

The Implementation of ISLA in MALL Technology: An Investigation into the Potential Effectiveness of Duolingo

Dustin Crowther, Kathy MinHye Kim, Shawn Loewen

Second Language Studies, Michigan State University

crowth14@msu.edu

Abstract

Following the increased implementation of mobile learning across the globe, specifically in the area of mobile-assisted language learning (MALL; Burston, 2015; Duman et al., 2015), the current paper provides an evaluation of the highly popular MALL application Duolingo. Specifically, this evaluation targets how effectively instructed second language acquisition (ISLA) research and theory has been implemented by Duolingo programmers. While current frameworks for the evaluation of MALL technology (e.g., Reinders & Pegrum, 2015) place a significant focus on the learning affordances available, less emphasis has been placed on the implementation of ISLA theory. As such, Chapelle's (2001) evaluation framework, originally developed for computer-assisted language learning programs, is revisited due to its basis in ISLA theory. Six criteria thus serve as the basis of this evaluation: Language Learning Potential, Meaning Focus, Authenticity, Learner Fit, Positive Impact, and Practicality. While certain benefits of Duolingo as a language learning tool are discussed, overall the evaluation indicates that the benefit of Duolingo is more likely as a learning support app than as the sole tool for autonomous learning.

Keywords: Duolingo, Evaluation, ISLA, MALL, Technology

The use of mobile technology globally, across developed and developing countries, has made significant strides over the last 15 years, with internet-enabled mobile devices surpassing the number of desktop and laptop computers as of 2013 (Pegrum, 2014). Accordingly, there has been increased focus on mobile learning (m-learning), in which teaching and learning is facilitated through the use of mobile technologies, including smart phones and tablets (Duman, Orhan, & Gedik, 2015). One area of empirical inquiry that has shown significant growth in this area is mobile-assisted language learning (MALL; Burston, 2015; Duman et al., 2015). Though MALL could be seen simply as an extension of computer-assisted language learning (CALL), there are affordances of MALL that distinguish it from CALL beyond a simple mobile versus

fixed distinction (Kukulka-Hulme, 2012; Pegrum, 2014; Reinders & Pegrum, 2015). Kukulka-Hulme (2012) lists ready access to help and information, flexibility in time and space, adaptation to personal habits, and continuity between learning in different settings as a few key benefits. Kukulka-Hulme and Shield (2008) also note the spontaneity of access and interaction across different contexts of use. Reinders and Pegrum (2015) highlight the ability to link between local and global, episodic and extended, and personal and social learning (among other affordances). Though existing MALL research is severely limited by a lack of objective, quantifiable measures of learning outcomes, when such measures are present, MALL technology has been shown to promote learning for reading, listening, and speaking (see Burston, 2015, for an overview of existing research).

Within MALL, Reinders and Pegrum (2015) distinguish between *mobile materials* and *mobile activities*. While the latter focuses on the use of mobile-based tasks within a larger learning context, mobile materials refer to web services and applications (apps) that include built in language learning content and pedagogy. Despite perceived limitations, including a tendency to rely on a behaviorist, teacher-centered approach towards language instruction (Reinders & Pegrum, 2015), such apps have proven to be quite popular for autonomous language learning. One such app that has gained global recognition is Duolingo, the Apple iPhone App of the Year (<https://itunes.apple.com/app/duolingo-learn-spanish-french/id570060128?mt=8>).

Launched in November 2011 (Robertson, 2011), Duolingo self-describes as a free science-based language education platform created by Luis von Ahn and Severin Hacker. In the initial two years of operation, Duolingo saw over 120 million users register for their language courses (www.duolingo.com/press). While course availability is currently geared towards first-language (L1) English learners of second-languages (L2s) (e.g., German, French, Spanish), an increasing number of courses are being put online for speakers of various other L1s (e.g., Korean [L2 English], Russian [L2 English, French, German], Turkish [L2 English, French, Russian]). Though also accessible via desktop computer (duolingo.com), one of Duolingo's biggest affordances is that it provides worldwide mobile access. The Duolingo app is downloadable for all tablets and mobiles using iOS, Android, and Windows operating systems (www.duolingo.com/press), with learner progress on one platform synced with the others.

As previously discussed, MALL provides affordances for language learning beyond those traditionally considered within the L2 classroom. In an effectiveness study funded by Duolingo, Vesselinov and Grego (2012) proposed that for L1 English speakers with no previous knowledge of Spanish, an average of 34 hours of Duolingo usage would be equivalent to a full 11-week university semester of study. This claim, which has served as a primary selling point (<https://www.youtube.com/watch?v=8OebgtUjLg4>), presents Duolingo as more effective than a university language course, because such courses usually require beyond 34 hours of total work (von Ahn, 2013). In essence, Duolingo claims that their software provides advantages to learners beyond what they can accomplish within the L2 classroom. However, Vesselinov and Grego cautioned that even though it would be fair to expect similar results across other languages, such a claim could not be made without further empirical study. Despite this limitation, which stems from their own self-commissioned paper, Duolingo does not make this distinction when promoting their program.

One conceivable way to determine the potential generalizability of Vesselinov and Grego's (2012) findings is to consider how effectively Duolingo makes use of empirically

researched theories of L2 learning. However, literature targeting this relationship is currently sparse. Krashen (2014), in responding directly to Vesselinov and Grego, highlighted how Duolingo utilizes language instruction that promotes conscious learning, which he argued is less efficient in developing language competence compared to methods that promote subconscious language acquisition. Additionally, Wagner and Kunan (2013) stressed that the primary tasks “harken back to the 1950s, when audiolingualism was the dominant theory in language learning” (p. 330), highlighting a lack of tasks that target communicative or interactional competence. These criticisms put forth by Krashen and Wagner and Kunan would simultaneously appear to advocate for a greater focus on language teaching methods promoted within instructed second language acquisition (ISLA) research. Such research may serve as an appropriate means in which to evaluate how effective Duolingo can be as a language learning tool.

ISLA considers “how the systematic manipulation of the mechanisms of learning and/or the conditions under which they occur enable or facilitate the development and acquisition of a language other than one’s first” (Loewen, 2015, p. 2). Though empirically based approaches that define ISLA inquiry have primarily been conducted within the L2 classroom (Loewen, 2015), the same systematic manipulation that defines ISLA is necessary in the programming of MALL-based apps such as Duolingo, indicating potential alignment between the two types of instruction. How to measure the potential of MALL-based apps as language learning tools requires a range of considerations, from educational affordances to affective design (Reinders & Pegrum, 2015), though at the heart of such an evaluative framework must remain a theoretical approach towards language instruction.

While several ISLA-based evaluation frameworks for CALL exist (e.g., Chapelle, 2011; Hubbard, 2006), Reinders and Pegrum (2015), to our knowledge, is the lone framework addressing MALL. Within Reinders and Pegrum’s (2015) framework, the authors place a primary emphasis on learning affordances MALL technology can provide. Examples of which include how well the technology can:

- link local (interaction within local environments) with global (connections with global networks of people and resources) learning
- allow for both episodic (bite-sized learning at users’ convenience) and extended (continuation between session) learning
- promote both personal (self-tailored learning) and social (community-based) learning.

Such MALL technology potentially advocates for both autonomous and networked learning. While these characteristics should not be devalued, there is limited overlap with an ISLA perspective. While Reinders and Pegrum include meaning-focused constructs (e.g., input, output, negotiation of meaning), very few form-focused considerations (e.g., corrective feedback) are provided. Considering the dual importance of meaning- and form-focused instruction in ISLA (Loewen, 2015), a more comprehensive evaluative approach is necessary (Hubbard, 2006). Though a handful of CALL evaluation rubrics have been proposed (e.g., Chapelle, 2001; Egbert & Hanson-Smith, 1999; Hubbard, 2006; Underwood, 1984), to our knowledge only Chapelle’s (2001) ISLA-inspired framework has been practically utilized to empirically evaluate CALL tasks (Jamieson, Chapelle & Preiss, 2004, 2005).

Though not designed specifically for MALL technology and subsequently lacking the affordances that are highlighted in Reinders and Pegrum’s framework, Chapelle’s (2001) ISLA-

informed criteria, which places a greater emphasis on the systematic manipulation of input and learner engagement (Loewen, 2015), allow for a more in-depth analysis of the learning potential MALL-based apps such as Duolingo may provide. Chapelle (2001) introduced six criteria to guide an analysis of CALL materials as an appropriate and potentially effective language learning tool. The criteria included: (a) Language Learning Potential, (b) Meaning Focus, (c) Authenticity, (d) Learner Fit, (e) Positive Impact, and (f) Practicality.

Using Chapelle's (2001) framework, the rest of this paper will provide an evaluation on the extent to which Duolingo adheres to ISLA-informed research theory, and relatedly, what this adherence may tell us of Duolingo's potential as a language learning tool. We begin with an operational description of the Duolingo program, followed by an in-depth overview of the six key criteria to Chapelle's framework. Once the framework is established, we will apply the framework to the program, to address the following two research questions.

1. In what ways does Duolingo adhere to ISLA-informed research theory?
2. Based on the level of adherence, how effective as a language learning tool across multiple target languages can Duolingo be predicted to be?

Evaluation Methodology

The current analysis is based on 34 hours of Duolingo usage by the authors, who were learning Turkish as part of a larger research study investigating the learning gains and experiences of nine ab initio learners, focusing on how far a group of motivated, highly linguistically aware students could progress. All nine participants were either graduate students or instructors in an MA or PhD level second language studies program.

Operational Description of Duolingo

Before attending to how Chapelle's (2001) framework, we offer an operational description and an evaluation of the features and activities available in Duolingo. These features will then be used to determine whether Duolingo implements ISLA theories. Based on the work of Burston (2003) and Hubbard (1988, 1996, 2006), three criteria were used to guide this evaluation: (1) technological features, (2) activity types, and (3) presentational schemes.

Technological features. Technological features are concerned with the technological aspects of a software program. Considering features such as ease of installation or platform compatibility, users can judge whether the language learning program is readily accessible and, thus, user-friendly (see Burston, 2003 for a full list of features). Duolingo has great merit in both ease of installation and platform compatibility, as web browser and mobile versions are offered free of charge with the latter available on both Android and iOS. While web browsers require Internet connection, a handful of lessons can be completed via mobile devices through an offline mode, with the only disadvantage being that progress made during offline sessions is not counted towards overall progress.

Activity types. Philips (1985) lists six key activity types: game, quiz, text reconstruction, text construction, problem solving, and exploratory (refer to Phillips, 1985, for a detailed breakdown). Of these six, Duolingo employs two formats: *quiz* and *text reconstruction*. In quiz activity type, learners are provided items that target grammar, vocabulary or pronunciation, and

which employ a mechanical drill that is guided by a stimulus-response interaction. Text reconstruction requires learners to decompose a text manipulated by the program. From these activity types, Duolingo generates seven types of questions in four forms:

- multiple choice (vocabulary-picture matching, select correct translation)
- matching (tap the correct translation pairs)
- ordering (unscramble sentence translation)
- direct stimulus-response items (oral repetition, transcribe what you hear, translate sentence/word)

Presentational schemes. Presentational schemes consider the way in which an activity type is presented to the user (Hubbard, 1988). Though many schemes exist, the seven most prominent ones incorporated by Duolingo include interface, timing, control options, feedback, user input, input judging, and help options.

The *interface* or *screen layout* addresses the presentation of the materials on screen. Components falling under this category include, but are not limited to, font size, spacing, location, and quality of animation/graphics/text. In Duolingo, while certain components of screen layout are consistent across question types (e.g., visualized progress bar, feedback with different colors) a number of layouts are relevant only to certain questions (e.g., the dual input of visual [written or image] and audio, slow-paced repetition). *Time limitations*, a second factor related to presentational schemes, appears only twice. First, regarding specific question types, only the oral repetition task provides a time limit for users to respond. For all other question types, users have full control over the time necessary to submit their responses. However, when completing a Strengthen Skills review lesson (a series of questions targeting a specific, previously encountered skill), users have the option of a timed (thirty seconds, with time added for each correct response) or an untimed session. The third factor is *control* which can be imposed by three entities: the instructor, the program, and the learner. When considering Duolingo as an autonomous learning tool, only the latter two are relevant, specifically in controlling the pace of and access to materials. Lesson sequence is one element strictly controlled by Duolingo, as users proceed through a predetermined path of lessons, with new lessons unlocked only as previous lessons are completed. One benefit for users is the option to “Test Out” of any specific skill, which, if successful, unlocks all lessons up to the point. Other features of which learners have control include skipping individual questions, quitting lessons, and seeking additional input, all of which can be performed using the “Skip”, “Quit”, and “Tips and Notes” functions, respectively. The fourth factor, *feedback*, relates to “information the program communicates in response to specific input” (Hubbard, 1998, p. 59). Duolingo provides automatic feedback to all responses with corrected forms, an occasional metalinguistic note, and different sound effects in response to correct and incorrect responses (see Figure 1). This feedback occurs in response to *user input*, in which the user responds to a given item (e.g., typing a translation, pronouncing a sentence, selecting a multiple-choice response).

The type of previously referenced feedback is determined through *input judging*, in which Duolingo decides if the input received for a specific item matches the expected input for that item. The extent of correctness impacts the amount and type of feedback provided. The last component of the presentational scheme is the *help options*, which indicate whether the software provides assistance to learners in completing activities. Duolingo features that fall under this factor include direct translations that are provided when the user places the cursor above an unknown word, an adjusted speed option for the “Type What You Hear” questions, and Tips and

Notes, a minimal description of the target grammatical feature, available every session, though only when using a PC.

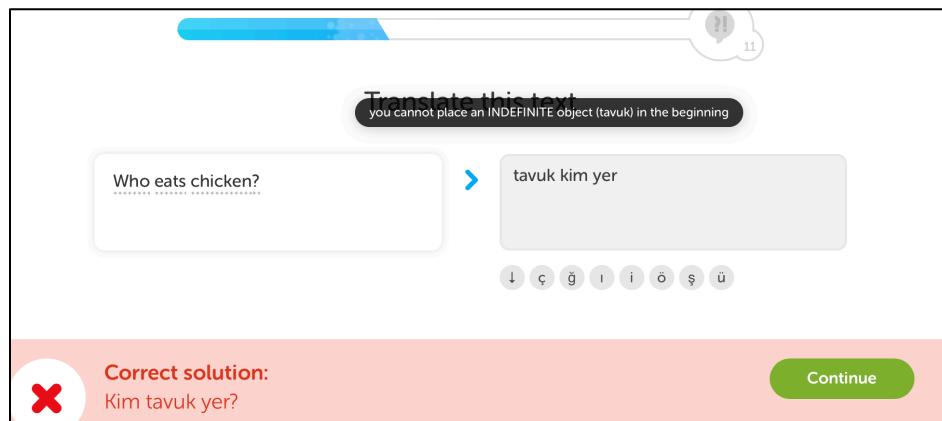


Figure 1. Example of metalinguistic feedback.

Tools of Analysis: Chapelle's (2001) framework

Having described the operationalization of Duolingo, Chapelle's (2001) six criteria (Language Learning Potential, Meaning Focus, Authenticity, Learner Fit, Positive Impact, and Practicality) are described as the guidelines for our evaluation. After, we will consider how such criteria are or are not realized within Duolingo.

Language Learning Potential. Chapelle defined Language Learning Potential (LLP) as "the extent to which the activity can be considered to be a language learning activity rather than simply an opportunity for language use" (p. 55). *Language learning activity* here aligns closely with focus-on-forms, instructional approaches incorporating variations of explicit instruction (e.g., consciousness raising, PPP; Loewen, 2015). Such approaches place an emphasis on linguistic form that may not be appropriately addressed within *language use*. Given the difficulty of attaining high levels of linguistic accuracy without attending to specific linguistic items, this criterion is important for evaluating CALL (and for our purposes, MALL) materials for language learning purposes. The primary guiding question put forth by Chapelle was whether the "task conditions present sufficient opportunity for beneficial focus on form" (p. 59).

To allow for easier evaluation, Jamieson et al. (2004) divided LLP into three sub-categories: input enhancement, interaction, and production. These three were chosen as targets because they had "a strong theoretical, pedagogical, and empirical base" (p. 399). In evaluating a series of ESL/EFL online materials, Jamieson et al. defined and implemented each of these targets. The evaluation presented below follows these revised guidelines.

Input enhancement. The evaluation of input enhancement targeted techniques employed to draw learners' attention to targeted linguistic features. Specifically, Jamieson et al. (2004) identified *salience* (e.g., visually marking a grammatical form), *modification* (e.g., altering initial input using linguistic and non-linguistic means to increase understanding), and *elaboration* (e.g., making grammatical additions to increase understanding). Examples of how these were operationalized included highlighting, animation, and repetition (salience), simplification, image/video, and L1 translation (modification), and adding grammatical phrases or clauses to

texts (elaboration).

Interaction. Three types of interaction are proposed in Jamieson et al. (2004): between people, between a person and computer, and between the self and the mind. Interaction *between people* allows opportunities for "negotiation for meaning, co-construction of meaning, and prompting their [the learner] attention to form" (p. 404). When *interacting with a computer* interface, there is potential for enhanced input, such as the use of hyperlinks to access online supplementary material. Finally, interaction *between the self and mind* occurs when a learning task requires the consideration of multiple linguistic options, thus prompting learners to engage in deeper cognitive processing of the linguistic input (focus on linguistic form).

Production. The type of output elicited can be evaluated from three perspectives, derived from cognitive, sociocognitive, and interactionist frameworks (Jamieson et al., 2004). The amount of *planning time* (cognitive) and the amount and type of *correction* (interaction) provided are seen as enhancing the benefits of language production. The *co-construction of meaning* with help from an interlocutor (sociocognitive) is acknowledged to push the learner beyond what they are able to produce on their own.

Meaning Focus. Meaning Focus targets how "primary attention is directed toward the meaning of the language that is required to accomplish a task" (Chapelle, 2001, p. 56). Specifically, within a Meaning Focus task, the learners are expected to complete an objective through the use of the target language, with examples including making a decision on a specified issue or an exchange of information. As discussed by Chapelle, Meaning Focus is not limited to oral communication, but includes both writing and reading tasks, assuming language is used purposefully to construct and interpret meaning.

Authenticity. Authenticity considers the practical, real-world implications of both LLP and Meaning Focus. Authenticity begins to move beyond the conditions believed to be relevant for language acquisition, with a larger focus on how L2 learning tasks reflect the language likely to be encountered beyond the L2 classroom (Chapelle, 2001). While not explicitly stated in Chapelle's initial definition, authenticity appears to have a strong alignment with a task-based teaching philosophy, where tasks are designed using needs analysis to target what learners need to do in real-world contexts (Long, 2015). This real-world connection is clear through the guiding questions provided by Chapelle, which ask whether there is a strong correspondence between CALL and real-world tasks and if learners are able to recognize this relation.

Learner Fit. Chapelle's (2001) definition of Learner Fit is relatively vague, though implies a need to take into account differences in linguistic ability of learners along with differences in non-linguistic characteristics. Hubbard (2006) provides a more in-depth description of learner fit, highlighting the necessity to link syllabus choices to individual learner needs. Specifically, Hubbard lists seven considerations: learning style, classroom management, linguistic objectives, language skills, language difficulty, program difficulty, and content. Table 1 provides a brief overview of each consideration.

Table 1.

Seven Considerations of Learner Fit (Hubbard, 2006)

Consideration	Descriptor
Learning Style	Cognitive style, preferred learning strategies, motivational orientations
Classroom Management	Use of hardware/software in regards to individual versus paired/group work
Linguistic Objectives	Discourse/text, syntax, lexis, morphology, and phonology/graphology
Language Skills	Reading, Writing, Grammar, Pronunciation
Language Difficulty	Level of linguistic challenge
Program Difficulty	Level of difficulty associated with using the program
Content	Consistency between course targets and course objectives

Positive Impact. With Positive Impact, Chapelle's (2001) criteria begins to move beyond a strictly language learning-based evaluation, focusing on how the instructional approach employed benefits the learner beyond linguistic knowledge. Examples provided include accountability within the learner for their own study, a willingness to interact with and within the target culture, and pragmatic ability. Beyond these potential non-linguistic benefits, Positive Impact asks whether the learner will have a positive learning experience with the technology they interact with.

Practicality. Though not directly related to theories of ISLA, practicality is inseparable from any evaluation of CALL/MALL software, as it relates directly to the "adequacy of resources to support the use of the CALL activity" (Chapelle, 2001, p. 55). While for CALL such resources often include the availability of and access to necessary hardware and software, along with knowledgeable personnel for on-site assistance, the readily available nature of MALL apps such as Duolingo require different considerations in regards to adequacy. Of importance to these additional affordances are the potential costs, and the availability of online connectivity during usage. In essence, practical MALL apps should be easy to implement in various language learning settings, compatible to different platforms, and available to as many people as possible with sufficient technical support.

An Evaluation of Duolingo Following Chapelle (2001)

Language Learning Potential.

Input enhancement. The enhancement of linguistic input is one of the more developed features that Duolingo provides. Of the different types of typographical effects used to increase *saliency* (color, bold, italic, underlining, and capitalization), color manipulation is most frequently employed, serving multiple purposes. For instance, orange is applied to highlight newly introduced vocabulary and/or linguistic features (see Figure 2). Color and cross out techniques are simultaneously provided to create a stronger effect on drawing users' attention to erroneous output. Such negative evidence is accompanied by positive evidence in which input is highlighted in red bars to indicate correct forms. Despite the use of various attention-drawing devices, Duolingo fails to draw selective attention to critical linguistic features of the target language. Figure 1 exemplifies a typical textual enhancement of a newly introduced linguistic feature, the present continuous in Turkish. Duolingo highlights *biliyorsun* (you want), despite the central importance being the morpheme *-iyor*. In fact, *biliyorsun* encompasses three linguistics features, the stem *bil-* (want), the present continuous marker *-iyor-*, and the second person singular marker *-sun*, but by enhancing the whole word, users' attention disperses to features that are not necessarily the focus of the current lesson.

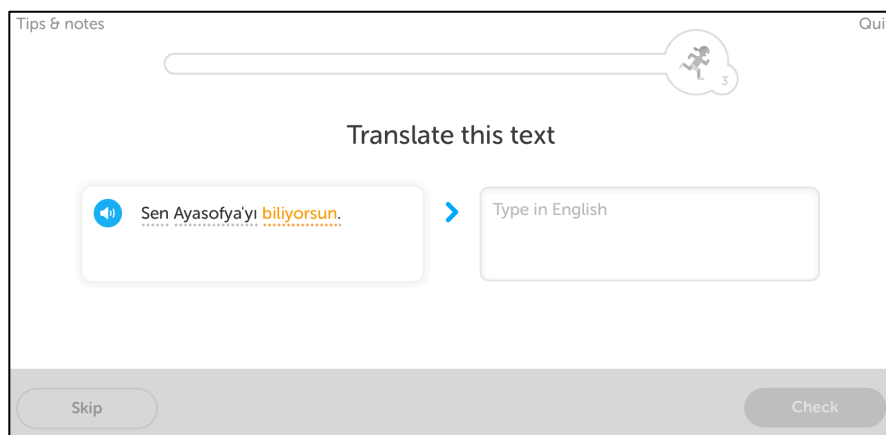


Figure 2. Example of enhanced vocabulary input.

Audio repetition is another function of input enhancement. Duolingo accompanies most visual input with audio representation, which allows for multiple, multimodal repetitions of the input. Though the number of repetitions is user-dependent, the potential exists for a frequency effect on the formation of stronger associations between words, as well as picking up on targeted linguistic features.

Three forms of *modifications* are present that enhance access to target language meaning (images, hypertext translations, and slow-paced repetitions). Mostly relevant to vocabulary learning, both visual and verbal information (see Figure 3) are simultaneously provided to promote users' understanding and retention of a new item (Paivio, 1986; Sternberb, 2003).

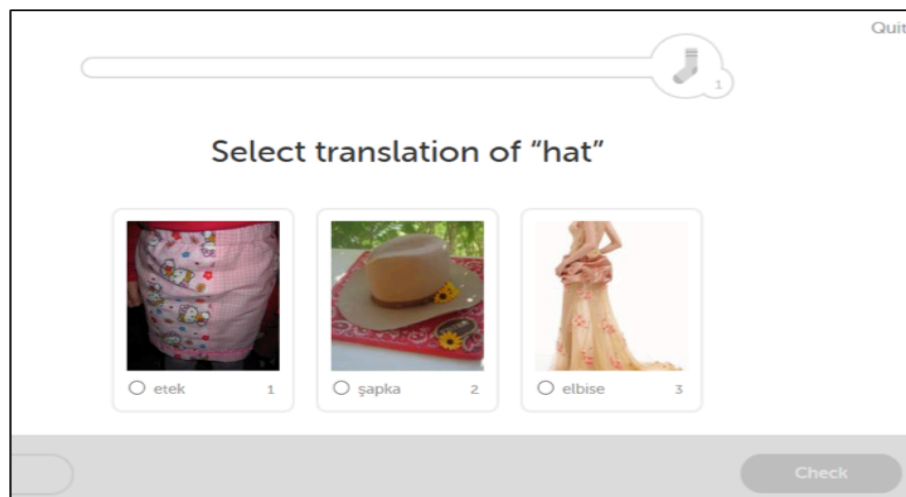


Figure 3. Example of modified vocabulary learning through image.

In a similar vein, both visual (hypertext) and aural (slow-paced repetition) modifications are linguistic forms of modifications provided to aid in comprehension of input. As shown in Figure 4, the hypertext facilitates comprehension as it provides a direct translation of the target items. In the case of slow-paced repetition, users listen to an audio version of the visual input multiple times at a slower rate. The adjusted pace function, mostly present for the “Type What You Hear” task, aids comprehension by adjusting the speed to a slower pace. Access to the hypertext and slow-paced repetition is voluntary and the extent to which the modification benefits learning is an empirical question depending on various factors (e.g., proficiency level, seriousness about the task, etc.).

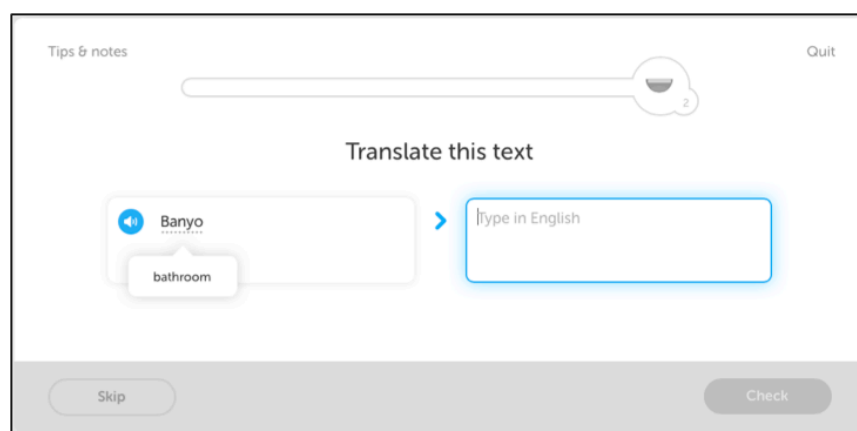


Figure 4. Example of modified vocabulary learning through hypertext.

Of the three enhanced input techniques, *elaboration* is the least developed feature available to users. In Duolingo, absence of contextualization is a key limitation (see Authenticity below); therefore, finding a lack of elaborated input, which is intended to enrich users’ understanding and to clarify textual meaning, is not surprising but rather expected. However, as will be discussed later, it might be the case that the inclusion of phrases and clauses, both

examples of elaboration, is intentionally avoided in the beginning to intermediate units, as these may be seen as complex linguistic features.

Interaction. Duolingo does allow for interaction between people, but is limited to those who choose to partake in online discussion board conversations, which are often in English. For each question, users can access a discussion board to post their comments, concerns, difficulties related to the item, and subsequently respond to any previous comments (while this option is available on the PC and Android platforms, it is oddly absent from mobile iOS platforms). Though such discussions provide limited opportunities for negotiation for/co-construction of meaning, they do provide an opportunity for greater attention to form. However, whether users make use of the discussion board is self-dependent, and thus the potential benefits may not be generalizable across users.

Surprisingly, a primary area where users could benefit from Duolingo's mobile, online status is relatively underused. By interacting with a computer (or tablet, mobile device), there is potential to link users' learning experience to resources/supplementary resources from across the Internet. However, Duolingo remains completely self-contained, with the only support provided being the aforementioned Duolingo-moderated discussion board.

The presence of the final interaction of interest, between self and mind, is debatable. Through the use of multiple-choice questions, Duolingo does indeed present the user with multiple linguistic options to consider, though the depth of cognitive processing involved is questionable. For example, in Figure 5, the user is required to choose the correct translation(s) of the English phrase "Come to the bar with me". However, of the three options, only one includes the Turkish translation for "the bar" (bara). For learners with strong lexical knowledge, there is no need to process the linguistic form at a deeper cognitive level, as lexical knowledge is all that is needed to complete the task.

In sum, interactive limitations are noticeable throughout the multiple-choice based items. ISLA research has focused considerably on the benefits of human interaction (Gass & Mackey, 2015). Considering this importance of interaction, the above limitations greatly impede Duolingo's potential effectiveness.

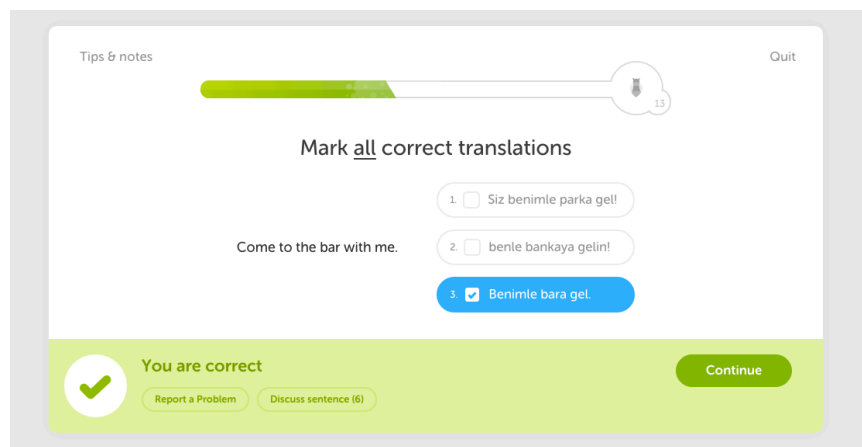


Figure 5. PC-based Duolingo Turkish selection task.

Production. It is in this area of production that Duolingo demonstrates its most significant limitation. For speaking, production is limited to a speech repetition task (with textual representation provided, see Figure 6), a task available only through PC. Similarly, written production is limited to translating sentences from English to the target language (see Figure 7). For neither productive skill is there a task that requires the user to produce spontaneous output. Considering the vital role that researchers argue output plays in language acquisition (Swain, 1995), the lack of productive tasks in Duolingo is a serious limitation. Without opportunities for language production, users are denied the potential acquisition benefits of planning time, correction, and co-construction of meaning identified by Jamieson et al. (2004).

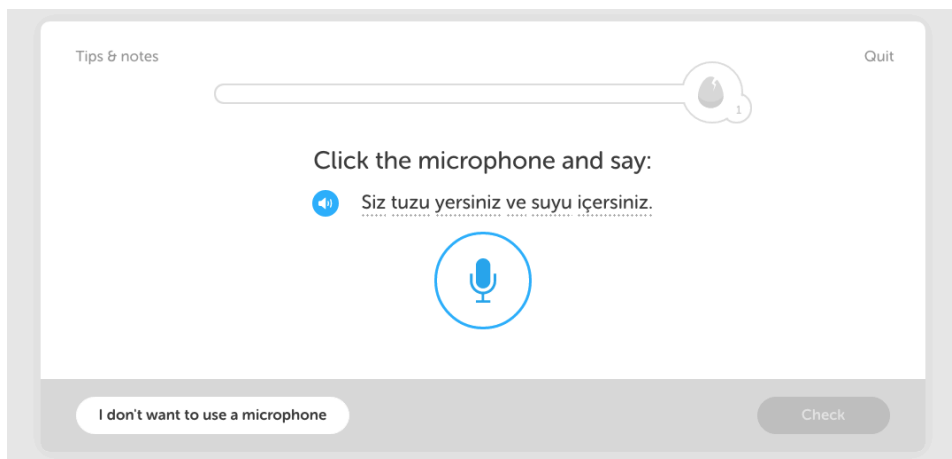


Figure 6. PC-based Duolingo Turkish pronunciation task.

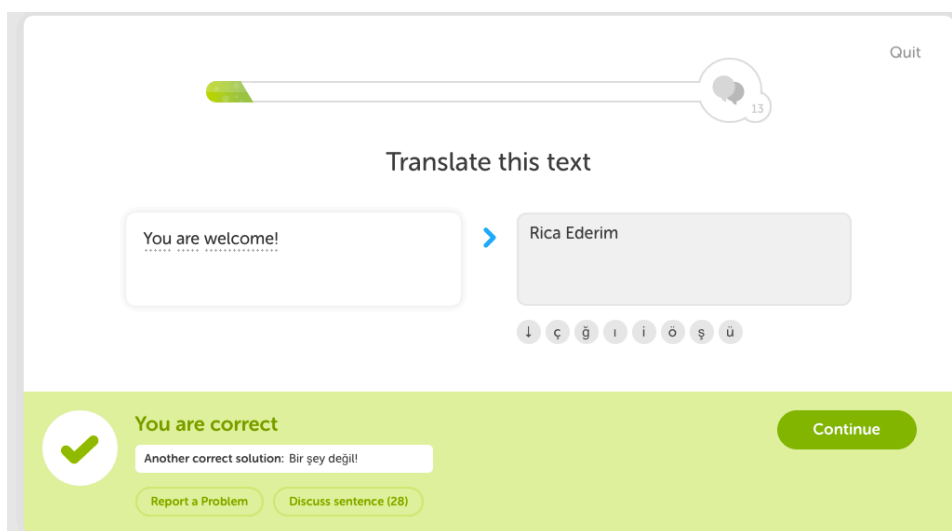


Figure 7. PC-based Duolingo English to Turkish translation task.

Meaning Focus. A key limitation to Duolingo, as echoed in other critiques of MALL technology, is a focus on a behaviorist, teacher-centered approach towards language instruction, which ignores active and collaborative modes of learning shown to be beneficial to acquisition (Reinders & Pegrum, 2015). Whether targeting receptive (reading, listening) or productive

(writing, speaking) skills, Duolingo learning activities place primary focus on lexical and grammatical learning with a limited focus on pronunciation, but no focus on meaning construction and interpretation. For example, one common translation activity provides users with a sentence in the target language and three potential meanings in their L1. Learners are then required to choose all correct responses (see Figure 5). Another activity presents a sentence in the target language and asks the user to provide a translation (see Figure 2). A final example requires users to repeat aloud a sentence in the target language (see Figure 6). None of the above activities require the user to produce the language in such a way that requires them to construct meaning, and while interpretation is present in the translation activities, it is limited to word level understanding, and devoid of any contextual meaning. Such a lexical and grammatical focus permeates the remaining Duolingo activities (e.g., transcribe what you hear, vocabulary-picture matching). As opposed to attention being drawn to the meaning of the language needed to accomplish an activity (Chapelle, 2001), Duolingo uses activities to draw attention to specific lexical items and syntactic forms.

Authenticity. Following the lack of a Meaning Focused approach towards language learning, it is unsurprising to see a lack of authenticity in Duolingo. As already described, there is a significant emphasis on lexical and grammatical learning, which rarely moves beyond the sentence level. Furthermore, many of the sentences are unlikely to be encountered outside of Duolingo (e.g., The rabbit reads the book), and while it could be argued that units targeting everyday phrases (e.g., how are you?, I'm fine thank you) have immediate usage potential, such phrases are presented in isolation, even separated from their traditional speech act partner (e.g., *how are you* and *I'm fine thank you* are never presented as a pair). Even if not subscribing to the strong sense of task-based language learning described in Long (2015), Duolingo still makes little attempt to relate the language learned to authentic usage in real-world contexts. This subsequently limits the potential beneficial relationship between technology and task previously indicated in language learning literature (Ziegler, 2016).

Learner Fit. Hubbard (2006) identified seven primary components for the analysis of Learner Fit, each italicized below. Progression through Duolingo follows a relatively linear path, with new lessons becoming available as current lessons are completed. As such, Duolingo proves quite easy to navigate, limiting any concerns in regards to *Program Difficulty*. Such an approach makes Duolingo accessible for a range of potential users. However, despite this range, there is little consideration for variations in *Learning Style* across users. Specific learning strategies users may employ are required to meet the needs of the program, which places a heavy emphasis on developing explicit over implicit knowledge. How this impacts *Language Difficulty* is a more subjective category, depending on individual users. Duolingo, though, seems to subscribe to a cumulative approach where linguistic forms and vocabulary items can often be seen re-emerging in later units. The primarily autonomous nature of Duolingo usage limits *Classroom Management* considerations, though Duolingo has recently launched “Duolingo for Schools”, creating a more relevant environment for such consideration (though not one considered in this review). One area of concern is in *Linguistic Objectives* and *Learning Skills*, where users have little control over the syntax, lexis, morphology, and phonology addressed, how each linguistic dimension is addressed, and the proportion of reading, writing, grammar, and pronunciation practice they receive. This is not to say that strict adherence to a predetermined path of development is a limitation, though, the path chosen has placed a primary focus on explicit knowledge over meaning construction, thus limiting the potential overall effectiveness of

Duolingo as a learning tool (discussed later).

Positive Impact. For Duolingo, the extent of Positive Impact can be seen in how the gamification of the learning process (Werbach, 2014) promotes accountability within the user. Duolingo allows users to set personal goals for study in the form of experience points (XP) earned (e.g., 50 XP would be equivalent to completing 5 lessons of study), and through the use of online connectivity, to compare their total XP gained to that of their peers. Users are also given the option to choose between strengthening previously completed units, advancing to new units (though following a predetermined path), and strengthening their knowledge as a whole (e.g., progress quiz). All of the above options contribute to how quickly a learner can “level up” in Duolingo. Beyond this gamified accountability, though, the Positive Impact of Duolingo is limited. There is little opportunity to interact with and within the target language culture, which is in contrast to the global affordances that MALL technology can provide (Pegrum, 2014), nor does Duolingo address pragmatic knowledge in the target language. Further, though it is unrealistic to objectively determine whether a user's learning experience with Duolingo will be positive or not, potentially informative in this regard is that within Vesselinov and Grego's (2012) effectiveness study, only 42% (66/156) completed the entire 34-hour training course, and 25% of the entire sample completed less than 8 hours (Krashen, 2014). While these statistics by themselves cannot be taken as definitive evidence of a lack of positive learning experience in Duolingo, combined with the lack of cultural and pragmatic knowledge, such figures raise concerns about the overall Positive Impact Duolingo may have on the user.

Practicality. The mobile nature of MALL technology allows users ready access to language learning apps in a wide array of settings. As previously indicated, Duolingo is accessible by PC via duolingo.com, but is also available as an app for tablet and mobile device operating systems (e.g., iOS, Android), with user progress synced across the multiple platforms. The only caveat to this readily available access is the need for an Internet connection, as Duolingo has limited mobile storage capacity when offline (<https://support.duolingo.com/hc/en-us/articles/204567584-Do-I-need-the-internet-to-use-Duolingo->). This one caveat aside, the free-of-charge nature of Duolingo combined with relatively wide-ranging accessibility provides users with a learning tool that appears quite accessible for a variety of learning needs and contexts.

Discussion

Two research questions guided this evaluation of the MALL-based app Duolingo. The first related to how strongly Duolingo adhered to ISLA-informed theory, the second to how effective Duolingo could be predicted to be based on this theoretical adherence (in absence of other non-design-based considerations which were not measureable within the current analysis). To address these questions, a CALL-inspired evaluative framework developed by Chapelle (2001) was utilized, with evaluation based on the researchers' combined experience learning Turkish (34 hours of study each). Table 2 provides a summary of this evaluation.

Table 2

Summary of Duolingo Evaluation (Chapelle, 2001; Jamieson et al., 2004, 2005)

Criteria	Evidence	Judgment
Language Learning Potential	<i>Input Enhancement</i> Saliency: Color; cross out techniques; lack of direct focus in attention drawing techniques	Fair
	Modification: Images; hypertext; slow-paced aural repetitions	Fair
	Elaboration: Clarification clauses, phrases; lack of context	Absent
	<i>Interaction</i> Discussion board; self-contained; items with multiple possible responses, choice of distractors too simple	Poor
	<i>Production</i> English to target language translation; spoken repetition task	Absent
Meaning Focus	Lexical and grammatical focus; no interpretation beyond word level meaning; no construction of meaning	Poor
Authenticity	No real-world context; lexical and grammatical focus; phrases presented out of context	Poor
Learner Fit	“Test Out” option; limited consideration of individual differences; need to consider in classroom setting	N/A
Positive Impact	Daily reminders; goal setting; limited cultural and pragmatic learning; overall learning experience unknown	Poor
Practicality	Free; Available across multiple platforms; Requires Wi-Fi	Good

Following Chapelle's (2001) ISLA-informed criteria, there is an apparent bias towards form-focused instruction, with the benefits of Duolingo being found in its ability to make salient and elaborate on lexical and grammatical elements of the target language. This approach is defined by the incorporation of techniques from both the explicit (consciousness raising) and implicit (input enhancement) ends of the form-focused instruction continuum, indicating a willingness to utilize both focus-on-forms and focus-on-form (Loewen, 2015). However, focus-on-form occurs within meaning-focused instruction, key aspects of which (i.e., interaction and production) are absent from the Duolingo instructional approach. A primary reason that multiple instructional approaches exist is to account for the limitations of the other. While form-focused instruction has been critiqued for potentially leading only to explicit knowledge, meaning-focused instruction has similarly been critiqued for potential shortcomings in linguistic development (Loewen, 2015). Based on our analysis, Duolingo prioritizes explicit knowledge, devoid of contextual meaning. Of considerable concern, then, is whether Duolingo promotes the implicit knowledge necessary for productive L2 communication or merely “language-like behavior” (Ornstein, Ewton Jr., & Mueller, 1971, p. 57), in which users complete programmed instruction without developing underlying competence (Hoopingarner, 2009).

This final point becomes vital when considering Duolingo's claim, based on Vesselinov & Grego's (2012) effectiveness study, that 34 hours of Duolingo usage is equivalent to (or more beneficial than) a semester of university language study. However, it is important to consider the type of knowledge measured in Vesselinov and Grego's study, in which they used the Web Based Computer Adaptive Placement Exam (WebCAPE; <http://www.perpetualworks.com/webcape/overview>). While commonly used in higher education settings (Lin & Warschauer, 2015), WebCAPE primarily addresses vocabulary, grammar, and reading knowledge/ability (Lin & Warschauer, 2015) assessed via multiple-choice questions, essentially placing form before meaning (Krashen, 2014). It should not come as a surprise to see Duolingo users demonstrate significant gains in knowledge on an assessment that allows them to call upon the explicit knowledge that Duolingo prioritizes. So, while 34 hours of Duolingo usage may parallel the features of form-focused instruction addressed during a one-semester language program, the extent to which it compares to the benefits of meaning-focused instruction is unknown (see Krashen, 2013, for a similar argument regarding Rosetta Stone). Between the lack of a meaning-focused knowledge assessment in Vesselinov and Grego's effectiveness study and the poor integration of productive and interactive tasks identified in our evaluation, Duolingo would appear to have a long way to go before it can effectively meet the real-world needs of teachers and learners alike.

The above evaluation is not meant to dismiss Duolingo as a language-learning tool outright. Form-focused instruction indeed plays a vital role in language acquisition, and though limited as a sole tool for language learning, Duolingo could potentially serve as a support tool in a classroom setting. In fact, founder Luis von Ahn has argued that Duolingo was never intended as a replacement for the language teacher, but as a tool to motivate learners to continue their learning beyond the classroom (“Interview with founder of Duolingo,” 2016). In support of this claim, Duolingo launched “Duolingo for Schools” in 2015 allowing teachers to track student progress, identify individual learner patterns and address individual- and group-level difficulties (<https://schools.duolingo.com/>). However, despite this argument put forth by von Ahn, supported by Duolingo for Schools, Duolingo still markets itself as a private tutor (<https://www.duolingo.com/info>), advocating the Vesselinov & Grego (2012) effectiveness study,

without making clear the potential limitations of a solely autonomous approach to learning.

It should be noted that the above evaluation is limited in several ways. A key limitation is that only the Turkish curriculum was considered, and only through 34 hours of usage. It cannot be claimed that progression through other language courses parallels that of Turkish, or that the type of instruction and tasks offered beyond 34 hours are the same as those offered before. An additional limitation of the current evaluation is that it focused specifically on how ISLA-based theory is implemented within Duolingo, leaving aside sociocultural and technological affordances that have come to define MALL technology (Reinders & Pegrum, 2015). This evaluation was carried out under the belief that a theoretical approach to language learning should underlie any instructional design; therefore the additional affordances that MALL technology can provide, such as access to local and global, episodic and extended, and personal and social learning (Reinders & Pegrum, 2015), are left unaddressed, even though they may provide new ways in which such language learning theory can be implemented with MALL technology. Finally, the current study considered only the potential of Duolingo as a language learning tool, in relationship to how well it adhered to ISLA-based theory, and provides no actual measures of acquisition. Therefore, future effectiveness research accounting for learners' outcome is encouraged. Based on the limitations of the WebCAPE assessment used in Vesselinov and Grego (2012) to address the effectiveness of Duolingo, there is a need for a more in-depth holistic assessments of Duolingo as a language learning tool, especially given its current worldwide popularity.

References

- Anderson, K. T. (2009). Applying positioning theory to the analysis of classroom interactions: Mediating micro-identities, macro-kinds, and ideologies of knowing. *Linguistics and Education*, 20(4), 291-310.
- Barkhuizen, G. P. (2013). *Narrative Research in Applied Linguistics*. Cambridge: Cambridge University Press.
- Bodycott, P. (2009). Choosing a higher education study abroad destination: What mainland Chinese parents and students rate as important. *Journal of Research in International Education*, 8(3), 349-373.
- Chacon, E. (1999). *Survey of International Students at the University of Alberta* (unpublished master's thesis). University of Alberta, Alberta.
- Davies, B., & Harré, R. (1990). Positioning: The discursive production of selves. *Journal for the Theory of Social Behaviour*, 20, 43-63.
- De Costa, P. I. (2011). Using language ideology and positioning to broaden the SLA learner beliefs landscape: The case of an ESL learner from China. *System*, 39(3), 347-358.
- Duff, P.A. (2008). *Case study research in applied linguistics*. New York, NY: Lawrence Erlbaum.
- Gieve, S., & Clark, R. (2005). 'The Chinese approach to learning': Cultural trait or situated response? The case of a self-directed learning programme. *System*, 33(2), 261-276.
- Clark, R., & Gieve, S. N. (2006). On the Discursive Construction of 'The Chinese Learner'. *Language, Culture and Curriculum*, 19(1), 54-73.
- Gu, Q., & Schweisfurth, M. (2006). Who Adapts? Beyond Cultural Models of 'the' Chinese Learner. *Language, Culture and Curriculum*, 19(1), 74-89.
- Hafner, C. A., Chik, A., & Jones, R. H. (2013). Engaging with digital literacies in TESOL. *TESOL Quarterly*, 47(4), 812-815.
- Hatch, E. M. (1978). *Second language acquisition: A book of readings*. New York: Newbury House Publishers.
- Huang, J., & Brown, K. (2009). Cultural factors affecting Chinese ESL students' academic learning. *Education*, 129(4), 643-653.
- Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Karoma Pub.
- Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis. *Studies in Second Language Acquisition*, 16(01), 73-98.
- International Studies and programs: Office for International Students and Scholars (2013). *Statistical Report*. <http://oiss.isp.msu.edu/about/statistics.htm>
- Jackson, J. (2011). Mutuality, engagement, and agency: Negotiating identity on stays abroad. *Identity formation in globalizing contexts*, 127-146.

- Jin, L., & Cortazzi, M. (2006). Changing Practices in Chinese Cultures of Learning. *Language, Culture and Curriculum*, 19(1), 5-20.
- Kanno, Y., & Norton, B. (2003). Imagined communities and educational possibilities: Introduction. *Journal of Language, Identity, and Education*, 2(4), 241-249.
- Kayi-Aydar, H. (2013). 'No, Rolanda, completely wrong!' Positioning, classroom participation and ESL learning. *Classroom Discourse*, 4(2), 130-150.
- Kayi-Aydar, H. (2014). Social Positioning, Participation, and Second Language Learning: Talkative Students in an Academic ESL Classroom. *TESOL Quarterly*, 48(4), 686-714.
- Kayi-Aydar, H. (2015). Teacher agency, positioning, and English language learners: Voices of pre-service classroom teachers. *Teaching and Teacher Education*, 45, 94-103.
- Kinginger, C. (2008). Language learning in study abroad: Case studies of Americans in France. *The Modern Language Journal*, 92(s1), 1-124.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590-619.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge university press.
- Lehto, X., Cai, L., Fu, & Chen, Y. (2014). Intercultural interactions outside the classroom: Narratives on a US campus. *Journal of College Student Development*, 55(8), 837-853.
- Liou, T. (2014). Chinese-speaking international students' experiences and perspectives of ESL university writing programs. *Dissertation Abstracts International*, 74, 7-A(E).
- Liu, C. (2013). From Language Learners to Language Users: A Study of Chinese Students in the UK. *International Journal of Applied Linguistics*, 23(2), 123-143.
- Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal*, 92(1), 71-86.
- Lu, Y., & Hsu, C. (2008). Willingness to communicate in intercultural interactions between Chinese and Americans. *Journal of Intercultural Communication Research*, 37(2), 75-88.
- Luxin, Y. (2010). Doing a group presentation: Negotiations and challenges experienced by five Chinese ESL students of Commerce at a Canadian university. *Language Teaching Research*, 14, 141-160.
- Miller, L. (2014). Academic listening in a second language context: An Asian perspective. *The Asian Journal of Applied Linguistics*, 1, 64-75.
- Norton, B. (2000). *Identity and language learning: Gender, ethnicity and educational change*. Buenos Aires: Editorial Dunken.
- Norton, B. (2013). *Identity and language learning: Extending the conversation*. Bristol: Multilingual Matters.
- Office for International Students and Scholars. (2014). International Students by Country of

- Origin: Asia. In *2014 Statistical Report*. Retrieved June 16, 2015, from <http://oiss.isp.msu.edu/documents/statsreport/14pdfs/Asia2014.pdf>
- Ranta, L., & Meckelborg, A. (2013). How Much Exposure to English Do International Graduate Students Really Get? Measuring Language Use in a Naturalistic Setting. *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 69(1), 1-33.
- Rastall, P. (2006). Introduction: The Chinese learner in higher education – transition and quality issues. *Language, Culture and Curriculum*, 19(1), 1-4.
- Talmy, S. (2010). Qualitative interviews in applied linguistics: from research instrument to social practice. *Annual Review of Applied Linguistics*, 30, 128-148.
- Thorne, S., Sauro, S., & Smith, B. (2015). Technologies, identities, and expressive activity. *Annual Review of Applied Linguistics*, 35, 215-233.
- Valdés, G. (1998). The world outside and inside schools: Language and immigrant children. *Educational Researcher*, 4-18.
- Willett, J. (1995). Becoming first graders in an L2: An ethnographic study of L2 socialization. *TESOL Quarterly*, 29(3), 473-503.