

Review paper

A review of mobile language learning applications: trends, challenges and opportunities

Catherine Regina Heil*, Jason S. Wu**, Joey J. Lee***
Teachers College, Columbia University, USA

Torben Schmidt****
Leuphana University of Lüneburg, Germany

*crh2139@tc.columbia.edu | **jasonwu@columbia.edu | *** j13471@tc.columbia.edu |
**** torben.schmidt@leuphana.de

Abstract

Mobile language learning applications have the potential to transform the way languages are learned. This study examined the fifty most popular commercially-available language learning applications for mobile phones and evaluated them according to a wide range of criteria. Three major trends were found: first, apps tend to teach vocabulary in isolated units rather than in relevant contexts; second, apps minimally adapt to suit the skill sets of individual learners; and third, apps rarely offer explanatory corrective feedback to learners. Despite a pedagogical shift toward more communicative approaches to language learning, these apps are behaviorist in nature. To better align with Second Language Acquisition (SLA) and L2 pedagogical research, we recommend the incorporation of more contextualized language, adaptive technology, and explanatory feedback in these applications.

Keywords: Mobile-Assisted Language Learning (MALL), Communicative Language Teaching (CLT), adaptive learning, vocabulary instruction, grammar instruction, corrective feedback, assessment.

1. Introduction

A remarkable number of people are turning to their mobile devices to learn a foreign language. The global market for digital English language learning products, for example, reached \$1.8 billion in 2013. Revenues are projected to surge to over \$3.1 billion by 2018, with a compound annual growth rate (CAGR) over a five-year period of 11.1% (Adkins, 2008). Language learning apps like *DuoLingo* are immensely popular, with over 70 million sign-ups (Hickey, 2015). Mobile language learning approaches are clearly in demand and will continue to grow in use as more people turn to smartphones or tablets as a primary computing device.

The rise of mobile app usage for language learning raises an important question: are current commercial mobile language learning apps *effective* tools for language learners, based upon what we know about research in L2 pedagogy, pedagogical design, and Second Language Acquisition (SLA) research? And further, given this information, how can the state of commercial applications inform academic research and vice versa? While the pedagogical uses and new opportunities of mobile technology for language learning have been studied in academic contexts, existing commercial mobile language learning apps have not been systematically evaluated and characterized.

In this paper, we conduct and provide a comprehensive and systematic review of the fifty most popular language learning apps available for iOS and Android phones as of Spring 2015. This sampling provides a broad characterization of the state of apps that are being used for mobile language learning. An analytical protocol was developed to

investigate the following questions regarding areas of instruction, assessment, and feedback. Specifically, we investigated:

- What are the primary pedagogical focuses of popular language learning apps?
- Do apps adapt to individual needs, language proficiency levels, and styles of learning?
- How is corrective feedback employed in these apps?

Before attempting to answer these questions, we begin with a brief review of existing literature and our theoretical framework. We then describe our methodology for sampling and analytical coding. Finally, we present our results with a discussion of major trends and our recommendations for the field.

2. Literature review

Research in MALL has largely been mediated by technological development. Early applications made use of portable audio devices such as the Sony Walkman or Apple iPod (Godwin-Jones, 2011). Early internet-capable devices such as cell-phones and personal digital assistants (PDAs) made basic use of email and web browsing for language learning (Chinnery, 2006). Pedagogical approaches were fairly limited on these devices, constraining most applications to one-way content delivery with little peer-to-peer communication or interaction (Kukulska-Hulme & Shield, 2007; Kukulska-Hulme & Shield, 2008).

Published MALL studies increased dramatically in 2008 (Duman, Orhon, & Gedik, 2015). Coinciding with the emergence of smartphone technology, applications began to make greater use of web-based activities (e.g. Nah, White, Rol, & Sussux, 2008; Stockwell, 2008). Since then, mobile technology has grown in sophistication, resulting in the release of a large amount of language-learning software. There are over a million apps available to users in both the Google Play and Apple iTunes app stores; educational apps comprise 9.95% of this total (Statista Inc., 2015). The number of language learning apps has been estimated to be as high as 1,000 to 2,000 in total (Sweeney & Moore, 2012).

Despite rapid growth in app numbers, MALL research has been criticized for a lack of objective, quantifiable learning outcomes. Burston (2015) conducted a meta-analysis of 291 MALL studies spanning 20 years, and found only 35 were of sufficient duration (1 month) and involved a minimal number of subjects –ten. Burston also noted that many of the studies were afflicted by inadequate research design due to failure to address confounding variables that exist outside of the device itself –novelty effects, content, the instructor, etc.– perhaps due to an overly “technocentric” approach that overemphasizes the role technology plays in learning.

Shortcomings aside, the positive reports of many of these MALL studies support the notion that mobile devices are efficacious learning tools - in particular for vocabulary instruction. In Duman, Orhon and Gedik’s (2015) literature review of research trends in MALL from 69 studies from 2000-2015, “teaching vocabulary” was the most popular topic, addressed by 28 of those studies; conversely, only one study examined grammar instruction and writing. Likewise, Burston (2015) noted that 58% of the 291 MALL studies examined were concerned with vocabulary acquisition, most of which reported positive learning outcomes (2015, p. 12). Burston also noted positive reports for vocabulary learning, reading competency, listening, and speaking skills across the studies.

An important concept that has emerged recently is the notion of *adaptive learning*, which uses computers as personalized teaching devices. Adaptive learning proposes a softer version of the artificial intelligence driven systems proposed by early research in Computer Assisted Language Learning (ICALL), developments that would heavily rely on improved natural language processing, and the computer’s ability to extrapolate meaning from speech (Warschauer & Healy, 1998). Kerr (2013) predicts a move away from traditional textbooks and towards interactive adaptive learning platforms (p. 18), with both an incorporation of more gamified elements and the use of big data and analytics to store content about users.

3. Theoretical framework

In making sense of what types of instructional design are most effective, the contributions of SLA and pedagogical research are indispensable. As Kukulska-Hume and Bull (2009) observe, "There is a large body of research on many aspects of second language learning, but often much of the relevant theory and empirical findings are overlooked by developers of language learning technology support" (p. 1). Reinders and Pegrum's (2016) framework for evaluating mobile apps notes the importance of discussing findings of both SLA and pedagogy when evaluating applications. SLA has core requirements: "the need for comprehensible input, comprehensible output, negotiation of meaning in interaction, and noticing of new language, the last of which can be promoted through effective feedback" (p. 6). Without these rudimentary components, it is challenging for learners to truly gain communicative competence in the target language.

Theoretical models of language knowledge (e.g. Canale and Swain, 1980; Bachman and Palmer, 1996; Purpura, 2004) tease apart the differing components into a number of categories, such as grammatical knowledge, pragmatic knowledge, discourse knowledge, functional knowledge, and sociolinguistic knowledge, among others. To gain communicative competence in a language, one must develop a multifaceted range of knowledge; simply knowing words is insufficient. Pedagogical approaches to app development ought also to take this into consideration when determining what content to include, and how to assess learners, especially if the intention is to teach learners language and not just to teach learners words.

Classical methodologies for classroom language teaching, such as the grammar translation method popular in the 1950s, have been characterized as *behaviorist* in nature, as they call upon skills such as memorization, drilling practice, and repetition (Brown, 2007). The behaviorist model posits that learning occurs as a result of stimulus-response associations, which build in learners a repository of knowledge that can be strengthened or weakened based on the frequency of reinforcement or inattention (Fosnot & Perry, 1996). Language knowledge is objectively attainable, and exists outside of the learner; the role of the teacher is to help to develop and strengthen associations to words and grammatical rules. Though behaviorism has seen a resurgence in popularity and is certainly not without its merits, especially in language learning, it may be, on its own, insufficient to characterize how language is learned. "Missing from this perspective [...] is any treatment of the underlying structures or representations of mental events and processes and the richness of thought and language" (Pellegrino, Chudowsky, & Glaser, 2001, p. 62). Behaviorism misses the social element, the notion that language use is a fundamentally communicative act.

In contrast to behaviorism, a *constructivist* theory of learning, often attributed to thinkers including John Dewey, Lev Vygotsky and Jean Piaget, rejects the idea that "human knowledge is a direct reflection of an objective reality" (Blyth, 2007, p. 3). In other words, constructivism is rooted in an epistemological framework that denies the existence of a singular, objective truth that can somehow be transmitted from teacher to student. Knowledge is acquired by processes that blend the learner's pre-existing knowledge framework, acquired through years of development and experience, with that encountered in social contexts; "The individual learns by being part of the surrounding community and the world as a whole" (Oxford, 1997, p. 445). As such, learning a language is viewed as a social activity.

This study emphasizes the notion that language is a tool for communication with instrumental rather than ends-based value. Simply knowing words and structures does not itself enable a learner; rather, it is one's ability to *use* them meaningfully that makes them valuable. This idea, often referred to as the learner's *communicative competence* (Hymes, 1972), can be thought of "in terms of the expression, interpretation, and negotiation of meaning" (Sauvignon, 2002, p. 1) rather than mastery of words and forms. Or as Ur (2013) states, it requires a focus on "use" and not only "usage" (p. 2). This important distinction guides much of our analysis and discussion.

With this in mind, we consider what values are embodied by the apps that are easily accessible on mobile phones. There are many ways to learn a language, and varying degrees and definitions of what it means to be “proficient.” Many language learners find that a combination of drilling and communicative practice lead to communicative competence. Other learners may not intend to be fluent in a language, but perhaps only intend to learn some vocabulary. Our aim is to characterize apps currently available and to make recommendations that may help guide their future development.

4. Methodology

4.1. Research design

This study examined fifty of the top commercial apps for Apple iOS and Google Android mobile phones, employing an *exploratory-qualitative-interpretive* approach (Grohtjahn, 1987). According to this approach, apps were selected and coded according to a grounded set of criteria, and data were analyzed to determine the most relevant trends and characteristics.

4.2. Selection of apps

Fifty apps were selected on the basis of their rankings on Google Play and in the Apple iTunes App Store by searching for the key phrase “language learning”. App rankings were used for selection as they represent a metric for the most popular apps a typical user might find upon searching for “language learning.” While the exact algorithms used by Google and Apple to calculate these rankings are not disclosed to the public, they are roughly based on the total number of downloads, reviews, and income earned from sales (Edwards, 2014).

The app analytics engine App Annie (App Annie, 2015) was used to identify and compile a list of the top 50 apps in both stores as of March 2015. App Annie, though not directly affiliated with Apple or Google, collects information from users and uses it to estimate rankings of apps. Apps holding multiple rankings for different languages were considered as a single app and were only included once. Some apps were excluded due to irrelevance to the research questions, such as those that teach computer programming languages or those that focused solely on translation. A full list of apps included in this survey may be found in *Appendix A*.

4.3. Instrument design and coding

The survey instrument was carefully constructed during initial testing in order to answer our primary research questions. Questions on the survey were designed to capture a broad range of aspects. Topics covered included: languages taught, operating system, monetization, areas of assessment, modes of grammar instruction, corrective feedback, and types of input and output to the device. The final instrument resulted in 24 questions covering 149 subcriteria using selected-response checkboxes.

It is important to note that subcriteria were not typically mutually exclusive, allowing for multiple selection of subcriteria under a particular question. For example, a single app may be coded for both implicit and explicit grammar instruction, if it contains features of both. However, when an app is coded for “None” as a subfeature, it was not coded for any additional features.

An overview of the questions and subcriteria are presented in Table 1. The survey instrument is presented in *Appendix B*.

4.4. Data collection and reliability

Prior to data collection, a norming session was held to ensure coders were selecting criteria in a similar fashion. Four coders in total examined the apps. During the process of data collection, the coders met on a weekly basis to discuss any issues related to coding. Eleven apps were randomly selected for coding by two raters, providing a sample for reliability analysis. Cohen’s Kappa was calculated and questions with low reliability ($\kappa < .60$) were not included for analysis. For the questions presented here, Kappa ranged from $\kappa = .629$ ($p < .015$) to $\kappa = 1$ ($p < .0005$) with an average of $\kappa = .865$.

Question Topic	Subcriteria & Explanation
Languages	Languages supported by the app were manually entered by the coders.
Platforms	Possible platforms: Android, iOS, Windows Phone, Blackberry
Monetization	<p>None - No apparent monetization scheme</p> <p>Pay to Unlock - User pays a flat fee to access languages or levels</p> <p>Subscription - User pays a recurring fee to access content</p> <p>In-App Advertisements - Advertisements placed throughout the app</p>
User Input to Device	<p>Touch Gestures - User touches the device to provide input</p> <p>Writing on Keyboard - User writes on the device keyboard</p> <p>Speaking into Microphone - User speaks into the microphone on the device</p>
Areas of Instructional Assessment	<p>The areas of instruction were examined based upon areas of language ability that were <i>assessed</i> by the application. Thus, the user would need to be tested on their ability to use the following features when interacting with the application.</p> <p>Vocabulary in Isolation - User ability to select, write, or speak individual words without placing them into the context of other words</p> <p>Vocabulary in Context - User ability to select, write, or speak words or sentences that have been placed into the context of other words</p> <p>Grammatical Form - User demonstrates knowledge of morphosyntactic form and/or sentence structure in clauses</p> <p>Pragmatics - User demonstrates understanding of situational use of certain expressions over others</p> <p>Pronunciation - User demonstrates ability to appropriately pronounce words</p> <p>No Assessment - No explicit measures taken to assess learner input to device</p>
Modes of Grammar Instruction	<p>Implicit - User must deduce understanding of grammatical forms. No explicit coverage of grammar or metalinguistic terminology included</p> <p>Explicit - Grammar Presentation - Grammar explicitly referenced by the app in the form of explanations about grammatical features prior to assessment</p> <p>Explicit - Grammar Feedback - Grammar explicitly referenced in feedback provided to learners during interaction</p> <p>None - Grammar addressed neither explicitly nor implicitly; apps teach words in isolation, therefore do not address grammar</p>
Corrective Feedback	<p>Sound Effects - A sound indicates correctness of answer</p> <p>Visual Feedback - A visual stimulus indicates correctness of answer</p> <p>Textual Corrections - A short textual correction is provided when an answer is incorrect</p> <p>Textual Explanations - A textual explanation indicating rationale for correctness of answer is provided</p> <p>None - No feedback is provided on correctness of answer</p>
Listening, Reading, & Writing	<p>Output and input in the form of letters and text, as read, written (either by selecting or typing on the keyboard), or heard by the user. Textual input and output were categorized according to length and type:</p> <p>Letters - Individual letters</p> <p>Words - Individual words and phrases not in sentences</p> <p>Sentences - Complete sentences</p> <p>Passages - Any text a paragraph or longer</p> <p>Dialogues - A conversation between two or more speakers</p> <p>Songs - Any text set to music</p>

Table 1. Overview of question topics and subcriteria assessed with survey instrument.

5. Results

Below we highlight findings which provide an overview of currently available language-learning apps and address our three primary research questions.

5.1. Languages supported

Most of the selected apps taught multiple languages. The top ten languages taught were English (36 of 50 apps, 72%), French (36 of 50 apps, 72%), Spanish (34 of 50 apps,

68%), German (33 of 50 apps, 66%), Chinese (28 of 50 apps, 56%), Italian (27 of 50 apps, 54%), Japanese (25 of 50 apps, 50%), Portuguese (21 of 50 apps, 42%), Russian (21 of 50 apps, 42%), and Arabic (19 of 50 apps, 38%). Twelve of the selected apps taught only a single language; one app taught a maximum of 200; the mean number of languages taught per app was 15.1.

5.2. Platforms supported

While 25 of the apps selected were from the Apple Store (for iOS) and 25 were from the Google Play store (for Android), some of these apps were compatible with multiple platforms. Many Android apps were also available for iOS and vice versa. The total percentages were: iOS (40 of 50 apps, 81%), Android (34 of 50 apps, 69%), Windows Phone (5 of 50 apps, 8%), and Blackberry (2 of 50 apps, 3%).

5.3. Monetization

The majority of apps (29 of 50 apps, 64%) included a “pay to unlock” feature requiring users to pay a flat fee to access additional levels or languages. Other forms of monetization included a subscription payment system (7 of 15 apps, 15%) and in-app advertisements (11 of 50 apps, 23%). Only a minority of apps (6 of 50 apps, 14%) had no apparent monetization scheme.

5.4. User input

While all apps used touch gestures, 16 of 50 (32%) included writing words using an onscreen keyboard and 12 of 50 (24%) allowed the user to speak into the device using the microphone.

5.5. Assessment and instructional focus

Our first research question asks about the focus of instruction in individual apps. In order to determine intended instructional focus, we examined which language areas were being assessed by each app. Our rationale is that assessment reveals which aspects of language are being taught and emphasized (Figure 1). We looked at a variety of models of L2 communicative language ability (Canale and Swain, 1980; Bachman and Palmer, 1996; Purpura, 2004), and found that areas assessed could be divided into vocabulary instruction (whether isolated or in context), grammatical form, pragmatics, and pronunciation.

The majority of apps (42 of 50, 84%) included a focus on vocabulary items as isolated units, that is, as individual words without context. Just over half of the apps (23 of 50, 53%) assessed vocabulary in context. Other apps focused on grammatical form, pragmatics, and pronunciation. 5 of 50 apps (10%) did not offer a formal means of assessment; rather, they focused only on delivering content, either in the format of written phrasebooks or audio lessons.

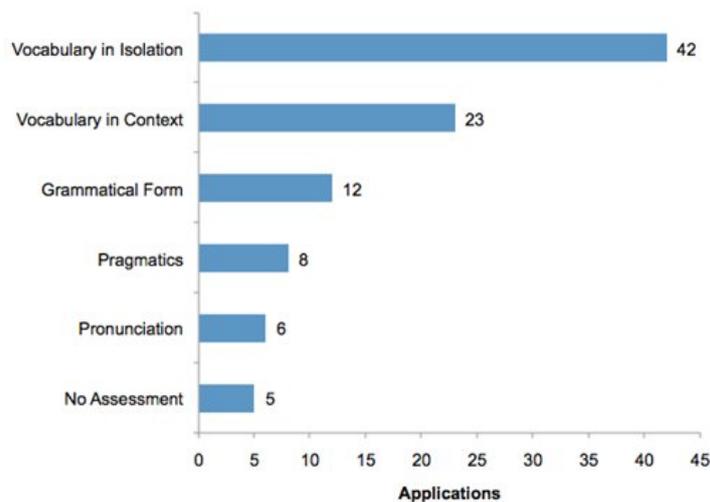


Figure 1. Areas of assessed instructional focus in language learning apps.

5.6. Implicit and explicit grammar instruction

Implicit grammar requires users to make inferences about grammatical form and meaning without the use of any metalinguistic terminology. Explicit grammar instruction was classified as either direct presentation of grammatical rules to the user, or corrective feedback that made explicit references to grammatical errors made by the user (Figure 2).

In many apps (21 of 50, 42%), no grammar instruction was evident; this typically occurs when apps assess individual vocabulary items without context. In the remaining 29 apps that did include grammatical instruction, feedback was coded as implicit or explicit. Some apps were coded for both as they contained both implicit and explicit styles of instruction. A sizeable group (19 of 50, 38%) included an implicit grammar instruction approach. A smaller number of apps (10 of 50, 20%) provided an explicit grammatical presentation to users, whereas only 3 of 50 apps (6%) provided feedback that made explicit reference to specific grammatical errors made by the user.

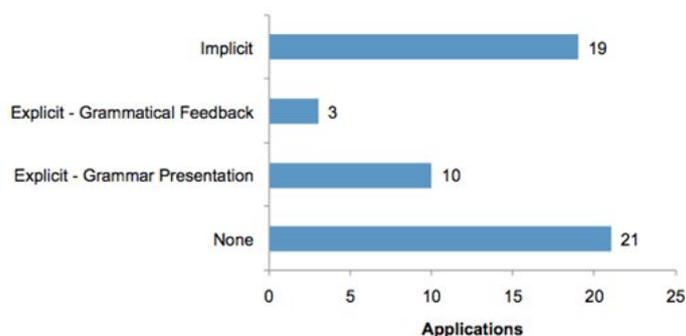


Figure 2. Implicit versus explicit instruction in language learning apps.

5.7. Corrective feedback

Corrective feedback occurs when an app assesses the user's language input and provides correction when necessary (Figure 3). The most common types of feedback given are visual (41 of 50, 82%) or sound effects (32 of 50, 64%). Some apps (14 of 50, 28%) offered simple textual corrections (i.e. providing the correct answer in the place of the wrong answer), yet only 3 of 50 apps (6%) provided any explanation as to why certain mistakes that were made were incorrect.

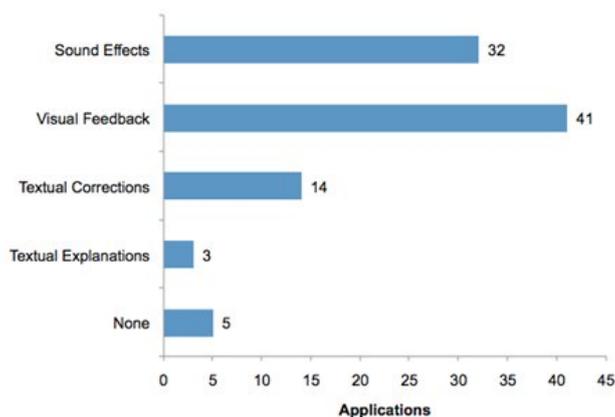


Figure 3. Corrective feedback in language learning apps.

5.8. User interaction - listening, reading and writing

We also examined the frequency and types of user interaction (listening, reading, or writing) with the apps, and categorized these by the level of language involved (e.g. words, sentences or passages) (Figure 4). Writing included typing via onscreen keyboard, selecting letters to form words, and words to form sentences.

Users most often interact with language on the word or sentence level when listening, reading, and writing on a mobile device. Writing is the most underutilized skill in comparison to listening and reading. In a small number of apps emphasizing spelling, letters were occasionally targeted for listening, reading, or writing. Longer forms of input and output, such as songs, dialogues, and passages, were very rare in all skill areas. Apps tended to focus on receptive skills such as listening or reading combined with simple activities like fill the blank or drag & drop, rather than productive skills, like speaking or text production. Open-ended activities were rare, and written or spoken production was generally limited to very simple one word utterances, allowing for the app to easily assess input and provide corrective feedback.

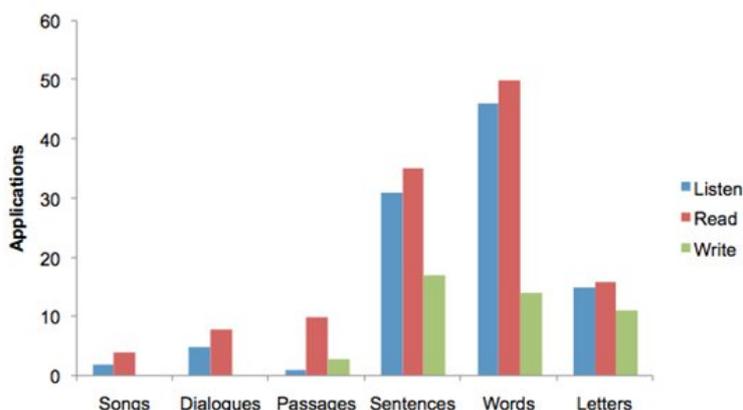


Figure 4. User interaction – listening, reading and writing.

6. Discussion

From our analysis, three major trends were found. First, the majority of apps tend to teach vocabulary units in isolated chunks rather than in relevant contexts. Second, many apps tend not to adapt to suit the skill sets of individual learners. Third, current apps tend to offer minimal explanatory corrective feedback to learners. These findings provide areas of focus for next-generation language learning apps.

6.1. Vocabulary instruction

Our results showed that vocabulary instruction was the main instructional focus of apps –and in some cases, the only instructional focus. In 84% of apps (42 out of 50), vocabulary was taught in isolation, while only 23 of 50 apps (53%) taught vocabulary in context. An example contrasting *vocabulary units in isolation* versus *vocabulary units in context* is depicted in Figure 5. A common activity used to assess vocabulary in isolation was to match images to meanings of words. Oftentimes these activities were gamified through time constraints or aesthetics, such as an activity from Mindsnacks Spanish (Figure 5, left). In this activity, the user must fill up a frog’s belly by identifying the image that matches a given word in order to provide the frog with a snack. In contrast, activities such as the “cloze” test from DuoLingo (Figure 5, center) and Voxy (Figure 5, right) assess vocabulary in context. While the Mindsnacks game combines visually-appealing images with music and sound, the user is not provided any textual environment for the words, but rather matches words to pictures.

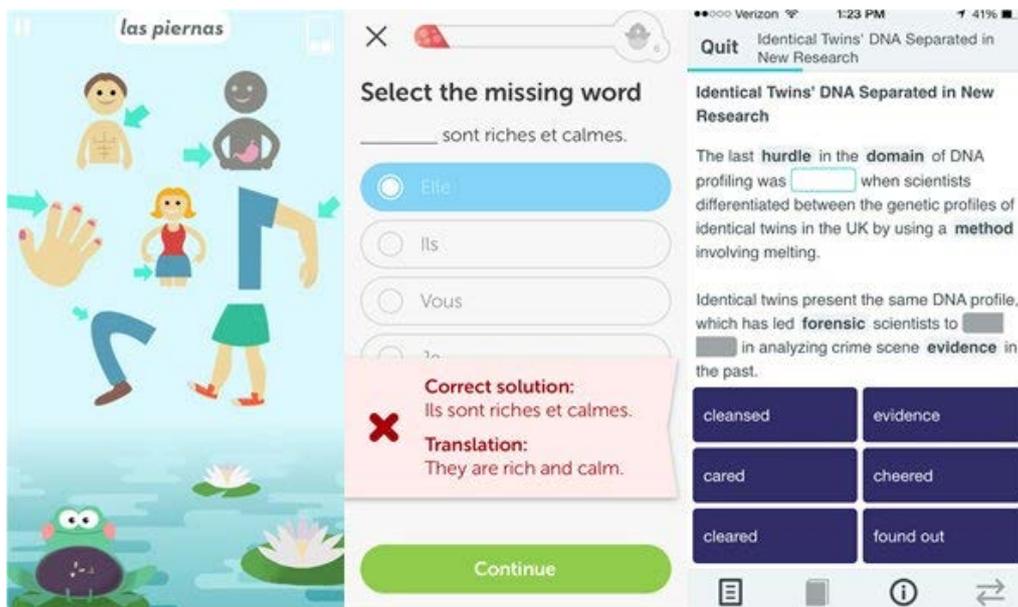


Figure 5. Exercises contrasting vocabulary in isolation, as in in MindSnacks Spanish (left), versus vocabulary in context, as in Duolingo (center) and Voxy (right).

Context plays an important role in language learning. New contexts for lexical items allow learners to enrich knowledge of that word by understanding varied senses of meaning. The more times one comes across a word in a different context, the better understanding one has of both the immediate and extended senses of the word. Additionally, Nation (2015) has noted that vocabulary knowledge is a function of the number of times one is exposed to a word as well as the quality of each meeting. The attention given to the word can either be incidental or deliberate. While all of these apps draw deliberate attention to the vocabulary units in question, context provides additional means for learners to strengthen their vocabulary knowledge through incidental repeated exposure to new words.

Many of the reading contexts were limited to sentences and not full reading passages. Only 8 of 50 apps, (16%) called for the user to read dialogues and only 10 of 50 apps (20%) included reading passages (textual content longer than a sentence), such as the one from Voxy (Figure 5, right). While some developers might dismiss the idea of including longer reading passages due to limited attention span of users related to the portable nature of phones, positive learning outcomes have been reported by users (Wang and Smith, 2013; Chen & Hsu, 2008; Wu et. al. 2011). Such activities are encouraged as they would provide learners with a means to situate vocabulary in authentic and meaningful texts, and thus be able to recognize when and how to apply them in the future.

When vocabulary is taught in a flashcard style –matching word to meaning (whether represented textually, or visually, as in the Mindsnacks game above)– learners may improve their knowledge of the immediate or central sense of a word, the literal, or lexical meaning (Purpura, 2003). However, the interactional or pragmatic meaning of the word is not addressed, meaning that learners will not fully understand the appropriate contexts for use of the word. Additionally, a focus on literal meaning means that users will miss out on understanding other senses of the word, such as the morphosyntactic form, which includes “articles, prepositions, pronouns, affixes, syntactic structures, word order, simple, compound, and complex sentences, mood, voice, and modality” as well as the morphosyntactic meaning, which allows us to understand the word in relation to time, negation, to show focus, contrast, and attitude (Purpura, 2003, p. 94). A user may know a verb, but have no idea how to conjugate it or put it in a sentence.

By teaching vocabulary in context, some grammatical information is typically deduced rather than taught explicitly. In the example of Duolingo (Figure 5, center), the user is asked to select the appropriate pronoun to complete the sentence from a list of options.

This task additionally assesses understanding of grammatical form by requiring user knowledge of subject-verb agreement. However, the user still has to make inferences about the correspondence of pronouns in French and pronouns in English. The user must be able to infer that “they” is the third person plural; this information is not explicitly stated.

In contrast, apps such as Babbel provide more explicit grammar instruction, where users are given metalinguistic information about words as they are acquired. While learning the personal pronoun “tú,” for instance, the user is provided some clues: “sg., informal.” in 38 of 50 apps, 42% of cases, no grammar instruction was evident at all, either implicit or explicit, and most often this was because of a lack of context for words due to a vocabulary-drill-only approach.

Of the apps that did include a focus other than vocabulary instruction, 18 of 50 (36%) of apps included an *implicit* grammar instruction approach, and 12 of 50 (24%) provided *explicit* instruction, in which users were coached to understand grammatical meaning. The remaining 20 apps were coded as having no grammatical instruction. There are benefits and drawbacks to both approaches, and learning style will no doubt factor into a preference for inductive or deductive learning. While implicit grammar instruction may be beneficial in that it allows learners to take ownership of their learning discoveries, it may also cause learners to make incorrect assumptions about grammar. Explicit grammar instruction is challenging given the constraints of the mobile device, such as screen and file sizes, but it may detract learners from a focus on fluency. It is likely that a combined approach is most ideal.

Ultimately, a design focused solely on drilling isolated vocabulary units represents a one-dimensional approach to language learning. There is wide recognition that vocabulary is only one component in models of language ability (e.g. Canale and Swain, 1980; Bachman and Palmer, 1996; Purpura, 2004). Therefore, if these apps intend to instruct in a more holistic way, it is essential to move beyond vocabulary drilling.

6.2. Adaptation

One of the greatest advantages of software-as-teacher, as compared to human-as-teacher, is that software possesses the potential to record complex user input in a precise, reliable manner. While a teacher may not remember every error that a student generates, software, if developed properly, could provide invaluable formative information that would otherwise be too substantial for a human to plausibly record. This ability for software to automatically update its functionality based on input received or data processed is known as adaptive learning. While growing in popularity, it is still a largely unexplored arena in mobile language learning applications.

Machine learning has been incorporated into the field of educational technology via Intelligent Tutoring Systems (ITS), or more specifically, Intelligent Language Tutoring Systems (ILTS), which offer users a way to interact with a computer by individually adjusting the sequence of instruction based upon user input (Gamper & Knapp, 2002; Moundridou & Virvou, 2003, Stockwell, 2007). An ITS system would be able to make “intelligent” decisions, such as adjusting the level placement of the user based on their performance, determining which areas require additional exercise to compensate for weaknesses, modifying settings to appropriately scaffold content based on the skill level of the user, or even changing visual cues in order to better motivate. The screenshots from Mondly, Memrise, and Mindsnacks shown in Figure 6 display performance analyses shown upon user completion of levels. In some instances, these data are used to motivate the user to improve their performance, but are only minimally used to adjust the level of gameplay to match the level of the user. For instance, in Mondly (Figure 6, left), the user obtains experience points (XP) for completing levels, and users can log in via Facebook to compare their XP level to other users. This allows progress to be tracked from level to level, but nonetheless the path from level to level remains the same regardless of the user.

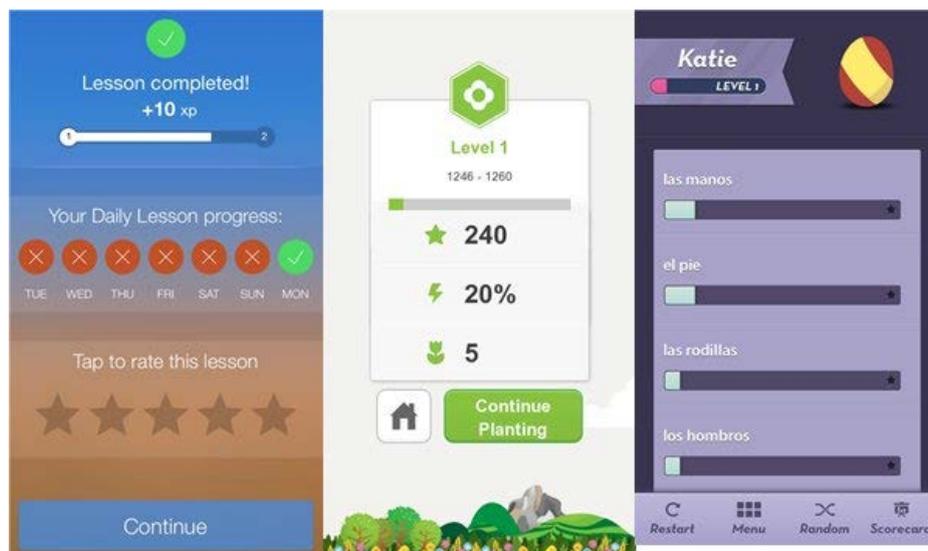


Figure 6. Performance analyses provided by Mondly (left), Memrise (center), and Mindsnacks (right).

We believe that the information collected by apps ought to be used *formatively*, rather than displayed as a summative performance analysis. Just as teachers adjust their explanations to suit the needs of their students, apps should adjust their content to suit the needs of users. To accomplish this, results ought to be used by machine learning algorithms to adjust functionality accordingly. By coding into language apps the types of grammar mistakes that users make while practicing on the app, it would be possible to identify the frequency of different types of learner errors. Presenting this information to the learner could lead them to notice mistakes that would otherwise go undetected; for instance, they might realize that they frequently replace present perfect for past tense forms, or that they tend to drop certain endings. Using machine learning algorithms, apps could adjust activities based upon the rate of various errors present, allowing users to spend more time practicing those forms that are appropriately challenging to the learner, making gameplay more intriguing, less routine, and more likely to result in learning outcomes.

While this feature was not readily apparent in any of the apps that were coded for grammatical instruction, a similar adaptive feature was noted in apps that teach vocabulary. For instance, both Memrise (Figure 6) and Mindsnacks (Figure 6, right), apps for vocabulary instruction, exemplify adaptive learning in vocabulary instruction. These two apps determine mastery based upon how many times a user has answered a question containing a given vocabulary word correctly. Memrise uses machine learning technology to continue asking the user questions on words that have not yet been mastered. In Mindsnacks, a series of bars indicating the user's mastery of a list of words is displayed on the screen at the end of each level. The program then increases the frequency of the most challenging words for the user in future tasks.

This movement from simple to complex tasks (or an increase in the frequency of challenging words) is compatible with both behaviorist and constructivist approaches, with a caveat. While a behaviorist approach might emphasize strengthening through repetition and increases in frequency, a constructivist approach would emphasize strengthening through *understanding* of ideas. As constructs have social origins, and "people construct experience according to the organization of the cognitive system [...]" A corollary is that ICALL must teach learners all the metacognitive tools necessary for appropriate self-regulation" (Oxford, 1998, p. 362). Combining this adaptability with better feedback, which will be described in the next section, is more likely to provide learners with the necessary tools to understand and improve their performance.

6.3. Feedback

While there is much debate about the best way to deliver feedback to learners, many studies in second language acquisition have revealed the efficacy of explicit metalinguistic feedback (e.g. Carrol & Swain, 1993; Lyster & Ranta, 1997; Ellis,

Loewen, & Erlam, 2006). Knowing that an utterance is ungrammatical (i.e. having “negative evidence”) is important, but knowing *why* this is the case further enables the learner to avoid making these mistakes in the future, and also avoids the pitfalls of the behaviorist tendency to essentialize and overlook the quality of knowledge gained. As Pellegrino, Chudowski, and Glaser (2001) have noted: “Whereas [...] the behaviorist approach focuses on how much knowledge someone has, cognitive theory also emphasizes what type of knowledge someone has. An important purpose of assessment is not only to determine what people know, but also to assess how, when, and where they use what they know” (p. 62).

Typically, it was found that feedback in apps was most often given through visual clues such as color changes or highlights (40 of 50 apps, 82%), or through the use of sound effects (31 of 50 apps, 63%). Only 14 of 50 apps (28%) offered any textual feedback, and an even lower 3 of 50 apps (6%) offered explanations to users about why their choices may be incorrect. Our analysis revealed that apps have done a poor job at providing useful feedback to users. Without additional information from apps about why users are making mistakes, the likelihood that these activities will result in learning is diminished.

Many ITS systems include an NLP pipeline in which different modules are systematically executed –such as tokenization, part-of-speech tagging, lemmatization, parsing, etc.– in order to interpret user input to the device. This functionality would equip apps with the power to make better decisions based upon the text –for instance, knowing that the user has typed the correct word, but perhaps the wrong form. Or at the sentential level, knowing that the user has typed the correct words, but, for example, has placed an adjective in the incorrect place with respect to the noun it modifies. If the computer is able to actually comprehend and process user input, it would be much easier to provide feedback that is uniquely tailored to users and their particular types of errors.

Without the ability to parse words, the skill of writing is generally neglected in comparison to listening and reading. Only 13 of the 50 apps (26%) allowed users to write full words. We would recommend the incorporation of more adaptive technology that can understand what types of mistakes users are making, and thus provide more intelligent, personalized feedback.

7. Conclusion

Our review has shown that, in the commercial app space, there is a predominant focus on teaching language as isolated vocabulary words rather than contextualized usage. Most use drill-like mechanisms and offer very little explanatory corrective feedback, and there is little adaptation to the needs of individual learners. Despite advances in language teaching that have stressed the importance of communicative competence in language learning, MALL technology is still primarily utilized for vocabulary instruction rather than fluency-building.

This paper examined commercial applications; nonetheless, given the influence of academic research on commercial MALL application, the relevancy of these suggestions need to be considered. The focus on vocabulary instruction is prevalent in MALL research, as noted, but more focus on adaptive learning and intelligent design features in applications –especially those which highlight learning outcomes– would be useful target areas for future research.

Overall, there is great opportunity to leverage emerging technologies for language learning; we suggest a stronger emphasis on intelligent commercial app design. By providing more contextualized, authentic written input, users will begin to process more than individual words and basic vocabulary. The incorporation of more adaptive learning features would provide a more personalized experience, both in terms of content delivered during instruction as well as feedback. NLP technologies could allow for more accurate recognition of written text. Such a design methodology would teach authentic usage of language with an end-goal focus of making learners communicatively competent in the language they intend to learn. In this way, language educational technology can move past “drill and kill” behaviorist-style instruction that has long-since been abandoned in language classrooms, and turn toward a more communicative,

holistic model that reflects our current understanding of language ability and acquisition.

References

- Adkins, S. S. (2008). The US Market for Mobile Learning Products and Services: 2008-2013 Forecast and Analysis. *Ambient Insight*, 5.
- App Annie (2015). <http://www.appannie.com/top/>.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Blyth, C. (2007). A constructivist approach to grammar: Teaching teachers to teach aspect. *The Modern Language Journal*, 81(1), 50-66. Retrieved from <http://www.jstor.org/stable/329160>.
- Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy*. White Plains, NY: Pearson Education.
- Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL*, 27(01), 4-20.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback. *Studies in second language acquisition*, 15(03), 357-386.
- Chen, C-M and Hsu, S-H. (2008) Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education*, 51(2), 624-645.
- Chinnery, G. (2006). Going to the MALL: Mobile Assisted. *Language Learning, Language Learning & Technology*, 10(1), 9-16.
- Duman, G., Orhon, G., & Gedik, N. (2015). Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL*, 27(02), 197-216.
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in second language acquisition*, 28(02), 339-368.
- Edwards, J. (2014, February 12). A 'Dark Pattern' In Flappy Bird Reveals How Apple's Mysterious App Store Ranking Algorithm Works. *Business Insider*. Retrieved from <http://www.businessinsider.com/how-apple-app-store-ranking-algorithm-works-2014-2>
- Fosnot, C. T., & Perry, R. S. (1996). Constructivism: A psychological theory of learning. *Constructivism: Theory, perspectives, and practice*, 8-33.
- Gamper, J., & Knapp, J. (2002). A review of intelligent CALL systems. *Computer Assisted Language Learning*, 15(4), 329-342.
- Godwin-Jones, R. (2011). Emerging technologies: Mobile apps for language learning. *Language Learning & Technology*, 15(2), 2-11. Retrieved from <http://llt.msu.edu/issues/june2011/emerging.pdf>
- Grohtjahn, R. (1987). On the methodological basis of introspective methods. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (54-82). Clevedon, England: Multilingual Matters.
- Hickey (2015, March 8). Learning the Duolingo – how one app speaks volumes for language learning. *The Guardian News and Media Limited*. Retrieved from <http://www.theguardian.com/business/2015/mar/08/learning-the-duolingo-how-one-app-speaks-volumes-for-language-learning>
- Hymes, D.H. (1972) "On Communicative Competence" In: J.B. Pride and J. Holmes (eds.) *Sociolinguistics. Selected Readings*. Harmondsworth: Penguin, pp. 269-293.

- Kukulska-Hulme, A., & Bull, S. (2009). Theory-based support for mobile language learning: Noticing and recording. *International Journal of Interactive Mobile Technologies*, 3, 12-18.
- Kukulska-Hulme, A., & Shield, L. (2007). An Overview of Mobile Assisted Language Learning: Can mobile devices support collaborative practice in speaking and listening. In *conference EuroCALL'07 Conference Virtual Strand*.
- Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, 20(03), 271–289.
- Lee, Kwang-wu. "English teachers' barriers to the use of computer-assisted language learning." *The Internet TESL Journal* 6.12 (2000): 1-8.
- Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19(xx), 37-66.
- Jurafsky, D., & Martin, J. H. (2008). Speech and language processing: An introduction to speech recognition. *Computational Linguistics and Natural Language Processing*. Prentice Hall.
- Kerr, P (2013). *A Short Guide to Adaptive Learning in English Language Teaching*. The Round. Retrieved from <http://the-round.com/wp-content/uploads/downloads/2014/07/A-Short-Guide-to-Adaptive-Learning-in-English-Language-Teaching2.pdf>.
- Lee, J.F. & VanPatten, B. (2003). *Making communicative language teaching happen* (2nd ed.). New York: McGraw-Hill.
- Moundridou, M., & Virvou, M. (2003). Analysis and design of a web-based authoring tool generating intelligent tutoring systems. *Computers & Education*, 40(2), 157–181.
- Nah, K. C., White, P., & Sussex, R. (2008). The potential of using a mobile phone to access the Internet for learning EFL listening skills within a Korean context. *ReCALL*, 20(03), 331-347.
- Nation, P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, 27(1), 136.
- Oxford, R. L. (1997). Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. *The Modern Language Journal*, 81(4), 443–456.
- Oxford, R. L. (1995). Linking theories of learning with intelligent computer-assisted language learning (ICALL). In V.M. Holland, M.R. Sams, & J.D. Kaplan (Eds.) *Intelligent language tutors: Theory shaping technology* (359-369). Lawrence Erlbaum Associates, Inc.: Mahwah, New Jersey.
- Pellegrino, J., Chudowski, N., and Glaser, R. (2001). *Knowing what students know: the science and design of assessment*. National Academies Press.
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge University Press.
- Reinders, H., & Pegrum, M. (2016). Supporting Language Learning on the Move: An Evaluative Framework for Mobile Language Learning Resources. In B. Tomlinson (Ed.), *Research and Materials Development for Language Learning* (pp. 219-232). New York, NY: Routledge.
- Sauvignon, S. J. (2002). *Interpreting communicative language teaching: Contexts and concerns in teacher education*. New Haven: Yale University.
- Statista Inc. (2015). *Number of apps available in leading app stores as of July 2015*. Retrieved from <http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/>
- Stockwell, G. (2008). Investigating learner preparedness for and usage patterns of mobile learning. *ReCALL*, 20(3), 253-270.

Sweeney, P. & Moore. C. (2012). Mobile Apps for Learning Vocabulary: Categories, Evaluation and Design Criteria for Teachers and Developers. *International Journal of Computer-Assisted Language Learning and Teaching*, 2(4), 1-16, October-December 2012.

Ur, P. (2013). *The communicative approach revisited*. Cambridge: Cambridge University Press. Retrieved from <http://www.cambridge.com.mx/pennyur/Penny-TCAR.pdf>

Wang, S, and Smith, S. (2013) Reading and grammar learning through mobile phones. *Language Learning & Technology*(17)(3): 117-134.

Warschauer, M. (1996b). Computer-assisted language learning: an introduction. In S. Fotos (Ed.), *Multimedia language teaching*, 3–20. Tokyo: Logos.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language teaching*, 31(02), 57–71.

Wu, T. Sung, W. and Burston, J (2011) Reexamining the effectiveness of vocabulary learning via mobile phones. *Turkish Online Journal on Educational Technology*, 10(3), 203-214.

Appendix A. Selection of 50 Language Learning Apps.

App Name	App Store Ranking	Google Play Ranking
Duolingo	1	1
Rosetta Stone	2	3
Memrise	3	17
PenyoPal	13	
Learn English (Anspear)	15	
Mindsnacks	16, 27, 33, 43	
Learn [Language] with Lingo Arcade	18, 46	
Speak American English FREE (Mondly, ATi Studios)	20	
Innovative Language 101	21	
Busuu	22, 42, 54, 59, 62	6, 57, 67, 69, 90
ChineseSkill	23	
Vocabulary and Grammar! (TribalNova)	24	
Japanese!! (Square Poet)	35	
Translate Keyboard Pro	36	
Human Japanese Lite (Brak Software)	45	
Spanish by Living Language (Random House Inc)	47	

English with LinguaLeo	49	
Salsa - Spanish Language Learning (Mobile Madness)	50	
Learn Phrasebook (Codegent)	53	28, 37, 61, 63, 84
Speak Spanish - For Survival (Brainscape)	56	
Voxy	57	46
Voxy	57	
Fit Brains Language Trainer (Rosetta Stone)	28	
Phrasebook (Bravolo!)	29	
Learn & Play Languages (CoolForest Publishing)	30	
Learn Spanish - Brainscape	32	
FREE 24/7 Language Learning	4, 6, 14, 19, 34, 55	
Language Learning Games for Kids (StudyCat Limited)	40, 43, 51	33
Learn Japanese/Chinese/English Easily (Wan Peng)	7, 38, 41	
Hangman for Spanish Learners		22
Learn Arabic (AppVerx Limited)		32
Learn English Conversation Free (rwabee)		48
Learn English, Speak English (Speaking Pal)		49
Learn Languages: English (Jose Ortega)		21
Learning Japanese (Ignatius Reza)		27
Babbel - Learn Language		7, 14, 24
Byki Mobile		51
Easy Language Learning (PinDropApps)		9, 19, 59, 68, 100
English Podcast for Learners (tidahouse)		12
English-App: Learn English (Culture Alley)		42
HelloTalk Language Exchange		53
Learn 50 Languages		2, 31, 54, 60, 66, 75, 81, 88, 89
Learn 6,000 Words (Fun Easy Learn)		13, 36, 38, 39, 50, 58, 76

Learn English (Rwabee)		29
Learn English Kids Languages (Pinfloy Mobile Games)		16
Learn English/ Korean/ Portuguese/ Chinese/ etc. (bravolol)		5, 44, 78, 80, 87
Learn Languages Free (Murat)		35
Learning Japanese (sagetsang)		8
Lerni. learn languages		26
Mango Languages		25
Play & Learn LANGUAGES (Shift Interactive Party Ltd)		34
Sight Words Learning Games		47
Tourist Language Learn & Speak		4

Appendix B. Survey Instrument.

<p>Q1. Name of the App</p> <p>Q2. Possible reason for deletion</p> <p>Q3. Rater</p> <p>Q2. Languages Supported</p> <ol style="list-style-type: none"> 1) English 2) German 3) French 4) Spanish 5) Italian 6) Japanese 7) Portuguese 8) Russian 9) Turkish 10) Arabic 11) Chinese 12) Polish 13) Thai 14) Swedish 15) Hindi/Urdu 16) Bengali 17) Korean 18) Swahili 19) Finnish 20) Greek 21) Other: _____ <p>Q3. Platforms Supported</p>	<p>Q10. Implicit/Explicit Grammar Instruction</p> <ol style="list-style-type: none"> 1) Implicit 2) Explicit - grammar presentation (rules explained prior to activity) 3) Explicit - feedback (rules explained when you make a mistake) 4) None (words taught in isolation) 5) Other: _____ <p>Q11. Types of Feedback</p> <ol style="list-style-type: none"> 1) None 2) Non-corrective (sound effects, visuals) 3) Corrective feedback but no editing of mistake required by the user 4) Corrective feedback with editing of mistake required by the user 5) Other: _____ <p>Q12. Types of Feedback</p> <ol style="list-style-type: none"> 1) None 2) Sound effects 3) Visual feedback (colors, icons, etc.) 4) Simple textual feedback (Corrections) 5) Textual explanation 6) Other: _____ <p>Q13. Types of Feedback</p> <ol style="list-style-type: none"> 1) No editing (moves onto the next question) 2) Editing required by process of elimination 3) Hint or suggestion provided
---	---

- 1) iOS
- 2) Android
- 3) Windows Phone
- 4) Blackberry
- 5) Other: _____

Q4. Monetization

- 1) None
- 2) In-app ads
- 3) Pay to unlock levels
- 4) Subscription
- 5) Power-ups
- 6) Upgrades
- 7) Pay to unlock languages
- 8) Other: _____

Q5. Gamification

- 1) Lives or health
- 2) Positive/Negative reinforcement
- 3) Time limits
- 4) Progress indication
- 5) Cumulative point system
- 6) Achievements/Badge/Accomplishments
- 7) Missions/Quests/Tasks
- 8) Random rewards (same each time)
- 9) Fixed rewards (same each time)
- 10) New daily content
- 11) Unlocking levels
- 12) Win condition
- 13) 2D world
- 14) 3D world
- 15) Narrative
- 16) Avatar - representation of self
- 17) Other: _____

Q6. User level placement (i.e. How does the app know the user's level)

- 1) None
- 2) Preliminary testing
- 3) Option to test out of activities/levels
- 4) Manual level selection
- 5) Other: _____

Q7. Audio Requirements

- 1) None
- 2) Speaker
- 3) Microphone
- 4) Other: _____

Q8. User Input to Device

- 1) Keyboard (writing)
- 2) Touch gestures (tapping, swiping)
- 3) Speaking (microphone)

- 4) Copy correct answer
- 5) Other: _____

Q14. Game Mechanics

- 1) Selection - pick the correct answer
- 2) Matching image to meaning
- 3) Matching/selecting/writing L2 word(s) to correspond with L1 meaning (translation)
- 4) Matching/selecting/writing L2 word(s) to correspond with L2 meaning (definition)
- 5) Cloze
- 6) Other: _____

Q15. Visual Input

- 1) Words
- 2) Images
- 3) Videos
- 4) Animations
- 5) Other: _____

Q16. Listening

- 1) None
- 2) Listen to letters
- 3) Listen to words
- 4) Listen to sentences
- 5) Listen to dialogues
- 6) Listen to passages
- 7) Listen to songs
- 8) Other: _____

Q17. Reading

- 1) Read letters
- 2) Read words
- 3) Read sentences
- 4) Read passages
- 5) Read dialogues
- 6) Read songs
- 7) None
- 8) Other: _____

Q18. Writing

- 1) Write letters on keyboards
- 2) Write words on keyboards
- 3) Write sentences on keyboards
- 4) Write passages on keyboards
- 5) Moving/selecting words to form sentences
- 6) Moving/Selecting letters to form words
- 7) None
- 8) Other: _____

Q19. Speaking

- 1) Repetition
- 2) Reply
- 3) None

4) Other: _____

Q9. Elements of language instruction (NOTE: code ONLY if element is assessed in app)

- 1) Vocabulary - isolated units
- 2) Vocabulary - in context
- 3) Grammar (sentence construction, verb tenses, etc.)
- 4) Pragmatics (usage/appropriacy)
- 5) Pronunciation

4) Other: _____

Q20. Social Integration

- 1) Peer review
- 2) Tutoring services
- 3) Chatting
- 4) Native speaker review
- 5) None
- 6) Other: _____

Q21. Comments