

## Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020

Mitchell Shortt, Shantanu Tilak, Irina Kuznetcova, Bethany Martens & Babatunde Akinkuolie

**To cite this article:** Mitchell Shortt, Shantanu Tilak, Irina Kuznetcova, Bethany Martens & Babatunde Akinkuolie (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, 517-554, DOI: [10.1080/09588221.2021.1933540](https://doi.org/10.1080/09588221.2021.1933540)

**To link to this article:** <https://doi.org/10.1080/09588221.2021.1933540>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 05 Jul 2021.



[Submit your article to this journal](#)



Article views: 53820



[View related articles](#)








[View Crossmark data](#)



Citing articles: 29 [View citing articles](#)

# Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020

Mitchell Shortt<sup>a</sup> , Shantanu Tilak<sup>b</sup> , Irina Kuznetcova<sup>a</sup> ,  
Bethany Martens<sup>c</sup>  and Babatunde Akinkuolie<sup>d</sup> 

<sup>a</sup>Learning Technologies, The Ohio State University, Columbus, OH, USA; <sup>b</sup>Educational Psychology, The Ohio State University, Columbus, OH, USA; <sup>c</sup>Teaching and Learning: Foreign, Second, & Multilingual Language Education, The Ohio State University, Columbus, OH, USA; <sup>d</sup>Research Laboratory for Digital Learning, The Ohio State University, Columbus, OH, USA

## ABSTRACT

More than 300 million people use the gamified mobile-assisted language learning (MALL) application (app) Duolingo. The challenging tasks, reward incentives, systematic levels, and the ranking of users according to their achievements are just some of the elements that demonstrate strong gamification elements within this popular language learning application. This application's pervasive reach, flexible functionality, and freemium business model has brought significant attention to gamification in MALL. The present systematic review aims to summarize different methods, frameworks, settings, and research samples used to assess Duolingo's design and impact on various learning outcomes. We carried out a complete database search for articles focused on the issues of design, application, and pedagogies in the use of Duolingo. Three hundred and sixty-seven records were initially found, and 35 of those were selected for final inclusion based on language choice, theoretical frameworks, design, sampling, data collection, and analyses (see [Appendix 1](#) for full list). The results indicated that the majority of research from 2012–2020 was design-focused, quantitative in nature, and used non-probability sampling methods. The focus on app design marks an emphasis on the creation of tools rather than the process and outcomes of language learning from using these tools. Additional results revealed preferences for performance-based research questions, for English as language of choice in research, and for the USA as the most prominent context for Duolingo research studies. Furthermore, our review highlights research gaps specific to Duolingo, yet generalizable to other MALL applications. The results are useful to researchers seeking to assess, evaluate, and understand MALL, gamification, and Duolingo as well as to practitioners interested in utilizing MALL in formal and informal learning environments.

## KEYWORDS

Systematic review;  
Duolingo; CALL; MALL;  
language learning;  
design

**CONTACT** Mitchell Shortt  [shortt.10@osu.edu](mailto:shortt.10@osu.edu)

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

## Introduction

Duolingo is one of the most dominant and influential mobile language learning applications (apps) on the market today (Duolingo Help Center, 2020, para. 1). Its popularity among millions of users has been determined, at least in part, by its free-of-charge access model and gamified features (Duolingo About us: Mission, 2021). In the field of Mobile-Assisted Language Learning (MALL), Duolingo is generally seen as a strong representation of gamification in MALL applications (Huynh, Zuo, & Iida, 2016, 2018). Gamification refers to the use of game-based elements to engage individuals, motivate action, promote learning, and solve problems (Kapp, 2012). Challenging tasks, incentive rewards, systematic levels, and the ranking of users based on achievements are just some gamification elements Duolingo utilizes. Most MALL tools use gamification elements to some extent, making it a prevailing approach to content delivery in such environments. Duolingo stands as one of the most gamified MALL apps (Govender & Arnedo-Moreno, 2020).

As smartphones, tablets, and similar devices become more widespread and powerful (Silver, 2019), at times rivaling computers, and gamified MALL platforms see more popularity and acceptance, understanding the effectiveness, impact, and guidelines for implementation of such MALL applications is becoming more and more important for language learning researchers, practitioners, and educators (Burston, 2014).

To keep abreast of research trends in MALL and gamification, analyze new developments and persisting gaps in the field, and provide recommendations for future research and educational practice, we conducted a systematic review of gamified MALL applications with a focus on Duolingo (as one of the most representative gamified MALL platforms), surveying studies ranging from its public release in January 2012 to April 2020.

## Mobile-assisted language learning

Mobile Learning is a type of learning activity that is mediated through mobile devices which does not require the learner to be tied to a particular geographical location (Wu et al., 2012). Mobile-Assisted Language Learning is one of its subcategories focusing specifically on the context of language learning.

MALL has been reported to enhance teacher-student and student-student communication, speaking, and listening skills (Golonka, Bowles, Frank, Richardson, & Freynik, 2014; Hwang, Chen, & Chen, 2011; Kondo et al., 2012; Toland, Mills, & Kohyama, 2016). Some of the most recent systematic reviews for MALL involve learning in authentic environments (Shadiev, Liu, & Hwang, 2020) and collaborative

language learning (Kukulska-Hulme & Viberg, 2018). Shadiev et al. (2020) found that task-based learning and communicative language teaching were the most common occurrence, and questionnaires, pretest/posttest, and interviews were the most frequent data collection methods used. They concluded that researchers need to explicitly state what approaches were utilized, expand data collection beyond questionnaires and pretest/posttest, critically address issues with MALL, and employ more context-based approaches to explore social communication experiences. Kukulska-Hulme and Viberg (2018) similarly concluded that second language acquisition (SLA) principles and frameworks should be clearly revealed in empirical studies, in addition to exploring social interaction aspects in MALL. This is especially true when researchers attempt design-focused experiments and studies.

### **Duolingo and gamification in MALL**

Language learning can be challenging, stressful, and anxiety-inducing (Akbari, 2015; Iaremenko, 2017; Rafek, Ramli, Iksan, Harith, & Abas, 2014). Furthermore, learning a new language is time-consuming and requires perseverance to keep practicing; without adequate motivation, students are likely to quit (Han, 2015; Turan & Akdag-Cimen, 2019). Game-like elements present a motivating environment (Hanus & Fox, 2015; Kapp, 2012; Lui, 2014; Munday, 2016) that can increase language accuracy and confidence (Castañeda & Cho, 2016). Adding gamification elements to MALL platforms not only has shown to positively affect student behavior, commitment, and motivation (Huang & Soman, 2013), it has also provided a visible trace of the language learning process itself. In their systematic review on the effects of digital gamification in language learning since 2015, Dehganazadeh and Dehganazadeh (2020) found that most studies reported increases in student motivation and engagement in gamified environments.

Gamification involves using game-based mechanics, aesthetics, and game thinking to engage people, motivate action, promote learning, and solve problems (Kapp, 2012). Gamification implements game elements and ideas in contexts other than games themselves to increase commitment and influence participant behavior (Marczewski, 2013). While the topic of gamification has seen significant increase in research and practical use over the last several years, scholars still debate the definition, necessity, and value of gamification. In his piece, *'Redefining Gamification'*, Werbach (2014) advocates for gamification to be understood as 'the process of making activities more game like' (p. 266). Nah, Telaprolu, Rallapalli, and Venkata (2013) defined 5 elements of gamification. These include (1) Goal orientation or setting an objective; (2) Achievement

or the experience of success; (3) Reinforcing certain behaviors in response to outcomes; (4) Competition to encourage performance motivation, and (5) Fun orientation to ease stress and increase engagement. There is much potential for mobile applications to adhere to these elements.

Gamification in MALL is still a relatively new field (Bunchball, 2010; Dehganazadeh & Dehganazadeh, 2020; Garland, 2015; Shadiev et al., 2020). Yet, there is a fast-growing interest in the application and implications of gamified language learning. Rapid technological development further complicates both research and practice in this realm. As such, there still appear to be gaps in relation to mobile-learning platforms, design, and pedagogies used to study gamification in MALL. For example, Dehganazadeh, Fardanesh, Hatami, Talaei, and Noroozi (2019) conducted a review of gamification in the field of learning English as a Second Language (ESL) and discovered that publications report positive outcomes related to motivation, engagement, and enjoyment, but do not highlight the specific gamification aspects that benefit these learning outcomes. Specific learner characteristics were also neglected. Thus, the interactions between gamification elements and learning outcomes at the micro-level are still understudied. Dehganazadeh and Dehganazadeh (2020) found that Duolingo was the platform that researchers investigated the most, indicating a significant interest in exploring this platform for practitioners and researchers.

## Duolingo

At the time of the present systematic review, Duolingo offers approximately 95 languages to learn, and users can additionally learn in languages other than English (Duolingo About us: Approach, 2021; Viberg & Grönlund, 2012). Users start out by choosing a desired target language and can take a placement quiz if they already have some background knowledge. They set a certain amount of experience points as a daily goal and get bonuses for achieving it. Completing one lesson per day (achievement) adds one day to the Streak, which gets completely reset to zero if no lessons are completed on any given day (reinforcement). The application sometimes offers optional challenges to the user, such as maintaining the streak for several more days, comparing experience points against others in different leagues, or offering a reward on challenge completion (fun orientation and competition) (Nah et al., 2013).

The lesson system is built around specific topics such as family, food, and travel; each topic introduces some grammar and cultural concepts with very limited explanations but the lessons themselves focus mostly on introducing new vocabulary and drills. The exercises offered include translation, multiple-choice word recognition questions, and spelling. Incorrect answers are handled in two ways. Some users have a 'heart'

system, in which a certain number of mistakes leads to losing one out of 5 hearts (Figure 1). When all 5 hearts are lost, the user cannot practice until they recover at least some hearts. On some devices, mistakes trigger more repetition and drills and a slightly lower amount of experience or achievement gained upon completing the lesson. It appears that the ‘heart’ feature has been administered for some accounts but not others and is absent on the web version of the app. Mistakes are usually accompanied by a short comment and providing correct answers gives a bonus at the end of a lesson and a short form of positive feedback. In both scenarios, users can go to a forum thread dedicated to each given question and engage with other learners.

While other gamified MALL apps like Babbel and Busuu offer somewhat similar experiences, Duolingo is significantly more widespread due to the freemium business model and non-English learning options (Loewen et al., 2019). Govender and Arnedo-Moreno (2020), in their survey of gamification elements in MALL applications identified 22 gamification elements in Duolingo’s design, including (but not limited to) progress indicators (daily goal and experience points, unlocking levels), feedback (correct/incorrect answer), fixed reward schedule (experience points), time-dependent rewards (streaks), customization (buying

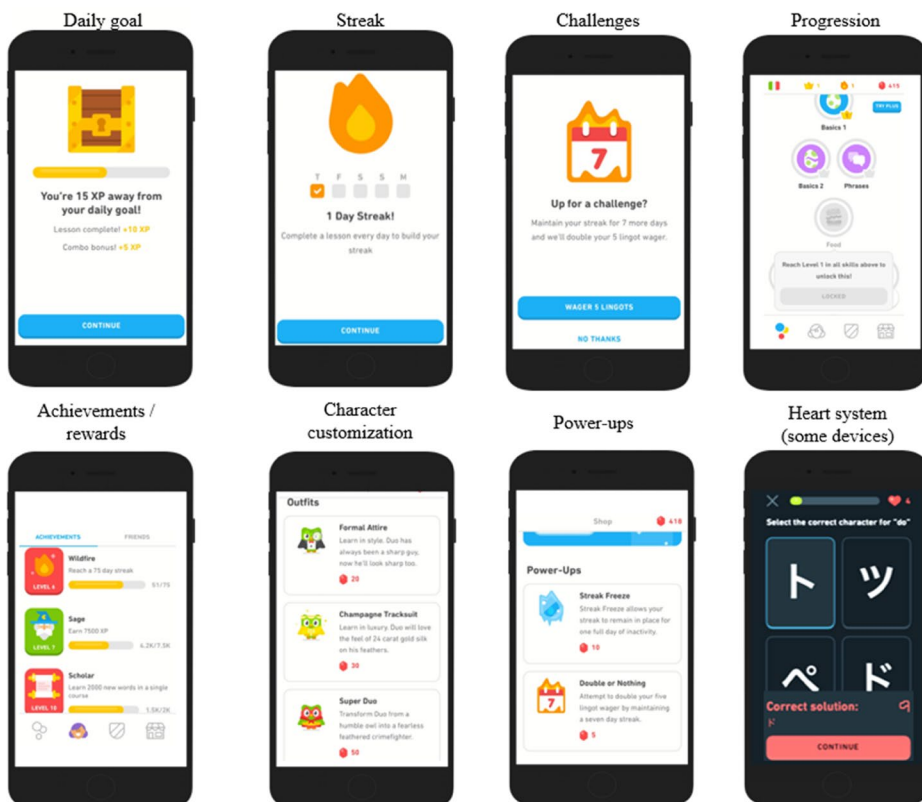


Figure 1. Duolingo's gamification elements.

outfits for the owl mascot), challenges, knowledge sharing (forums), leaderboards, badges and achievements, and virtual economy (lingots and gems) (Figure 1). They noted that Duolingo was one of the most gamified applications out of the 20 MALL applications they analyzed, further reinforcing the idea that this platform is representative of implementing gamification in this learning medium.

## **The present study and research questions**

The present systematic review aims to provide an overview of research focused on Duolingo as a representation of gamified MALL tools to address the lack of an up-to-date, detailed examination of current research trends, challenges, and directions in the field. To achieve this goal, the following research questions guided the study:

**RQ1.** What are the trends in design, tools and methodology employed in recent research on Duolingo?

**RQ2.** What are the trends in the findings from recent research on Duolingo (specifically, in terms of its impact on language performance, users' attitudes and motivation, and design features)?

**RQ3.** What are the implications of these trends for researchers and practitioners in the field of gamified MALLs?

## **Method, data collection and analysis**

### ***Article search and selection***

An adapted version of the PRISMA coding scheme guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) was applied to select articles (Moher et al., 2009). PRISMA is an internationally accepted and validated set of guidelines used for conducting systematic reviews (Bacca, Baldiris, Fabregat, Graf, & Kinshuk, 2014; Moher et al., 2015; Sønderlund, Hughes, & Smith, 2019). These guidelines can be successfully applied to a range of disciplines and subjects, including clinical medical trials, social sciences, and language learning (e.g. Shadiev et al., 2020). Other guidelines, such as the Consolidated Standards of Reporting Trials (CONSORT), are less applicable to conducting systematic reviews (Johansen & Thomsen, 2016) and/or require multiple extensions to be successfully deployed for educational studies (Grant, Mayo-Wilson, Melendez-Torres, & Montgomery, 2013). Although the guidelines are from 2009 and in the process of updating/validating, PRISMA has been formally endorsed by approximately 170 editorial organizations, including the World Health Organization (Moher et al., 2015; Sønderlund et al., 2019). By virtue of its validity, transparency,



and adaptability, we elected to use PRISMA guidelines for the foundations of this systematic review.

We searched for articles in major databases including EBSCOHost (All selected), Web of Science, Scopus, ProQuest, JSTOR, and Electronic Journal Center. Using the Boolean terms ‘Duolingo’ OR ‘Duo-Lingo’ OR ‘DuoLingo’ AND ‘language learning’ OR ‘second language acquisition’ OR ‘second language instruction’ OR ‘ESL’ OR ‘foreign language education’ OR ‘CALL’ OR ‘mobile-assisted language learning’ generated 443 results in total. We chose not to use more generic search terms such as ‘gamification’ or ‘gam\*’ because they could result in reduced precision, overly constricted results/recall, increased potential for Boolean errors, or excluding a record that should have been included (Salvador-Oliván, Marco-Cuenca, & Arquero-Avilés, 2019).

We limited the date range to 2012–2020 inclusively, as 2012 is Duolingo’s global launch year (Loeb, 2018). After removing duplicate cases (which reduced the record count from 443 to 367), two researchers read the article titles and abstracts for relevance. The collected records (Figure 2) were then screened by the research team using the following inclusion/exclusion criteria.

## Inclusion

Must be either a peer-reviewed journal article or dissertation (dissertations were included to account for newer, more recent research trends);

1. Must include Duolingo in the analysis and results;
2. Must be published in the following languages which our team could understand: English, Spanish, Russian, Portuguese, Mandarin (Simplified/Traditional Chinese), or Hindi.

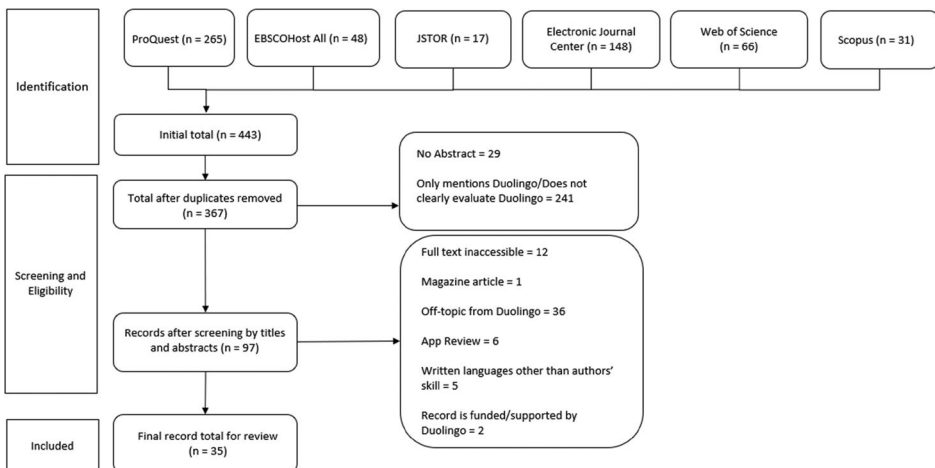


Figure 2. Flowchart of selection process.



## Exclusion

Articles that merely mentioned Duolingo, such as citing Duolingo as an example, rather than empirically analyzing the app;

3. App review articles/studies;
4. Studies that declared that they were funded or commissioned by Duolingo (to eliminate bias). For example, articles that appear on Duolingo's website: Research – Duolingo (2021) were excluded;
5. Studies published in languages outside of our team's language skills.

As stated earlier, these criteria take into account that gamification and MALL are both embedded into Duolingo, meaning that one cannot separate gamification from Duolingo as this is the primary technique (Govender & Arnedo-Moreno, 2020).

After scanning for records appropriate for our research questions, a total of 97 articles were selected. The research team then read the full texts for further adherence to the inclusion/exclusion criteria. From this initial repository of 97 pieces, 35 articles were short-listed. Excluded articles lacked focus on Duolingo or appropriate methodology. For example, Fennell, Zuo, and Lerman (2019) tested a large-scale data collection algorithm they developed that models/tracks human behavior from a statistical approach on various platforms such as Duolingo rather than empirically assessing the app. Another example is Chik (2020) where the author only collected natural interaction data among users on Duolingo from an observational ethnographic perspective instead of conducting an empirical analysis of Duolingo. Figure 2 below shows a PRISMA-style flowchart of our systematic review process.

## *The coding process and the analysis*

The research team deductively created codes for the 35 short-listed articles, reaching consensus through weekly meetings. Authors took notes based on several aspects that ran across most articles – the theories, methods and analysis used and the results. Based on these initial universal elements, we discussed how this information could be classified in a more uniform way, developing a provisional qualitative codebook featuring labels that could be ascribed to elements across the 35 articles we chose (Saldaña, 2016). 9 codes were deductively created: theoretical framework, study design, sampling method, sample characteristics, research questions, data collection, analysis, research site (country), and language(s) taught. These were further diversified into detailed subcategories (see Appendix 2; Saldaña, 2016). All categories were multi-coded to ensure the most accurate representation of the data. For

example, research articles often asked several research questions (RQs), necessitating multi-codes. All information was organized in a collaboratively edited spreadsheet. Any discrepancies/questions were brought to the group, and then discussed and resolved with all authors present.

We then tested the reliability of the coding scheme through interrater reliability analysis to ensure it would accurately capture trends in data. Subsequently, we calculated frequencies of the qualitative codes in each category using binary codes: 0 for absence and 1 for presence. This produced descriptive data regarding prevalence of certain types of literature related to Duolingo, adding a visual and computational supplement to our comprehensive review.

After iteratively developing our coding scheme using the PRISMA (Moher et al., 2009) coding scheme for literature reviews, we categorized selected studies based on theoretical framework, study design, research questions, sampling method and covariates/sample characteristics, data collection methods, and analysis type. These were further divided into specific categories (see Results and Discussion section), and dichotomously coded for absence (0) or presence (1) in a piece of literature.

The subcodes for each category were created through both inductive and deductive reasoning. The research team searched for specific information stating theoretical frameworks, sampling methods, etc., in the selected articles (inductive bottom-up approach) and, when such information was implicit, assigned subcodes based on the deductive (top-down) approach (Saldaña, 2016). The results from both approaches were considered together to create the final subcodes.

Theoretical frameworks were divided into 5 subcodes. Social frameworks were defined by their focus on social interactions as the basis of learning (e.g. the communicative focus of Rolando, Gabriela, Carolina, and Antonio (2019)); cognitive frameworks – by their focus on cognitive processes in language learning (e.g. the cognitive theory of listening comprehension in Bustillo, Rivera, Guzmán, and Acosta (2017)); social-cognitive frameworks – by their focus on the elements of both cognition and social interactions (e.g. the theory of self-efficacy and Zone of Proximal development in Rachels and Rockinson-Szapkiw (2018)). Theoretical frameworks focusing on gamification and application design were labeled as design-focused (e.g. Izzyann, Huynh, Xiong, Aziz, & Iida, 2018). When no theoretical framework was explicitly stated or could be inferred, the article was assigned a subcode ‘none’.

Study designs were coded based on the most common types of research design (Creswell, 2014): quantitative, qualitative, and mixed methods. Three more categories of subcodes were obtained in a similar way: the types of collected data (quantitative for surveys, tests and eye-tracking data; qualitative for observations and interviews; both for

the combination of the two types); sampling methods (probability sampling – simple random, systematic, stratified, cluster – and non-probability – convenience, purposive, snowball); and the type of analysis (qualitative for grounded theory and analytic induction, quantitative for descriptive, correlations, regression, ANCOVA, t-tests, SEM). Since the study designs were very diverse, we also added a comparative code that further specified each design type (Saldaña, 2016). It included quasi-experimental, pre-experimental, pretest/posttest, and experimental (Glen, 2021). This subcode always accompanied another study design subcode.

The types of research questions were coded into three subcodes: performance (learning outcomes such as vocabulary acquisition), attitudes and motivation (e.g. user and teacher experience), and design (e.g. reviewing specific features of Duolingo design). Sample characteristics were divided into demographic (student status, major, age, gender, country of origin, socioeconomic status, right/left-handed) and background (prior knowledge and performance, gaming experience) subcodes. Each category had a 'none' subcode to indicate studies in which corresponding information was not stated and not possible to infer.

Two raters applied the coding scheme to a randomly selected section of the data (14 articles or 40% of the sample considered) and categorized their content. Crosstabs were created in SPSS, and Cohen's kappa score was calculated. Kappa achieved between the two raters ( $K=0.86$ ) indicated high level of reliability. After ascertaining reliability, descriptive statistics were computed to understand the frequency or prevalence of the different types of literature published regarding the language learning application, Duolingo, which forms the primary focus of our study (see [Appendix 3](#) for the full list of values; see [Appendix 2](#) for coding examples).

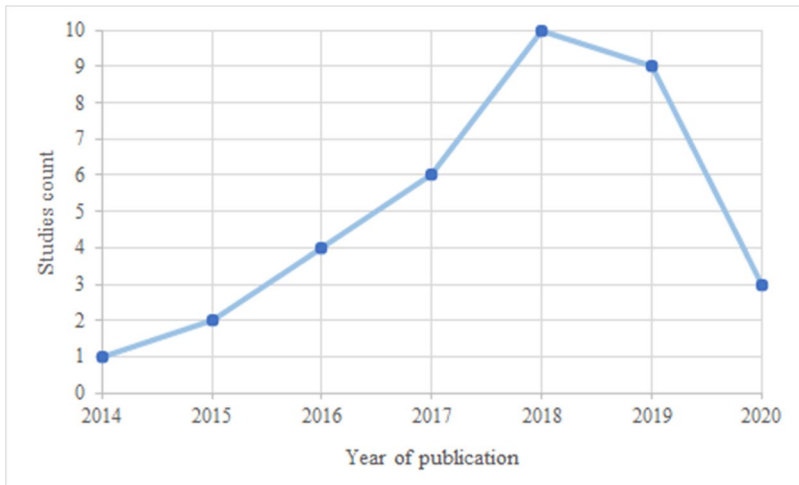
## Results and discussion

### *The analysis of the context and methodology of the reviewed studies*

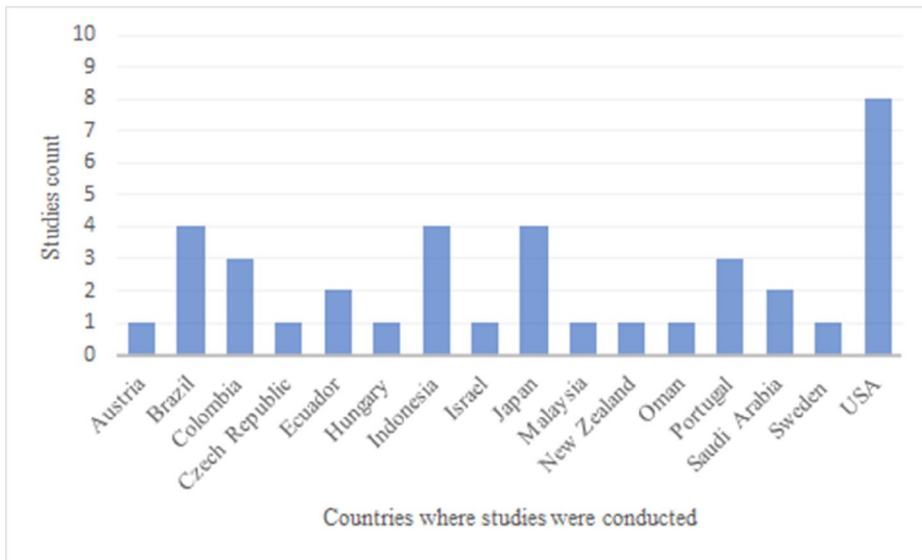
#### *Article distribution by year*

[Figure 3](#) reveals that Duolingo research articles did not appear until 2014, two years after Kapp introduced gamification and shortly after clear elements of gamification were outlined (Nah et al., 2013). The count steadily increased until peaking in 2018 ( $n=10$ ), which is in line with the findings of Dehganzadeh and Dehganzadeh (2020). After 2018, however, the count declined by only one in 2019 ( $n=9$ ). Early 2020 already has 3 studies, all of which are based in Indonesia (Ajisoko, 2020; Fadhli, Sukirman, Ulfa, Susanto, & Syam, 2020; Pramesti, 2020).

Given the numbers above, the trend appears to be an increase in analyzing Duolingo. With an increase in remote learning due to the global pandemic in 2020 (Schaffhauser, 2020), it is likely that research



**Figure 3.** Article count per year.



**Figure 4.** Article count by the country of research site.

on gamification applications in MALL will continue to rise in the 2020s as practitioners search for ways to engage students online.

#### ***Article distribution by country***

Three studies were conducted in multiple research sites simultaneously. **Figure 4** reflects this in the country distribution. Studies conducted in the USA have the highest representation (e.g. James & Mayer, 2019), followed by a three-way tie between Brazil, Indonesia, and Japan. Although Asia ( $n=9$ ), South America ( $n=9$ ), and North America ( $n=8$ )

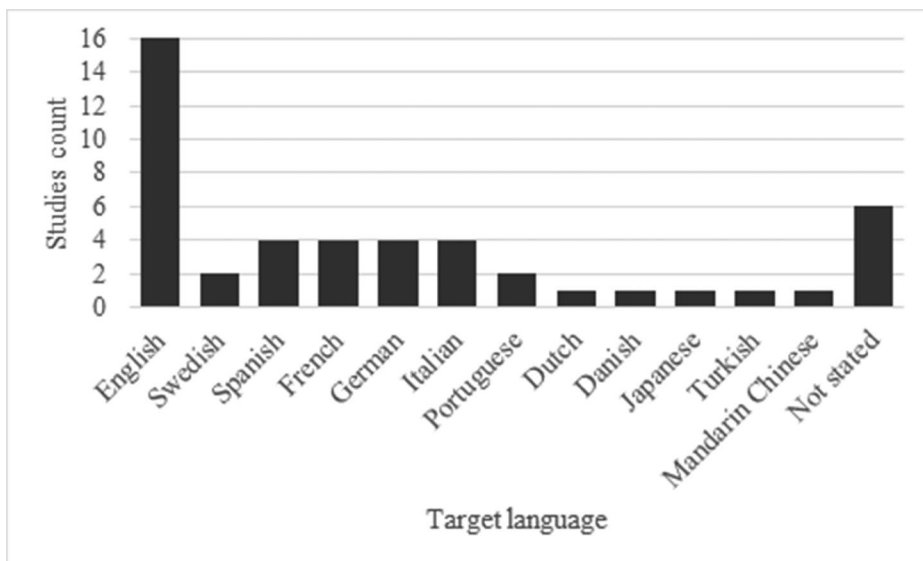
had the highest counts per continent, we detected no studies conducted in Africa. Interestingly, gamification is becoming so popular in Asia that Kennedy and Chi-Kin Lee (2018) conducted a systematic review of game-based learning and gamification research in Asia, finding that Asian educational game research is still undergoing a transition from content-focused on to context-focused design. The costs of incorporating MALL elements in educational contexts might have contributed to the global representation of studies included.

The USA might also have had the most studies because Duolingo is an American-based application. It began using English for instruction and started adding other languages relatively recently. As Duolingo adds more languages and the interest in technology-assisted language learning grows, we can expect to see the trend slowly increase internationally, as seen in Indonesia throughout 2020 (Fadhli et al., 2020). Figure 4 also indicates that Duolingo, MALL, and gamification is of increasing global interest.

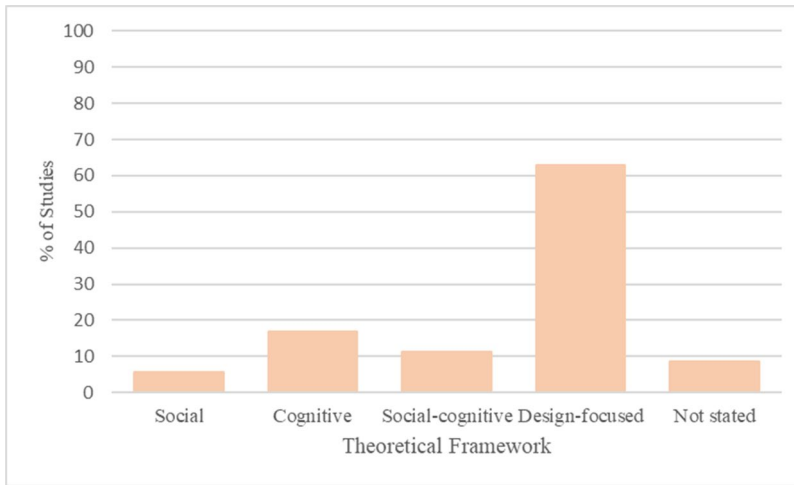
#### *Article distribution by target language*

Some studies used multiple target languages, though English is the most common language being taught in our final records ( $n=16$ ). Italian ( $n=4$ ), French ( $n=4$ ), Spanish ( $n=4$ ), and German ( $n=4$ ) are tied for second most frequent. Turkish, Mandarin Chinese, and Japanese are the lowest with one case each (Figure 5; Izzyann et al., 2018; Loewen et al., 2019; Yao, 2018).

The prevalence of English is not surprising as it is one of the most widespread global languages. Many non-English speaking countries



**Figure 5.** Article count by the target language.



**Figure 6.** The distribution of reported theoretical frameworks.

place importance on English Foreign Language (EFL), and this trend does not appear to be waning (Ghosh, 2020). There are six *Not Stated* cases, all of which are either theoretically framed in design, have design-focused RQs, or both. Target language is valuable even in design analysis since some design variables vary depending on the target language. For example, some languages on Duolingo have an interactive story feature while others do not.

### **Theoretical frameworks**

Figure 6 displays the theoretical frameworks used in each article (See Appendix 2 for coding categories). Research involved design-focused frameworks by a large margin (e.g. Huynh et al., 2018), followed by cognitive approaches (e.g. Bustillo et al., 2017).

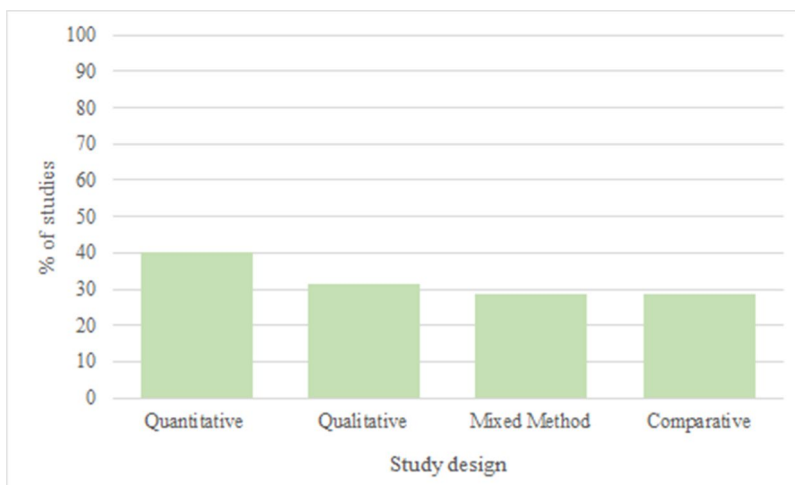
More articles included design-focused frameworks than all the other categories combined. The design-focused studies focused on trying interventions or analyzing Duolingo's features rather than exploring the application as a language learning tool. This is problematic because design features by themselves do not ensure learning, although they might create a more or less productive and motivating learning setting (Glassman, 2016). Analyzing Duolingo's design separately from their contribution to the teaching and learning process without tapping into cognitive and social theories of learning does not provide sufficient insights into the interplay of gamification, technology, teaching, and learning processes.

Additionally, given that language learning is often likened to a social learning experience (Block, 2003; Lantolf, Thorne, & Poehner, 2015;

Vygotsky, 1997), it is somewhat surprising that cognitive theories surpass both social and social-cognitive frameworks. As Teske (2017) points out, this can be attributed to perceptions of Duolingo as a language tutor, which provides an explanation for the drill-focused instructional methods within the application. The prevalence of design-focused studies might also stem from researchers' interests in exploring Duolingo's congruence with MALL, CALL, and gamification frameworks as an app rather than measuring language learning through social/cognitive theoretical frameworks. Very few studies assessed learners' vocabulary development and listening comprehension, as well as speaking and pronunciation despite Duolingo's abundant pronunciation examples. Most activities in Duolingo represent behaviorist approaches (Catania & Harnad, 1988) to language learning, without taking meaning or content into consideration. Behaviorist approaches have been widely criticized for encouraging output that is repetitive and not indicative of real-life contexts. Teske (2017) additionally points out the lack of activities centered on pragmatic or cultural skills in Duolingo, which can be problematic to instructors teaching from sociocultural perspectives. It is crucial to de-emphasize research on the app design features and to focus more on social and social-cognitive frameworks to gain a better sense of Duolingo effectiveness for authentic language learning.

### **Study design**

Reviewed articles were generally equally distributed in terms of reported study designs, with quantitative designs (e.g. Luke, Wiharja, & Sidupa, 2018) slightly more common than qualitative, mixed methods, and comparative designs (Figure 7). While quantitative and comparative (e.g.



**Figure 7.** The distribution of reported study designs.

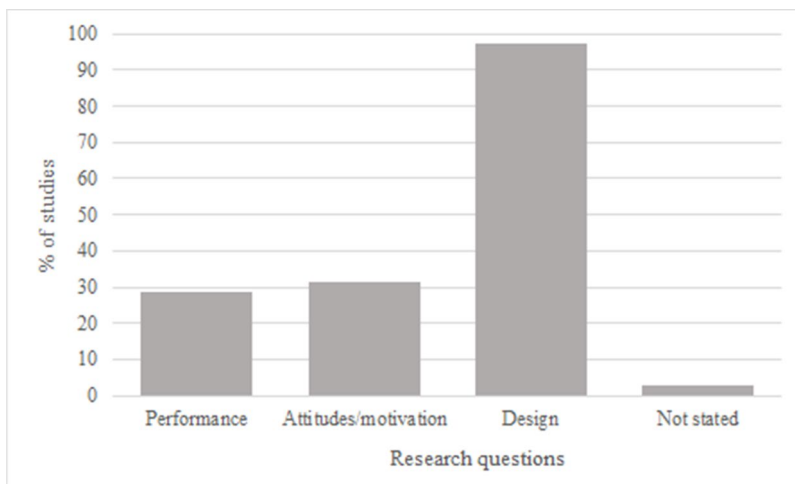


pre- and post-test and experimental) studies are valuable in assessing participants' learning outcomes, learning (and language learning in particular) is a function of many variables, of which performance is but one. Learners' background, the way they engage with the platform, their motivation for learning a language and using Duolingo, and their technology experience can influence the learning process and its result (Dörnyei, 2014). More qualitative and mixed method studies could help tackle various aspects of learning with Duolingo as a process and start developing theories of learning with gamified MALL applications in general. This aligns with Werbach's (2014) call for gamification to be understood as 'the process of making activities more game like' (p. 266). By redefining gamification as a process, there are more opportunities to focus on the types of experiences applications seek to create, with an additional more inclusive focus on the participant, the context, and the application itself.

### **Research questions type per study**

Commensurate with theoretical frameworks, Figure 8 reveals that the vast majority of articles (over 90%) presented design-focused research questions. For example, Huynh et al. (2016) design-focused question was: 'How do game elements make an effect when applying it into an education situation?' (p. 269). Fewer studies (around 30%) focused on performance and attitudes in their questions. Only one study tapped into the impact of Duolingo's use on learners' beliefs (Rachels and Rockinson-Szapkiw (2018), included self-efficacy beliefs as a variable in the analysis).

The prevalence of design-focused research questions may be caused by Duolingo's emphasis on gamification which makes it look different

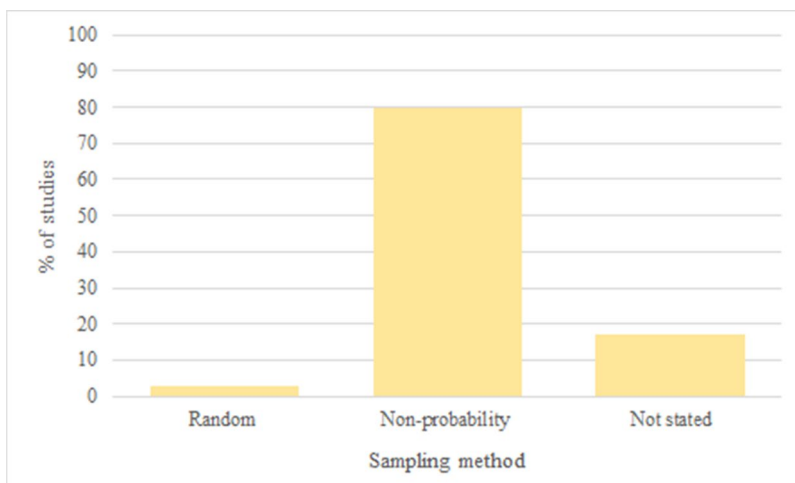


**Figure 8.** The distribution of research question types.

from traditional classroom instruction formats and the relative novelty of the application. However, as important as it is to understand the design features of and learners' attitudes toward a learning medium, focusing on these variables obscures more important questions. How does learning happen in gamified apps like Duolingo? What aspects of such platforms contribute to learning gains the most? What are the intended goals of the platform's gamified features? Do these features achieve their intended goals (e.g. do lingots in Duolingo motivate students to study more often)? Are the intended goals pedagogically sound in the first place (e.g. what is more important – fostering extrinsic motivation to study every day or encouraging deep information processing, authentic language use and communication)? Does gamification support productive learning mechanisms or is it just the 'chocolate' that covers the 'broccoli' of learning (Laurel, Crisp, & Lunenfeld, 2001)? To advance the field of MALL further, it is critical to start asking such questions.

### ***Sampling method***

As shown in Figure 9, nearly all reviewed articles ( $n=28$ ) used non-probability sampling, with most utilizing either convenience or classroom samples. Only one study used random sampling (Luke et al., 2018). Non-probability samples raise questions regarding sampling bias and margins of error (Vehovar, Toepoel, & Steinmetz, 2016) and biases in favor of students already enrolled in language-learning classes or people already pursuing language learning (Creswell & Clark, 2018). Participant students who are recruited for a Duolingo study, for example, might



**Figure 9.** The distribution of reported sampling methods.

already possess highly positive attitudes towards language learning, which might skew results (Saldaña, 2016). Without more studies employing random-sampling methodology, results cannot be generalized, limiting our knowledge about Duolingo's effectiveness and impact on language learning. Of course, random sampling can be expensive, time-consuming, and is not always possible. Nevertheless, considering that only one out of the 35 studies conducted over the span of 8 years utilized probability sampling (Luke et al., 2018; although the authors did not collect any demographic data such as age and gender), it is important to highlight the need for more studies employing probability sampling methods.

### **Data collection types**

Figure 10 shows that quantitative data collection (80% of studies) prevailed over qualitative (about 30%) and quantitative and qualitative (about 10%) data collection types. Questionnaires, surveys, and test scores were the most common instruments of data collection (e.g. Marques-Schafer & da Silva Orlando, 2018). Quantitative data is an important source of information but in isolation from the larger narrative of the learning process mediated by Duolingo, it is not enough to uncover how such learning happens. The focus on quantitative data in the reviewed Duolingo research parallels the most common type of their research questions: Does Duolingo (do X)? Does it work? Do learners like it? Is it easy to use in class? The numbers that help answer these questions do not provide sufficient insight into the micro-processes that shape learners' experiences. Utilizing more qualitative data collection methods such as interviews, observations, and participant journals (in



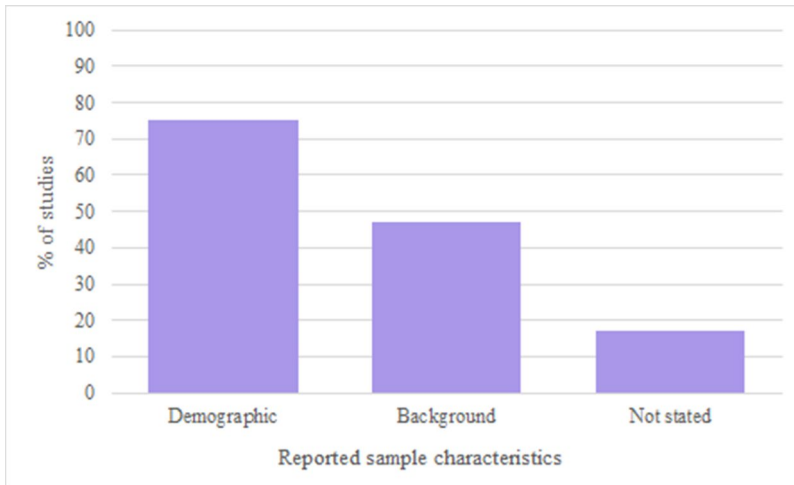
**Figure 10.** The distribution of data collection types.

conjunction with quantitative methods; Creswell, 2014) is the next step to deepening our understanding of the research on this topic.

### ***Sample characteristics***

When determining language learning outcomes, knowing learners' demographic information (such as gender, age, and major) and background (including prior knowledge, first language, knowledge of other languages, technology skills, etc.) are critical as individual differences can have a differential impact on learning processes and results (Skehan, 2014). Language background, in particular, is very important to consider (Celce-Murcia, Brinton, Goodwin, & Griner, 2010), especially for phonology, because of native-language interference (or negative transfer). For example, a student who speaks German as their first language (L1) might have an easier time learning English due to the similarities between German and English than a student whose L1 is Japanese (Kubota, 2003; Kubota & Lin, 2009). From Figure 11, most reviewed studies (75%) reported demographic information such as age, student status, and major, but less than 50% of them stated background information. Assessing learning experiences that lie on the intersection of language learning, mobile technology and gamification requires taking into account multiple demographic and background variables. Learners of different ages process language and approach language learning differently (Chen, 2014); similarly, attitudes toward and the use of technology vary among different age groups (Charness & Boot, 2009); language study strategies may also differ based on gender (Liyanage & Bartlett, 2012). Learners' past language experiences and schema shape individual learning experience (Hinkel, 2016), and technology self-efficacy predicts learners' engagement with mobile learning technologies and the outcomes of mobile learning (Menekse, Anwar, & Purzer, 2018). Because Duolingo is also a gamified platform, learners' experience with games and gamified applications might also have a role in their learning experience and perceptions.

Overlooking these and other similar demographic and background variables leads to inaccurate representation of studied target populations or, for case studies, miss important aspects of their bounded system (Creswell & Poth, 2016). Moreover, without investigating potential mediating and moderating effects of such variables, it is difficult to make claims about the effectiveness of Duolingo in the first place. Considering the costs associated with conducting random sampling or in-depth mixed method inquiries, it is imperative to glean as much background information from a sample as possible (Creswell & Clark, 2018; Saldaña, 2016).



**Figure 11.** The distribution of reported sample characteristics.

### ***Analysis type***

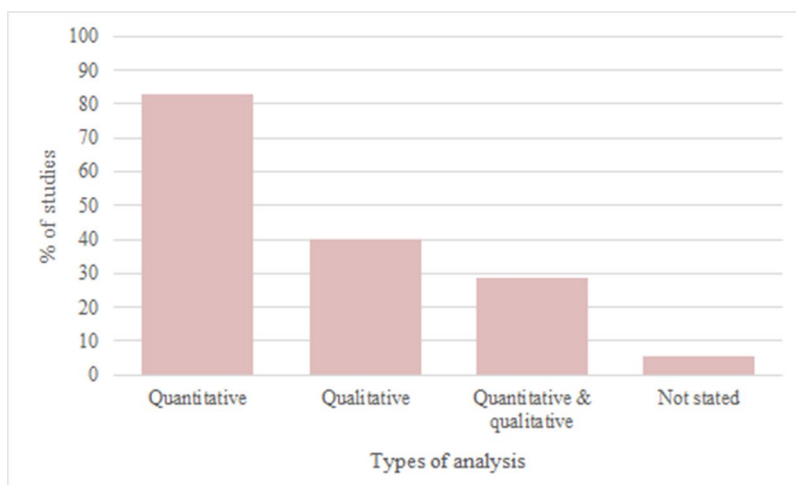
Figure 12 indicates that quantitative-only analyses were the most prevalent in selected articles (over 80%) while qualitative (40%) and the combination of quantitative and qualitative analyses (about 30%) were less common. It is important to note that articles stating both a quantitative and a qualitative analysis method (e.g. Guaqueta & Castro-Garces, 2018) were coded separately from only quantitative or only qualitative methods.

Many studies used descriptive statistics and means comparisons to analyze data. Descriptive statistics are very limited and are typically more conducive to summarizing data rather than making inferences (Creswell & Poth, 2016). This, combined with prevailing non-probability sampling methods and lack of controlling for demographic and background variables, suggests that the majority of Duolingo studies might not be accurately representing target populations. Despite this, some authors still made claims about the effectiveness of Duolingo based on descriptive or limited inferential statistics (e.g. Pramesti, 2020). As shown in the next section, such results should be interpreted with much caution and cannot be generalized.

## **The analysis of the findings of the reviewed articles**

### ***Duolingo & foreign language performance***

Overall, research findings indicate a positive correlation between the use of Duolingo and foreign language performance. For example, Abaunza, Martinez-Abad, and Conde-Rodriguez (2019) found that using Duolingo improved academic achievement in English (although the



**Figure 12.** The distribution of data analysis types.

authors did not clearly define how they measured it); Ajisoko (2020) and Guaqueta and Castro-Garces (2018) reported improvement in English vocabulary mastery, Bustillo et al. (2017) – in English listening skills, and Rolando et al. (2019) – in English communicative skills. Loewen et al. (2019) discovered a correlation between studying Turkish with Duolingo and improved performance on a holistic test assessing listening, speaking, writing, reading and lexicogrammatical knowledge of Turkish but noted that only one participant received a passing score based on the university standards (after 34 hours of study with Duolingo). Duolingo was also shown to be beneficial for the writing skills of children with Down syndrome (Salcedo, Fernandez, & Duarte, 2018). On the other hand, Rachels and Rockinson-Szapkiw (2018) reported no significant differences in elementary students' Spanish achievement and self-efficacy after 12 weeks of Duolingo use in the classroom. Similarly, no significant differences were found by James and Mayer (2019) between college students who learned Italian at home using Duolingo during 7 sessions and those who learned it using an online slideshow ( $n=64$ ).

Although most studies report a positive impact of Duolingo on various language competencies, the accuracy of their results is rather debatable. Sample sizes ranged from 1 to 44 (for studies that assessed achievement) with the exception of the study by Rachels and Rockinson-Szapkiw (2018) that had 167 participants. Many potential confounding variables (such as motivation, language background, the use of other class activities, demographics such as gender and ethnicity) were not considered when analyzing changes in participants language performance; for example, Ajisoko (2020), Rolando et al. (2019) and Guaqueta and Castro-Garces (2018) analyzed the difference between participants' pre- and post-test

scores without controlling even for prior performance, and the only study that featured a larger sample size and a more rigorous instrumentation and methodology (Rachels & Rockinson-Szapkiw, 2018) only controlled for prior performance.

With the lack of studies that employed a relatively substantial sample size, controlled for confounding variables (at least for prior performance), the claims about Duolingo's effectiveness leave much to question.

### ***Duolingo's design & usability***

Participants in the reviewed studies highlighted the interactive and engaging gamified nature of Duolingo's design (Gadanecz, 2018), chunked presentation of information (Lotze, 2019), flexibility of use (Loewen et al., 2019) and ease of access (free of charge and cross-platform; Marques-Schafer & da Silva Orlando, 2018). Some gamification elements were generally perceived positively. For example, badges and streaks were linked to students' motivation (Huynh et al., 2018); experience points and leader boards created a community-oriented learning environment (although one based in competition rather than collaboration), which participants perceived as a motivator (Loewen et al., 2019); the use of lingots also contributed to participants' engagement with the app (Marques-Schafer & da Silva Orlando, 2018). Moreover, participants appreciated the feedback on activities and mistakes (Marques-Schafer & da Silva Orlando, 2018).

At the same time, some users cited concerns over the distractions that come together with using a mobile device for learning (Gafni, Biran Achituv, & Rahmani, 2017). They found the types of activities too repetitive and over-reliant on translation and receptive skills (listening and reading) as opposed to productive skills (writing and speaking) (Loewen et al., 2019). The usefulness of feedback is limited by lack of grammatical explanations, and only few participants knew about and used Duolingo's discussion forums (Marques-Schafer & da Silva Orlando, 2018).

These findings indicate that while Duolingo's users appreciate some of its gamified design elements, those features do not make up for lack of detailed explanations, activities to practice productive language skills, and meaningful social engagement. Interestingly, very few articles mention Duolingo's discussion forums – the only aspect of the application that allows freestyle writing practice and socializing with other learners collaboratively. On mobile devices, discussions can only be accessed by clicking the forum button in the feedback to each activity, which prevents users from accessing other types of discussions; in the browser version, all discussions are accessible in a separate tab. The reasoning



behind this design decision is not apparent but it might suggest designers' focus on gamified competition rather than collaborative reflection.

### ***Duolingo & users' attitudes and motivation***

Overall, Duolingo users reported a relatively high level of satisfaction and enjoyment of using the app (e.g. Carvalho & Oliveira, 2017; James & Mayer, 2019; Marques-Schafer & da Silva Orlando, 2018), positive perceptions of Duolingo as a helpful tool (Bustillo et al., 2017), and some participants felt willing to engage in similar learning experiences in the future (James & Mayer, 2019).

However, a few participants reported more negative experiences with the application. Some students in the Loewen et al. (2019) study felt their motivation diminish throughout the learning process (34 hours of Duolingo use in total), although the authors acknowledge that the students lacked initial investment and motivation in learning Turkish. The students also found it difficult to interpret their Duolingo learning progress in terms of real-life language use and felt that their language skills were not adequate outside of the application tasks. Marques-Schafer and da Silva Orlando (2018) noted that some of their participants stopped using Duolingo because it became too repetitive and boring, and Huynh et al. (2016) cautioned against using Duolingo with more advanced learners who might find it less interesting than beginner students.

To sum up, these results paint a mixed (and sometimes negatively skewed) picture of Duolingo's effectiveness in improving foreign language performance and generating sustained engagement and motivation. Gamified presentation seems to be more enjoyable (at least initially) than typical classroom text-based presentation of learning materials; yet, once the novelty effect wears off, the gamification elements cannot compensate for the design decisions prioritizing competition over collaboration, repetition and translation over meaningful feedback and context, and passive receptive skills (listening and reading) over active productive skills (speaking and writing). Moreover, due to small samples and lack of reported demographic and background sample characteristics, generalizing these results to larger populations is not possible. In other words, 8 years after the start of research on Duolingo, we still have very little conclusive evidence about its effectiveness and role in the language learning process.

### **Conclusions**

The results of this review revealed that overall research surrounding Duolingo is design-focused, non-probability, and quantitative in nature, with pre- and post-test and questionnaire data being the dominant

choice of data collection methods. The USA was the most frequent location for Duolingo research, and English was the most common target language; however, some studies did not explicitly state a target language. Most categories, in fact, contained several studies that did not state or were unclear regarding theoretical frameworks, sample characteristics, etc. (see [Appendix 3](#)). As Shadieff et al. (2020) point out, researchers need to be as clear and descriptive as possible because readers require essential context-specific information to better design their own research and practice. Without clarity, evaluating the effectiveness of Duolingo, and therefore gamification in MALL, becomes significantly more complex (Golonka et al., 2014).

We expected that literature related to Duolingo would measure the learning outcomes and contextual understandings of language that learners achieve from using a gamified platform; however, the focus was on the creation of tools rather than the understanding of human agency associated with the use of these tools. Habermas (2006) noted that tools are but one aspect of technology. The way in which we direct our agency to gain knowledge using technology needs further inquiry to understand the implications that such mediating tools have for our learning. It is important to start asking ‘how’ and ‘why’ questions as opposed to (or at least in conjunction with) conducting ‘Does Duolingo...?’ types of inquiries. At the same time, the current evidence about the latter type of questions is limited at best due to lack of controlling for demographic and background variables, describing the study context, limited sample sizes, and non-probability sampling. Understanding the ‘Does...?’ using methodologically sound approaches is an essential first step, but it should be followed up by asking deeper, critical questions about the learning process mediated by Duolingo, its design and use in the classroom – while considering the differential impact of learners’ individual differences. While random sampling is often prohibitive due to its time and financial requirements, recruiting larger samples and describing sample characteristics and study context can be a big step forward in obtaining more generalizable and accurate results.

Finally, none of the reviewed studies attempted to challenge or reimagine the standard use of Duolingo – that is, as a largely behaviorist learning tool focused on rote learning, translation, competition and extrinsic rewards – which stands in contrast to a widely accepted view on language learning as process rooted in social and cultural exchange (Lantolf et al., 2015). The only feature of Duolingo rooted in the sociocultural view of language learning – discussion forums – was barely mentioned in only few articles, and the platform was used ‘as is’ instead of incorporating it into larger contextual activities. While technology design creates certain constraints and shapes the learning setting, it does not have to be the sole factor dictating how

we use it (Glassman, 2016). Therefore, experimental studies can focus on using the platform in communicative and social ways, and design-focused studies can highlight potential activities conducive to such integrations and investigate how design and gamification elements can support it.

## Future implications and suggestions

Based on the identified trends and weaknesses in the reviewed studies on Duolingo, we propose a set of guidelines that can direct researchers to deeper, more meaningful and methodologically sound inquiries. These guidelines can also be used to study other gamified MALL tools.

### Theoretical framework

- Decide what theoretical perspectives will inform your inquiry and clearly state both specific theories and more general theoretical frameworks they fit in (when possible). For example, if you are investigating the use of Duolingo from the perspective of Zone of Proximal Development (Vygotsky, 1978), make the larger framework (sociocultural theory) clear to the reader.
- Explain why the theory is suitable for the chosen study context.

### Sampling method and sample characteristics

- When possible, opt to use random sampling and larger sample sizes.
- If random sampling is not possible, focus on describing the context of the study and sample characteristics in as much detail as possible. Report sample characteristics even in studies focused on the application design or participants' attitudes.
- Consider including the following variables to control for (the list is not exhaustive):
  - Demographic information: gender, age, major, occupation, socioeconomic status
  - Background information: first language, other languages known, prior performance in the target language (TL), TL level, technology self-efficacy, academic self-efficacy (and other types of self-efficacy if appropriate), types and length of participants' engagement with the target platform, motivation for studying the TL, experience with games and gamified platforms.

### Study design, data collection and analysis

- When possible, choose mixed methods methodology for your study, using quantitative and qualitative methods to capture various aspects of the learning process.

- Have a comparable control condition.
- Consider the effect of the intervention length (if applicable).
- Alternatively, consider supplementing your qualitative or quantitative study design with other types of data (e.g. include open-ended questions in surveys, conduct observations along with measuring performance through tests, etc.).
- Focus on inferential statistics (when possible) while controlling for possible confounding, mediating or moderating variables. When focusing on descriptive statistics, make sure to limit generalizations and state appropriate limitations of the results.

### Research questions

The following potential questions can direct future research on the topic.

1. Duolingo efficiency/performance questions:
  - Does Duolingo improve language performance outcomes controlling for participants' demographic and background variables as compared to learners who participate in comparable instruction without Duolingo?
    - i. If yes, what aspects of Duolingo contribute to improving learning outcomes the most?
    - ii. How do outcomes vary based on how learners engage with the platform?
  - How does Duolingo impact language performance when used as a stand-alone vs. a supplemental mode of instruction?
  - How does Duolingo impact language learning performance in informal language learning/
2. Students' motivation and attitudes:
  - How do learners of different backgrounds (e.g. language levels, ages, genders, different levels of technology self-efficacy, different experiences of gaming and gamified applications) perceive the use and usefulness of Duolingo in their learning?
  - Why do learners and teachers choose to use this platform and what do they expect from it? Do their expectations align with what Duolingo promises and how it is designed?
  - How does using Duolingo shape/affect learners' motivation, especially over an extended period of time?
3. Duolingo's design and gamified features:
  - Does Duolingo's design align with the type of learning and outcomes promised by the Duolingo developers?

- What types of learning does the app design facilitate and what is the role of gamification elements in the facilitation?
- What are the intended goals of the platform's gamified features and can they facilitate deep, meaningful learning?
- What gamification features have the most impact on learning outcomes (if at all)?

## Contributions

The present study aimed to increase our understanding surrounding gamification in MALL through a systematic review of Duolingo and contributes several salient points to current literature. First, this review had a thorough range for article selection (from the release of Duolingo to early 2020) and included articles not published in English. Many systematic reviews do not include non-English entries despite the prevalence of EFL in non-English journals. Second, the last review-oriented study on Duolingo was published in 2017 (Nushi & Eqbali, 2017). In the field of MALL, a small timespan can produce a trove of new insights, owing to rapid developments in the Information Age. Our comprehensive typology of research is well-founded in this rapidly developing landscape, and provides a peek into future possibilities associated with approaches to MALL on Duolingo. Third, this review critically evaluates the methodology and the results of the selected articles and provides an actionable set of guidelines for researchers to conduct deeper, more critical, and more methodologically sound inquiries in the field of gamified MALL applications.

## Limitations

Even though we were able to review articles in languages other than English, there might still have been relevant publications in other languages. Another limitation was the decision to observe Duolingo as a representation of MALL and gamification. There are other mobile apps that incorporate gamification aspects, such as Babbel, but since Duolingo is viewed as one of the most representative gamification-in-MALL platforms (Dehganazadeh & Dehganazadeh, 2020), it was chosen as the focus for this study. At the time of this review, this is a highly comprehensive and salient review that presents a path to extend language-learning inquiry; due to fast technological advancements, Duolingo is likely to keep changing. This might render some results outdated.

## Disclosure statement

No potential conflicts of interest were reported by the authors.

## Compliance with ethical standards

**Research involving human participants and/or animals:** No animals/human participants were involved

**Informed consent:** No human participants were involved.

## Notes on contributors

*Mitchell Shortt* is a doctoral candidate at The Ohio State University. Mitchell is also a first-generation college student of color. He earned his M.S. in International Management and Marketing from the University of Sheffield (United Kingdom) and subsequently his M.A. in Teaching English to Speakers of Other Languages from Saint Michael's College. His research interests include student motivation in digital environments, teaching strategies, game-based learning, and computer-assisted language learning.

*Shantanu Tilak* is a PhD student of Educational Psychology at The Ohio State University. His work focuses on creating a cybernetic conceptualization of psychology in the Information Age, aiming to facilitate an understanding of ideas related to the mind and skill acquisition through the metaphor of machines applied to social systems.

*Irina Kuznetcova* loves games, technology, teaching and language learning - and her research lies on the intersection of all of them. In her early academic career, she focused on using Multi-User Virtual Environments (MUVEs) to create Deweyan, democratic classrooms, as well as the social network theory. Later, she shifted to researching mobile Virtual Reality and visuospatial thinking. Her dissertation combined all of the above, looking into developing visuospatial thinking training through Deweyan serious games in middle school students. As a linguist by her undergraduate training and an avid language learner, she seeks to apply her serious games and technology expertise in the field of language teaching and learning, particularly focusing on language learning communities and language learning application design.

*Bethany Martens* is an instructor and PhD student of Teaching and Learning: Foreign, Second, and Multilingual Language Education at The Ohio State University. She has an M.A. in TESOL Education from MidAmerica Nazarene University and has taught English as a Second Language in South Korea and China for over 7 years. Her current research interests include Teacher Education, CALL, and Linguistic Landscape pedagogy in TESOL education.

*Babatunde Akinkuolie* is a Doctoral Student of Learning Technologies at the Ohio State University. He earned his bachelor's degree in Information Technology from the University of Belize in Belize City, Belize. He subsequently received his master's degree in Computer Engineering from National Chiao Tung University, Taiwan and an M.B.A from Southern New Hampshire University, New Hampshire. His research interests include Computer-Aided Instruction, Distance Learning, Student-centered learning in digital environments and Technology Integration in Teaching and Learning.

## ORCID

Mitchell Shortt  <http://orcid.org/0000-0001-5594-0356>

Shantanu Tilak  <http://orcid.org/0000-0001-5264-2652>

Irina Kuznetcova  <http://orcid.org/0000-0001-6843-1905>

Bethany Martens  <http://orcid.org/0000-0003-3255-6954>

Babatunde Akinkuolie  <http://orcid.org/0000-0001-7287-686X>

## References (\* denotes an included record for review)

- \*Abaunza, G., Martinez-Abad, F., & Conde-Rodriguez, M. J. (2019, October). *Web applications in the EFL class in contexts rural school Colombian: Aplicaciones web en la clase de EFL en contextos de escuela rural colombiana*. In Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality, TEEM 2019, pp. 1–6. doi:10.1145/3362789.3362879
- \*Ahmed, H. (2016). Duolingo as a bilingual learning app: A case study. *Arab World English Journal*, 7(2), 255–267. doi:10.24093/awej/vol7no2.17
- \*Ajisoko, P. (2020). The use of Duolingo apps to improve English vocabulary learning. *International Journal of Emerging Technologies in Learning*, 15(07), 149–155. doi:10.3991/ijet.v15i07.13229
- Akbari, Z. (2015). Current challenges in teaching/learning English for EFL learners: The case of Junior High School and High School. *Procedia - Social and Behavioral Sciences*, 199, 394–401. doi:10.1016/j.sbspro.2015.07.524
- \*Al-Sabbagh, K. W., Bradley, L., & Bartram, L. (2019). Mobile language learning applications for Arabic speaking migrants – A usability perspective. *Language Learning in Higher Education*, 9(1), 71–95. doi:10.1515/cercles-2019-0004
- \*Alsaif, S. A. M., & Farhana, D. D. (2019). Vocabulary learning through Duolingo mobile application: Teacher acceptance, preferred application features and problems. *International Journal of Recent Technology and Engineering*, 8(2s9), 79–85. doi:10.35940/ijrte.B1017.0982S919
- Bacca, J., Baldiris, S., Fabregat, R., Graf, S., & Kinshuk. (2014). Augmented reality trends in education: A systematic review of research and applications. *Journal of Educational Technology & Society*, 17(4), 133–149. <http://www.jstor.org/stable/jeductechsoci.17.4.133>
- Block, D. (2003). *The social turn in second language acquisition*. Edinburgh: Edinburgh University Press. doi:10.3366/j.ctvxcrwd8
- Bunchball, I. (2010). Gamification 101: An introduction to the use of game dynamics to influence behavior [White paper]. Retrieved from <http://jndglobal.com/wp-content/uploads/2011/05/gamification1011.pdf>
- Burston, J. (2014). MALL: The pedagogical challenges. *Computer Assisted Language Learning*, 27(4), 344–357. doi:10.1080/09588221.2014.914539
- \*Bustillo, J., Rivera, C., Guzmán, J. G., & Acosta, L. R. (2017). Benefits of using a mobile application in learning a foreign language. *Sistemas y Telemática*, 15(40), 55–68. doi:10.18046/syt.v15i40.2391
- \*Campos, A. A. M. (2017). *Adopting smartphone applications for second language acquisition: Investigating readiness and acceptance of mobile learning in two higher education institutions* (Doctoral dissertation). Universidade NOVA de Lisboa (Portugal), ProQuest Dissertations Publishing.
- \*Carvalho, M., & Oliveira, L. (2017). Emotional design in web interfaces. *Observatorio (OBS\*)*, 11(2), 14–34. doi:10.15847/obsOBS1122017905
- Castañeda, D. A., & Cho, M. H. (2016). Use of a game-like application on a mobile device to improve accuracy in conjugating Spanish verbs. *Computer Assisted Language Learning*, 29(7), 1195–1204. doi:10.1080/09588221.2016.1197950
- Catania, A.C., & Harnad, S. (1988). *The operant behaviorism of B. F. Skinner: Comments and consequences*. New York, NY: Cambridge University Press.
- Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). New York, NY: Cambridge University Press.



- Charness, N., & Boot, W. R. (2009). Aging and information technology use: Potential and barriers. *Current Directions in Psychological Science*, 18(5), 253–258. doi:10.1111/j.1467-8721.2009.01647.x
- Chen, M. L. (2014). Age differences in the use of language learning strategies. *English Language Teaching*, 7(2), 144–151. doi:10.5539/elt.v7n2p144
- Chik, A. (2020). Humorous interaction, language learning, and social media. *World Englishes*, 39(1), 22–35. doi:10.1111/weng.12443
- \*Corrêa, C. R. (2019). A gamificação e o ensino/aprendizagem de segunda língua: Um olhar investigativo sobre o Duolingo. *Revista Linguagem & Ensino*, 22(4), 1020–1039. doi:10.15210/rle.v22i4.16471
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.
- Creswell, J. W., & Clark, V. L. P. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Thousand Oaks, CA: Sage Publications.
- \*Davis, D. J. (2015). *Mapping student activity data to the visual design of online learning environments* (Doctoral dissertation). Georgetown University.
- Dehganazadeh, H., & Dehganazadeh, H. (2020). Investigating effects of digital gamification-based language learning: A systematic review. *Journal of English Language Teaching and Learning*, 12(25), 53–93.
- Dehganazadeh, H., Fardanesh, H., Hatami, J., Talaei, E., & Noroozi, O. (2019). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning*, 1–24. doi:10.1080/09588221.2019.1648298
- Dörnyei, Z. (2014). *The psychology of the language learner: Individual differences in second language acquisition*. New York, NY: Routledge.
- Duolingo. (2021). *About us: Mission*. Retrieved from <https://www.duolingo.com/info>
- Duolingo. (2021). *About us: Approach*. Retrieved from <https://www.duolingo.com/approach>
- Duolingo Help Center. (2020). *What is Duolingo?* <https://support.duolingo.com/hc/en-us/articles/204829090-What-is-Duolingo->
- \*Eisenlauer, V. (2014). Multimodality in mobile-assisted language learning. *Communications in Computer and Information Science*, 479, 328–338. doi:10.1007/978-3-319-13416-1\_32
- \*Fadhli, M., Sukirman, S., Ulfa, S., Susanto, H., & Syam, A. R. (2020). Gamifying children's linguistic intelligence with the Duolingo app: A case study from Indonesia. In S. Papadakis, & M. Kalogiannakis (Eds.), *Mobile learning applications in early childhood education* (pp. 122–135). Hershey, PA: IGI Global. doi:10.4018/978-1-7998-1486-3.ch007
- Fennell, P. G., Zuo, Z., & Lerman, K. (2019). Predicting and explaining behavioral data with structured feature space decomposition. *EPJ Data Science*, 8(1), 1–27. doi:10.1140/epjds/s13688-019-0201-0
- \*Gadanecz, P. (2018). *The nature of positive emotions via online language learning*. In 2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), 197–204. IEEE. doi:10.1109/CogInfoCom.2018.8639965
- \*Gafni, R., Biran Achituv, D., & Rahmani, G. (2017). Learning foreign languages using mobile applications. *Journal of Information Technology Education: Research*, 16(1), 301–317. doi:10.28945/3855

- Garland, C. M. (2015). *Gamification and implications for second language education: A meta analysis* (Master's thesis). Cloud State University, St. Retrieved from [https://repository.stcloudstate.edu/cgi/viewcontent.cgi?article=1043&context=engl\\_etds](https://repository.stcloudstate.edu/cgi/viewcontent.cgi?article=1043&context=engl_etds)
- Ghosh, I. (2020, February 15th). Ranked: The 100 most spoken languages around the world. *Visual Capitalist*. Retrieved from <https://www.visualcapitalist.com/100-most-spoken-languages/>
- Glassman, M. (2016). *Educational psychology and the internet*. New York, NY: Cambridge University Press.
- Glen, S. (2021). Experimental design. *StatisticsHowTo.com*. Retrieved from <https://www.statisticshowto.com/experimental-design/>
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70–105. doi:10.1080/09588221.2012.700315
- Govender, T., & Arnedo-Moreno, J. (2020). *A survey on gamification elements in mobile language-learning applications*. In Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'20), October 21–23, 2020, Salamanca, Spain. ACM, New York, NY. doi:10.1145/3434780.3436597
- Grant, S. P., Mayo-Wilson, E., Melendez-Torres, G. J., & Montgomery, P. (2013). Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PLoS One*, 8(5), e65442. doi:10.1371/journal.pone.0065442
- \*Guaqueta, C. A., & Castro-Garces, A. Y. (2018). The use of language learning apps as a didactic tool for EFL vocabulary building. *English Language Teaching*, 11(2), 61–71. doi:10.5539/elt.v11n2p61
- Habermas, J. (2006). Political communication in media society: Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. *Communication Theory*, 16(4), 411–426. doi:10.1111/j.1468-2885.2006.00280.x
- Han, Y.J. (2015). Successfully flipping the ESL classroom for learner autonomy. *NYS TESOL Journal*, 2(1), 98–109.
- Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, 80, 152–161. doi:10.1016/j.compedu.2014.08.019
- Hinkel, E. (Ed.). (2016). *Teaching English grammar to speakers of other languages*. New York, NY: Routledge.
- Huang, W. H. Y., & Soman, D. (2013). Gamification of education. *Report Series: Behavioural Economics in Action*, 1–29.
- \*Huynh, D., Zuo, L., & Iida, H. (2016). Analyzing gamification of “Duolingo” with focus on its course structure. In *International Conference on Games and Learning Alliance* (pp. 268–277). Cham: Springer. doi:10.1007/978-3-319-50182-6\_24
- \*Huynh, D., Zuo, L., & Iida, H. (2018). *An assessment of game elements in language-learning platform Duolingo*. In 2018 4th International Conference on Computer and Information Sciences (ICCOINS) (pp. 1–4). doi:10.1109/ICCOINS.2018.8510568
- Hwang, W. Y., Chen, C. Y., & Chen, H. S. (2011, October). *Facilitating EFL writing of elementary school students in familiar situated contexts with mobile devices*. In 10th World Conference on Mobile and Contextual Learning, 18–21, October 2011, Beijing, China: MLearn2011 Conference Proceedings, 15–23.

- Iaremenko, N. (2017). Enhancing English language learners' motivation through online games. *Information Technologies and Learning Tools*, 59(3), 126–133. doi:[10.33407/itlt.v59i3.1606](https://doi.org/10.33407/itlt.v59i3.1606)
- \*Iida, H. (2017). *Serious games discover game refinement measure*. In 2017 International Conference on Electrical Engineering and Computer Science (ICECOS), 1–6. doi:[10.1109/ICECOS.2017.8167112](https://doi.org/10.1109/ICECOS.2017.8167112)
- \*Izzlyann, F., Huynh, D., Xiong, S., Aziz, N., & Iida, H. (2018). Comparative study: Case study in analyzing gamification between mind-snacks and Duolingo. *JP Journal of Heat and Mass Transfer*, SV2018(1), 101–106. doi:[10.17654/HMSI118101](https://doi.org/10.17654/HMSI118101)
- \*James, K. K., & Mayer, R. E. (2019). Learning a second language by playing a game. *Applied Cognitive Psychology*, 33(4), 669–674. doi:[10.1002/acp.3492](https://doi.org/10.1002/acp.3492)
- Johansen, M., & Thomsen, S. F. (2016). Guidelines for reporting medical research: A critical appraisal. *International Scholarly Research Notices*, 2016, 1–7. doi:[10.1155/2016/1346026](https://doi.org/10.1155/2016/1346026)
- Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. San Francisco, CA: John Wiley & Sons.
- Kennedy, K., & Chi-Kin Lee, J. (2018). *Routledge international handbook of schools and schooling in Asia*. New York, NY: Routledge.
- Kondo, M., Ishikawa, Y., Smith, C., Sakamoto, K., Shimomura, H., & Wada, N. (2012). Mobile assisted language learning in university EFL courses in Japan: Developing attitudes and skills for self-regulated learning. *ReCALL*, 24(2), 169–187. doi:[10.1017/S0958344012000055](https://doi.org/10.1017/S0958344012000055)
- Kubota, R. (2003). New approaches to gender, class, and race in second language writing. *Journal of Second Language Writing*, 12(1), 31–47. doi:[10.1016/S1060-3743\(02\)00125-X](https://doi.org/10.1016/S1060-3743(02)00125-X)
- Kubota, R., & Lin, A. M. (Eds.). (2009). *Race, culture, and identities in second language education: Exploring critically engaged practice*. New York, NY: Routledge.
- Kukulska-Hulme, A., & Viberg, O. (2018). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology*, 49(2), 207–218. doi:[10.1111/bjet.12580](https://doi.org/10.1111/bjet.12580)
- Lantolf, J., Thorne, S. L., & Poehner, M. (2015). Sociocultural theory and second language development. In B. van Patten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 207–226). New York: Routledge.
- Laurel, B., Crisp, D. G., & Lunenfeld, P. (2001). *Utopian entrepreneur*. Cambridge, MA: MIT Press.
- Liyanage, I., & Bartlett, B. J. (2012). Gender and language learning strategies: Looking beyond the categories. *The Language Learning Journal*, 40(2), 237–253. doi:[10.1080/09571736.2011.574818](https://doi.org/10.1080/09571736.2011.574818)
- Loeb. (2018, June 22nd). When Duolingo was young: The early years. *Vator, Inc.* Retrieved from <https://vator.tv/news/2018-06-22-when-duolingo-was-young-the-early-years>
- \*Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, 31(3), 293–311. doi:[10.1017/S0958344019000065](https://doi.org/10.1017/S0958344019000065)
- \*Lotze, N. (2019). Duolingo: Motivating students via m-homework. *TESOL Journal*, 11(1), 1–3. doi:[10.1002/tesj.459](https://doi.org/10.1002/tesj.459)
- Lui, S. (2014). Use of gamification in vocabulary learning: A case study in Macau. In *4th CELC Symposium Proceedings*, 90–97.
- \*Luke, J. Y., Wiharja, C. K., & Sidupa, C. (2018). *The effectiveness level and positive values of practicing translation using mobile app DUOLINGO for Indonesian freshmen*

- students. In Proceedings of the 2nd International Conference on E-Society, E-Education and E-Technology, 26–29. doi:[10.1145/3268808.3268834](https://doi.org/10.1145/3268808.3268834)
- \*Macalister, J. (2017). Language learning principles and MALL: Reflections of an adult learner. *The TESOLANZ Journal*, 25, 12–24. <https://doi.org/10.1177/0033688218771385>.
- Marczewski, A. (2013). *Gamification: A simple introduction*. Surrey, UK: Andrzej Marczewski (self-published via [kdp.amazon.co.jp](http://kdp.amazon.co.jp)).
- \*Marques-Schafer, G., & da Silva Orlando, A. A. (2018). Languages learning conceptions and Duolingo: A critical analysis on its proposals and learners experiences. *Texto Livre - Linguagem e Tecnologia*, 11(3), 228–251.
- \*Martinelli, M. (2016). *Effectiveness of online language learning software (Duolingo) on italian pronunciation features: A case study* (Doctoral dissertation). University of Hawaii, Manoa. Retrieved from [https://shareok.org/bitstream/handle/11244/54566/Martinelli\\_okstate\\_0664M\\_14766.pdf?sequence=1&isAllowed=y](https://shareok.org/bitstream/handle/11244/54566/Martinelli_okstate_0664M_14766.pdf?sequence=1&isAllowed=y)
- Menekse, M., Anwar, S., & Purzer, S. (2018). Self-efficacy and mobile learning technologies: A case study of CourseMIRROR. In C. Hodges (Ed.) *Self-efficacy in instructional technology contexts*. Cham: Springer. doi:[10.1007/978-3-319-99858-9\\_4](https://doi.org/10.1007/978-3-319-99858-9_4)
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. doi:[10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1–9. doi:[10.1186/2046-4053-4-1](https://doi.org/10.1186/2046-4053-4-1)
- Munday, P. (2016). The case for using DUOLINGO as part of the language classroom experience. *RIED. Revista Iberoamericana de Educación a Distancia*, 19(1), 83–101. doi:[10.5944/ried.19.1.14581](https://doi.org/10.5944/ried.19.1.14581)
- Nah, F. F. H., Telaprolu, V. R., Rallapalli, S., & Venkata, P. R. (2013, July). Gamification of education using computer games. In *International Conference on Human Interface and the Management of Information*, 99–107. Springer.
- \*Northrop, L., & Andrei, E. (2019). More than just word of the day: Vocabulary apps for English learners. *The Reading Teacher*, 72(5), 623–630. doi:[10.1002/trtr.1773](https://doi.org/10.1002/trtr.1773)
- \*Notaro, G. M., & Diamond, S. G. (2018). *Development and demonstration of an integrated EEG, eye-tracking, and behavioral data acquisition system to assess online learning*. In Proceedings of the 10th International Conference on Education Technology and Computers, 105–111. doi:[10.1145/3290511.3290526](https://doi.org/10.1145/3290511.3290526)
- Nushi, M., & Egbali, M. H. (2017). Duolingo: A mobile application to assist second language learning. *Teaching English with Technology*, 17(1), 89–98. <https://files.eric.ed.gov/fulltext/EJ1135889.pdf>
- \*Pramesti, A. S. (2020). Students' perception of the use of mobile application Duolingo for learning English. *International Journal of Scientific and Technology Research*, 9(1), 1800–1804. Retrieved from <http://www.ijstr.org/final-print/jan2020/Students-Perception-Of-The-Use-Of-Mobile-Application-Duolingo-For-Learning-English.pdf>
- \*Rachels, J. R., & Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. *Computer Assisted Language Learning*, 31(1-2), 72–89. doi:[10.1080/09588221.2017.1382536](https://doi.org/10.1080/09588221.2017.1382536)
- Rafek, M. B., Ramli, N. H. L. B., Iksan, H. B., Harith, N. M., & Abas, A. I. B. C. (2014). Gender and language: Communication apprehension in second language learning. *Procedia - Social and Behavioral Sciences*, 123, 90–96. doi:[10.1016/j.sbspro.2014.01.1401](https://doi.org/10.1016/j.sbspro.2014.01.1401)

- Research – Duolingo. (n.d). *Research - Duolingo*. Retrieved from <https://research.duolingo.com/>
- \*Rolando, A. P. P., Gabriela, R. E. A., Carolina, M. S. E., & Antonio, A. R. M. (2019). Incidencia de Duolingo en el desarrollo de las habilidades comunicacionales verbales del idioma ingles a nivel de educación superior. *European Scientific Journal ESJ*, 15(16), 29–44. doi:10.19044/esj.2019.v15n16p29
- \*Salcedo, S. P., Fernandez, F. H., & Duarte, J. E. (2018). Mejora de la habilidad de escritura en inglés en niños con Síndrome de Down con el apoyo de nuevas tecnologías. *Revista Espacios*, 39(10), 18–31.
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- \*Salomão, R. C. S., Rebelo, F., & Rodríguez, F. G. (2015). Defining Personas of university students for the development of a digital educational game to learn Portuguese as a foreign language. *Procedia Manufacturing*, 3, 6214–6222. doi:10.1016/j.promfg.2015.07.941
- Salvador-Oliván, J. A., Marco-Cuenca, G., & Arquero-Avilés, R. (2019). Errors in search strategies used in systematic reviews and their effects on information retrieval. *Journal of the Medical Library Association*, 107(2), 210–221. doi:10.5195/jmla.2019.567
- Schaffhauser, D. (2020). Remote learning will continue growing over the next three years. *The Journal*. Retrieved from <https://thejournal.com/articles/2020/10/29/remote-learning-will-continue-growing-over-the-next-three-years.aspx>
- Shadiev, R., Liu, T., & Hwang, W. Y. (2020). Review of research on mobile-assisted language learning in familiar, authentic environments. *British Journal of Educational Technology*, 51(3), 709–720. doi:10.1111/bjet.12839
- Silver, L. (2019, February 5th). Smartphone ownership is growing rapidly around the world, but not always equally. *Pew Research Center*. Retrieved from <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/>
- Skehan, P. (2014). *Individual differences in second language learning*. New York, NY: Routledge.
- \*Škuta, P., & Kostolányová, K. (2016). The Inclusion of gamification elements in the educational process. *DIVAI*, 2016, 421–429. doi:10.13140/RG.2.1.3026.7766
- Sønderlund, A. L., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, 50(5), 2594–2618. doi:10.1111/bjet.12720
- Teske, K. (2017). Duolingo. *CALICO Journal*, 34(3), 393–401. doi:10.1558/cj.32509
- Toland, S. H., Mills, D. J., & Kohyama, M. (2016). Enhancing Japanese university students' English-language presentation skills with mobile-video recordings. *The JALT CALL Journal*, 12(3), 179–201. doi:10.29140/jaltcall.v12n3.207
- Turan, Z., & Akdag-Cimen, B. (2019). Flipped classroom in English language teaching: A systematic review. *Computer Assisted Language Learning*, 33(5-6), 590–606. doi:10.1080/09588221.2019.1584117
- Vehovar, V., Toepoel, V., & Steinmetz, S. (2016). Non-probability sampling. In C. Wolf, D. Joye, T. W. Smith, & Y. Fu, (Eds.), *The Sage handbook of survey methods* (pp. 329–345). Thousand Oaks, CA: Sage Publications.
- Viberg, O., & Grönlund, Å. (2012). *Mobile assisted language learning: A literature review*. In *11th World Conference on Mobile and Contextual Learning*. Helsinki, Finland.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

- Vygotsky, L. S. (1997). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky: Vol. 1. Problems of general psychology* (pp. 39–285). New York, NY: Plenum Press. doi:[10.1007/978-1-4613-1655-8](https://doi.org/10.1007/978-1-4613-1655-8)
- Werbach, K. (2014, May). (Re)defining gamification: A process approach. In A. Spagnolli, L. Chittaro, & L. Gamberini (Eds.), *International conference on persuasive technology* (pp. 266–272). New York, NY: Springer. doi:[10.1007/978-3-319-07127-5\\_23](https://doi.org/10.1007/978-3-319-07127-5_23)
- Wu, W. H., Wu, Y. C. J., Chen, C. Y., Kao, H. Y., Lin, C. H., & Huang, S. H. (2012). Review of trends from mobile learning studies: A meta-analysis. *Computers & Education*, 59(2), 817–827. doi:[10.1016/j.compedu.2012.03.016](https://doi.org/10.1016/j.compedu.2012.03.016)
- \*Yao, W. (2018). *Narrative casual learning: Explore a new way of learning a language through a Game: Mandarin mystery* (Doctoral dissertation). Northeastern University.

Appendix 1: Table of Primary Records

Citation	Theoretical framework	Study design	Sampling	Sample characteristics	RQs	Data collection (instruments and methods)	Analysis	Language Taught
Abounza, G., Martínez-Abad, F., & Conde-Rodríguez, M. J. (2019, October)	design	comparative, quantitative	non-probability	demographic	performance, design, attitudes and motivation	quantitative/mixed	quantitative	English
Ajisoko, P. (2020)	Cognitive	comparative, quantitative	non-probability	demographic, background	performance, design	quantitative/mixed	quantitative	English
Al-Sabbagh, K. W., Bradley, L., & Bartram, L. (2019)	design	mixed methods	non-probability	demographic, background	attitudes and motivation, design	quantitative/mixed, qualitative	quantitative, qualitative	Swedish (for Arabic speakers)
Alsaif Sarah Ali, M., & Farhana Deris, D. (2019)	design	mixed methods	non-probability	demographic	attitudes and motivation, performance, design	quantitative/mixed	quantitative, qualitative	English
Bahjet Essa Ahmed, H. (2016)	NA	mixed methods	NA	demographic, background	design, performance	quantitative/mixed, qualitative	quantitative, qualitative	Spanish, English
Bustillo, J., Rivera, C., Guzmán, J. G., & Acosta, L. R. (2017)	Cognitive	quantitative, comparative	non-probability	demographic, background	design, performance	quantitative/mixed	quantitative	English
Campos, A. A. M. (2017)	design	mixed methods	non-probability	demographic	design, performance	quantitative/mixed	quantitative	English
Carvalho, M., & Oliveira, L. (2017)	design, Social	quantitative	non-probability	demographic	design, attitudes and motivation	quantitative/mixed	quantitative	English
Corréa, C. R. (2019)	design	qualitative	NA	NA	performance, design	NA	None	French, German, English
Davis, D. J. (2015)	Social Cognitive design	mixed methods	non-probability	NA	attitudes and motivation, design	quantitative/mixed	quantitative	Unspecified
Eisenlauer, V. (2014)		qualitative	NA	NA	performance, design	quantitative/mixed	quantitative, qualitative	German
Fadhli, M., Sukirman, S., Ulfa, S., Susanto, H., & Syam, A. R. (2020)	cognitive, design	qualitative	non-probability	demographic	design, attitudes and motivation, performance	qualitative	qualitative	English
Gadanecz, P. (2018)	design, Social Cognitive	mixed methods	non-probability	demographic, background	design, attitudes and motivation	quantitative/mixed, qualitative	quantitative	Swedish, Italian, Portuguese, Dutch, German, Danish, French, Spanish
Gafni, R., Achituv, D. B., & Rachmani, G. J. (2017)	design	qualitative	non-probability	demographic, background	design, attitudes and motivation, design	quantitative/mixed	quantitative, qualitative	Unspecified
Guqueta, C. A., & Castro-Garces, A. Y. (2018)	Cognitive , design	mixed methods	non-probability	demographic	design	quantitative/mixed	quantitative, qualitative	English

(Continued)





### Appendix 1: (Continued)

Huynh, D., Zuo, L., & Iida, H. (2016)	design	quantitative, comparative	non-probability	background	design	quantitative/mixed	quantitative	Unspecified
Huynh, D., Zuo, L., & Iida, H. (2018)	design	Quantitative, comparative	non-probability	demographic, background	design	quantitative/mixed	quantitative	Spanish, French, Italian
Iida, H. (2017)	design	Quantitative, comparative	non-probability	background	design	quantitative/mixed	quantitative	Unspecified
Izzyann, F., Huynh, D., Xiong, S., Norshakirah, A., & Iida, H. (2018)	design	comparative, quantitative	non-probability	background	design	quantitative/mixed	quantitative	Japanese
James, K. K., & Mayer, R. E. (2019)	Social Cognitive design	comparative, quantitative mixed methods	non-probability	demographic, background	design	quantitative/mixed	quantitative	Italian
Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019)	design	quantitative	non-probability	demographic	design	qualitative	quantitative, qualitative	Turkish
Lotze, N. (2020)	design	qualitative	non-probability random	demographic	design	quantitative/mixed	quantitative	English
Luke, J. Y., Wiharja, C. K., & Sidupa, C. (2018)	design	quantitative	NA	NA	design	qualitative	quantitative	French
Macalister, J. (2017)	design	qualitative	non-probability	demographic, background	attitudes/ motivation, design	quantitative/mixed	quantitative, qualitative	Unspecified
Marques-Schafer, G., & da Silva Orlando, A. A. (2018)	Cognitive	mixed methods	non-probability	demographic, background	design	quantitative/mixed	quantitative	Italian
Martinelli, M. (2016)	Cognitive	quantitative	NA	NA	design	quantitative/mixed	quantitative	English
Northrop, L., & Andrei, E. (2019)	NA	Quantitative	non-probability	demographic, background	design	quantitative/mixed	quantitative	German
Notaro, G. M., & Diamond, S. G. (2018)	Cognitive, design	qualitative	non-probability	demographic, background	NA	qualitative	quantitative, qualitative	English
Pranesti (2020)	social Cognitive	comparative, quantitative, comparative	non-probability	background	performance, attitudes/ motivation, design	quantitative/mixed	quantitative	Spanish
Rachels, J. R., & Rockinson-Szapkiw, A. J. (2018)	social Cognitive	quantitative, comparative	non-probability	demographic	design	quantitative/mixed	quantitative	English
Rolando, A. P. P., Gabriela, R. E. A., Carolina, M. S. E., & Antonio, A. R. M. (2019)	design	qualitative	Non-probability	demographic	design	qualitative, quantitative/ mixed	quantitative	English
Salcedo, S. P., Fernandez, F. H., & Duarte, J. E. (2018)	design	qualitative	non-probability	demographic, background	design	qualitative	quantitative	Portuguese
Salomão, R. C. S., Rebelo, F., & Rodriguez, F. G. (2015)	design	qualitative	non-probability	demographic, background	design	qualitative	quantitative	Unspecified
Skuta, P., & Kostolányová, K. (2016)	design	qualitative mixed methods	NA	NA	design	qualitative	None	Mandarin Chinese
Yao, W. (2018)	NA	mixed methods	non-probability	demographic, background	design, attitudes/ motivation	quantitative/mixed	quantitative, qualitative	

Appendix 2: Coding Legend

Frameworks	Examples	Study Design	Examples	Sampling	Examples	Sample Characteristics	Examples
Social	Emotional Design Vygotksy (sociocultural) Csikzentmihalyi (flow) Connectivism	Mixed Qualitative Quantitative Comparative	Mixed Case study Correlational Quasi-experimental, Pretest/Posttest Experimental	Probability sampling	Simple random Systematic Stratified Cluster	Demographics	Student status Major/work Age Gender
Cognitive	Vocabulary learning Listening comprehension Theory of perception Emergentist Bandura (social-cognitive)			Non-probability sampling	Convenience Purposive Snowball		Country of Origin Socioeconomic status Right/Left handed Down Syndrome diagnosis Prior knowledge Gaming experience Game-design experience
Social Cognitive	Constructivism User centered design (including testing and prototyping)					Background	
Design	Technology Acceptance model 1 & 2 L2 human-computer interaction Multimodality Gamification Game Refinement Theory Language Learning Apps (LLAs) CALL/MALL Blended Learning						

Research Questions	Examples	Data Collection Methods	Examples	Analysis	Examples
Performance Attitudes/ motivation	Learning outcomes User experience Teacher experience	Quantitative/ Mixed	App data Surveys/ Questionnaires Test scores Eye-tracking Brain-mapping Interview Observation Auto-ethnography	Qualitative	Grounded Theory Analytic Induction Wireframe analysis Descriptives Correlation Regression ANCOVA T-test SEM
Study Behaviors/Habits Design	Parent opinion Behavioral App design	Qualitative		Quantitative	Eye-tracking GRT formula

### Appendix 3: Component Statistics (Multi-coded Components)

*Note.* Percent (%) was calculated by dividing the number of records where category presence was detected by the total number of records collected (ex., for Design research questions:  $34/35 = 97.14$ ). We used 0 and 1 for absence and presence of a category respectively (see page 13). Multi-codes were necessary for categorization (See [Appendix 1](#)).

Component	Subcomponents	Percent (%)
Theoretical framework	Social framework	5.71
	Cognitive framework	20.00
	Social cognitive framework	11.43
	Design focused framework	68.57
	None	8.57
Study design	Quantitative design	40.00
	Qualitative design	31.43
	Mixed Method design	28.57
	Comparative	28.57
Research questions	Performance	28.57
	Attitudes/motivation	31.43
	Design	97.14
	None	2.86
Sampling method	random sampling	2.86
	non-probability sampling	80.00
	None	17.14
Sample characteristics	Demographic	75.00
	Background	47.22
	None	17.14
Data collection type	quantitative	80.00
	qualitative	28.57
	Both	11.11
	None	2.85
Types of analysis	Quantitative	82.85
	Qualitative	40.00
	Both	28.57
	None	5.72