structured written follow-up interviews	To compare overall evaluations of different mechanics and explain major differences in user perceptions
RQ2: Which gamification mechanics do users perceive as most supportive of sustained engagement?	Questionnaire ratings and ranking questions; structured written follow-up interviews	To identify mechanics most strongly associated with continued use and explore why users perceive them as motivating
RQ3: Which gamification mechanisms are perceived as contributing to pressure, fatigue, or superficial engagement?	Questionnaire items on pressure and discomfort; structured written follow-up interviews	To identify negative or ambivalent experiences and examine how particular mechanics may encourage pressure or minimum-effort participation
RQ4: How do users evaluate the extent to which different gamification mechanisms support their perceived language learning?	Structured written follow-up interviews	To examine perceived learning value and distinguish mechanics that support learning from those that mainly encourage continued platform use

4 Results and Analysis

This chapter presents the results and analysis of Duolingo's gamification mechanics and users' perceptions of them. It integrates three sources of evidence: a design analysis of Duolingo's visible gamification mechanics, an online survey of 155 users, and seven structured written follow-up interviews. The analysis examines how Duolingo's mechanics appear to encourage repeated engagement, how users evaluate these mechanics, and how users describe their potential influence on continued participation and perceived learning value.

The chapter distinguishes between continued app engagement and meaningful learning engagement. This distinction is important because repeated app use does not necessarily indicate equivalent depth or quality of learning. Previous research has identified cases in which users become increasingly focused on gamification objectives rather than learning objectives, including the pursuit of points, rankings, and other rewards (Mogavi et al., 2022). More broadly, research on gamified educational software has identified potential negative outcomes associated with points, badges, leaderboards and competition, including motivational difficulties, reduced performance, limited understanding and gaming of the system (Almeida et al., 2023).

Sections 4.1 – 4.3 present the empirical findings, while Sections 4.4 – 4.5 integrate these findings and interpret them in relation to the theoretical framework and previous research.

4.1 Design Analysis of Duolingo Gamification Mechanics

The design analysis identified a range of visible gamification mechanics used in Duolingo during the 2024–2025 study period, including streaks, XP, daily goals, leaderboards, Hearts, badges, notifications, progress indicators, widgets, and streak freeze. Table 2 summarises the principal functions, forms of feedback, behaviours encouraged, and potential limitations associated with these mechanics.

Table 2 Analysis of Duolingo's Visible Gamification Mechanics

Mechanic	Function	Feedback	Behaviour	Potential Limitation
Streak	Encourage continuity and regular practice	Visual and numerical	Daily return and routine formation	Pressure or anxiety associated with losing the streak; minimum-effort completion
XP system	Make activity measurable and reward task completion	Quantitative	Complete lessons and other activities	Attention may shift from learning quality to point accumulation
Daily goal	Divide participation into manageable short-term targets	Goal/progress feedback	Meet a recurring daily target	Goal fatigue or completion-focused behaviour

Leaderboard	Introduce competition and social comparison	Rank position	Earn XP and compare performance with others	Stress, discouragement, or excessive competition
Heart system	Attach a consequence to incorrect responses	Negative/limiting feedback	Answer more carefully and monitor mistakes	Frustration, pressure, or interruption of a learning session
Streak freeze	Protect accumulated continuity after a missed day	Protective feedback	Return after interruption without losing the streak	May reduce the perceived cost of missing a day
Reminders	Provide an external cue to return	Push notification/visual cue	Return to the app regularly	Notification fatigue or perceived intrusiveness
Badges	Mark milestones and achievements	Visual achievement feedback	Complete challenges or reach milestones	Limited motivational value when achievements are not personally meaningful
Progress indicators	Visualise advancement through learning content	Spatial and progress feedback	Continue along the learning path	Possible overemphasis on completion rather than mastery
Widgets	Keep learning status visible outside the app	Persistent visual reminder	Return to the app and maintain a streak	Pressure, annoyance, or reminder fatigue

Source: Author's analysis based on observed Duolingo interfaces during the 2024–2025 study period and Duolingo product documentation.

This table shows that Duolingo's gamification system combines multiple forms of behavioural and informational feedback. Streaks and daily goals emphasise continuity and routine; XP and badges make activity and achievement visible; leaderboards introduce competition; progress indicators represent advancement; while hearts and streak loss create potentially aversive consequences. These mechanics are also interconnected. A single lesson may simultaneously extend a streak, generate XP, contribute towards a daily goal, alter leaderboard position and move the learner forward on the learning path.

This layered structure is important because the motivational effect of a mechanic cannot necessarily be understood in isolation. Previous gamification research suggests that different game elements may influence motivational, behavioural and cognitive outcomes differently rather than producing a single uniform "gamification effect" (Sailer and Homner, 2020).

4.1.1 Streak

The streak is one of the most visible gamification mechanics in Duolingo. It records the number of consecutive days on which a user has completed learning activity and converts continuity into a visible numerical record. Duolingo itself describes the streak as a habit-oriented feature intended to encourage regular practice and

acknowledges that longer streaks may increasingly be protected because users become reluctant to lose accumulated progress (Duolingo, 2022).

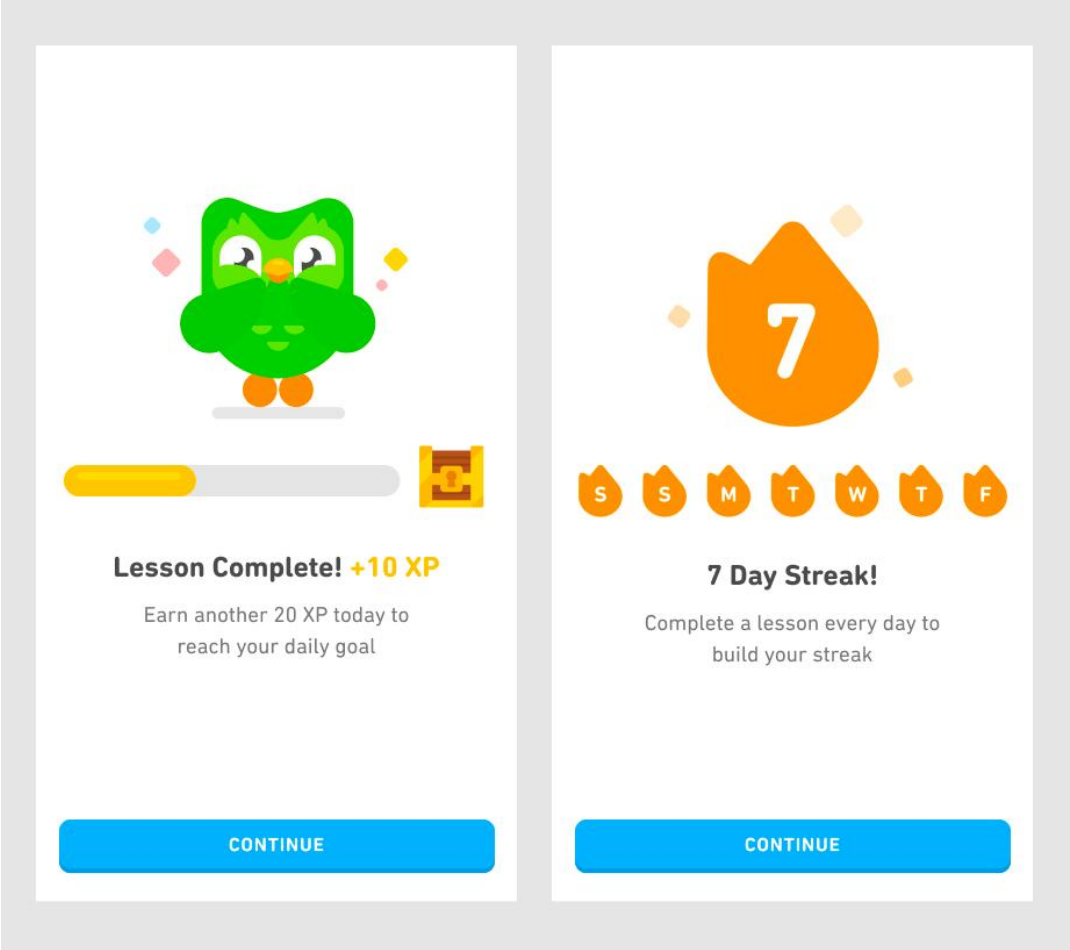

Figure 2 Duolingo's XP & Streak Interface

Source: Author's screenshot from Duolingo app, collected during the 2024–2025 study period.

From the user's perspective, the streak can provide a continuing sense of achievement. Extending the streak gives immediate confirmation that the day's activity has preserved a longer pattern of participation. At the same time, the mechanism may create pressure because failing to complete an activity can threaten an accumulated record. Duolingo's own discussion of the feature recognises both sides of this mechanism: the company links streaks to regular practice but also acknowledges that losing a streak may be demotivating and that concern about losing it can influence behaviour (Duolingo, 2022).

The streak therefore has a potentially dual role. It may support routine formation by giving users a reason to return even on days when their motivation is relatively low, but it may also redirect attention from the learning activity to preservation of the record itself. This distinction is particularly relevant to the present study because previous qualitative research on Duolingo has identified forms of "gamification

misuse" in which users become preoccupied with game-related objectives rather than learning objectives (Mogavi et al., 2022).

4.1.2 XP, Daily Goals, and Progress System

The XP system quantifies activity by awarding points for lessons and other tasks. XP is connected with other parts of Duolingo's gamification structure, particularly leaderboards and some goal-oriented activities. Progress indicators and the learning path provide a different form of feedback by showing where users are positioned within the course. Duolingo's product documentation explicitly describes XP as a basis for leaderboard competition and its course progression tools as mechanisms for visualising advancement (Duolingo, 2024a).

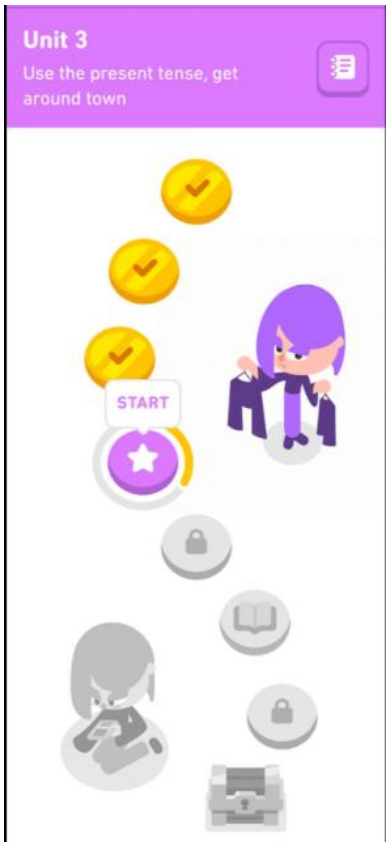

Figure 3 Duolingo progress bar and learning path Interface

Source: Author's screenshot from Duolingo app, collected during the 2024–2025 study period.

These mechanics may contribute to perceived competence by making activity and progress visible. Users can observe immediate outcomes of their actions through points, completed activities and movement through the course. From a Self-Determination Theory perspective, competence can be supported when individuals receive information that helps them experience themselves as effective (Ryan and Deci, 2000). However, game elements such as points and leaderboards should not

automatically be treated as evidence of increased intrinsic motivation; experimental research has shown that such elements can increase performance or function as progress indicators without necessarily increasing autonomy, competence satisfaction or intrinsic motivation in every context (Mekler et al., 2017).

A further limitation is that measurable activity and learning quality are not identical. Duolingo's own 2024 discussion of its Time Spent Learning Well metric acknowledged that counting sessions could favour users repeatedly completing shorter and easier activities and that XP-focused systems could create opportunities for "grinding" rather than progression through more demanding learning content (Duolingo, 2024b). This company-authored evidence is particularly relevant to the distinction made in the present study between visible engagement and meaningful learning engagement.

4.1.3 Leaderboard

The leaderboard introduces explicit social comparison and competition. Users accumulate XP over a weekly period and are ranked against other users, with higher-performing users able to move between leagues. Duolingo describes XP accumulation and leaderboard competition as connected elements of its gamification system (Duolingo, 2024a).

For users who enjoy competition, the leaderboard may provide an additional reason to practise and an external goal beyond individual course progression. However, the same comparative structure may produce pressure or discouragement among users who prefer self-paced learning or who perceive other users' XP totals as difficult to match.

This variability is consistent with the broader literature on educational gamification. Almeida et al. (2023), for example, identified leaderboards and competition among the game design elements frequently associated with reported negative effects in educational software. At the same time, this does not mean that leaderboards are inherently harmful; rather, their effects depend on implementation, context and user characteristics.

Duolingo's own product analysis similarly acknowledges that leaderboard competition can produce XP grinding and may appear unfair to users who prioritise learning content rather than accumulating large quantities of XP (Duolingo, 2024b). This makes the leaderboard a particularly useful mechanic for examining the tension between engagement quantity and learning-oriented participation.

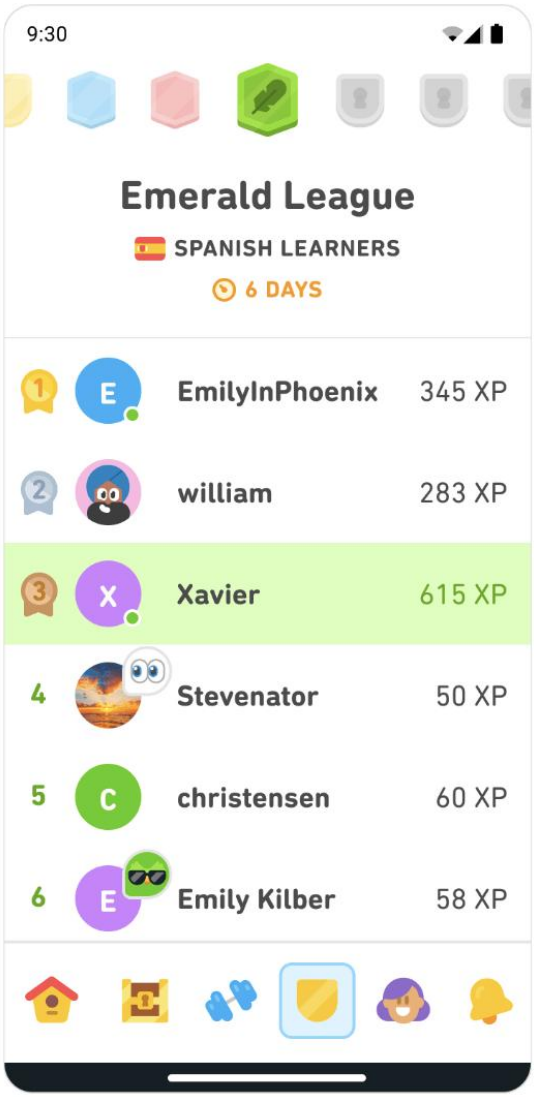

Figure 4 Duolingo Leaderboard and League Tiers

Source: Author's screenshot from Duolingo app, collected during the 2024–2025 study period.

4.1.4 Hearts, Streak Freeze, Reminders, and Badges

Several other mechanics support or constrain engagement in different ways. During the main data-collection period, the heart system linked mistakes to continued access to lessons for free users: incorrect responses reduced the number of available hearts, and users who exhausted them could be required to practise, wait, or otherwise replenish them before continuing. Duolingo's December 2024 product guide explicitly described the system in these terms, while subscribers had access to unlimited hearts (Duolingo, 2024a).

The heart system therefore differs from reward-oriented mechanics such as XP or badges. It may encourage more careful responding, but it may also make mistakes feel costly. This distinction is important in a learning context, where errors can provide useful information for subsequent learning rather than simply representing failure.

Streak freeze performs an almost opposite function. It protects an existing streak when a user misses a day. Duolingo has explicitly described this feature as a way to introduce flexibility into the streak system and reported that allowing greater streak flexibility was associated with continued participation in its internal product data (Duolingo, 2022).

Reminders and widgets provide external cues to return. Duolingo describes its widget as a persistent visual reminder intended to encourage users to open the application and complete a lesson, particularly when a streak is at risk (Duolingo, 2023). Badges, meanwhile, mark achievements and milestones and are explicitly presented by Duolingo as a way of recognising continued activity and accomplishments (Duolingo, 2023b).

Together, these mechanics demonstrate that Duolingo does not rely on a single gamification strategy. Instead, the visible system combines rewards, progress information, habit cues, competition, protective mechanisms and penalty-like consequences. The following survey and written follow-up findings examine how users perceived these different forms of gamification.

4.2 Survey Findings

The online survey collected 155 valid responses from Duolingo users. The purpose of the survey was to examine users' perceptions of different gamification mechanics, including their perceived helpfulness for continued engagement and their potential association with pressure or discomfort.

Because the survey relied on self-reported perceptions and descriptive statistics, the findings should not be interpreted as causal evidence that a particular mechanic objectively increases long-term retention or language proficiency. Instead, they show how participants evaluated the mechanics and how these evaluations varied across the sample.

4.2.1 Sample Overview

The respondents represented different levels of experience with Duolingo, although the sample was weighted towards relatively new users. Among the 155 respondents, 78 participants (50.32%) had used Duolingo for less than one month. A further 37 participants (23.87%) had used it for one to three months, while 10 participants (6.45%) had used it for three to six months. The remaining 30 participants (19.35%) had used Duolingo for more than six months.

Overall, 125 respondents (80.65%) had used Duolingo for six months or less. This distribution is important when interpreting questions related to long-term engagement. The full sample is therefore better understood as providing evidence about perceived support for continued engagement, rather than direct evidence of demonstrated long-term retention.

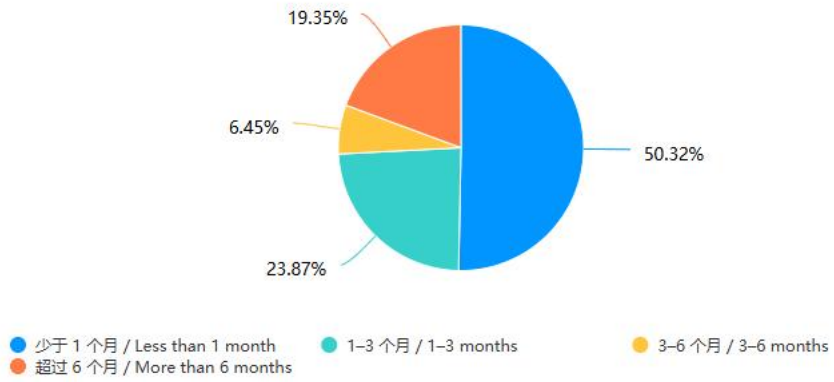

Figure 5 Duration of Duolingo Use Among Participants

Source: Survey data collected by the author.

The frequency data show that most respondents used Duolingo regularly, although daily use was less common. A total of 23 respondents (14.84%) reported daily use, while 87 respondents (56.13%) used Duolingo a few times a week. The remaining 45 respondents (29.03%) reported using the app less than once a week. Thus, more than half of the sample used Duolingo several times per week, whereas daily use characterised a smaller proportion of participants.

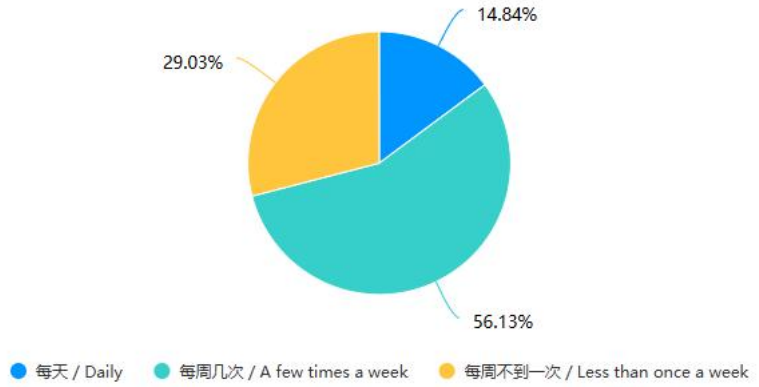

Figure 6 Frequency of Duolingo Use Among Participants

Source: Survey data collected by the author.

Most respondents used the free version of Duolingo. Specifically, 137 respondents (88.39%) reported that they were not paying users, whereas 18 respondents (11.61%) reported using a

paid plan. The survey therefore predominantly reflects the experience of users of Duolingo's free version.

Payment status is especially relevant when interpreting responses to the heart system. During the period represented by the study, free users could lose hearts after incorrect answers, whereas subscribers had access to unlimited hearts (Duolingo, 2024a). Consequently, participants were not necessarily exposed to the heart mechanic in exactly the same way.

4.2.2 Streak-Driven Behaviour

One survey question asked participants whether they had ever opened Duolingo mainly to avoid losing their streak. Of the 155 respondents, 76 (49.03%) answered "yes", 35 (22.58%) answered "no", and 44 (28.39%) selected "not sure".

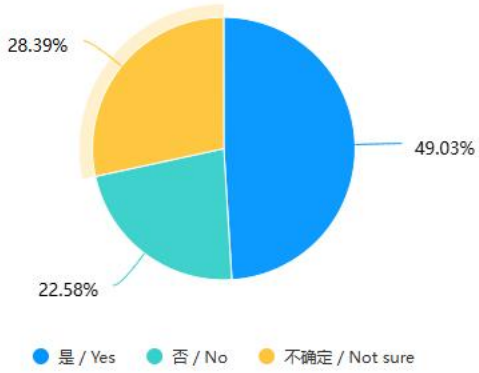

Figure 7 Logging in to Maintain a Streak Among Participants

Source: Survey data collected by the author.

The results indicate that streak-maintenance behaviour was common within the sample. Almost half of respondents explicitly reported having opened Duolingo primarily to avoid losing their streak. This does not demonstrate that the streak caused long-term retention, but it does indicate that preserving the streak was an important self-reported reason for returning to the app.

This finding is consistent with the visible design logic of the streak. Duolingo itself describes the mechanism as a way to encourage consistent practice and has acknowledged that the prospect of losing an accumulated streak can become motivational as the record grows (Duolingo, 2022).

The 44 "not sure" responses also deserve attention. Users may not always separate streak preservation from language-learning intentions. Completing a daily lesson can simultaneously represent genuine learning activity, fulfil a daily routine and protect a

streak. The distinction between "learning because I want to practise" and "learning because I do not want to lose the streak" may therefore be blurred in everyday use.

This ambiguity is central to the study. The streak may support consistency while also making maintenance of the gamified record an objective in its own right.

4.2.3 Feature Helpfulness Ratings

Participants were asked to rate how helpful different gamification mechanics were for keeping them engaged with Duolingo. A five-point response scale was used, ranging from 1 ("not helpful at all") to 5 ("extremely helpful").

The mean scores were relatively close, ranging from 2.76 to 3.03.

Figure 8 Mean Helpfulness Ratings of Gamification Mechanics

Source: Survey data collected by the author.

The streak received the highest mean rating (M = 3.03), followed by streak freeze (M = 2.97), daily goal (M = 2.93), leaderboard (M = 2.90), XP rewards (M = 2.89), widgets or lock-screen display (M = 2.87), notifications (M = 2.85), learning path and progress indicators (M = 2.84), and the heart system (M = 2.76).

These results should be interpreted descriptively. The differences between the mean ratings are small, and no inferential statistical test was conducted to establish whether the differences were statistically significant. It would therefore be inappropriate to claim that the streak was statistically more effective than the other mechanics.

Nevertheless, the descriptive pattern provides useful information about participants' perceptions. The three highest mean ratings were associated with continuity-oriented mechanics: streak, streak freeze and daily goal. The heart system, by contrast, received the lowest mean rating.

The response distributions provide additional context. A total of 47 respondents (30.32%) rated the streak as either "very helpful" or "extremely helpful". Streak freeze received 46

such responses (29.68%), the leaderboard received 43 (27.74%), and the daily goal received 42 (27.10%). The heart system had the lowest overall mean and the largest number of "not helpful at all" responses, selected by 25 respondents (16.13%).

The results therefore suggest that participants did not evaluate gamification as a single, uniform experience. Instead, their evaluations varied across individual mechanics. Continuity-supporting features were descriptively evaluated somewhat more positively, while the mechanic based most directly on limiting users after mistakes received the lowest mean rating.

4.2.4 User Preference Rankings

A separate ranking question asked respondents to order selected gamification mechanics according to how helpful they considered them for maintaining engagement. The ranking data were summarised using a weighted scoring system, with higher-ranked mechanics receiving more points.

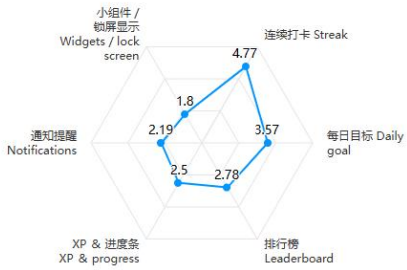

Figure 9 Ranking of Selected Mechanics for Maintaining Engagement

Source: Survey data collected by the author.

The streak received the highest average ranking score at 4.77 and was ranked first by 90 respondents (58.06%). Daily goal ranked second, with an average score of 3.57, and was ranked first by 30 respondents (19.35%).

The remaining mechanics received lower average ranking scores. Leaderboard received an average score of 2.78, the combined XP and progress option received 2.50, notifications received 2.19, and widgets or lock-screen display received 1.80.

The ranking data therefore provide a clearer indication of relative preference than the mean helpfulness scores. Although the differences in mean helpfulness ratings were small, the streak was substantially more likely to be selected as the highest-priority mechanic when participants were required to compare selected features directly.

Daily goals also appeared relatively important, although they were less dominant than streaks. Leaderboards occupied a middle position, suggesting that competition was valued by some users but not prioritised as consistently. Notifications and widgets received lower ranking scores, indicating that external reminders may contribute to return behaviour without necessarily being perceived as central sources of motivation.

It should be noted that the ranking question did not reproduce the helpfulness-rating item in exactly the same format: for example, XP and progress were grouped together in the ranking option. The two questions should therefore be treated as complementary rather than directly equivalent measures.

4.2.5 Pressure and Discomfort

The survey also asked whether any gamification mechanics made respondents feel stressed, uncomfortable or pressured. Participants could select more than one option, meaning that percentages in this question are not mutually exclusive and do not sum to 100%.

The leaderboard was the most frequently selected source of pressure or discomfort, chosen by 70 respondents (45.16%). The streak was selected by 62 respondents (40.00%), followed by notifications by 61 respondents (39.35%), widgets or lock-screen display by 52 respondents (33.55%), daily goal by 41 respondents (26.45%), and XP and progress by 39 respondents (25.16%). A further 17 responses (10.97%) were recorded under “other”.

These results complicate a simple positive-versus-negative interpretation of gamification. The leaderboard, for example, was not the lowest-rated mechanic in the helpfulness question, yet it was the most frequently selected source of pressure. Similarly, the streak was the most strongly prioritised mechanic in the ranking question, while 40.00% of respondents also associated it with pressure or discomfort.

The important pattern is therefore not that an individual mechanic is either helpful or harmful. Rather, the same mechanic can be perceived as motivational and pressuring at the same time. This is consistent with research demonstrating that gamification can generate beneficial and adverse experiences depending on the element, user and context (Almeida et al., 2023).

For the streak, repeated use may arise partly from a desire to preserve accumulated continuity. For the leaderboard, additional participation may be associated with competition and social comparison. In both cases, increased willingness to return to the app does not necessarily imply that the resulting experience is entirely enjoyable or learning-oriented.

The heart system was not included as a standard response option in this pressure/discomfort question. It would therefore be methodologically inappropriate to use this item to claim that a particular proportion of the survey sample experienced pressure from hearts. The heart system is instead interpreted through its helpfulness rating and the structured written follow-up responses.

4.2.6 Long-Term User Segment

Because a substantial proportion of the full sample consisted of relatively new users, participants who had used Duolingo for more than six months were examined

separately as an exploratory subgroup. This subgroup consisted of 30 respondents, representing 19.35% of the full sample.

Because the subgroup is small and was not drawn from a probability sample, the following comparisons are descriptive and exploratory. They should not be interpreted as representative estimates of all long-term Duolingo users.

Within this subgroup, 11 respondents (36.67%) reported daily Duolingo use, 15 (50.00%) used it a few times a week, and 4 (13.33%) used it less than once a week. Thus, 26 of the 30 longer-term users (86.67%) reported using Duolingo at least a few times per week. Descriptively, daily use was more common in this subgroup than in the full sample.

Longer-term users were also more frequently paying users. Thirteen respondents (43.33%) in this subgroup reported using a paid plan, compared with 17 (56.67%) who used the free version. In the full sample, only 18 of 155 respondents (11.61%) were paying users. This descriptive association may reflect greater investment among longer-term users, or paid users may be more likely to continue using the application; however, the cross-sectional survey cannot establish the direction or cause of this relationship.

Streak-maintenance behaviour was also relatively common among longer-term users. Eighteen respondents (60.00%) reported having used Duolingo mainly to avoid losing their streak, seven (23.33%) answered “no”, and five (16.67%) selected “not sure”. Compared with the full-sample percentage of 49.03%, this descriptive pattern suggests that accumulated streaks may become especially salient for some users with longer experience.

The helpfulness ratings also differed descriptively. Among longer-term users, the streak received the highest mean rating (M = 3.53), followed by leaderboard (M = 3.43), streak freeze (M = 3.37), daily goal (M = 3.17), XP rewards (M = 3.10), widgets or lock-screen display (M = 3.07), notifications (M = 3.00), learning path and progress indicators (M = 2.90), and the heart system (M = 2.73).

Several mechanics therefore received higher descriptive ratings among longer-term users than in the full sample, particularly streaks, leaderboards and streak freeze. However, the pressure data reveal an important counter-pattern. Among the 30 longer-term users, the leaderboard was identified as a source of pressure or discomfort by 20 respondents (66.67%), the streak by 14 (46.67%), notifications by 10 (33.33%), XP and progress by 9 (30.00%), widgets or lock-screen display by 8 (26.67%), and daily goal by 7 (23.33%).

The subgroup therefore strengthens the interpretation of gamification as a potentially dual experience. Some mechanics were rated relatively positively by longer-term users while simultaneously being associated with pressure. In particular,

leaderboards and streaks appear to combine motivational value with obligation, competition or concern about losing accumulated progress.

These findings also reinforce the need to define long-term engagement carefully. In the context of this study, engagement is not treated simply as repeated app use. Rather, perceived long-term engagement may involve a combination of habit, motivation, progress preservation, competition and, for some users, pressure.

4.3 Findings from the Structured Written Follow-up Interviews

Seven participants completed the structured written follow-up interviews. To protect participants' identities, they are referred to throughout this section using the labels R1–R7.

The written responses provided further explanation of several patterns identified in the questionnaire. In particular, participants described experiences relating to habit formation, streak-related pressure, minimum-effort participation, competition and leaderboard fatigue, frustration with the Heart system, and perceived learning value.

The qualitative findings are used to provide explanatory context rather than statistical generalisation. Table 3 summarises the principal themes identified in the structured written follow-up responses.

Table 3 Summary of Themes from the Structured Written Follow-up Interviews

Theme	Description	Relevance to research question
Streak as a habit-supporting mechanism	Participants described the streak as a reason to return regularly	Helps explain the relatively positive streak ratings and rankings
Streak-related pressure	Participants described concern or pressure associated with losing a streak	Helps explain why a highly valued mechanic was also associated with discomfort
Minimum-effort completion	Participants sometimes completed short or easy tasks mainly to preserve progress	Illustrates the distinction between app engagement and learning depth
Leaderboard motivation	Competition encouraged additional participation for some users	Helps explain positive perceptions of leaderboards
Leaderboard fatigue	Competition was tiring, stressful or unimportant for some users	Helps explain the high pressure/discomfort response for leaderboards

Theme	Description	Relevance to research question
Heart-system frustration	Losing hearts could interrupt a learning session	Provides qualitative context for the relatively low helpfulness rating of hearts
Learning as continued exposure	Participants saw Duolingo as useful for maintaining regular contact with a language	Clarifies the perceived learning value of continued engagement

4.3.1 Streak as a Habit-Supporting Mechanism and a Source of Pressure

The written responses were consistent with the survey finding that the streak was highly salient to many participants. Several participants described it as a visible record of continuity and as a reason to return even when their immediate motivation to study was relatively low.

R1, who had used Duolingo for more than three years and maintained a streak of over 1,000 days, explained:

> "Many times I did not really want to study, but when I saw that the streak was already so long, I thought, 'fine, I will just do one lesson.'"

This response illustrates the mechanism's dual role. The streak helped sustain behavioural continuity even in the absence of strong immediate learning motivation. At the same time, the reason for returning had partly shifted from interest in the lesson itself to preservation of an accumulated record.

Other participants similarly described pressure associated with the possibility of losing a streak. This suggests that the same feature can support consistency while simultaneously creating a sense of obligation.

This interpretation is consistent with Duolingo's own description of the streak, which explicitly links longer streaks with increasing reluctance to lose accumulated progress, and with Mogavi et al.'s (2022) observation that gamified objectives can sometimes become goals in their own right.

4.3.2 Minimum-Effort Completion

A recurring issue in the written responses was minimum-effort completion. Some participants described selecting short, easy or familiar activities when their immediate objective was primarily to maintain a streak or satisfy a gamified requirement.

R4 described using the application “more to keep the streak going” than specifically for learning. R1 similarly described occasions when they opened the application late at night and selected a very easy activity simply to avoid losing the streak.

These accounts reveal an important distinction between behavioural continuity and depth of learning engagement. A gamification mechanic may successfully encourage a user to perform an action every day while having much less influence over the cognitive quality of that action.

This pattern is especially important because it is not unique to the present participants. Mogavi et al. (2022) describe gamification misuse in Duolingo as occurring when attention becomes excessively focused on game-related goals at the expense of learning. Duolingo’s own internal product discussion has also acknowledged that simple session-count metrics can reward repeated shorter or easier activities and that XP-focused systems may encourage grinding behaviour (Duolingo, 2024b).

The present study does not measure whether minimum-effort sessions resulted in lower language proficiency. It can, however, show that some users themselves distinguished between “opening and completing something” and “engaging seriously with learning content”.

4.3.3 Leaderboard Motivation and Fatigue

The structured written responses also revealed contrasting attitudes towards leaderboards. For some participants, competition provided an additional challenge and encouraged more frequent participation. For others, ranking was tiring, stressful or largely irrelevant to their learning goals.

R2 described the leaderboard as one of the reasons for continuing to use Duolingo, particularly when competition with friends created a shared topic of conversation. The participant stated that:

> “Seeing my ranking stay ahead was a happy thing.”

R5 similarly reported paying considerable attention to the leaderboard during the first two months of use and returning to the app when another user overtook them.

These responses illustrate how competitive feedback can provide an additional participation goal beyond course progression. However, other respondents described competition as tiring or felt that XP accumulation redirected attention away from language learning.

The qualitative data therefore help explain why the survey results for the leaderboard appeared contradictory. It was rated moderately positively as a helpfulness

mechanism while simultaneously being the most frequently selected source of pressure or discomfort.

Rather than categorising the leaderboard as either beneficial or harmful, the results suggest that its perceived value is highly dependent on the individual user. This interpretation is consistent with wider research showing that leaderboards and competition can produce both desirable and undesirable outcomes in gamified educational environments (Almeida et al., 2023).

4.3.4 Hearts and Interrupted Learning Flow

The heart system emerged in the written responses as a source of frustration for some participants. Users who discussed hearts focused particularly on the possibility that repeated mistakes could interrupt an ongoing learning session.

R5 described the system as strict because a limited number of hearts meant that multiple incorrect answers could prevent the user from continuing normally. When the material became difficult, this participant sometimes stopped the session after losing too many hearts.

R7 described a similar experience, explaining that several incorrect answers could disrupt the rhythm of the session and lead them to leave the application rather than continue.

These responses provide qualitative context for the heart system's relatively low mean helpfulness rating in the survey. They do not demonstrate that hearts reduce learning performance, but they show that some participants experienced the mechanic as an interruption rather than as helpful learning feedback.

The distinction is pedagogically relevant. Feedback on an incorrect answer can help a learner recognise an error and adjust future performance, but a penalty associated with that error may be experienced differently. Hattie and Timperley (2007) emphasise that the impact of feedback depends on its form and function rather than simply on the existence of feedback itself.

Notably, Duolingo itself later identified frustration with its old heart system as a product-design problem. In July 2025, the company announced testing of an Energy system intended to replace Hearts and stated that beginners were particularly likely to run out of hearts during lessons (Duolingo, 2025). This product change occurred during the broader period in which the present study was conducted and therefore provides useful contextual support for interpreting the participants' experiences, although it is not part of the survey evidence itself.

4.3.5 Perceived Learning Value and Daily Language Exposure

Participants generally described Duolingo as useful for maintaining regular contact with a language, although some distinguished this benefit from deeper or more comprehensive language development.

R1 stated that Duolingo helped maintain daily contact with English but did not necessarily produce substantial improvements in speaking or writing. The participant characterised the application as:

> "a tool for maintaining language exposure rather than a main systematic learning tool."

R7 made a similar distinction, explaining that gamification was useful for developing a habit but that they sometimes needed to remind themselves not to use the application only for maintaining a streak.

These accounts help clarify the type of learning benefit that participants associated with gamification. Gamification can lower the threshold for beginning a learning activity and can help make language contact more regular. However, the participants did not necessarily equate frequent use with comprehensive language development.

This distinction is consistent with earlier reviews of Duolingo research, which have noted that studies often focus on engagement, design and measurable learning outcomes but that the relationship between gamified participation and broader language-learning goals remains complex (Shortt et al., 2023).

The evidence in the present study therefore supports a cautious conclusion: Duolingo's gamification mechanics appear capable of supporting continued participation and language exposure, but the study does not demonstrate that such participation necessarily produces equivalent gains in language proficiency.

4.4 Integrated Analysis

Taken together, the design analysis, survey findings and structured written follow-up interviews indicate that Duolingo's gamification mechanics are perceived in different ways and can combine motivational benefits with negative or pressuring experiences.

Table 4 Integrated Comparison of Selected Gamification Mechanics

Mechanic	Design analysis	Survey finding	Interview finding	Interpretation
Streak	Makes continuity visible and encourages regular return	Highest mean helpfulness rating and clearly highest ranking; also associated with pressure	Frequently described as motivational, but also associated with obligation and minimum-effort use	Most salient perceived engagement-supporting mechanic, but clearly dual-sided
Daily goal	Provides manageable recurring targets	Relatively positive descriptive rating	Not a dominant standalone interview theme	Appears to support routine, but evidence is mainly survey-based
Streak freeze	Protects continuity after missed days	Second-highest mean helpfulness rating	Not a dominant standalone interview theme	May reduce the cost of interruption and support return
XP	Quantifies activity and provides immediate reward information	Moderate descriptive helpfulness rating	Discussed indirectly in relation to completion-oriented behaviour	Useful for making activity visible, but may redirect attention towards measurable output
Leaderboard	Introduces social comparison and competition	Moderately rated but most frequently selected source of pressure	Motivating for some participants and tiring or stressful for others	Highly dependent on user preference and competitive orientation
Hearts	Links mistakes with limits on continued activity	Lowest mean helpfulness rating	Participants described session interruption and frustration	Potentially problematic when corrective feedback is experienced as punishment
Notifications	Provide external return cues	Lower ranking and frequently associated with pressure	Limited standalone qualitative evidence	Useful as reminders but potentially intrusive
Widgets	Maintain visible reminders outside the app	Low ranking; frequently associated with pressure/discomfort	Limited standalone qualitative evidence	May support return behaviour without being a central motivational mechanism

The most consistent finding across the three data sources concerns the streak. In the design analysis, it is structured around continuity and repeated return. In the survey, it received the highest mean helpfulness rating and was overwhelmingly the most frequently ranked first. In the written follow-up responses, it was repeatedly described as a reason to return even when immediate motivation was low.

However, the streak also illustrates the central tension identified by the study. Forty per cent of the survey sample associated it with pressure or discomfort, and participants described occasions on which preserving the streak became more important than the content of the lesson.

The appropriate conclusion is therefore not that the streak is objectively the mechanic most strongly perceived as supporting engagement. Rather, it was the most strongly perceived engagement-supporting mechanic in this sample, while simultaneously displaying a clear potential to produce pressure and minimum-effort behaviour.

Daily goals and streak freeze showed a somewhat different pattern. Their survey ratings were relatively positive, but neither generated the same level of qualitative tension as the streak or leaderboard in the follow-up data. Streak freeze is particularly interesting because it preserves the motivational value of accumulated continuity while reducing the immediate consequence of missing a day. Duolingo's own product research similarly describes streak freeze as a mechanism designed to introduce flexibility into streak maintenance (Duolingo, 2022).

Leaderboards produced the clearest conflict between motivation and negative emotional experience. They were not rated particularly poorly in the helpfulness item, and competitive participants described them positively. Nevertheless, the leaderboard was the most frequently selected source of pressure in both the full survey sample and the longer-term subgroup. This suggests that a feature can stimulate participation without necessarily improving users' subjective quality of experience.

The heart system showed a different pattern. It received the lowest mean helpfulness score, and follow-up participants described occasions on which losing hearts disrupted an ongoing learning session. This finding should remain carefully bounded because hearts were not included as a standard option in the survey pressure item. Nevertheless, the combination of the low helpfulness rating and qualitative frustration provides reasonable evidence that some users perceived the mechanic negatively.

The broader integrated finding concerns the distinction between engagement with the application and engagement with language learning. Some mechanics clearly gave participants reasons to return, complete activities and maintain continuity. Yet the follow-up responses show that these actions could sometimes be accomplished through the shortest or easiest available activity.

This distinction is supported particularly strongly by two external sources. Mogavi et al. (2022) identified gamification misuse on Duolingo when game objectives displaced learning objectives, while Duolingo's own 2024 product analysis acknowledged that metrics based simply on numbers of sessions could favour

repeated completion of shorter and easier activities and that XP-based competition could create grinding behaviour.

These findings do not imply that short sessions are inherently educationally ineffective or that gamified motivation is undesirable. Instead, they demonstrate that frequency of participation and quality of participation should be analytically separated.

4.5 Theoretical Interpretation

The findings can be interpreted through the three theoretical perspectives established in Chapter 2: Self-Determination Theory, Reinforcement Theory, and research on feedback mechanisms. These frameworks are used as interpretive lenses rather than as variables directly tested by the study.

4.5.1 Self-Determination Theory

Self-Determination Theory proposes that the quality and sustainability of motivation are influenced by the satisfaction of three basic psychological needs: autonomy, competence and relatedness (Deci and Ryan, 1985; Ryan and Deci, 2000).

Several Duolingo mechanics provide information that may contribute to perceived competence. Progress indicators, XP, completed goals, badges and streaks make users' activity visible and provide repeated signals that something has been accomplished. However, visible achievement should not automatically be interpreted as competence satisfaction. Research on individual gamification elements has shown that points, levels and leaderboards can influence performance without necessarily producing equivalent increases in intrinsic motivation or psychological-need satisfaction (Mekler et al., 2017).

Autonomy provides a particularly useful interpretation of the streak results. A streak may support self-organised routine when users personally value daily practice. However, the same mechanism may feel controlling when a user returns primarily because they are afraid of losing an accumulated record. The survey finding that the streak was both highly prioritised and frequently associated with pressure is consistent with this distinction between voluntarily endorsed and externally or internally pressured behaviour.

Daily goals can similarly support autonomy when users experience them as manageable, self-endorsed targets, but may be experienced as controlling when they become obligations. It would therefore be too strong to state that daily goals inherently increase autonomy.

Relatedness requires equally careful interpretation. Leaderboards place users within a visible social environment, but social comparison is not identical to psychological relatedness. For some participants, competition with friends created shared experiences and motivation, as illustrated by R2. For others, leaderboard comparison generated pressure or disengagement. Thus, competitive social mechanisms may support connection for some users while failing to satisfy, or potentially undermining, relatedness for others.

Self-Determination Theory therefore helps explain why the same mechanic may produce different experiences across users. The central issue is not merely whether a reward, goal or comparison exists, but whether the user experiences the resulting participation as competent, socially meaningful and self-endorsed.

4.5.2 Reinforcement Theory

Reinforcement Theory provides a behavioural perspective on repeated participation. In Skinner's account of operant behaviour, consequences following behaviour can influence the probability of future behaviour (Skinner, 1953).

Duolingo contains multiple reward contingencies that are consistent with this perspective. Completing a lesson may produce XP, extend a streak, contribute towards a daily goal, maintain leaderboard position and advance visible progress. In this way, a single learning action can be followed by several immediate consequences.

It is important, however, not to describe these mechanics as experimentally demonstrated "positive reinforcers" in the present study. Demonstrating reinforcement in a strict behavioural sense would require evidence that a specific consequence increased subsequent behaviour. The present survey measures user perceptions rather than experimentally manipulating consequences. It is therefore more accurate to state that these mechanics can be interpreted as reward contingencies consistent with reinforcement principles.

The streak findings are particularly relevant. Almost half of the full survey sample reported having opened Duolingo primarily to avoid losing a streak, and the proportion was 60.00% among the longer-term subgroup. The follow-up responses also provided examples of participants completing minimal activities to preserve the streak. The consequences associated with continuity therefore appear closely connected to repeated use in participants' own accounts.

At the same time, reinforcement of activity does not necessarily imply reinforcement of deeper learning. A system may successfully increase the probability that a user completes some activity without determining whether that activity is challenging,

meaningful or pedagogically appropriate. The minimum-effort pattern observed in the written responses illustrates this distinction.

Penalty-like consequences are also relevant. Hearts and potential streak loss differ from XP and badges because they involve avoiding or experiencing an undesirable outcome. The follow-up responses suggest that these mechanisms may increase careful attention for some users but can also produce frustration or withdrawal. Streak freeze partly modifies this dynamic by reducing the consequence of a missed day.

Reinforcement Theory therefore helps explain why Duolingo can be effective at generating repeated behavioural responses while leaving open a separate question about the educational quality of those responses.

4.5.3 Feedback Mechanisms

Feedback research provides the most direct theoretical framework for understanding how Duolingo communicates progress and performance. Hattie and Timperley (2007) argue that feedback can support learning when it helps learners understand their current performance, their goals and the steps needed to reduce the gap between the two. However, feedback is not uniformly beneficial; its effectiveness depends on the type, level and context of the information provided.

Duolingo provides at least two analytically distinct forms of feedback. The first can be described as learning-related feedback, such as whether an answer is correct, where an error occurred and how the learner is progressing through course content. The second can be described as gamification-related feedback, such as XP totals, streak length, badges, goal completion and leaderboard position.

Both forms make activity visible, but they direct attention towards different targets. Corrective feedback can direct attention towards language performance, whereas XP or leaderboard feedback may direct attention towards activity quantity, status or competition.

This distinction helps explain several findings in the present study. Progress indicators may provide useful orientation because they tell users where they are within the course. XP makes effort immediately visible but may encourage users to evaluate progress in terms of points. Leaderboards transform XP into comparative feedback, which can motivate competitive users but also generate pressure. Streaks provide feedback about behavioural consistency rather than language mastery.

The heart system is especially important from this perspective. Corrective information following an error can potentially support learning, but when the same

error is also associated with losing a limited resource, the feedback becomes partly punitive. Participants who described stopping after losing hearts illustrate how a feedback system can communicate both information and consequences.

The theoretical distinction between learning-related and gamification-related feedback also helps explain the central finding of the chapter. Users may receive abundant feedback showing that they are active—through streaks, points, rankings, badges and goal completion—without receiving equivalent evidence that their language competence has improved to the same degree.

This does not make gamification feedback inherently ineffective. Rather, the educational value of such feedback depends on whether it remains aligned with the underlying learning goal. This interpretation is consistent with broader gamification evidence showing positive average effects while also indicating substantial variation across designs and outcomes (Sailer and Homner, 2020).

4.6 Summary of Analysis

The results show that users did not perceive Duolingo's gamification mechanics uniformly. Instead, different mechanics were associated with different combinations of motivation, routine, pressure and perceived learning value.

In relation to RQ1, the overall evaluations varied across mechanics, although the mean helpfulness ratings were relatively close. The streak received the highest descriptive mean rating, while the heart system received the lowest. The ranking task produced a clearer distinction, with the streak substantially more likely to be ranked first than the other selected mechanics.

In relation to RQ2, the streak emerged as the mechanic most strongly perceived as supporting continued engagement. Daily goals and streak freeze were also evaluated relatively positively. The longer-term user subgroup showed particularly high descriptive ratings for streaks, leaderboards and streak freeze, although these subgroup results remain exploratory.

In relation to RQ3, the findings demonstrate that engagement-supporting mechanics may also produce negative experiences. Leaderboards were the most frequently reported source of pressure or discomfort, while streaks were both highly valued and frequently associated with pressure. The written follow-up responses further identified minimum-effort completion, leaderboard fatigue and frustration associated with the heart system.

In relation to RQ4, participants generally perceived Duolingo as useful for maintaining regular language exposure and learning routines, but the qualitative

responses also distinguished repeated app use from deeper language learning. Some participants reported completing easy or minimal activities primarily to protect gamified progress rather than to engage deeply with the learning content.

Overall, the findings suggest that Duolingo's gamification system is perceived as particularly effective in making continued participation easier to initiate and maintain. However, perceived engagement should not be treated as equivalent to learning quality. Streaks, XP, goals, rankings and other visible feedback may successfully organise repeated behaviour while sometimes redirecting attention towards the gamified representation of progress.

The central finding of this chapter is therefore not that gamification is simply beneficial or harmful. Rather, its effects appear to be mechanic-specific and user-dependent. The same feature can support consistency while generating pressure, and the same participation can represent either meaningful practice or minimum-effort maintenance. This dual character is essential for understanding how gamification shapes long-term engagement in language-learning applications.

5 Conclusion

5.1 Overview

The findings indicate that participants perceived Duolingo's gamification mechanics differently in terms of their support for continued engagement and perceived learning value. The streak emerged as the mechanic most strongly perceived as supporting continued engagement in this sample, while daily goals and streak freeze were also evaluated relatively positively. At the same time, both streaks and leaderboards were associated with pressure or discomfort, illustrating that mechanics perceived as motivating may also create unwanted forms of obligation or competition.

The findings further suggest an important distinction between continued app engagement and meaningful learning engagement. Participants described cases in which they returned to Duolingo or completed short activities primarily to preserve a streak or other gamified progress. Such behaviour still represents continued participation, but it should not automatically be interpreted as evidence of deeper language learning or improved proficiency.

Overall, this study suggests that the perceived value of gamification depends on the specific mechanic and on how users experience it. Continuity-oriented mechanics may support regular language exposure and routine formation, while competitive or restrictive mechanics may also create pressure, fatigue, or interruption. Because the study relied on cross-sectional self-reported data, these findings should be interpreted as evidence of perceived sustained engagement rather than causal effects on long-term retention or language proficiency.

The study therefore suggests that gamification can support sustained participation in language learning, and that continuity of practice remains an important component of the learning process. However, engagement generated by gamification does not necessarily translate directly into stronger learning outcomes. Ultimately, gamification can make language learning more attractive, structured, and easier to sustain, but it cannot substitute for deeper and more effective engagement with the language itself.

5.2 Limitations

Several limitations of this study should be acknowledged.

First, the study relied primarily on self-reported data. Survey respondents and participants in the supplementary written follow-up provided accounts of their own

experiences, attitudes, and perceptions, but these responses were not verified against platform-level behavioural data from Duolingo. The findings should therefore be interpreted primarily as evidence of perceived engagement rather than objectively measured long-term behaviour or retention. Self-reported responses may also be affected by factors such as recall and response bias.

Second, the survey adopted a cross-sectional design. Participants reported their experiences at a particular point in time rather than being followed repeatedly over an extended period. Consequently, the study cannot determine how individual users' motivation, attitudes towards gamification mechanics, or patterns of engagement change over months or years. It also cannot establish whether mechanics that users initially perceive as motivating remain effective after prolonged use.

Third, the composition and sampling strategy of the survey limit the generalisability of the findings. Participants were recruited through purposive online sampling, rather than probability sampling, and the sample contained a relatively high proportion of newer Duolingo users. Of the 155 respondents, 125 (80.65%) had used Duolingo for six months or less, while only 30 respondents (19.35%) had used the app for more than six months. Although the long-term user subgroup provided useful additional evidence, the overall sample cannot be treated as representative of the wider population of Duolingo users. The findings regarding long-term engagement should therefore be interpreted as users' perceptions of factors that may support continued use rather than direct evidence of long-term retention.

Fourth, the supplementary written follow-up sample was small. Seven participants provided detailed qualitative responses, which helped explain patterns identified in the survey, particularly those concerning streak pressure, minimum-effort participation, leaderboard competition, and frustration with the heart system. However, these responses were intended to provide interpretative depth rather than statistical generalisation. They cannot represent the full range of experiences among Duolingo users.

Fifth, the quantitative analysis was primarily descriptive. Percentages, mean ratings, ranking scores, and subgroup patterns were used to identify tendencies within the sample, but no inferential statistical tests were conducted to determine whether observed differences between mechanics or user groups were statistically significant. The comparisons presented in this study should therefore be interpreted as descriptive patterns rather than confirmed population-level differences.

Sixth, the study did not directly measure language-learning outcomes. No objective assessments of vocabulary acquisition, grammar development, pronunciation, speaking ability, retention, or overall language proficiency were conducted. The study can therefore discuss users' perceived learning value of different gamification mechanics, but it cannot determine whether particular mechanics objectively

improved language proficiency. This distinction is important because greater engagement with an application does not necessarily correspond directly to greater learning gains.

Finally, the study focused on visible gamification mechanics encountered by users during normal use of Duolingo. It did not examine Duolingo's internal algorithms, recommendation systems, user-level behavioural data, or the full pedagogical and curriculum structure underlying the platform. In addition, Duolingo is a continuously updated digital platform, and individual mechanics may be modified, replaced, or experienced differently across versions and subscription plans. The findings should therefore be understood in relation to the version of Duolingo observed during the 2024–2025 data-collection period rather than as a permanent description of all versions of the platform.

Taken together, these limitations mean that the present research should be understood as a small-scale, user-centred case study of Duolingo's visible gamification mechanics and perceived long-term engagement, rather than as a causal evaluation of Duolingo's effects on long-term retention or language proficiency.

5.3 Future Research

Future research could build on this study in several ways.

First, longitudinal research could follow Duolingo users over several months or years and collect repeated measurements of their engagement and perceptions. Such a design would make it possible to examine whether mechanics such as streaks, daily goals, streak freeze, and leaderboards remain motivating over time, become habitual, or gradually lose their motivational effect. It could also reveal how users' relationships with particular mechanics change as their experience with the platform increases.

Second, future studies could combine self-reported perceptions with objective behavioural data, where access is ethically and technically feasible. Measures such as login frequency, lesson completion, streak length, session frequency, time spent in the app, and changes in usage over time could be compared with users' reported motivation and attitudes. This would provide a stronger basis for distinguishing between users' perceptions of engagement and their actual patterns of behaviour.

Third, future research could recruit a larger and more diverse sample. A broader sample would allow more systematic comparisons between different groups, including short-term and long-term users, frequent and occasional users, free and paying users, highly competitive and non-competitive users, and users who have discontinued Duolingo. Such comparisons could help determine whether particular gamification mechanics function differently for different types of learners.

Fourth, future studies should investigate the relationship between gamification mechanics and objectively measured language-learning outcomes. Pre- and post-study assessments could be used to examine changes in areas such as vocabulary retention, grammatical knowledge, listening comprehension, speaking ability, or overall language proficiency. This would allow future research to distinguish more clearly between mechanics that encourage users to remain active in the app and mechanics that contribute to measurable language development.

Fifth, experimental or quasi-experimental research could provide a more direct assessment of the effects of individual gamification mechanics. Rather than examining the complete gamified system as a single intervention, future studies could compare conditions in which particular mechanics are present, absent, or modified. For example, researchers could compare learning experiences with and without leaderboards, different levels of streak pressure, or supportive versus penalty-based feedback. Such designs would make it easier to estimate the specific contribution of individual mechanics to engagement, pressure, and learning behaviour.

Sixth, future design research could investigate more adaptive and personalised forms of gamification. Users may respond differently to competition, reminders, rewards, and loss-based mechanics, and allowing greater control over features such as leaderboard participation, notification intensity, daily goals, or streak-related pressure may help maintain motivation while reducing unwanted stress. Collaborative, mastery-oriented, or individually configurable mechanics could also be examined as alternatives to strongly competition-based designs.

Overall, future extensions of this research should increase both the volume of data and the scope of factors considered, and provide more quantitative estimates of the extent to which different gamification mechanics influence user engagement and learning outcomes.

6 Acknowledgements

First and foremost, my deepest gratitude goes to my boyfriend, Chenyang Yu. Without him, this thesis would never have become what it is today.

He has shown me more patience than I could ever have asked for. He read through this thesis carefully, word by word, helped me proofread it, and gave me thoughtful and constructive suggestions throughout the process. At a time when I had almost lost hope in this thesis myself, he continued to believe that I could finish it. He supported me not only academically, but also emotionally. He encouraged me to face my fears instead of avoiding them, patiently guided me through moments of frustration and self-doubt, and kept reminding me to take one step at a time until the thesis was finally completed. At the same time, he took care of me in the most tangible and comforting way — by cooking countless delicious meals for me throughout this difficult period. His patience, encouragement, care, and companionship carried me through some of the hardest moments of this journey. I truly could not have done this without him.

I am also deeply grateful to my family for their unconditional support throughout my studies. Studying abroad has required a significant financial commitment, and my path toward graduation has not always gone as smoothly as expected. Yet, even during the periods when I struggled or fell behind, they never placed pressure on me. Instead, they continued to give me the freedom, trust, and support I needed to find my own way forward. I am profoundly thankful for everything they have done for me.

I would also like to express my sincere gratitude to my examiner, Tarja Susi, for her patience in reading and reviewing this thesis through its many iterations. Her detailed, concrete, and constructive feedback helped me see the problems in my work more clearly and guided me toward improving it step by step. I am equally grateful to Per Backlund for his guidance and support throughout this process.

Finally, I would like to thank Sweden. Living and studying here has allowed me to experience a life very different from the one I knew in my home country — with its own sweetness, bitterness, joys, frustrations, and surprises. Studying abroad on my own has not always been easy, but it has become an irreplaceable part of my life. The challenges I faced, the people I met, and the experiences I had here have given me memories that I will carry with me for many years to come.

This thesis marks not only the end of an academic journey, but also the end of an unforgettable chapter of my life. To everyone who accompanied me along the way, thank you.

References

Almeida, C., Kalinowski, M., Uchôa, A. and Feijó, B. (2023) 'Negative effects of gamification in education software: Systematic mapping and practitioner perceptions', *Information and Software Technology, 156*, 107142. doi: 10.1016/j.infsof.2022.107142. https://doi.org/10.1016/j.infsof.2022.107142

Andrade, C. (2021) 'The inconvenient truth about convenience and purposive samples', *Indian Journal of Psychological Medicine, 43*(1), pp. 86–88. doi: 10.1177/0253717620977000. https://doi.org/10.1177/0253717620977000

Barchard, K.A. and Williams, J. (2008) 'Practical advice for conducting ethical online experiments and questionnaires for United States psychologists', *Behavior Research Methods, 40*(4), pp. 1111–1128. doi: 10.3758/BRM.40.4.1111.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology, 3*(2)*,* pp. 77–101. doi: 10.1191/1478088706qp063oa.

British Educational Research Association (BERA) (2024) *Ethical Guidelines for Educational Research. 5th edn*. London: British Educational Research Association. Available at: https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-fifth-edition-2024-online

British Psychological Society (BPS) (2021) *Ethics Guidelines for Internet-Mediated Research*. Leicester: British Psychological Society. doi: 10.53841/bpsrep.2021.rep155. https://doi.org/10.53841/bpsrep.2021.rep155

Buchanan, E.A. and Hvizdak, E.E. (2009) 'Online survey tools: Ethical and methodological concerns of human research ethics committees', *Journal of Empirical Research on Human Research Ethics, 4*(2)*,* pp. 37–48. doi: 10.1525/jer.2009.4.2.37. https://doi.org/10.1525/jer.2009.4.2.37

Consalvo, M. and Dutton, N. (2006) 'Game analysis: Developing a methodological toolkit for the qualitative study of games', *Game Studies, 6(1)*.

Creswell, J.W. and Creswell, J.D. (2018) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* 5th edn. Los Angeles, CA: SAGE. Available at: https://collegepublishing.sagepub.com/products/research-design-5-255675

Deci, E.L. and Ryan, R.M. (1985) *Intrinsic Motivation and Self-Determination in Human Behavior.* New York: Plenum Press. doi: 10.1007/978-1-4899-2271-7. https://doi.org/10.1007/978-1-4899-2271-7

Deterding, S., Dixon, D., Khaled, R. and Nacke, L. (2011) 'From game design elements to gamefulness: Defining "gamification"', in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (MindTrek '11). New York, NY: Association for Computing Machinery, pp. 9–15. doi: 10.1145/2181037.2181040. https://doi.org/10.1145/2181037.2181040

Dicheva, D., Dichev, C., Agre, G. and Angelova, G. (2015) 'Gamification in education: A systematic mapping study', *Educational Technology & Society, 18*(3), pp. 75–88. Available at: https://www.jstor.org/stable/jeductechsoci.18.3.75

Dillman, D.A., Smyth, J.D. and Christian, L.M. (2014) *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method.* 4th edn. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9781394260645. https://doi.org/10.1002/9781394260645

Duolingo (2022) 'The Duolingo streak uses habit research to keep you motivated', *Duolingo Blog*. Available at: https://blog.duolingo.com/how-duolingo-streak-builds-habit/

Duolingo (2023a) 'How to install the Duolingo widget on iPhone or Android', *Duolingo Blog*. Available at: https://blog.duolingo.com/widget-feature/

Duolingo (2023b) 'How to earn achievement badges on Duolingo', *Duolingo Blog*. Available at: https://blog.duolingo.com/achievement-badges/

Duolingo (2024a) 'How to use Duolingo for language learning', *Duolingo Blog*. Available at: https://blog.duolingo.com/duolingo-101-how-to-learn-a-language-on-duolingo/

Duolingo (2024b) 'Understanding Duolingo's Time Spent Learning Well metric', *Duolingo Blog*. Available at: https://blog.duolingo.com/time-spent-learning-well/

Duolingo (2025) 'Why Duolingo switched from Hearts to a battery', *Duolingo Blog*. Available at: https://blog.duolingo.com/duolingo-energy/

Fetters, M.D., Curry, L.A. and Creswell, J.W. (2013) 'Achieving integration in mixed methods designs—Principles and practices', *Health Services Research, 48*(6 Pt 2), pp. 2134–2156. doi: 10.1111/1475-6773.12117.

Franzke, A.S., Bechmann, A., Zimmer, M., Ess, C.M. and the Association of Internet Researchers (2020) *Internet Research: Ethical Guidelines 3.0*. Association of Internet Researchers. Available at: https://aoir.org/reports/ethics3.pdf

Gill, P., Stewart, K., Treasure, E. and Chadwick, B. (2008) 'Methods of data collection in qualitative research: Interviews and focus groups', *British Dental Journal*, *204*, pp. 291–295. doi: 10.1038/bdj.2008.192. https://doi.org/10.1038/bdj.2008.192

Godwin-Jones, R. (2011) 'Mobile apps for language learning', *Language Learning & Technology, 15*(2), pp. 2–11. doi: 10.64152/10125/44244. https://doi.org/10.64152/10125/44244

Hamari, J., Koivisto, J. and Sarsa, H. (2014) 'Does gamification work?—A literature review of empirical studies on gamification', in *2014 47th Hawaii International Conference on System Sciences*. IEEE, pp. 3025–3034. doi: 10.1109/HICSS.2014.377. https://doi.org/10.1109/HICSS.2014.377

Hattie, J. and Timperley, H. (2007) 'The power of feedback', *Review of Educational Research, 77*(1), pp. 81–112. doi: 10.3102/003465430298487. https://doi.org/10.3102/003465430298487

Kukulska-Hulme, A. and Viberg, O. (2018) 'Mobile collaborative language learning: State of the art', *British Journal of Educational Technology, 49*(2), pp. 207–218. doi: 10.1111/bjet.12580. https://doi.org/10.1111/bjet.12580

Loewen, S., Crowther, D., Isbell, D.R., Kim, K.M., Maloney, J., Miller, Z.F. and Rawal, H. (2019) 'Mobile-assisted language learning: A Duolingo case study', *ReCALL, 31*(3), pp. 293–311. doi: 10.1017/S0958344019000065. https://doi.org/10.1017/S0958344019000065

Meho, L.I. (2006) 'E-mail interviewing in qualitative research: A methodological discussion', *Journal of the American Society for Information Science and Technology, 57*(10), pp. 1284–1295. doi: 10.1002/asi.20416.

Mekler, E.D., Brühlmann, F., Tuch, A.N. and Opwis, K. (2017) 'Towards understanding the effects of individual gamification elements on intrinsic motivation and performance', *Computers in Human Behavior, 71,* pp. 525–534. doi: 10.1016/j.chb.2015.08.048.

Mogavi, R.H., Guo, B., Zhang, Y., Haq, E.-U., Hui, P. and Ma, X. (2022) 'When gamification spoils your learning: A qualitative case study of gamification misuse in a language-learning app', in *Proceedings of the Ninth ACM Conference on Learning @*

Scale (L@S '22). New York, NY: Association for Computing Machinery, pp. 175–188. doi: 10.1145/3491140.3528274. https://doi.org/10.1145/3491140.3528274

Munday, P. (2016) 'The case for using DUOLINGO as part of the language classroom experience', *RIED. Revista Iberoamericana de Educación a Distancia, 19*(1), pp. 83–101. doi: 10.5944/ried.19.1.14581. https://doi.org/10.5944/ried.19.1.14581

Ryan, R.M. and Deci, E.L. (2000) 'Self-Determination Theory and the facilitation of intrinsic motivation, social development, and well-being', *American Psychologist, 55*(1), pp. 68–78. doi: 10.1037/0003-066X.55.1.68. https://doi.org/10.1037/0003-066X.55.1.68

Sailer, M. and Homner, L. (2020) 'The gamification of learning: A meta-analysis', *Educational Psychology Review, 32,* pp. 77–112. doi: 10.1007/s10648-019-09498-w. https://doi.org/10.1007/s10648-019-09498-w

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B. and Akinkuolie, B. (2023) 'Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020', *Computer Assisted Language Learning, 36*(3), pp. 517–554. doi: 10.1080/09588221.2021.1933540. https://doi.org/10.1080/09588221.2021.1933540

Skinner, B.F. (1953) *Science and Human Behavior*. New York: Macmillan.

Sullivan, G.M. and Artino, A.R. Jr. (2013) 'Analyzing and interpreting data from Likert-type scales', *Journal of Graduate Medical Education, 5*(4), pp. 541–542. doi: 10.4300/JGME-5-4-18.

Zhang, S. and Hasim, Z. (2023) 'Gamification in EFL/ESL instruction: A systematic review of empirical research', *Frontiers in Psychology, 13,* 1030790. doi: 10.3389/fpsyg.2022.1030790. https://doi.org/10.3389/fpsyg.2022.1030790.

Appendix A - Survey

Duolingo 用户体验调查问卷 | User Experience Survey – Duolingo Gamification Mechanics

数据使用声明 / Data Usage Statement
此问卷仅用于硕士毕业论文研究，数据将严格匿名处理，不涉及任何商业用途。
This survey is for academic research only. All responses are anonymous and will not be used for commercial purposes.

---

* 1. **Part 1. 基本信息 / Background Information**
您使用 Duolingo 多久了?
How long have you been using Duolingo?

- ◯ 少于 1 个月 / Less than 1 month
- ◯ 1–3 个月 / 1–3 months
- ◯ 3–6 个月 / 3–6 months
- ◯ 超过 6 个月 / More than 6 months

* 2. 您通常多久使用一次 Duolingo?
How often do you typically use Duolingo?

- ◯ 每天 / Daily
- ◯ 每周几次 / A few times a week
- ◯ 每周不到一次 / Less than once a week

* 3. 您是付费用户吗?
Are you a paying user?

- ◯ Yes
- ◯ No

* 4. **Part 2. 功能体验 / Gamification Mechanics Experience**

请根据您对以下功能的使用感受，选择最符合您看法的选项。

Please rate how much the following features help you stay engaged with Duolingo.

1：非常没有帮助 Not helpful at all； 5：极大帮助 Extremely helpful

	非常没有帮助 Not helpful at all	略有帮助 Slightly helpful	有帮助 Helpful	很有帮助 Very helpful	极大帮助 Extremely helpful
连续打卡 Streak	○	○	○	○	○
经验值 XP 奖励	○	○	○	○	○
每日目标 Daily goal	○	○	○	○	○
排行榜 Leaderboard	○	○	○	○	○
技能树 / 进度条 Skill tree & progress	○	○	○	○	○
红心机制 Heart system	○	○	○	○	○
通知提醒 Notifications	○	○	○	○	○
小组件 / 锁屏显示 Widgets / lock screen	○	○	○	○	○
连续打卡保护（药水）Streak freeze	○	○	○	○	○

* 5. **Part 3. 个人体验 / Personal Experience**

您是否有因为"不想断掉打卡"而使用 App?

Have you ever used Duolingo mainly to avoid losing your streak?

- ○ 是 / Yes
- ○ 否 / No
- ○ 不确定 / Not sure

* 6. **您认为下列功能中，哪些最有助于您保持长期使用 Duolingo？请根据对您个人的重要性，将它们从"最有帮助"到"最无帮助"进行排序。**

Which of the following features help you most in staying engaged with Duolingo? Please rank them from most helpful to least helpful based on your personal experience.

（请将选项按顺序排列，1 = 最有帮助，6 = 最无帮助）

(Rank all items: 1 = most helpful, 6 = least helpful)

【排序题】

- ☐ 连续打卡 Streak
- ☐ 排行榜 Leaderboard
- ☐ 每日目标 Daily goal
- ☐ XP & 进度条 XP & progress
- ☐ 通知提醒 Notifications
- ☐ 小组件 / 锁屏显示 Widgets / lock screen

* 7. 您是否对某些功能感到压力或反感？ （可多选）

Have any features made you feel pressure or discomfort? (Multiple choices allowed)

【多选题】

- ☐ 连续打卡 Streak
- ☐ 排行榜 Leaderboard
- ☐ 每日目标 Daily goal
- ☐ XP & 进度条 XP & progress
- ☐ 通知提醒 Notifications
- ☐ 小组件 / 锁屏显示 Widgets / lock screen
- ☐ 其他（请说明）： Other (please specify):

* 8. **Part 4. 后续研究 / Follow-up**

您愿意参加一次 15–20 分钟的线上访谈吗？ （自愿，匿名）

Would you be open to a short online interview (15–20 min)? Participation is voluntary and anonymous.

- ○ 愿意 / Yes

  如愿意，请留下您的联系方式（可选）：

  If yes, please leave your email (optional):
- ○ 不愿意 / No

Appendix B - Structured Written Follow-up Interview Questions

Overview:

These structured written follow-up interview questions are designed to complement the survey findings by exploring users' personal experiences with Duolingo's gamification mechanics. The interview focuses on perceived long-term engagement, motivation, emotional responses, and perceived learning value. All responses are voluntary, anonymous, and used only for academic purposes.

1. 您使用 Duolingo 多久了？通常多久使用一次？

How long have you been using Duolingo, and how often do you usually use it?

2. 您最初为什么开始使用 Duolingo？

Why did you start using Duolingo in the first place?

3. 哪个功能最能激励您持续使用 Duolingo？为什么？

Which feature motivates you most to continue using Duolingo, and why?

4. 您如何看待 streak 连续打卡功能？

How do you feel about the streak feature?

5. 有没有哪个功能让您感到压力、不舒服或失去兴趣？

Is there any feature that makes you feel pressured, uncomfortable, or less interested?

6. 您如何看待排行榜 leaderboard？

How do you feel about the leaderboard?

7. 您觉得这些游戏化功能真的帮助您学习了吗，还是主要只是让您登录？

Do you feel these gamification features actually help you learn, or do they mainly make you log in?

8. 总体来说，您认为 Duolingo 在帮助用户长期坚持学习方面做得怎么样？

Overall, how well do you think Duolingo supports long-term engagement in language learning?