Studies in Second Language Learning and Teaching
Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz
SSLLT 0 (0). 2024. 1-43. Published online: 28.10.2024
https://doi.org/10.14746/ssllt.39090
http://pressto.amu.edu.pl/index.php/ssllt

Developing explicit and implicit L2 knowledge with Duolingo's input-based features: Case studies of L2 Spanish learning

Anthony Brandy
Indiana University, Bloomington, United States
https://orcid.org/0000-0001-5675-3683
ajbrandy@iu.edu

Phil Hiver ✉
Florida State University, Tallahassee, United States
https://orcid.org/0000-0002-2004-7960
phiver@fsu.edu

Abstract

In this single-case study with a quasi-experimental design, we set out to examine the role of two input-based features of the language learning app Duolingo on learners' development of implicit and explicit second language (L2) knowledge. We investigated the effect of Duolingo's *Stories* feature and *Podcast* feature on two Spanish-L2 learners' performance on a battery of implicit and explicit L2 knowledge tests. Once a week, over a four-week period using these input-based features, both participants repeatedly performed oral narrative tasks, timed and untimed grammaticality judgment tasks, and metalinguistic knowledge tests. The participant assigned to the Podcast condition showed meager gains in productive implicit knowledge (measured by oral narrative tasks) but decreased in receptive explicit knowledge (measured by untimed grammaticality judgment tasks). The participant assigned to the Stories condition showed a large increase in receptive implicit knowledge of Spanish (measured by timed grammaticality judgment tasks). Neither participant showed meaningful gains in the measure of productive

explicit knowledge (measured by metalinguistic knowledge tasks). Although there was a slight increase in receptive explicit knowledge for the participant using Stories, the data from both participants discount the idea that app-based mobile-assisted language learning (MALL) prioritizes receptive explicit L2 knowledge, at least when considering these novel input-based features of Duolingo. These findings add nuance to previous research showing that the primary outcome of MALL app use is receptive explicit L2 knowledge.

Keywords: Duolingo; explicit L2 knowledge; implicit L2 knowledge; podcast feature; stories feature; Spanish learning

1. Introduction

Over the past several years, second language (L2) learning research has seen a large uptick in the number of studies investigating the benefits of computer- and mobile-assisted language learning (CALL and MALL, respectively) applications (apps) for L2 development (Chwo et al., 2018; Kukulska-Hulme et al., 2017). As recent advances in assistive technology and generative artificial intelligence (GenAI) indicate, language learning in the digital domain is poised for exponential growth (Meurers, 2021; Ruiz et al., 2021). However, important questions remain about how pedagogically sound such CALL and MALL platforms are, and to what extent they incorporate empirical insights from instructed second language acquisition (ISLA) research (Ruiz et al., 2024; Ziegler et al., 2017). This study adds to this growing body of scholarship by looking at the effect Duolingo has on implicit and explicit L2 knowledge development.

Of various features in Duolingo designed to assist language learners, the most often researched is its *Learn* feature (Jiang et al., 2020). The Learn feature is often assumed to be the default basis of Duolingo's curriculum (e.g., Crowther et al., 2017; Isbell et al., 2017; Rachels & Rockinson-Szapkiw, 2018). Indeed, new course pairings between the first language (L1) and any other language (LX) are only added to Duolingo's catalog when sufficient materials have been developed for the Learn feature. This feature contains a series of controlled-output exercises that prompt the learner to translate sentences, pair L2 words with their first language counterparts, and orally repeat L2 sentences verbatim (e.g., Savvani, 2019; Teske, 2017). However, this focus on the Learn feature in existing research is not without reason: Duolingo's other features are a recent development, having only been officially released for use in the past few years (Jiang et al., 2020; Savvani, 2019).

This inevitable focus on the primary feature available has created a blind spot in the CALL and MALL literature surrounding Duolingo. For instance, one of

the primary criticisms of Duolingo, its theoretical underpinnings, and associated learning techniques is the lack of meaningful input available to the user (Crowther et al., 2017; Krashen, 2014; Loewen et al., 2019). However, at least two recent additional features of Duolingo offer the user more meaningful input than the Learn feature: the *Stories* feature and the *Podcast* feature. Both features provide a potential antidote to these critiques since they focus almost entirely on providing the user with discourse-level input that is focused on meaning rather than the more form-focused instruction (FFI) found in the Learn feature. These under-researched features may provide Duolingo users with input that was previously unavailable to them, likely having implications for the type of L2 knowledge that second language learners can develop as a result of app use (Chen et al., 2022).

Across many domains of additional language learning and use, empirical work increasingly points to complex and pervasive interconnections between the various types of language knowledge that learners develop. Explicit knowledge is concrete, conscious, and verbalizable. Implicit knowledge, on the other hand, is subsymbolic, and remains tacit and unconscious (DeKeyser, 2017; Suzuki, 2017). While Duolingo has been theorized to primarily produce more explicit knowledge, no study has directly investigated this. Most, if not all, studies on Duolingo thus far have conducted surface-level analyses of learners' production, measuring coarse-grained learner gains using standardized language tests that are intended for placement or diagnostic purposes (e.g., Isbell et al., 2017; Loewen et al., 2019; Vesselinov & Grego, 2012), rather than examining the type(s) of L2 knowledge developed by the learner. Since different L2 knowledge types underlie distinct aspects of L2 performance (Ellis & Roever, 2021), it is important to examine more than surface features of language production to investigate the effects that Duolingo use may have on users' L2 development. The purpose of this study[1] was to investigate the effects of Duolingo's input-based features (e.g., the Stories and Podcast features) on the development of implicit and explicit L2 knowledge over the course of four weeks of use.

2. Literature review

2.1. App-based language learning and Duolingo

Broadly speaking, analyses and evaluations of MALL's potential for language learning have demonstrated the benefits of gamification through a social constructivist

[1] This study is based on the original, unpublished graduate thesis of the first author which can be accessed at: https://diginole.lib.fsu.edu/islandora/object/fsu:776815

view of learning (Chwo et al., 2018; Kukulska-Hulme & Viberg, 2018) and a more individualized approach to language learning (Teske, 2017). Despite these proposed benefits, one theoretical evaluation of Duolingo's Learn feature using an instructed SLA framework concluded that it lacked focus on meaning, authenticity, interaction, and positive impact (Crowther et al., 2017).

Several domain-general (i.e., non-SLA) studies on language learning app use have also compared performance of participants using the platform with students in university settings, attempting to quantify and compare the rates of learning between the two groups. Vesselinov and Grego's (2012) preliminary findings for Duolingo appeared to show the app was highly efficient, with learners achieving an average rate of increase of 8.1 points on the test for every one hour of study. The researchers extrapolated, based on this rate of increase, that 34 hours of Duolingo study on average were equivalent to one college course.

However, Vesselinov and Grego's (2012) Duolingo study has been critiqued by some instructed SLA scholars for its use of the Web Based Computer Adaptive Placement Exam (WebCAPE), a coarse-grained test which primarily measures explicit knowledge, as a measurement of learning gains (Crowther et al., 2017). In his critical appraisal, Krashen (2014) also disputed the claims made by Vesselinov and Grego (2012), arguing that the wide variability among participant gains actually resulted in a much lower points-per-hour-of-study rate than reported by the original authors. Nevertheless, subsequent Duolingo-funded research made stronger connections between Duolingo use and achievement in conventional university-level L2 courses, finding that participants who used Duolingo as their exclusive tool of second language instruction for an average of 141 hours performed roughly equivalently in tests of receptive language skills as university students in their fourth semester of the target L2 (Jiang et al, 2020).

As we elaborate below, however, the type of L2 knowledge typically tested on foreign language collegiate-level tests is receptive explicit knowledge, a type of L2 knowledge relied on for initial construction of form-meaning connections, concurrent processing, online encoding of input, and planned production (Rebuschat, 2013). This finding is in line with Loewen et al. (2020), who argued that receptive explicit knowledge is precisely the primary outcome of MALL app use due to the limited opportunities for productive skills and communicative ability practice, a finding also supported by Rachels and Rockinson-Szapkiw (2018), whose third- and fourth-grade Spanish-L2 participants (N = 79) reported learning gains similar to those of a control group (N = 88) on measures of receptive explicit knowledge.

Independent research on MALL, although meager, has provided insight into the effectiveness of these platforms in influencing language learning. The majority of this research has found a positive effect on the development of receptive knowledge of the L2 (Antonia & Pierpaolo, 2020; Burston, 2015; Lin &

Lin, 2019; Munday, 2016; Toto & Limone, 2019), especially in terms of vocabulary learning (Ajisoko, 2020; Al-Sabbagh et al., 2018).

2.2. Implicit and explicit L2 knowledge in MALL

Although the nature of the L2 knowledge developed by app-based language learning is central to the claims regarding its efficacy, little, if any, research has investigated this issue directly (Ruiz et al., 2024). Explicit knowledge takes the form of concrete, conscious, and verbalizable L2 knowledge and underlies the initial construction of form-meaning connections, online encoding of input, and planned production. Implicit knowledge, because it remains subsymbolic and unconscious, contributes to rapid, effortless retrieval of previously learned language or to spontaneous production (Andringa & Rebuschat, 2015). The distinction between explicit and implicit knowledge is not categorical (Godfroid, 2022). For instance, there are instances where learners rely on conscious knowledge about the language even if their access to it is rapid or automatic, and this is referred to as automatized explicit knowledge (Suzuki & DeKeyser, 2017). Additionally, the consensus in the field is that certain L2 targets (i.e., rule-like regularities vs. item-based exemplars) may be more responsive to certain types of learning than to others (i.e., explicit learning = deliberate, planned, and conscious vs. implicit learning = unplanned, unintentional, and without awareness) if the type of processing that those language targets rely on differs. L2 development ultimately requires both explicit and implicit modes of processing.

There are various tests and measures of L2 knowledge used by researchers to assess the quality of L2 learners' implicit and explicit knowledge of a target language (Rebuschat, 2013). These tests aim to gauge salient characteristics of L2 knowledge that learners possess including the degree of awareness (conscious awareness vs. intuitive awareness), its systematicity (systematic vs. inconsistent), its accessibility (conscious processing vs. automatic processing), whether it can be articulated (verbalizable vs. non-verbalizable), and when that knowledge is accessed (during planning and monitoring vs. during fluent performance) (Ellis & Roever, 2021). Some of the more established measures of L2 knowledge, ordered from more implicit to more explicit, include elicited imitation tasks, oral narrative tasks, timed and untimed grammaticality judgements, and metalinguistic knowledge tests (e.g., Ellis, 2005).

To date, no study has employed a full range of established measures of L2 knowledge in examining the efficacy of Duolingo use for language learning. Instead, experimental studies involving Duolingo use have more generally examined learners' controlled production (e.g., Rachels & Rockinson-Szapkiw, 2018) and

improvement in collegiate foreign language tests (e.g., Isbell et al., 2017; Krashen, 2014; Loewen et al., 2019; Vesselinov & Grego, 2012), interpreting these data as indicators of overall language proficiency. In contrast to these studies, Loewen et al. (2020) examined the effects of MALL on the development of oral communicative ability, a capacity thought to be supported by implicit L2 knowledge, finding modest gains specifically for users of Babbel. While oral communicative ability is by no means a one-to-one indicator of implicit knowledge and may reflect some automatized explicit knowledge (Kukulska-Hulme & Viberg, 2018; Loewen et al., 2020), this finding is related to questions regarding the type(s) of L2 knowledge developed through app-based MALL use since both explicit and implicit knowledge are constituents of learners' competence (Philp, 2009).

It is widely assumed that explicit knowledge is a byproduct of explicit processing/learning mechanisms, and that both implicit and explicit knowledge can result from implicit processing (Kang et al., 2019; Leow, 2019). Therefore, Duolingo's pivot from the more form-focused Learn feature to the more meaning-focused Stories and Podcast features may coincide with meaningfully different types of knowledge developed by users of these input-based features. No study has yet measured implicit and explicit knowledge as they are affected by Duolingo's various features. However, previous research on the affordances of Duolingo's features (e.g., Crowther et al., 2017) has evaluated only Duolingo's Learn feature, using an instructed language learning evaluation framework and concluding that Duolingo "prioritizes explicit knowledge, devoid of contextual meaning" (p. 35). We now turn to a brief description of these input-based features and the theoretical support for their use in language learning.

2.3. Modified and unmodified input: The Stories feature

Modified and unmodified input refer broadly to the types of language exposure that learners receive, each with a slightly different purpose (Long, 2020). Modified input refers to language that has been altered in one of several ways (e.g., language that is simplified or elaborated) with the aim of making the language more accessible to learners, enhancing their comprehension, or directing their attention to specific features of the language. For example, input enhancement (Sharwood Smith, 1993) is one way of modifying language to deliberately make certain language forms more noticeable to learners, thereby increasing the likelihood that those features will be processed and learned. Unmodified input, however, refers to unaltered language that is presented in ways that reflect how language is used in real-life contexts (e.g., authentic language). Its purpose is to enable learners to develop the ability to understand and use the language as it is naturally spoken and written.

While Duolingo's Learn feature provides users with some input enhancement in the form of glossing and underlining (Teske, 2017), Duolingo's other, less-studied features have potential to provide input that is more meaningful and comprehensible (Chen et al., 2022). The addition of features that provide more meaningful input to the learner also relates to Nation's (2007) concept of the *four strands* which posits that a language course should have an equal balance of "meaning-focused input, meaning-focused output, language-focused learning and fluency development" (p. 2). In both cases, the presence of additional input for the learner is fundamentally beneficial to their development (Piske & Young-Scholten, 2008).

The Stories feature consists of sets of short stories that provide visual and auditory input to the app user. With its simplified lines of dialogue and frequent comprehension and vocabulary checks, the Stories feature provides more pre-modified input than the Podcast feature. However, the short unmodified monologues in the podcast episodes are scripted and interspersed with context and linguistic explanations given in the L1 (Jiang et al., 2020); this input is not strictly unmodified or authentic because it is still designed with language learners in mind (Long, 2020).

Due to its relative novelty, the effect that Duolingo's Stories feature has on users' L2 learning has not been independently tested. Still, there exists a body of research on the effects of modifying input to make it more comprehensible to L2 users. According to Long (2020), one type of pre-modified input is elaborated modified input. This type of input, whether spoken or written, aims to add meaningful cues, elaboration, and intentional redundancy as a way of compensating for, rather than avoiding, certain grammatical constructions and low frequency lexical items, as is often the practice in simplified texts. Elaborated modified input employs the kinds of discourse moves and informational aids that improve comprehensibility and sustain conversation. This approach is also evident in Duolingo's Stories feature, which includes frequent spontaneous comprehension checks intermittently throughout the storyline that are typically characteristic of dialogic interactions. After being presented with several lines of dialogue, the learner is prompted with mid-story exercises that check for the reader's comprehension of the plot and word meanings.

Elaborated modified input has been found to promote incidental processing mechanisms, a key component of implicit knowledge development (see Long, 2020 for one review of this topic). Moreover, experimental studies on conventionally modified input have demonstrated large effects on incidental vocabulary learning. For example, one random-effects meta-analysis of studies reported an average increase of 1.05 standard deviations for participants that received modified spoken input (de Vos et al., 2018). It may be that the provision of additional meaningful input for the learner leads to the development of more

implicit knowledge (Reinders & Ellis, 2009) or automatized explicit knowledge (Mostafa & Kim, 2021).

2.4. Unmodified input: The Podcast feature

Studies of Podcast creation and consumption in the L2 classroom have generally found positive benefits from the incorporation of podcasts as a new source of input (e.g., Cross, 2014) and as opportunities for output (e.g., Lord, 2008). This niche in the literature, however, suffers from a false equivalency between two types of podcasting: student creation of podcasts and student consumption of podcasts. Research on the latter type of podcasting has been "scant" (Lomicka & Lord, 2010, p. 5). While some articles have demonstrated that the creation of podcasts by students in classroom settings has been beneficial for linguistic gains (e.g., pronunciation of fossilized features) (Fouz-González, 2019), below we focus on research related more specifically to podcast consumption.

Rosell-Aguilar (2007) proposed that listening to podcasts can be understood from a constructivist view of learning "where an individual representation of knowledge is constructed through active exploration, observation, processing and interpretation" of information encountered (2007, p. 479). Studies following Rosell-Aguilar's (2007) and Schmidt's (2008) pedagogical evaluations of podcasting for language learning have found positive benefits for podcast consumption related to L2 learners' improvement across language skills (Hasan & Hoon, 2013), positive podcast listener responses (O'Bryan & Hegelheimer, 2007; Rosell-Aguilar, 2013; Rostami et al., 2017), and self-regulated language learning (Cross, 2014). These initial indications of the potential of podcasts as a source of meaningful input provide a basis for the claim that the provision of additional, minimally modified or unmodified input can positively affect the development of implicit and even explicit knowledge.

2.5. The current study

Virtually no work has been done to investigate the input-based features of Duolingo and the types of L2 knowledge that language learners develop as a result of using these app features, and this constitutes a knowledge gap in the field. In essence, these under-explored features of meaningful input and associated meaning-based learning techniques provide a potential antidote to many of the criticisms of MALL when it is dominated by the default FFI features. Additionally, most, if not all, studies on Duolingo thus far have analyzed learners' language

production as the target outcome rather than examining the type(s) of L2 knowledge developed by the learner. Since different L2 knowledge types underlie distinct aspects of L2 performance, research is needed to examine more than just features of language production and to investigate the effects that Duolingo use has on users' L2 knowledge development. Consequently, the primary aim of this study was to investigate the effects of Duolingo's input-based features (e.g., the Stories and Podcast features) on the development of learners' implicit and explicit L2 knowledge over the course of four weeks of use. In the present study, the following research questions were investigated:

RQ1: What gains in implicit and explicit L2 knowledge are made by a learner over the course of one month of consistent use of the Duolingo Stories feature?
RQ2: What gains in implicit and explicit L2 knowledge are made by a learner over the course of one month of consistent use of the Duolingo Podcast feature?
RQ3: What are the differences in explicit and implicit L2 knowledge gains between these two learners?
RQ4: How stable is the knowledge that the learner in each condition (i.e., Stories, Podcast) develops?

We hypothesized that the learner assigned to the Podcast condition would experience greater gains in implicit knowledge compared to the participant assigned to the Stories condition because the primary focus in Duolingo's Podcast features is meaning, not form (Hasan & Hoon, 2013; Jiang et al., 2020). We also hypothesized that the Podcast participant's explicit knowledge would be affected very little or not at all due to the minimal focus on metalinguistic features and consciousness raising in this feature (Schmidt, 2008). Considering the translation exercises used at regular intervals in the Stories feature – and that translation exercises prioritize explicit knowledge development over implicit knowledge development – we hypothesized that the learner assigned to the Stories condition would experience a sharper increase in explicit knowledge than the Podcast participant (Crowther et al., 2017; Teske, 2017). Finally, we hypothesized that the Stories participant would make gains in implicit knowledge as well as explicit knowledge, though less pronounced than those of the learner assigned to the Podcast condition (Jiang et al., 2020).

3. Method

The case study design we adopted here places emphasis on understanding mechanisms and processes (Yin, 2018), and as such it can offer rich and detailed

insights into these two participants' experience with the specific features of this language learning app. However, the potential to make predictions and generalizations is limited in such a design due to heterogeneity between the cases and the apparent lack of experimental control. We attempted to mitigate this both through our purposeful sampling detailed below and by conducting a robust comparative analysis across the two learners' cases, evident for example in our research questions, to identify commonalities and differences with broader relevance. In this way, despite some participant heterogeneity, an analysis of two cases can contribute to a more nuanced understanding of app-based language learning.

3.1. Participants

Participation in this study was voluntary and uncompensated. Using targeted criterion sampling we recruited two participants who: (a) had prior experience learning the target L2 Spanish (but no study-abroad experience), (b) expressed interest in pursuing a sustained course of L2 study, and (c) had ample experience with autonomous language learning through various MALL apps. These criteria thus increased the probability of participant fidelity over the 4-week course of use. We assigned one participant to each condition (Cir = Podcast feature; Eric = Stories feature). The names are pseudonyms chosen by the participants. Their demographic information is summarized in Table 1. The Podcast participant, Cir, had had several years of formal Spanish instruction prior to this study, while the Stories participant, Eric, self-reported taking only formal high school-level Spanish classes but had been studying Spanish independently for the six months prior to the study using various self-study modalities.

Table 1 Participant demographic summary

Participant pseudonym	Age	Gender	Languages learned	Condition assigned
Cir	32	Male	15	Podcasts
Eric	22	Male	5	Stories

3.2. Instruments

Taking cues from previous work and existing methodological guidance (Ellis, 2005; Rebuschat, 2013), this study used a battery of four validated tests of L2 knowledge. The two implicit knowledge tasks were a timed grammaticality judgment task (TGJT) and an oral narrative task (ONT), while the two explicit knowledge tasks were an untimed grammaticality judgment task (UGJT) and a

metalinguistic knowledge task (MKT). All tests of L2 knowledge were administered on a computer.

3.2.1. Grammaticality judgment tasks

Benchmarking all data elicitation materials on Ellis (2005), the GJTs consisted of 64 sentences that were divided evenly between grammatical and ungrammatical utterances (see Appendix A) and were administered twice: once in a randomized order for the timed GJT and once more in a randomized order for the untimed GJTs (see also Plonsky et al., 2020; Spinner & Gass, 2019). The materials for the GJTs were derived from Torres et al. (2019), a study on Spanish features susceptible to L1-English influence. Items were adjusted so that half contained grammar items that would appear in a first semester Spanish university course, such as subject-verb agreement, distinguishing between which of Spanish's two copulas was appropriate, and distinguishing between using the Spanish verb *conocer* or *saber*. To accommodate any pre-existing differences in Spanish knowledge demonstrated by the two participants, whose baseline proficiency levels in Spanish were not controlled for in the study, the items included in the GJTs represented a variety of difficulty levels for learners. Participants indicated whether each sentence provided was grammatical or ungrammatical. Their responses were coded dichotomously, with 1 for a correct answer or 0 for an incorrect answer. Participants also received a 0 if they did not give an answer before the 7-second time limit expired in the TGJT. In addition to the time constraint in the TGJT, the UGJT also prompted the participants to indicate on a rule-versus-feeling (RVF) spectrum whether they made their judgment based on a rule they knew or a more intuitive feeling. A score closer to 0 was a rule-based decision, and one closer to 100 was a feeling-based decision.

3.2.2. Oral narrative task

Following Ellis (2005), the ONT involved reading a story twice, then prompting the participant to record themselves retelling the story. The Spanish language story used in this ONT was titled "El Secreto" from VanPatten's (2018) *Cuentos Cortos Volume 1*, which is designed for novice and intermediate learners of Spanish (see Appendix B). The data from the ONT were transcribed, coded, and scored using an obligatory occasion analysis. The set of target structures was determined by examining consistent morphosyntactic and phonological errors made by a given participant across their five retellings of this ONT. Data from the ONTs were coded by the researchers and confirmed by a Spanish L1 user

versed in metalinguistic terminology. These codes were assigned dichotomously (correct/incorrect suppliance) to each occasion of a target structure. Table B1 (see Appendix B) summarizes all of the target features examined in the ONT across participants and the type of linguistic error they represent.

3.2.3. Metalinguistic knowledge task (MKT)

The MKT was adapted from Roehr and Gánem-Gutiérrez (2009) and consisted of 20 separate questions (see Appendix C). In the MKT, participants were instructed to read a sentence or short dialog in Spanish that contained an underlined error, then give a correction of the error and an explanation. Participants were scored on whether they supplied the correct form, gave the correct explanation, and named the grammatical form using metalanguage. For instance, if the participant read the sentence

> *Juan se compró un coche blanca.*
> Juan purchased a.masc.sing car.masc.sing white.fem.sing
> *Juan purchased a white car*

and gave the correction *blanco* and an adequate explanation that mentioned either masculine adjectives or gender agreement or both, the participant would get a score of three for that one item. Each of the 20 sentences in the MKT included a maximum score of three points, for a total possible high score of 60.

3.3. Procedure

We first notified the participants of the purpose of the research, the schedule/duration and the types of activities we would ask them to complete, and the steps we would take to protect their privacy and the confidentiality of their responses and data. Both participants then gave their informed consent. Participants completed the initial battery of tests online prior to enrolling in the first week of Duolingo use. This served as the study's baseline pretest before any input-based Duolingo study commenced. They were then given instructions on the prescribed minimum amount of study time and advised to meet this minimum (see below) but also exceed it if possible, beginning the first week of the data collection process. At the end of each week, participants self-reported how many stories or episodes they respectively completed to the researchers and completed another round of tests (administered through Qualtrics) measuring

their implicit and explicit L2 knowledge. This sequence of using Duolingo during the weekdays and taking the battery of tests at the end of the week was repeated each week. Including the baseline pretest in the week prior to the beginning of the study, the two participants completed a total of five iterations of the battery of implicit and explicit knowledge tasks. All interaction between the researcher and the participants and all data collection were done remotely. All test items in the MKT were randomized by Qualtrics, and all items in the GJTs were pre-randomized by the researchers for ease of cleaning, coding, and scoring the data.

The participant assigned to the Podcast condition, Cir, was instructed to listen to at least three podcasts every week during the weekdays prior to taking the series of L2 knowledge tests during the weekends. Eric, the participant assigned to the Stories condition, was similarly required to complete at least three stories a day or 15 stories a week in Duolingo's Stories feature, before taking the battery of L2 knowledge tests each weekend. Both participants were encouraged to do more than the minimum if able. These two minimums were established through a pilot of the study by the researchers which determined that each Duolingo podcast was about 20 minutes in length, and one Duolingo story took approximately 4-5 minutes to complete. This minimum of three podcasts a week or 15 stories a week was roughly equal to one hour of Duolingo study every week, a benchmark established by other case studies investigating the efficacy of Duolingo (see e.g., Isbell et al., 2017; Loewen et al., 2019).

3.4. Data analysis

Descriptive statistics were calculated based on percentage score correct on each of the four tests over the five data collection points. Comparisons of average scores across participants and tests were also conducted. The data from both participants were normalized into *Z*-scores to make cross-measures more comparable and to standardize the comparisons in the four tests between the two participants. Due to the study's short-term longitudinal design, data were analyzed using min-max graphs, moving averages, and moving correlations (Verspoor et al., 2011). Moving correlations were calculated based on the *Z*-scored data and compared across and within participant performance.

4. Results

Over the course of this study, the Podcast participant (Cir) completed an average of 3.75 episodes each week (an average of 1.13 hours in duration per week),

totaling 15 podcast episodes over the four-week period (a total of 4.53 hours in duration). In the following sections we report our results sequentially, detailing the data analyses we performed for each separate participant, in relation to our specific research questions.

4.1. Podcast participant's implicit and explicit L2 knowledge

Related to RQ1, we examined the specific gains in implicit and explicit L2 knowledge that were made by the learner through consistent use of the Duolingo Podcast feature. As Figure 1 illustrates, the combined data for the Podcast participant offer two insights broadly related to this. First, they show that where increases in L2 knowledge were possible they happened, with the exception of the MKT. They also show that Cir's explicit and implicit L2 knowledge were already advanced even from the point of departure. This high score was maintained throughout the study, never dipping below 92% in any measure, except in the MKT (77% in the final week). This dip is likely a result of task repetition effects because it is improbable that Cir lost explicit knowledge as a result of using Duolingo's Podcast feature.

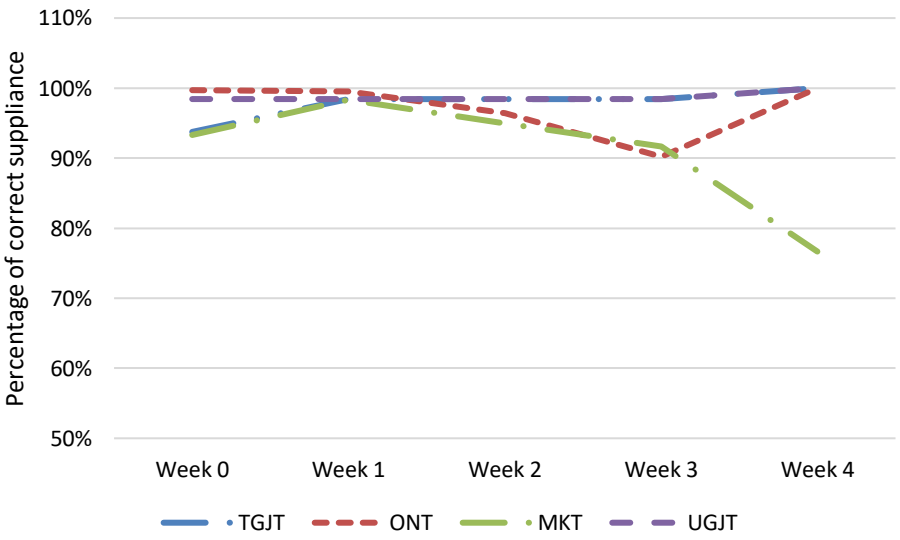


Figure 1 Percentage of correct suppliance across all tasks and weeks in the Podcast condition (TGJT = timed grammaticality judgment task; ONT = oral narrative task; MKT = metalinguistic knowledge task; UGJT = untimed grammaticality judgment task)

To further answer research questions 1 and 2, which asked what gains in implicit and explicit L2 knowledge are made by learners using the Duolingo Stories feature (RQ1) and the Duolingo Podcast feature (RQ2) consistently over one month, it was also important to examine each participant's implicit L2 knowledge in relation to their explicit L2 knowledge. One easily comparable measure is the two GJTs. Figure 2 presents a comparison of averages (shown as percentages) of the two GJTs for Cir.

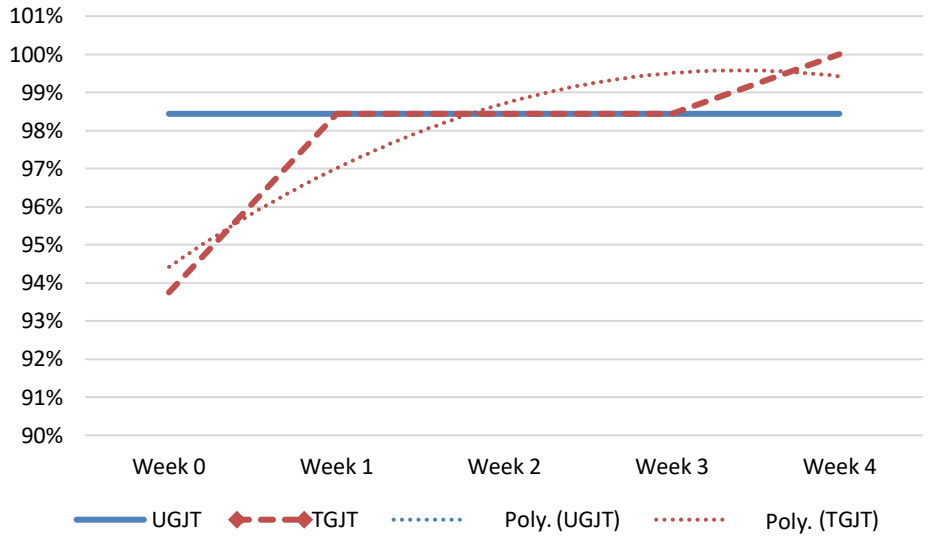


Figure 2 Comparison of averages across the two GJTs in the Podcast condition ("Poly." = polynomial smoothing of the data, which approximates the smoothed trajectory of the data over consistent time series intervals)

Figure 2 indicates that, while Cir's explicit L2 knowledge, as measured by the UGJT, remained constant throughout the data collection period, his implicit L2 knowledge, as measured by the TGJT, increased. This provides evidence for the learner's improvement on implicit knowledge, but not explicit knowledge, as a consequence of using Duolingo's Podcast feature.

Figure 3 presents composite scores from the two measures in each category of L2 knowledge. These are shown as an average score for Cir's implicit knowledge (blue line), represented by his ONTs and TGJTs, and his explicit knowledge (red line), which is a composite of his UGJTs and MKTs.

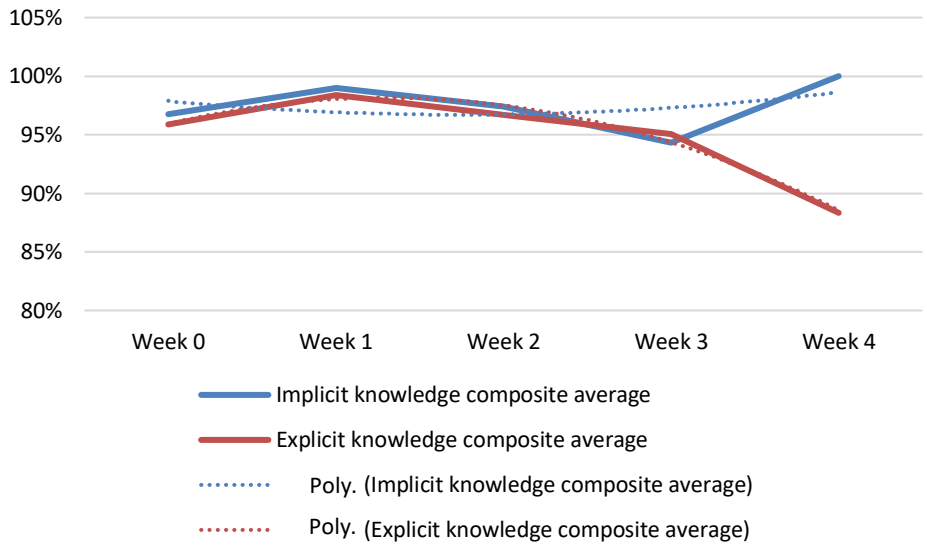


Figure 3 Comparison of composite implicit and explicit knowledge scores from the Podcast condition ("Poly." = polynomial smoothing of the data)

Juxtaposed this way, it appears that Cir's explicit and implicit L2 knowledge had a largely parallel trajectory throughout all weeks apart from the final week, in which his explicit knowledge clearly diverges from the implicit knowledge. As alluded to earlier, this dip was caused primarily by his low scores on the MKT.

4.2. Stories participant's implicit and explicit L2 knowledge

Eric, the Stories participant, self-reported completing three stories every weekday, equaling roughly 15 stories a week. However, due to an unforeseen change in the Stories feature, Eric was unable to complete all 15 stories for the final week of the study. Duolingo instituted a "Crown requirement" to unlock new sets of short stories in the Stories feature by completing units in the default Learn feature. Because Eric was explicitly required to use only the Stories feature, the limit of the stories he had unlocked with his current use of Duolingo was met in the middle of the final week (see discussion of limitations below). Figure 4 shows Eric's task performance on all tests of implicit and explicit knowledge across all weeks.

These data provide insight into the gains in implicit and explicit L2 knowledge made through the use of the Duolingo Stories feature. They show that the gains made by Eric in all of the tasks were meager, with the exception of the TGJT, which rose steadily from 55.56% in week 0 to 76.19% in week 4. The

other measures also provide evidence related to our first research question by showing meager gains and relative stability.

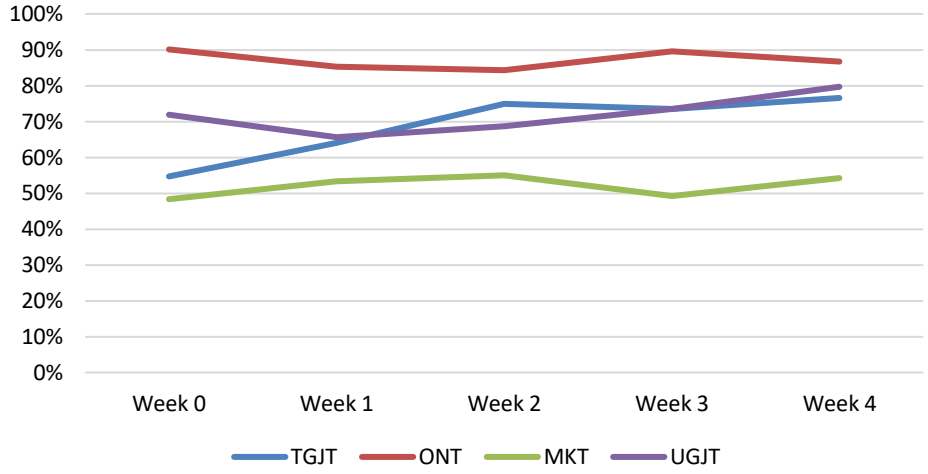


Figure 4 Percentage of correct suppliance across all tasks and weeks in the Stories condition

These data provide insight into the gains in implicit and explicit L2 knowledge made through the use of the Duolingo Stories feature. They show that the gains made by Eric in all of the tasks were meager, with the exception of the TGJT, which rose steadily from 55.56% in week 0 to 76.19% in week 4. The other measures also provide evidence related to our first research question by showing meager gains and relative stability.

According to his scores on the GJTs, Eric's implicit and explicit L2 knowledge were closely related during the data collection period. Figure 5 summarizes his averages on both GJTs over all weeks. These data provide important evidence for the gains in implicit and explicit L2 knowledge made through the use of the Duolingo Stories feature, as they demonstrate that, at least as measured by the GJTs, the learner's implicit and explicit knowledge types were affected in similar ways by the Stories feature. Similarly, Figure 6 provides a summary of the composite scores of Eric's explicit and implicit L2 knowledge derived from the averages of the two tasks in each respective category.

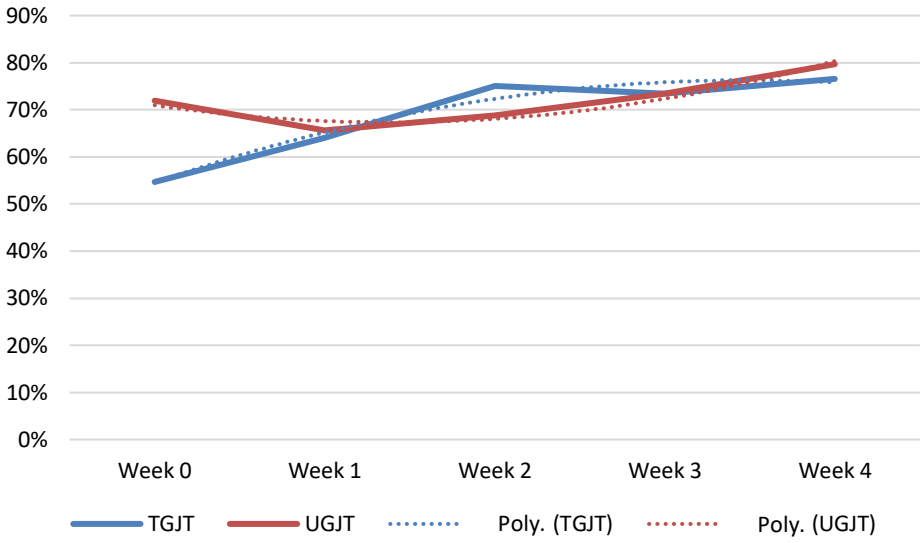


Figure 5 Comparison of averages across the two GJTs for Eric ("Poly." = polynomial smoothing of the data)

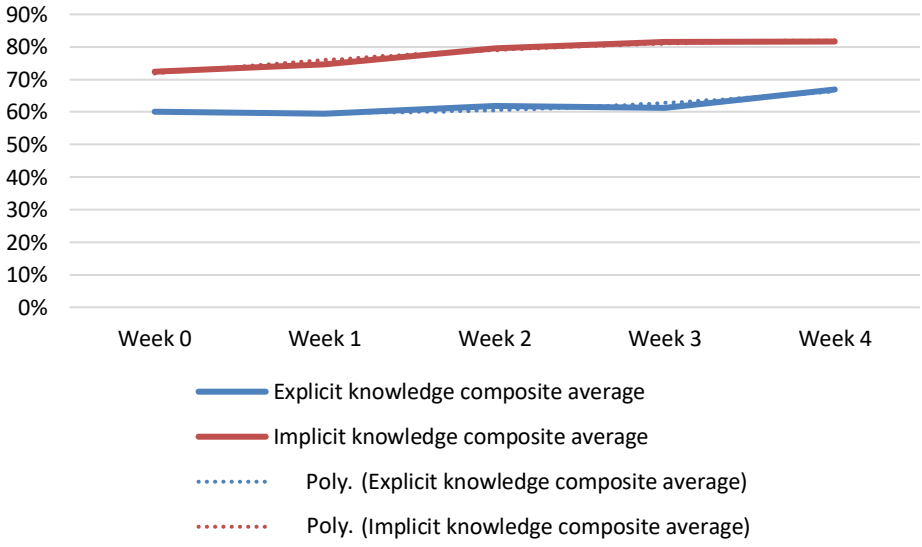


Figure 6 Comparison of implicit and explicit knowledge scores from the Stories condition ("Poly." = polynomial smoothing of the data)

Figures 5 and 6 provide data relating to the gains in implicit and explicit L2 knowledge made through the use of the Duolingo Stories feature, demonstrating

that Eric's implicit and explicit L2 knowledge ran narrowly parallel to each other, with similar trajectories and slight increases. Eric's implicit knowledge of Spanish was already higher than his explicit knowledge. This was an unexpected feature of Eric's L2 knowledge but was maintained throughout all weeks, with only a brief moment during week 3 where the difference was slightly larger than in other weeks.

4.3. Comparison of explicit and implicit L2 knowledge across conditions

Research question 3 asked what differences there might be in gains made by the two participants in all measures of explicit and implicit L2 knowledge. Figure 7 shows the composite averages of the two participants' explicit knowledge.

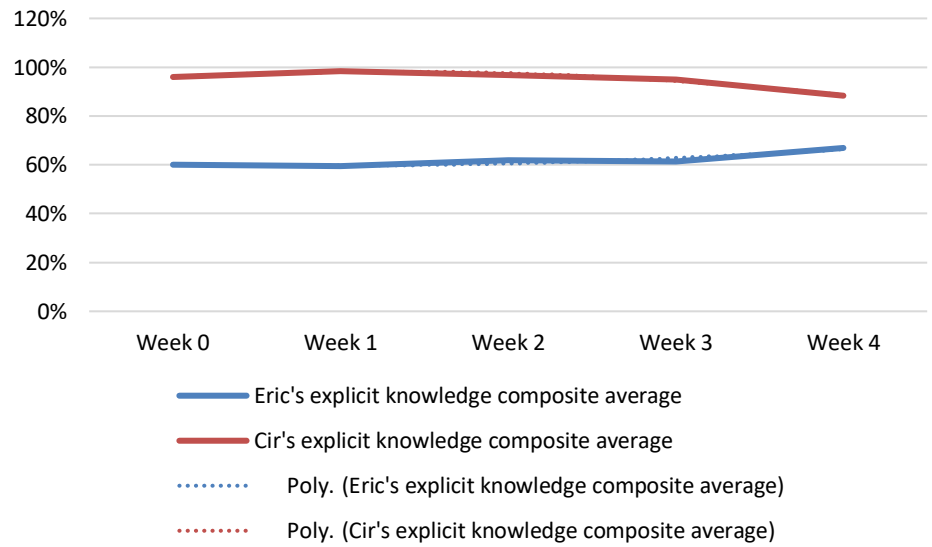


Figure 7 Comparison of explicit knowledge composite averages across participants ("Poly." = polynomial smoothing of the data)

Again, what is noticeable from the examination of these averages is the relative stability across all weeks and participants. With reference to RQ 3, examining the differences in L2 knowledge gains between the two participants, the difference between these two learners in terms of their explicit L2 knowledge remained roughly the same. The one exception is in week 4, where the Podcast participant's explicit L2 knowledge scores decreased while the Stories participant's scores increased. This data suggests that the Stories feature prioritizes explicit L2 knowledge development more than the Podcast feature does.

In Figure 8, similar composite averages were made by combining averages from the two implicit knowledge tests (the ONT and the TGJT). These offer a snapshot of the participants' overall implicit knowledge and how the two compare.

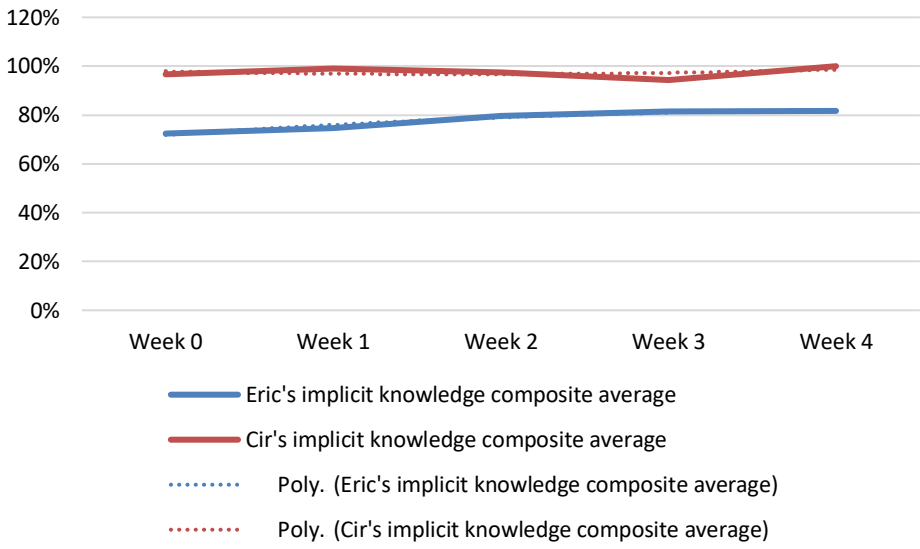


Figure 8 Comparison of L2 implicit knowledge across participants ("Poly." = polynomial smoothing of the data)

The two participants showed small, steady increases in their implicit knowledge over time, with the distance between their two scores tightening at points from weeks 2-4. This sheds more light on differences in gains made by the two participants in all measures of explicit and implicit L2 knowledge, as it shows that implicit L2 knowledge increased for the Stories participant slightly more than for the Podcast participant. The change in Eric's implicit L2 knowledge was marked only by increases, while Cir's occasionally decreased, particularly from weeks 1 to 2 and 2 to 3. Eric's implicit L2 knowledge increased by an average of 2% each week, whereas Cir's increased at a rate of 1% each week. These differences in gains are summarized in Table 2.

Table 2 Summary of gains by L2 knowledge type by participant condition

Participant condition	Knowledge type	Total change in knowledge from week 0 to week 4	Average weekly change in knowledge
Podcasts	Explicit	-8%	-2%
Podcasts	Implicit	3%	1%
Stories	Explicit	7%	2%
Stories	Implicit	9%	2%

4.4. Knowledge stability across participants

The fourth and final research question (RQ4) we had in this study was to examine the stability of the types of knowledge developed by each participant. Both participants presented stable knowledge gains in nearly all tasks. However, there were some instances where the knowledge gains were unstable. We present these selectively below. (For all figures and datasets, see supplementary online material.)

4.4.1. Podcast participant's L2 knowledge stability

First, Cir's grammaticality judgments for the UGJT remained the same throughout all five weeks, and his rule-versus-feeling (RVF) responses – a spectrum indicating whether a judgment was based on a rule he knew or an intuitive feeling – were markedly stable, with moving correlations at or above 0.96 for all test instances. Unlike Cir's UGJT scores, his answers on the MKT, as mentioned earlier, varied week-by-week due to decreasing specificity in the explanations he gave. This resulted in moving correlations for this one task that never reached above 0.54 and which also dipped below 0 in his answers from week 1 to week 2. As for his implicit knowledge stability, Cir's grammaticality judgments on the TGJT were initially unstable, with a moving correlation of 0.48 from week 0 to week 1. After this, though, his answers increased in stability and approached 1.0.

Additionally, Cir's productive implicit knowledge as measured by the ONT was also relatively stable, beginning with near-perfect suppliance of all features (99.72% in week 0) and ending with a perfect suppliance of features (100%) in week 4. This is evidence of a classic U-shaped learning curve. Evidently, Cir's implicit knowledge was strong from the beginning of the study, and that strength was maintained throughout. However, a necessary drawback of the ONT is that a participant may get a higher correct score simply through avoidance (i.e., more conservatively attempting a feature). Nevertheless, Cir's overall length and, therefore, attempts at various features grew over the weeks (see Appendix B, Table B2). These data offer a broad answer to our first research question: the knowledge gains for the Podcast participant were largely stable with the exception of the MKT.

4.4.2. Stories participant's L2 knowledge stability

Although the Podcast participant showed little to no variability across time in either his UGJT scores or his rule-versus-feeling (RVF) responses, the Stories participant demonstrated wide variability in his RVF scores. He showed a substantial

increase in the number of questions he answered by feel rather than rule over the five task instances. These data are shown in Table 3.

Table 3 Eric's RVF responses across time

	Week 0	Week 1	Week 2	Week 3	Week 4
Totals	1000	4100	5900	5700	6000
Averages (*SD*)	15.63 (36)	64.06 (48)	92.19 (27)	89.06 (31)	93.75 (24)
Total change	N/A	3100	1800	-200	300

Whereas the Podcast participant answered only one more item by feel rather than rule every two weeks, Eric answered, on average, 12.5 more items by feel rather than rule compared to the previous week. A Wilcoxon Signed Rank test across the five testing times revealed that the RVF increases from week 0 to week 1 as well as week 1 to week 2 were significantly different ($p < .001$), but the increases from week 2 to week 3 ($p = .480$) and week 3 to week 4 ($p = .366$) were not. This directly pertains to our first research question and demonstrates that, whereas Eric's explicit L2 knowledge showed meager gains on the UGJT, his use of feel or intuition, a more implicit mechanism, increased significantly. Moving correlations from the raw RVF scores are presented in Figure 9, which shows the relative stability of the change in RVF score over time.

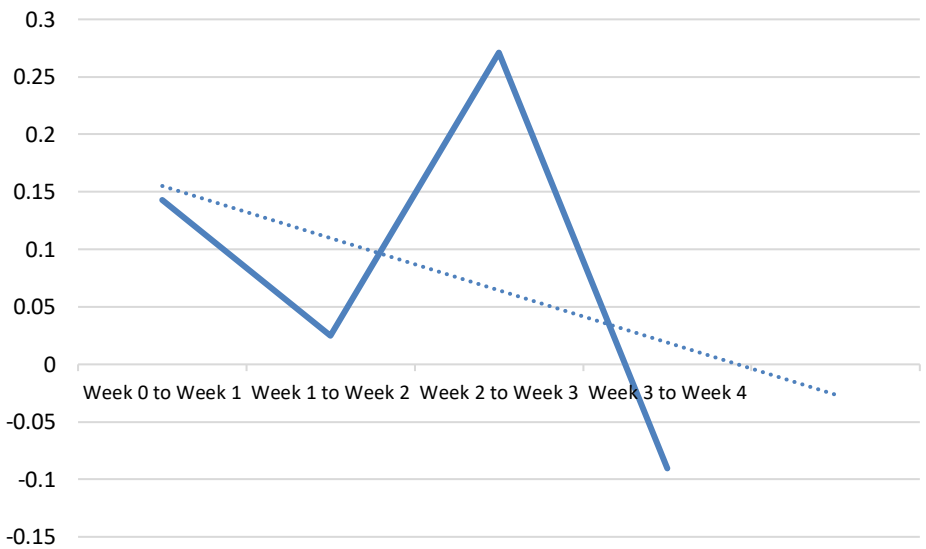


Figure 9 Moving correlation for UGJT's RVF scores for Stories participant

The above moving correlations reveal that Eric's RVF ratings varied greatly by week and by item. This is further demonstrated in Table 4 by a number of items where specific variability is evident.

Table 4 Illustrative UGJT features showing Eric's wide variability in RVF responses

Grammatical feature and condition	Week 0 RVF	Week 1 RVF	Week 2 RVF	Week 3 RVF	Week 4 RVF
Ungrammatical verbal aspect #3	0	0	100	0	100
Ungrammatical direct object marking #4	100	0	100	100	0
Grammatical *saber* vs. *conocer* #4	0	0	100	100	0
Ungrammatical subjunctive #4	0	0	100	0	100
Grammatical direct object marking #1	0	100	0	0	100
Ungrammatical direct object marking #3	0	100	100	0	100

Although the increase in RVF score over time was substantial for the Stories participant, the increase was not consistent, seemingly occurring at random or perhaps evidencing guesswork. This casts some doubt on the interpretation of his large RVF score increase as a measure of increasing implicit L2 knowledge. Such an interpretation makes sense considering his overall UGJT score did not increase significantly, despite the large RVF increase. Data from Eric's UGJT performance provide context to this answer to research question 1, showing that his explicit L2 knowledge did not benefit greatly from the use of Stories. Further, a lack of focus on form in the Stories feature may have been the cause of his guesswork in the latter three test instances, where his high RVF scores are maintained yet remain unstable or inconsistent.

5. Discussion

In this study, we wanted to know about: (1) participants' gains in explicit and implicit L2 knowledge as well as the stability thereof during one month of use of a single input-based feature of Duolingo and (2) the differences between those gains made across participants, weeks, and knowledge types. The MKT measured productive explicit knowledge, the UGJT measured receptive explicit knowledge, the TGJT measured receptive implicit knowledge, and the ONT measured productive implicit knowledge (Ellis, 2005). These established, validated measures of explicit and implicit L2 knowledge have not been used in the MALL literature despite specific claims about the type of L2 knowledge developed by using MALL apps. For example, receptive explicit L2 knowledge is posited by many as the main outcome of MALL app use given that there is little opportunity for oral communication in these apps (Loewen et al., 2020; Rachels & Rockinson-Szapkiw, 2018). The present study sought to address this gap by using a full range of established measures of L2 knowledge.

For the first two research questions, related to the gains in learners' implicit and explicit L2 knowledge from using the Duolingo Stories feature (RQ1) and the Duolingo Podcast feature (RQ2) consistently over one month, our data revealed a consistent, slight improvement across both participants and knowledge types with the exception of the Podcast participant's performance on the MKT. The data also showed strong stability across all knowledge types and participant conditions in the gains made. The changes participants made in their L2 knowledge were sustained, typically showing an increasing correlation as weeks progressed. One important exception to this pattern is the large variability in the Stories participant's RVF scores. His gains made on implicit knowledge tests (i.e., the ONT and the TGJT) demonstrated a sustained increase in implicit L2 knowledge over time, which may explain the large increase in RVF scores on his UGJT scores over time. Nevertheless, because both participants' data showed a general increasing correlation specifically with the gains made in the GJTs and the MKT, this indicates that the explicit and implicit L2 knowledge development of the two participants was largely stable. Participants made sustained increases across multiple L2 knowledge types, and this finding suggests that both implicit and explicit L2 knowledge can benefit from MALL use. Contrary to earlier claims that receptive explicit knowledge is the primary outcome of MALL app use (Loewen et al., 2020), learners can develop both types of L2 knowledge especially when the features used are more input-based.

For RQ3, we had four hypotheses. First, we hypothesized that the Podcast participant would have higher scores than the Stories participant in implicit knowledge measurements as a consequence of the Podcast feature's inherent focus on meaning (see e.g., Jiang et al., 2020; Savvani, 2019). Our results did not support this first hypothesis. Although the two participants made different gains in implicit knowledge, this cannot be seen as resulting directly from the use of the Podcast feature because the two participants had meaningfully different starting points in both types of knowledge. Another indication of the implicitness of the knowledge developed is the participants' RVF scores, which the Stories participant consistently rated higher than the Podcast participant despite having lower overall scores in the UGJT. These results indicate that the Stories feature had a greater effect on implicit knowledge development than the Podcast feature.

In addition to revealing that L2 implicit knowledge can be developed by input-based features in MALL apps (Loewen et al., 2020), these findings also specify which type of input-based feature is more beneficial for implicit L2 knowledge development at different proficiency levels. Specifically, implicit L2 knowledge increased the most for the lower proficiency Stories participant. This is not surprising given existing findings that show the largest gains through app-based language learning tend to be made by those who are at the beginning

stages of learning (Loewen et al., 2019, 2020). It may also be the case that the intermittent and interactive comprehension checks which the Stories feature has but the Podcast feature lacks served to reinforce the story and increase the comprehensibility of the input, thus improving the Stories participant's implicit L2 knowledge more than the Podcast participant. However, as with previous findings (Sudina & Plonsky, 2023), our study also shows that the gains made by using input-based features of MALL apps tend to be quite modest. This may also be because the measurements used in the present study were not sensitive enough to fully capture both participants' gains over the duration given potential ceiling effects.

We also postulated that the Podcast participant's explicit L2 knowledge would not be heavily affected by the use of the Podcast feature because consciousness raising and metalinguistic features are not prominent in the episodes (Hasan & Hoon, 2013). This was supported by our analyses, as explicit knowledge seemed unaffected in this participant. From a theoretical perspective, lower scores on explicit knowledge tests over time are not indicative of the explicit knowledge actually decreasing but more plausibly indicative of task repetition or practice effects. However, this may also be a result of the high degree of explicit knowledge that the Podcast participant had from the very beginning; his lack of explicit knowledge development may be a result of the relatively simple features that the MKT and the UGJT tested for. This consistency potentially demonstrates a departure from previous claims that MALL apps like Duolingo as a whole primarily benefit receptive explicit knowledge (Crowther et al., 2017; Loewen et al., 2020), although additional studies with a larger sample size and scope to improve on participants' explicit L2 knowledge are merited to investigate this more directly.

We further hypothesized that the Stories participant would show a higher increase in explicit knowledge over time than the Podcast participant because of the translation exercises at the end of each short story. This hypothesis was supported by our data, although the Stories participant's explicit knowledge increased alongside his implicit knowledge (see Table 2). In line with the broader conclusions of some previous studies, these data suggest that explicit knowledge development is benefited by the use of Duolingo's Stories feature more than by the Podcast feature (Crowther et al., 2017).

Our final hypothesis was that the Stories participant would increase both their implicit and explicit L2 knowledge because meaningful input is nevertheless present in these stories (Jiang et al., 2020). We hypothesized that these gains would be less pronounced than the Podcast participant. This hypothesis, which runs counter to previous findings that app-based MALL's primary outcome is receptive explicit knowledge (Loewen et al., 2020), was supported by our results, although the Podcast participant made smaller gains than the Stories participant in both knowledge types. Our results from the battery of measures administered to participants in two

separate conditions cast doubt on claims that Duolingo's content and interface primarily benefits receptive explicit knowledge (Crowther et al., 2017), especially when considering how the four tasks tapped into not only different L2 knowledge types but also different modalities such as productive and receptive skills.

Instead of increases on the UGJT in the present study, clear but subtle gains were seen in the ONT for both participants, as well as large gains made by both participants in the TGJT (RQ4), both of which are tests of implicit L2 knowledge (Ellis, 2005). Interestingly, neither participant made meaningful gains in the MKT, a measure of productive explicit L2 knowledge. This is consistent with the conclusions drawn from previous studies (Kukulska-Hulme & Viberg, 2018; Loewen et al., 2020), and it shows that productive explicit L2 knowledge in the present study was not a primary outcome for either the Stories or the Podcast features. Further, the increase in TGJT lends credence to the position that receptive implicit L2 knowledge, rather than receptive explicit L2 knowledge, could be the primary outcome of using the Stories and Podcast features.

These results can be seen as building on, and also somewhat of a departure from, the field's current understanding of app-based MALL. Efficacy research in MALL app use has focused heavily on comparing scores on collegiate foreign language tests that primarily measure receptive explicit L2 knowledge in order to provide insight on how closely MALL app use approximates university-level foreign language classes and corresponds with their instructional results and outcomes (e.g., Isbell et al., 2017; Krashen, 2014; Loewen et al., 2019; Vesselinov & Grego, 2012). While some MALL research has pointed out the narrow scope of these L2 assessment measures (Crowther et al., 2017; Loewen et al., 2020), our study is the first to employ an established battery of explicit and implicit L2 knowledge tests for MALL app users. Our data, therefore, advance the field by showing that app-based MALL does not result exclusively in receptive explicit L2 knowledge and, indeed, can benefit the development of other important L2 knowledge types.

6. Implications

There are several key implications from our study that we outline here. First, although the gains in implicit knowledge were modest, the use of the Podcast feature clearly provides meaningful input for learners to incorporate into their developing system as implicit knowledge. This has practical implications for language learners and second language educators alike. For the former, it indicates that input deliberately encountered in the form of audio podcasts can meaningfully influence L2 implicit knowledge development in an easily accessible manner and at a consistent rate of a 1% increase for every 1.13 hours of study. For

the latter, it suggests that assigning podcasts for L2 students to listen to can have a positive effect on students' productive and receptive implicit knowledge types. This implication also extends findings from other recent naturalistic experiments, in which Duolingo use resulted in some modest gains in written and oral production (e.g., Sudina & Plonsky, 2023). The form of meaning-focused input available in Duolingo through the Podcast feature is a welcome addition to the language teacher's toolkit and one that is readily accessible and available for free in a growing number of different L1-LX pairings, with new episodes released on a weekly basis.

On the other hand, the Stories feature presents more bite-sized chunks of meaning-based input for language learners that can still have a positive effect on their receptive implicit and explicit L2 knowledge. If the goal is to perform well on an exam geared towards productive explicit knowledge, however, the Stories feature may not be appropriate because articulating grammatical features is intentionally absent from the Stories feature curriculum. Language teachers whose students are primarily interested in receptive or productive explicit L2 knowledge should exercise caution when making use of these features of language learning apps as they do not appear directly appropriate for building this type of L2 knowledge. The Stories participant's data support DeKeyser's (2017) notions of knowledge proceduralization, wherein explicit knowledge is gradually automatized to become more implicit-like. This is especially the case if the TGJT, on which Eric made the largest gains, is interpreted as a measure of automatized explicit knowledge (Suzuki, 2017). As a result, the Stories feature may help solidify explicit L2 knowledge, nudging it toward implicitness. Finally, for both the Stories and Podcast features of app-based MALL, a major implication is that even relatively short durations of purposeful in-app engagement can lead to modest gains in L2 knowledge (Sudina & Plonsky, 2023).

7. Conclusion

The purpose of this study was to introduce nuance and clarity to the discussion on Duolingo's efficacy and influence on explicit and implicit L2 knowledge. As observed in some reviews, the effectiveness of many such language learning apps remains unclear since they tend to be evaluated primarily on their usability, functionality, and reliability (Ruiz et al., 2024). This study examined Duolingo's relatively novel, more input-based features: the Stories and the Podcast features. The data from two participants over the four-week target period of use showed that slight increases in productive implicit knowledge in the Podcast condition were possible after one month of use, but that this increase was not as evident in the Stories condition. Nevertheless, the Stories participant made

and sustained consistent gains in both types of knowledge, but this gain was not consistent across productive and receptive modalities. These findings suggest that the Stories feature may benefit receptive explicit and implicit knowledge alike, as measured by the two GJTs.

Several limitations remain. First, as with other apps whose functionality focuses on helping users learn language rather than toolkits to facilitate data collection or learning analytics, we had to administer the tests of L2 knowledge outside the Duolingo platform. These tests were administered repeatedly, and although the items on the MKT and GJTs were randomized each week, testing effects may also have played a role in the participants' gains over time since retrieval practice is known to result from repeated testing (Rawson & Dunlosky, 2011; Suzuki, 2024). Future studies should, therefore, consider employing more varied questions at each testing window that are randomized through various appropriate techniques.

As part of fidelity measures, we took care to ensure that participants utilized only their respective Duolingo feature during the four-week data collection period. Still, the Podcast participant's data revealed a ceiling effect that was a clear indicator of his prior explicit L2 knowledge (Roehr & Gánem-Gutiérrez, 2009). The Stories participant reported taking only one formal high school-level Spanish class but had also studied Spanish independently for the six months prior to this study. The question remains whether, despite our best efforts at recruiting, these participants did in fact belong to the population of learners who would benefit most from Duolingo's input-based features (Jiang et al., 2020). Previous research indicates that apps are especially good for learning unfamiliar lexical items, formulaic sequences, and multiword units and promoting other linguistic gains, but less effective at generating opportunities for extended input or for interactive use of the language. Consequently, MALL may be especially effective for taking learners from a level of "no real functional ability, to being able to communicate minimally by using a number of isolated words and memorized phrases" (Loewen et al., 2020, p. 227).

It is also important to note that this case study design and our cross-case comparison emphasized the mechanisms and processes of learning from MALL platforms, allowing us to particularize two participants' learning gains in one target language. The trade-off, however, is that our results cannot be generalized to larger populations, especially considering both participants' previous use of MALL platforms and their significant prior language learning experience. Given this limitation, future studies are needed to investigate whether these gains apply equally to learners with different socio-demographic profiles, including limited-proficiency learners and those with limited experience using MALL apps, who are attempting to learn other target languages.

Despite this study's limitations, these preliminary data indicate that Duolingo's Podcast or Stories features have nuanced effects on learners' receptive, productive, implicit, and explicit L2 knowledge. Productive implicit knowledge development is slow but can nonetheless be positively influenced by the use of the Podcast feature. Additionally, although the Stories feature had little effect on implicit productive knowledge, it had a large and sustained positive effect on receptive implicit knowledge development. In tandem, these two novel input-based features of Duolingo provide the learner with input that is practical, entertaining for the learner, and effective in positively influencing various knowledge types and modalities.

As with most studies, further research is needed to achieve a higher degree of empirical clarity and statistical power, to extend these findings, and to independently verify the bold claims made by purveyors of app-based L2 learning about its effectiveness. There are inherent design limitations when implementing and researching apps designed as closed, proprietary systems that do not allow for adaptive tailoring to meet the needs and levels of their users in ecologically valid ways (Ruiz et al., 2024). Due also to design constraints, this study did not meet Chwo et al.'s (2018) call for long-duration implementation of at least eight weeks in MALL research, and we focused on L2 knowledge types without tracking individual difference factors over time shown by some studies to be affected by MALL use (García Botero et al., 2019). Relatedly, it is clear that the two participants in this study had large differences in L2 proficiency, which likely contributed to the Podcast participant's ceiling effects. Future studies could sample based on proficiency to identify if the changes in explicit and implicit knowledge reported in the present study affect all levels in the same way. Similarly, since our study was not designed with a particular proficiency range in mind, the tasks we used to elicit L2 knowledge may not have been sensitive enough to detect proficiency level-related differences. Though we adopted established measures that have been used in previous empirical work, future research is needed to extend our findings with measures of L2 knowledge that represent the full range of L2 target structures.

References

Ajisoko, P. (2020). The use of Duolingo apps to improve English vocabulary learning. *International Journal of Emerging Technologies in Learning*, *15*(7), 149-155. https://doi.org/10.3991/IJET.V15I07.13229

Al-Sabbagh, K. W., Bradley, L., & Bartram, L. (2018). Mobile language learning applications for Arabic speaking migrants: A usability perspective. *Language Learning in Higher Education*, *9*(1), 71-95. https://doi.org/10.1515/cercles-2019-0004

Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning: An introduction. *Studies in Second Language Acquisition*, *37*(2), 185-196. https://doi.org/10.1017/S027226311500008X

Antonia, T. G., & Pierpaolo, L. (2020). E-learning, MALL and LMOOC or a new didactic tendency in foreign language learning: A meta-analysis from 2014 to 2018. *Psychology*, *11*(4), 581-593. https://doi.org/10.4236/psych.2020.114039

Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL*, *27*(1), 4-20. https://doi.org/10.1017/S0958344014000159

Chen, X., Meurers, D., & Rebuschat, P. (2022). ICALL offering individually adaptive input: Effects of complex input on L2 development. *Language Learning & Technology*, *26*(1), 1-21. https://hdl.handle.net/10125/73496

Chwo, G. S. M., Marek, M. W., & Wu, W. C. V. (2018). Meta-analysis of MALL research and design. *System*, *74*, 62-72. https://doi.org/10.1016/j.system.2018.02.009

Cross, J. (2014). Promoting autonomous listening to podcasts: A case study. *Language Teaching Research*, *18*(1), 8-32. https://doi.org/10.1177/1362168813505394

Crowther, D., Kim, K. M., & Loewen, S. (2017). The implementation of ISLA in MALL technology: An investigation into the potential effectiveness of Duolingo. *MSU Working Papers in SLS*, *8*, 20-39.

de Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, *68*(4), 906-941. https://doi.org/10.1111/lang.12296

DeKeyser, R. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15-32). Routledge.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. *Studies in Second Language Acquisition*, *27*(2), 141-172. https://doi.org/10.1017/S0272263105050096

Ellis, R., & Roever, C. (2021). The measurement of implicit and explicit knowledge. *Language Learning Journal*, *49*(2), 160-175. https://doi.org/10.1080/09571736.2018.1504229

Fouz-González, J. (2019). Podcast-based pronunciation training: Enhancing FL learners' perception and production of fossilized segmental features. *ReCALL*, *31*(2), 150-159. https://doi.org/10.1017/S0958344018000174

García Botero, G., Questier, F., & Zhu, C. (2019). Self-directed language learning in a mobile-assisted, out-of-class context: Do students walk the talk? *Computer Assisted Language Learning*, *32*(1-2), 71-97. https://doi.org/10.1080/09588221.2018.1485707

Godfroid, A. (2022). Hypotheses about the interface between explicit and implicit knowledge in second language acquisition. In A. Godfroid & H. Hopp (Eds.), *The Routledge handbook of second language acquisition and psycholinguistics* (pp. 294-307). Routledge.

Hasan, M. M., & Hoon, T. B. (2013). Podcast applications in language learning: A review of recent studies. *English Language Teaching*, *6*(2), 128-135. https://doi.org/10.5539/elt.v6n2p128

Isbell, D. R., Rawal, H., Oh, R., & Loewen, S. (2017). Narrative perspectives on self-directed foreign language learning in a computer- and mobile-assisted language learning context. *Languages*, *2*(2), 4. https://doi.org/10.3390/languages2020004

Jiang, X., Rollinson, J., Plonsky, L., & Pajak, B. (2020). Duolingo efficacy study: Beginning-level courses equivalent to four university semesters. *Duolingo Research Report DRR-20-04*. https://www.duolingo.com/efficacy

Kang, E. Y., Sok, S., & Han, Z. (2019). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research*, *23*(4), 428-453. https://doi.org/10.1177/1362168818776671

Krashen, S. (2014). Does Duolingo "trump" university-level language learning? *International Journal of Foreign Language Teaching*, *9*(1), 13-15.

Kukulska-Hulme, A., Lee, H., & Norris, L. (2017). Mobile learning revolution: Implications for language pedagogy. In C. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 217-233). Wiley-Blackwell.

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

Leow, R. (2019). ISLA: How implicit or how explicit should it be? Theoretical, empirical, and pedagogical/curricular issues. *Language Teaching Research*, *23*(4), 476-493. https://doi.org/10.1177/1362168818776674

Lin, J. J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878-919. https://doi.org/10.1080/09588221.2018.1541359

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

Loewen, S., Isbell, D. R., & Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign Language Annals*, *53*(2), 209-233. https://doi.org/10.1111/flan.12454

Lomicka, L., & Lord, G. (2010). Podcasting – past, present and future: Applications of academic podcasting in and out of the language classroom. In B. R. Facer & M. Abdous (Eds.), *Academic podcasting and mobile assisted language learning: Applications and outcomes* (pp. 1-20). IGI Global.

Long, M. H. (2020). Optimal input for language learning: Genuine, simplified, elaborated, or modified elaborated? *Language Teaching*, *53*(2), 169-182. https://doi.org/10.1017/S0261444819000466

Lord, G. (2008). Podcasting communities and second language pronunciation. *Foreign Language Annals*, *41*(2), 364-379. https://doi.org/10.1111/j.1944-9720.2008.tb03297.x

Meurers, D. (2021). Natural language processing and language learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley. https://doi.org/10.1002/9781405198431.wbeal0858.pub2

Mostafa, T., & Kim, Y. J. (2021). The effects of input and output based instruction on the development of L2 explicit and automatized explicit knowledge: A classroom based study. *Language Awareness*, *30*(1), 17-41. https://doi.org/10.1080/09658416.2020.1760292

Munday, P. (2016). The case for using Duolingo as part of the language classroom experience Duolingo como parte del curriculum de las clases de lengua extranjera. *RIED V*, *19*(1), 83-101.

Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, *1*(1), 2-13. https://doi.org/10.2167/illt039.0

O'Bryan, A., & Hegelheimer, V. (2007). Integrating CALL into the classroom: The role of podcasting in an ESL listening strategies course. *ReCALL*, *19*(2), 162-180. https://doi.org/10.1017/S0958344007000523

Philp, J. (2009). Pathways to proficiency: Learning experiences and attainment in implicit and explicit knowledge of English as a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 194-215). Multilingual Matters.

Piske, T., & Young-Scholten, M. (Eds.). (2008). *Input matters in SLA*. Multilingual Matters.

Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2020). A methodological synthesis and meta-analysis of judgment tasks in second language

research. *Second Language Research*, *36*(4), 583-621. https://doi.org/10.1177/0267658319828413

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. https://doi.org/10.1080/09588221.2017.1382536

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*(3), 283-302.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595-626. https://doi.org/10.1111/lang.12010

Reinders, H., & Ellis, R. (2009). Effects of two types of positive enhanced input on intake and L2 acquisition. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language* learning (pp. 281-302). Multilingual Matters.

Roehr, K., & Gánem-Gutiérrez, G. A. (2009). The status of metalinguistic knowledge in instructed adult L2 learning. *Language Awareness*, *18*(2), 165-181. https://doi.org/10.1080/09658410902855854

Rosell-Aguilar, F. (2007). Top of the pods – in search of a podcasting "podagogy" for language learning. *Computer Assisted Language Learning*, *20*(5), 471-492. https:// doi.org/10.1080/09588220701746047

Rosell-Aguilar, F. (2013). Podcasting for language learning through iTunes U: The learner's view. *Language Learning and Technology*, *17*(3), 74-93. https://doi.org/10125/44340

Rostami, M. A., Azarnoosh, M., & Abdolmanafi-Rokni, S. J. (2017). The effect of podcasting on Iranian EFL learners' motivation and attitude. *Theory and Practice in Language Studies*, *7*(1), 70. https://doi.org/10.17507/tpls.0701.09

Ruiz, S., Rebuschat, P., & Meurers, D. (2021). The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL. *Language Teaching Research*, *25*, 510-539. https://doi.org/10.1177/1362168819872859

Ruiz, S., Rebuschat, P., & Meurers, D. (2024). Supporting individualized practice through intelligent CALL. In Y. Suzuki (Ed.), *Practice and automatization in second language research: Perspectives from skill acquisition theory and cognitive psychology* (pp. 119-143). Routledge.

Savvani, S. (2019). State-of-the-art Duolingo features and applications. *Advances in Intelligent Systems and Computing*, *917*, 139-148. https://doi.org/10.1007/978-3-030-11935-5_14

Schmidt, J. (2008). Podcasting as a learning tool: German language and culture every day. *Die Unterrichtspraxis/Teaching German*, *41*, 186-194.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, *15*(2), 165-179. https://doi.org/10.1017/S0272263100011943

Spinner, P., & Gass, S. (2019). *Using judgments in second language acquisition research*. Routledge.

Sudina, E., & Plonsky, L. (2023). The effects of frequency, duration, and intensity on L2 learning through Duolingo: A natural experiment. *Journal of Second Language Studies*. https://doi.org/10.1075/jsls.00021.plo

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, *38*(5), 1229-1261. https://doi.org/10.1017/S014271641700011X

Suzuki, Y. (Ed.). (2024). *Practice and automatization in second language research: Perspectives from skill acquisition theory and cognitive psychology*. Routledge.

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747-790. https://doi.org/10.1111/lang.12241

Teske, K. (2017). Duolingo. *Computer Assisted Language Instruction Consortium*, *34*(3), 393-402. https://doi.org/10.1558/cj.32509

Torres, J., Estremera, R., & Mohamed, S. (2019). The contribution of psychosocial and biographical variables to heritage language learners' linguistic knowledge of Spanish. *Studies in Second Language Acquisition*, *41*(4), 695-719. https://doi.org/10.1017/S0272263119000184

Toto, G. A., & Limone, P. (2019). Contemporary trends in studies on mobile learning of foreign languages: A meta-analysis. *International Journal of Engineering Education*, *1*(2), 85-90. https://doi.org/10.14710/ijee.1.2.85-90

VanPatten, B. (2018). El Secreto. In *Cuentos Cortos Volume 1: Flash fiction in Spanish for novice and intermediate levels* (pp. 1-6). Input and More.

Verspoor, M. H., de Bot, K., & Lowie, W. (Eds.). (2011). *A dynamic approach to second language development*. John Benjamins.

Vesselinov, R., & Grego, J. (2012). *Duolingo effectiveness study*. Duolingo. 1-25. http://static.duolingo.com/s3/DuolingoReport_Final.pdf

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.

Ziegler, N., Meurers, D., Rebuschat, P., Ruiz, S., Moreno Vega, J. L., Chinkina, M., Li, W., & Grey, S. (2017). Interdisciplinary research at the intersection of CALL, NLP, and SLA: Methodological implications from an input enhancement project. *Language Learning*, *67*, 209-231. https://doi.org/10.1111/lang.12227

APPENDIX A

Timed and untimed grammaticality judgment materials

The following 64 sentences are divided evenly between grammatical and ungrammatical utterances and were administered twice: once in randomized order for the timed GJT and once more in randomized order for the untimed GJTs.

Direct object marker (*a* personal)

GR_dom01	Marcos ve al Dr. Sánchez mañana.
GR_dom02	María visita a su abuela cada semana.
GR_dom03	La compañía paga a los trabajadores cada viernes.
GR_dom04	Juan busca a su hija.
UG_dom01	Pedro ve el profesor el lunes.
UG_dom02	Teresa visita su tío en diciembre.
UG_dom03	Carlos paga los empleados cada semana.
UG_dom04	Pepe busca su hermano.

Noun/adjective gender agreement

GR_gndr01	Manuel vive en una casa roja.
GR_gndr02	Mi mejor amiga tiene pelo negro.
GR_gndr03	Los estudiantes viven en un apartamento pequeño.
GR_gndr04	El chico lleva una chaqueta blanca.
UG_gndr01	Ana tiene una falda negro.
UG_gndr02	El maestro tiene un bolígrafo roja.
UG_gndr03	Los niños quieren un regalo cara.
UG_gndr04	La chica tiene una pulsera negro.

Gustar

GR_gustar01	A los estudiantes no les gusta el examen.
GR_gustar02	A los niños les gusta el chocolate.
GR_gustar03	A mi hermano le gusta la pizza.
GR_gustar04	A ella no le gustan las clases de matemáticas.
UG_gustar01	Los chicos no les gusta la prueba.
UG_gustar02	Los niños les gusta los caramelos.
UG_gustar03	Mi hermana le gusta la televisión.
UG_gustar04	Ella no le gustan las cucarachas.

Aspect

GR_asp01	Ayer los niños jugaron hasta muy tarde.
GR_asp02	Mi padre estuvo en el hospital tres meses.
GR_asp03	Cuando era niño, iba al cine todas las semanas.
GR_asp04	Teresa vivió muchos años en México.
UG_asp01	Anoche los chicos estudiaban hasta las 3:00
UG_asp02	Mi hermano estaba en Venezuela seis meses.
UG_asp03	Cuando era joven, fui de compras todos los fines de semana.
UG_asp04	Mi amiga vivía diez años en Perú.

Subject-verb agreement

GR_sv01	Yo conozco a muchas personas interesantes.
GR_sv02	Ella come mucha hamburguesa.
GR_sv03	El problema es que no tengo dinero.
GR_sv04	La ventana está abierta.
UG_sv01	El perro estamos corriendo.
UG_sv02	Yo quiere un taco.
UG_sv03	Ellos creo que soy loco.
UG_sv04	Pájaros eres terribles.

Ser versus e*star*

GR_ser01	Yo soy un hombre muy fuerte.
GR_ser02	Ella está comiendo mucho
GR_ser03	Ellos están en el aeropuerto.
GR_ser04	Nosotros somos altos.
UG_ser01	Nosotros estamos muy fuertes.
UG_ser02	Yo soy corriendo ahora.
UG_ser03	Ellas están bajas.
UG_ser04	Ella es en el mercado.

Saber versus *conocer*

GR_svc01	Yo conozco al hombre.
GR_svc02	Ella sabe hacer su tarea.
GR_svc03	Ellos conocen la ciudad.
GR_svc04	Nosotros sabemos hablar español.
UG_svc01	Ella conoce a qué hora es la clase.
UG_svc02	Yo sé a David.
UG_svc03	Nosotros conocemos hablar inglés.
UG_svc04	El trabajador sabe a su jefe.

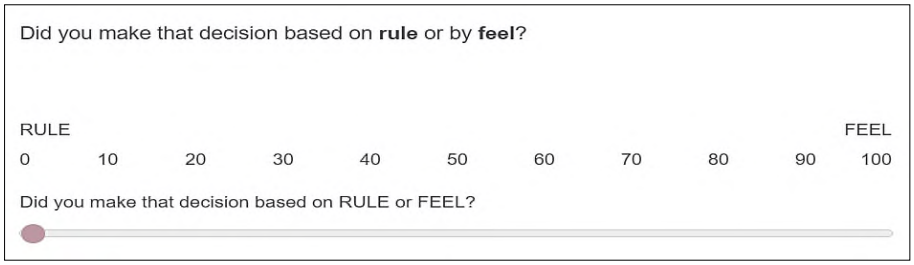
Present subjunctive with nominative clauses

GR_subnom01	Mamá quiere que Paco estudie mucho.
GR_subnom02	Papá pide que los niños hagan la tarea.
GR_subnom03	El maestro recomienda que los alumnos estudien mucho.
GR_subnom04	Mamá quiere que Susana limpie la casa.
UG_subnom01	Papá quiere que Elena estudia.
UG_subnom02	Mamá pide que hacemos la tarea.
UG_subnom03	La maestra no quiere que miramos la televisión.
UG_subnom04	Ana quiere que salemos con ella.

Untimed grammaticality judgment task

The UGJT differed from the TGJT by prompting participants after each judgment was made to indicate on a rule-versus-feeling (RVF) spectrum whether they made their judgment based on a rule they knew or a more intuitive feeling.

Rule vs feeling spectrum



APPENDIX B

Oral narrative task

In this task, you're going to read a story about a young man named Tomás two times. The story is told from his perspective. Once you've read the story twice, you will verbally tell the story in your own words.

You will be given some vocabulary to help you understand the story.

You will have to read through the story in its entirety before moving on.

Once you've read it a first time, click on the *next* button to read the story one more time.

Once you've read the story **two** times, click on the *next* button again and then you will be able to record yourself telling the story **in Spanish**.

For this task, you will be **recording yourself speaking**. Make sure you're in a place where your audio can be recorded easily and you won't be interrupted.

Don't make any notes. Try to tell the story from memory.

This test is supposed to get authentic, spontaneous, real-time Spanish from you, not prepared speech.

Here's some vocabulary to help you understand the story:

Spanish	English
Camisa	Shirt
Corbata	Tie
Traje	Suit
Relacion estrecha	Tight (close) relationship
Nos parecemos	We look alike

Now, read the story:

Me llamo Tomás. Tomás Rodríguez. Soy latino, de ascendencia mexicana, y tengo 19 años. Vivo en Sacramento, California. Sacramento es la capital de mi estado. Pero eso no es importante para mi historia.

Soy estudiante en la Universidad de California en Davis. Estudio negocios. Pero quiero cambiar mi campo a algo diferente. No me gustan los negocios y no qu iero trabajar donde tengo que vestir corbata, camisa blanca y traje todos los días. Pero eso tampoco es importante para mi historia.

Entonces, ¿Cuál es mi historia? Tengo un secreto. Tengo un secreto que nadie sabe. No lo saben mis padres. No lo sabe mi hermano, Carlos. Carlos tiene 18 años y va a la CSU Sacramento.

Tenemos una relación muy estrecha porque solo un año nos separa y somos los únicos hijos en la familia. También nos parecemos mucho. Muchas personas creen que somos gemelos pero no.

¿Y mis amigos? No tengo muchos. En la universidad no hablo mucho con otras personas. Solo con Ricky. Ricky *y* yo tenemos la misma clase de estadística. Es buen chico, muy estudioso. A veces comemos juntos y estudiamos juntos para los exámenes. Pero Ricky no sabe mi secreto.

Nadie sabe mi secreto.

No sé qué voy a hacer. Necesito hablar con alguien pero tengo miedo. ¿Por qué? Buena pregunta. Creo que tengo miedo de la reacción de otras personas.

De veras, no sé qué voy a hacer. Posiblemente escribo aquí mi secreto. Esta página es un espacio privado, ¿no? Pero, ¿qué pasa si alguien lee lo que escribo? ¿Qué pasa si alguien abre mi laptop y ve mi secreto? No. No lo voy a escribi r. No voy a revelar mi secreto aquí. Tengo miedo.

Es mejor esperar. ..

Here's the same vocabulary from before:

Spanish	**English**
Camisa	Shirt
Corbata	Tie
Traje	Suit
Relacion estrecha	Tight (close) relationship
Nos parecemos	We look alike

Now, read the story **a second time**:

Me llamo Tomás. Tomás Rodríguez. Soy latino, de ascendencia mexicana, y tengo 19 años. Vivo en Sacramento, California. Sacramento es la capital de mi estado. Pero eso no es importante para mi historia.

Soy estudiante en la Universidad de California en Davis. Estudio negocios. Pero quiero cambiar mi campo a algo diferente. No me gustan los negocios y no qu iero trabajar donde tengo que vestir corbata, camisa blanca y traje todos los días. Pero eso tampoco es importante para mi historia.

Entonces, ¿Cuál es mi historia? Tengo un secreto. Tengo un secreto que nadie sabe. No lo saben mis padres. No lo sabe mi hermano, Carlos. Carlos tiene 18 años y va a la CSU Sacramento. Tenemos una relación muy estrecha porque solo un año nos separa y somos los únicos hijos en la familia. También nos parecemos mucho. Muchas personas creen que somos gemelos pero no.

¿Y mis amigos? No tengo muchos. En la universidad no hablo mucho con otras personas. Solo con Ricky. Ricky *y* yo tenemos la misma clase de estadística. Es buen chico, muy estudioso. A veces comemos juntos y estudiamos juntos para los exámenes. Pero Ricky no sabe mi secreto.

Nadie sabe mi secreto.

No sé qué voy a hacer. Necesito hablar con alguien pero tengo miedo. ¿Por qué? Buena pregunta. Creo que tengo miedo de la reacción de otras personas.

De veras, no sé qué voy a hacer. Posiblemente escribo aquí mi secreto. Esta página es un espacio privado, ¿no? Pero, ¿qué pasa si alguien lee lo que escribo? ¿Qué pasa si alguien abre mi laptop y ve mi secreto? No. No lo voy a escribi r. No voy a revelar mi secreto aquí. Tengo miedo.

Es mejor esperar. ..

Now, using the button below, record yourself retelling the story out loud. Tell the story in Spanish. Try to tell the story in the *third-person*. Try to make mention of all of the characters, how they relate to Tomás, and some of the important details of the story. You may rerecord yourself as many times as you like if you get interrupted. Don't forget to press the "submit" button after you're done recording.

“El Secreto” from VanPatten’s (2018) Cuentos Cortos Volume 1

Table B1 Target features examined in the ONT

Feature	Which code it was present in	Type of error
Appropriate copula	both	Morphosyntactic
Gender agreement in NP	both	Morphological
Word choice	both	Lexicosemantic
Conjunction usage	Cir's	Syntactic
Preposition usage	Cir's	Syntactic
Reflexive pronoun	Cir's	Morphosyntactic
Stress features	Cir's	Phonological
Determiner use (non-gender related)	Eric's	Morphosyntactic
Object pronoun errors	Eric's	Syntactic
Pronoun dropping	Eric's	Syntactic
Radical-changing verb	Eric's	Morphological
Tense, aspect, mood, person, and number (TAMPN) agreement in verb phrase (VP)	Eric's	Morphological
Thematic vowel usage	Eric's	Morphological
Within-noun gender problem	Eric's	Lexical

Table B2 Summary of attempts by feature in Cir's ONTs

Feature	Week 0	Week 1	Week 2	Week 3	Week 4
Appropriate copula	1	1	4	2	3
Gender agreement in NP	6	8	12	20	21
Word choice	51	58	95	120	129
Conjunction usage	0	3	3	7	4
Preposition usage	4	1	10	7	7
Reflexive pronoun	1	1	1	3	3
Stress features (on a word-level)	51	58	95	120	129

Note. The categories of Word choice and Stress features (on a word-level) present the same number of attempts because they both pertain to word-level choices in production, whether suprasegmental or lexical. Correct suppliance, however, varied between the two features because they tested different constructs.

APPENDIX C

Metalinguistic knowledge task materials

1. Martha hablan tres idiomas: español, italiano y alemán.
   Correction:
   Explanation:

2. Juan no come carne, ella es vegetariano.
   Correction:
   Explanation:

3. Por favor señora, abrir la ventana pues hace mucho calor.
   Correction:
   Explanation:

4. Si tengo dinero, me compraría un coche.
   Correction:
   Explanation:

5. Quiero aprendo a nadar.
   Correction:
   Explanation:

6. - Ha salido Laura?
   - !Qué va!, lleva tres días de estudiar.
   Correction:
   Explanation:

7. La Ciudad de México está más contaminada de Londres.
   Correction:
   Explanation:

8. Hace un año, Andy pasaba seis semanas estudiando inglés en Colchester.
   Correction:
   Explanation:

9. At a doctor's surgery:
   Paciente: Buenos días.
   Recepcionista: Buenos días.
   Paciente: Sería posible hacer una cita con el Dr Ruíz.
   Recepcionista: Sí señora, ¿cómo te llamas?
   Paciente: Ana Martínez.
   Recepcionista: ¿Cuándo quiere la cita?
   Paciente: Tan pronto como sea posible.
   Correction:
   Explanation:

10. Mi hermano juga fútbol los domingos.
    Correction:

Explanation:

11. - ¿Quieres ir al mercado?
    - Ahorita no, está muy caliente. Prefiero quedarme en casa.
    Correction:
    Explanation:

12. La posibilidad de que hubo un cambio político en México provocó la intervención Estadounidense.
    Correction:
    Explanation:

13. Juan te ganó la lotería.
    Correction:
    Explanation:

14. Fuí al mercado para cuatro litros de leche.
    Correction:
    Explanation:

15. El domingo llevaré los niños al cine.
    Correction:
    Explanation:

16. The following exchange takes place after a meal in a restaurant:
    Mesero (waiter): Ha estado todo de su agrado señor.
    Comensal (customer): Sí, muchas gracias, me puede tener el billete, por favor.
    Correction:
    Explanation:

17. Ayer conocí a la esposa de mi jefe y me gustó bien.
    Correction:
    Explanation:

18. Los Beatles están muy famosos en todo el mundo.
    Correction:
    Explanation:

19. Christmas card: „Feliz Navidad y contento Año Nuevo"
    Correction:
    Explanation:

20. A mi novio le gusta los chocolates finos.
    Correction:
    Explanation: