37.

Landscape Analysis of Second Language Learning Games

Abstract

Games have the opportunity to provide language learners rich multimodal environments that ground language learning in a situated context. There are now a wide variety of second language learning games in a number of languages for different audiences. In this paper, we examine 68 games across different platforms to evaluate their approaches to pedagogy, proficiency, assessment, skills developed, and complexity. We describe our data collection and analysis procedures and then summarize the major trends in these areas. We found that most games take a didactic pedagogical approach, are targeted toward novices, incorporate assessment systems, focus on vocabulary development, and that average internal rating did not increase with the complexity of the learning within the game. The goal of this analysis is to inform and contextualize future potential efforts in this particular domain.

Introduction

Games are a natural environment for language learning. From simpler titles for young preliterate children like *Peekaboo Barn* to larger, online immersive social worlds for teenagers and young adults like *XENOS*, games provide learners rich multimodal environments that ground language learning in the situated context of its meaning and use (Gee, 2010), thereby enabling meaningful language acquisition to take place. This affordance of games to enable situated meaning construction and use is in close keeping with current paradigms for language learning instruction in both formal and informal classrooms (Young et al, 2012). Games allow the grounded use of language in a rich context, both virtually and, in some cases, socially. Thus, it is no surprise that games show dramatic gains in the area of language learning in meta-analyses and reviews to date (Peterson, 2010; Young et al, 2012; Wouters et al, 2013).

Games are also a natural environment for transformative assessment, providing learners with a rich digital context for task engagement on the front end and a large volume of clickstream user data on the back end. With the shift from a print-based educational ecosystem to a digital one, educational researchers and assessment experts can use the "data exhaust" of students' choices and performances in games and other digital tools to assess learning and improve instruction (Schoettler, 2012). Tracking and modeling student performance better enables the design and delivery of instruction and assessment (Shute & Becker, 2010). Games are no exception: they can be designed to align closely to content standards, give just-in-time feedback on performance, and present data on problem-solving in situ that is superior to those data gathered through traditional paper based measurement instruments (Levy & Mislevy, 2004). Recent developments in machine learning techniques, including both educational data mining and learning analytics, offer compelling means for measuring learning in game-based environments (Dangauthier et al, 2008).

What, then, is the current landscape of games for language learning and assessment? Here, we present a landscape analysis on the subject of games for second language (L2). Here, "second language (L2) learning" is defined to include acquisition of a second (non-native) language by speakers of school age years (6 years of age) through adulthood, including both oral and written communication. "Games" are defined as digital applications with game mechanics or gamification elements. We conducted a broad survey of existing L2 games on the current market and analyzed the corpus in terms of their core game mechanics and features (e.g. platform, languages covered, target age group), pedagogical elements (e.g. didactic, whole language, immersion), and assessment strategy (e.g. level-based summative, problem sequencing, challenge levels), and overall quality. In this report, we detail our data collection and analysis procedures and then summarize the major trends. The goal of this analysis is to inform and contextualize future potential efforts in this domain.

Methods

Data Collection

The games corpus use for this analysis was collected through broad search of academic print publications and online sources, triangulated against one another and recommendations from experts in the fields of games for learning, games for assessment, and L2 learning. Our final source set included 5 popular game distribution sites widely known to feature educational games (iOS App Education Store, Google Play, Steam, DBD Games, and CBeebies); 21 curated websites featuring games for learning, including both commercial sites (e.g. LeapFrog, BrainPOP) and philanthropic/community listings (e.g. Common Sense Media, Serious Games Directory BETA); and finally (4) 12 curated "best of" lists of games for language learning featured on heavily circulated blogs (e.g., Lifehacker.com, App Picker).

Using this source list, we then collected all L2 games described, reviewed, assessed, or referenced that were released after 2005 and developed on a platform still used (i.e. we eliminated DOS games). The final data corpus consisted of 68 game titles. The majority of games included were developed in United States or Europe and were made within the last five years.

Data Analysis

In order to analyze the game corpus, we developed a three-part framework for analysis. The first part of the framework consists of a simple coding scheme designed to capture descriptive characteristics of each title (title, platform, creator, year, cost, target age, source language, target language, genre). The second part of the framework includes an interpretive coding scheme focusing on the instructional nature of the game (pedagogy, proficiency level, skills developed, form of assessment used). The final part of the framework consists of a quality assessment rubric for gauging the overall value of the game; here, we note the public rating of each game in addition to our own 1-to-5 star rating assigned by the internal team.

Four research team members jointly developed the rubric and piloted it on 4 game titles jointly. Once agreement was established, we then piloted use of the framework on 4 games coded individually, then met and discussed the results. Adjustments were made to clarify more any ambiguous categories of the

rubric and then, once all existing uncertainties were addressed, we divided the remaining games corpus among the four researchers and completed all coding over a period of roughly four weeks. During this analysis phase, we met once a week to calibrate our assessments and resolve any additional questions or contentions that arose.

Results

Pedagogy

We classified the L2 games in terms of educational method or pedagogy used using three basic categories. Didactic games were games characterized by a focus on memorization and structure. Games that resemble flashcards are by and large didactic. Communicative games were games characterized by a focus on expressive language and task- based activities. Many educational role playing games that require players to converse with others as a means of completing tasks were categorized as communicative. Immersive games are games characterized by a "whole language" view of L2 learning which does not rely on the learner's native language. Games can and often do span more than one category, with about a fourth of the corpus (26%) falling into more than one classification. As Figure 1 shows, the majority of games were didactic in nature, featuring flash card like activities for rote memorization of vocabulary within the target language.

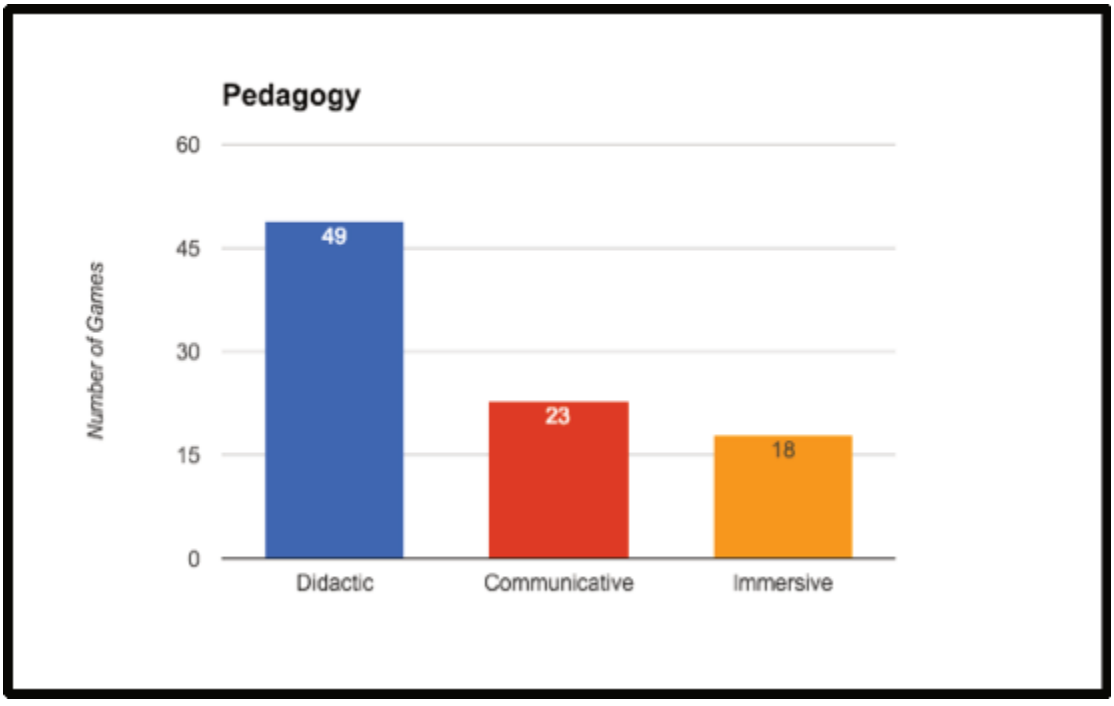

Figure 1. L2 learning games by pedagogy.

Proficiency Levels

L2 games on the current market differ in terms of the level of proficiency targeted (Figure 2). We

analyzed proficiency levels based on the American Council for the Teaching of Foreign Languages Proficiency Guidelines (2012).

Games often span more than one category; for example, when a given title covers material for beginning language learners up through intermediate. As Figure 2 shows, most L2 games focus only on rudimentary language skills (novice only), with rapidly decreasing representation in categories of high proficiency levels (intermediate and advanced). Games that attempt to take a learner from novice through advanced levels of language proficiency are ambitious and the exception. Noteworthy examples include expensive and well-established products like *Rosetta Stone* and its competitor *Voxy*, a subscription-based language teaching application.

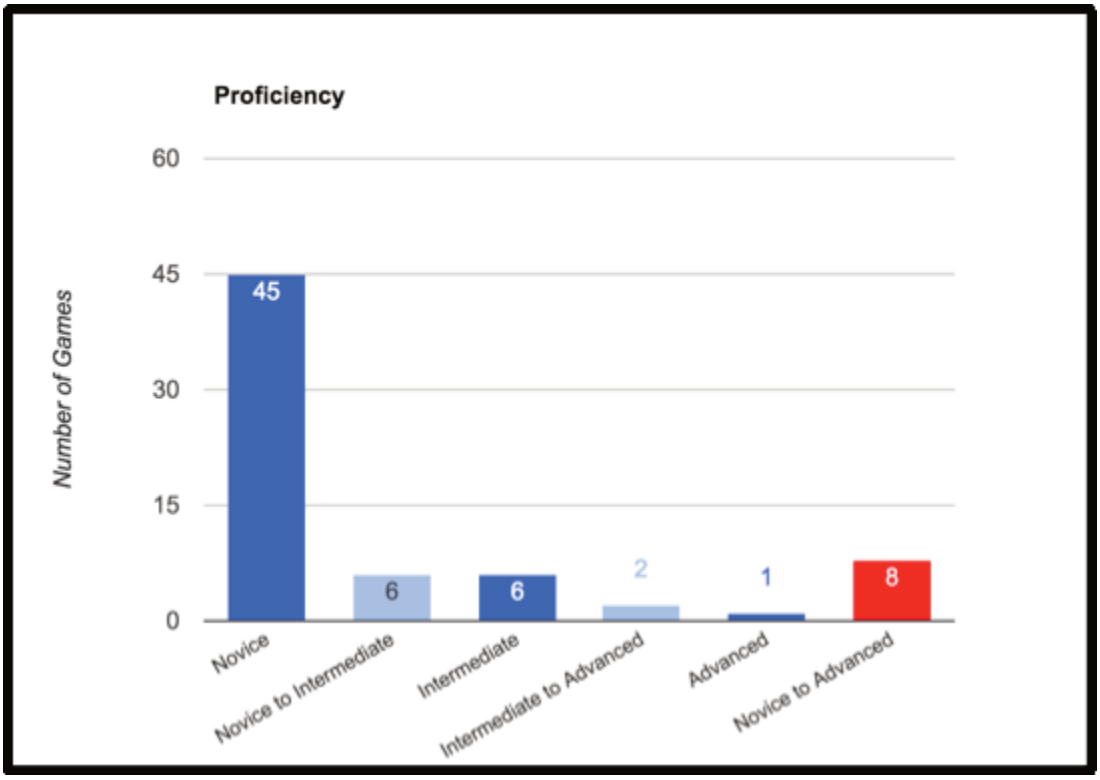

Figure 2. L2 learning games by proficiency level.

Assessment

For software to be interactive, it must take user input and provide some form of feedback or response. Thus, it is not surprising that nearly all L2 games on the market include some form of assessment of what the user does. After all, without at least tracking user progress through the presented content in some fashion, it would be difficult for a system to appropriate respond to what the user does. Thus, only 5% of the corpus contains no assessment model whatsoever. For the remaining game titles, the question is not whether they assess but to what degree. Figure 3 shows a breakdown of games by assessment type, with increasingly sophisticated models represented along the x-axis:

- **Tracks Progress** – the game tracks the amount of content completed
- **Tracks Performance** – the game tracks the amount of content not just completed but also to what

degree of success

- **Content Sequenced** – the game levels increase in difficulty with successful gameplay
- **Responds to Failure** – the game not only increases difficulty with success but also decreases difficulty with failure
- **Data Dashboard** – the game presents some aggregated representation of user data to the user (or their teacher/parent)

Games can and frequently do span more than one category of assessment – generally, in additive fashion with games in more advanced categories entailing all lower assessment forms. As one can see from Figure 3, as assessment method increases in sophistication, the number of L2 game titles featuring it declines in near linear fashion.

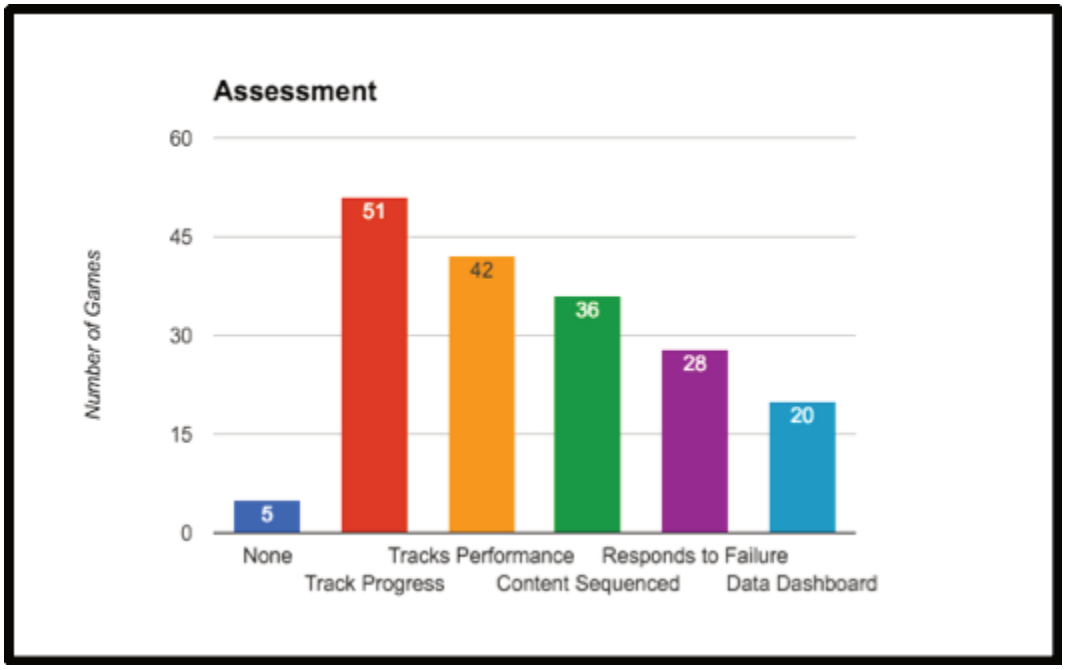

Figure 3: L2 learning games by assessment model used.

Skills Developed

Figure 4 shows the content L2 games are designed to teach based on their developer's claims and our internal play-thrus. Again, games may span multiple categories; for example, *Duolingo* teaches grammar, pronunciation, listening, reading and speaking. Note that skills whose instruction and assessment can be accomplished through little or no complex user input appear more frequently in the corpus than skills that require more complex user input: vocabulary is twice as frequent as grammar, listening is twice as frequent as speaking, and reading is twice as frequent as writing. General constraints on natural language processing may be the cause: automatically parsing simple strings of text is, after all, much more tractable than assessing complex paragraphs of prose. Advances in this area appear to be generally underleveraged in L2 games.

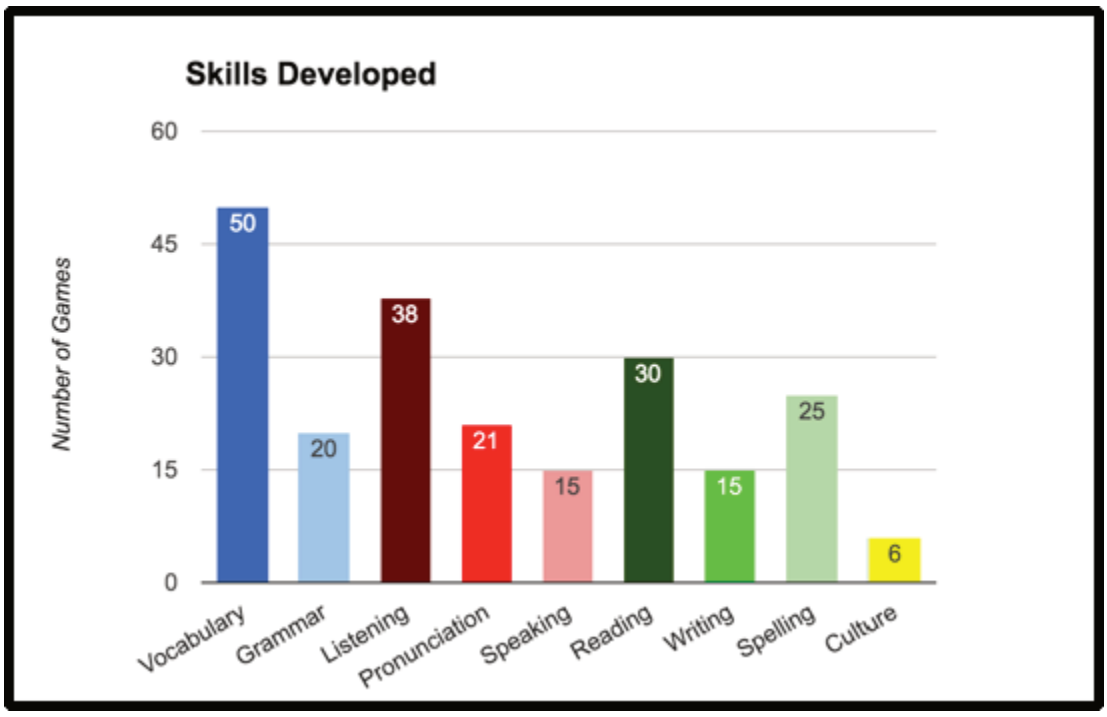

Figure 4. L2 learning games by skills developed.

Different Pedagogies for Different Content

Different pedagogical styles are used to target different learner groups, with far more didactic materials produced for novices than for more advanced learners (Figure 5). Didactic pedagogy frequently targets vocabulary, listening, and spelling skills (Figure 6).

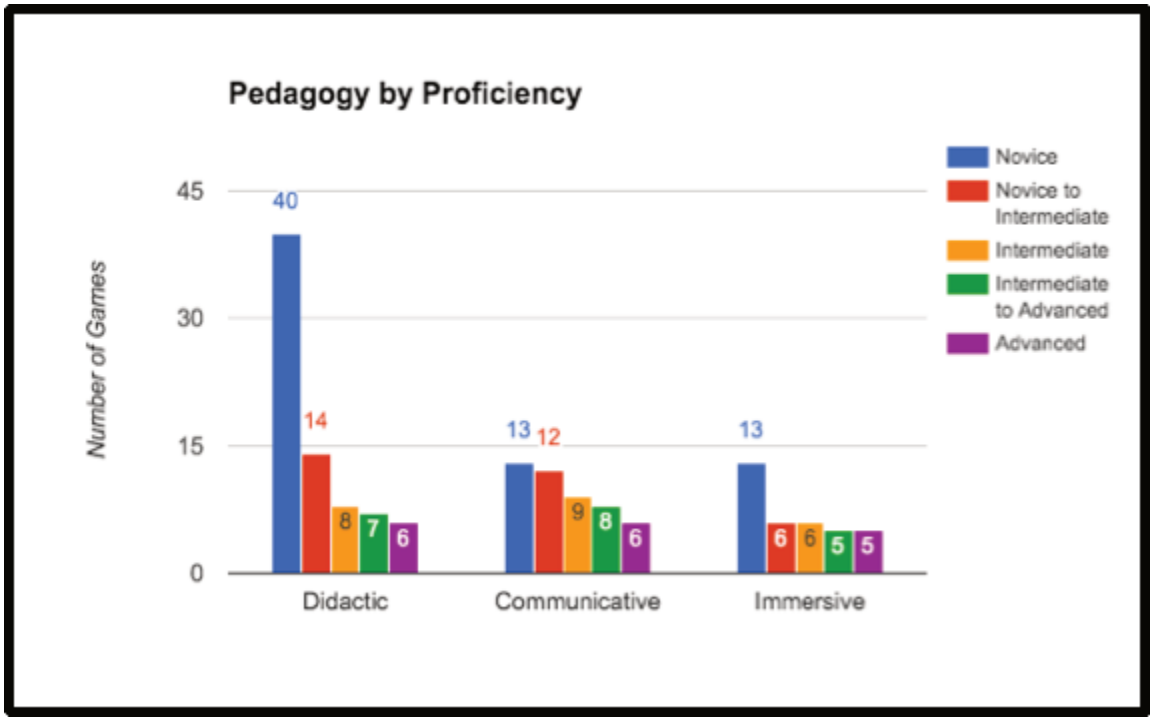

Figure 5. L2 game proficiency level by pedagogy.

As Figure 6 shows, when instead speaking and writing skills are of primary focus, you see a decline in didactic approaches to activities and a concomitant rise in communicative approaches. Both communicative and immersive pedagogies appear to target almost all skills examined for this analysis, often in combination as part of authentic tasks conducted in-game such as figuring out the appropriate level of formality when addressing a stranger in a business letter or interacting with a non-player character in the target language to order to buy coffee. There are also nuanced difference between communicative and immersive games: In immersive games, writing and culture is notably more frequently the focus while grammar and spelling is notably less frequently the focus. Such differences are in keeping with the whole language approach that marks immersive games overall.

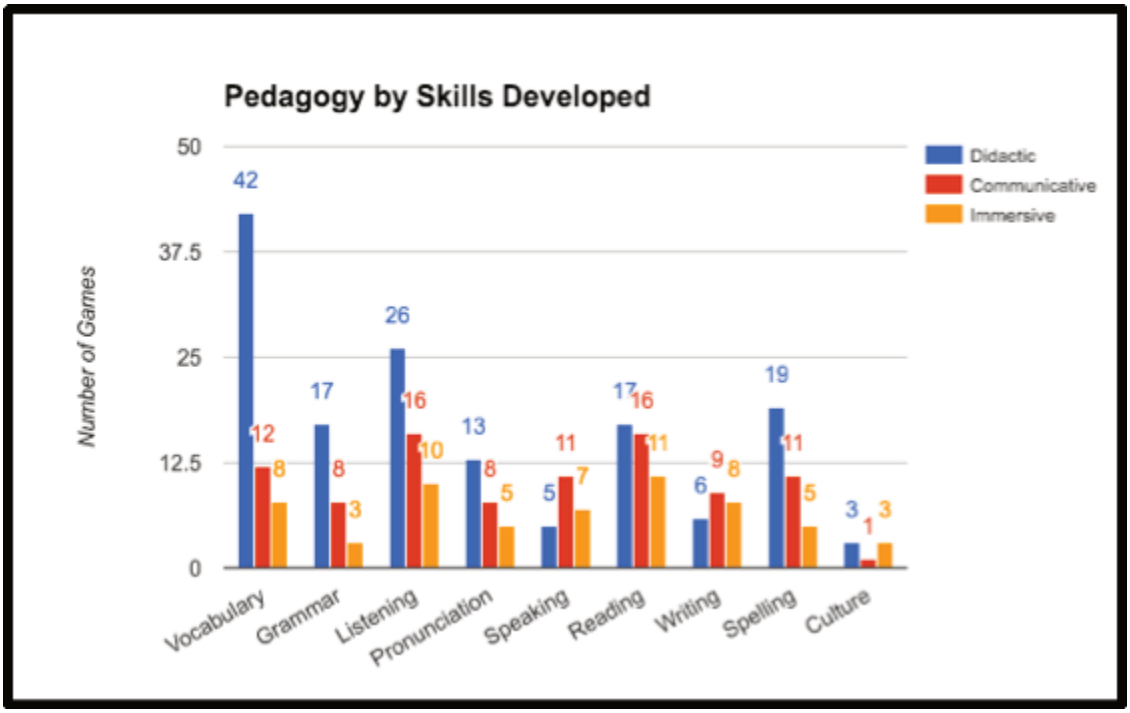

Figure 6. L2 game pedagogy by skills developed.

Complexity

Although our internal rating system, by definition, overtly paid special attention to comprehensiveness of content, more ambitious games did not necessarily earn higher ratings. Four of our codes consisted of ordinal variables that could be reduced down to a scaling degree of complexity: pedagogy, proficiency level, skills developed, and assessment. Figure 7 shows the average rating of games at different levels of complexity within each of the four variables, represented as separate lines in the graph. Viewed as such, it becomes clear that average rating is more or less the same regardless of complexity, essentially flatlining for all but lowest level on the assessment rubric (designating “none”).

Figure 7. L2 game ratings by category complexity.

In summary, games that integrate more complex mechanics and content do not necessarily equate better learning. Next, we will give an example of a game that integrates the mechanics that we have discussed thus far in a way that is both engaging and beneficial to the player.

Examples

Supiki (iOS platform) is an interactive textbook designed for English language learners ages thirteen and above who wish to practice English speaking in realistic situations. It offers 50 content units across ten levels of difficulty that spans all three proficiency levels from novice to intermediate to advanced for only 11.99 USD. *Supiki* tracks user progress throughout each unit and provides a summative "Big Quiz" at the end of each unit to assess learners' understanding of the expressions just learned. Its graphics are appealing, the interface is easy to use, and its content is thoroughly engaging.

Supiki's pedagogy spans both communicative and immersive, with each unit framed with a short video clip that introduces the narrative premise of the activity – the background and characters that contextualize the user's tasks. After watching the video, learners choose from one of two follow up extension activities: idioms or conversation (Figure 8). The idioms section provides definitions of the idioms from the video; users can review parts of the conversation again to see how specific expression are used in real time. They are asked to verbally produce a sentence that includes the selected idiom and then compare their utterance to a native speaker. The conversation section allows learners to practice speaking with characters on everyday topics related to the narrative premise and to receive instant responses to their verbal performances. Throughout both sections, the game content is practical and authentic, encouraging users to improve their fluency and gain more confidence in speaking.

One of *Supiki*'s more noteworthy features is its smart speech recognition software. While users are practice their language skills, Supiki makes them feel like they are talking with real people. For example, a character from Supiki asked a research team member, "What type of work do you do?" The researcher responded, "I'm a teacher." Unlike other games, *Supiki* continued the conversation in a way that built on her previous answer, responding " I think that Education is very important…." After conversing in this way with an in-game character, learners can review their recordings and those of other users, email their recordings to others, or share their scores on Facebook and Twitter (Figure 8).

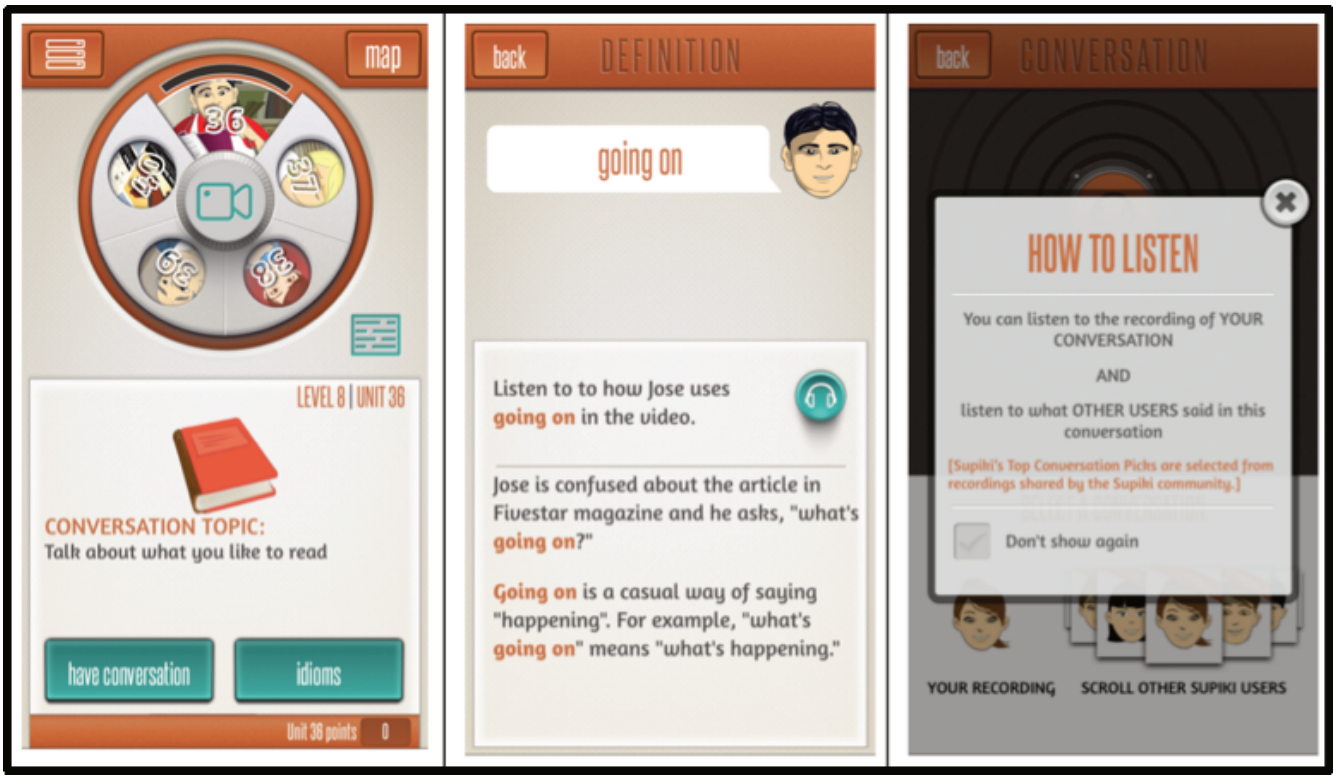

Figure 8. Supiki's interface.

Overall, *Supiki* is a beautifully designed and effective tool for practice speaking and listening language communication skills in a naturalistic context, with the game narrative and mechanics working in unison to emulate everyday interpersonal interaction and to codify the overall results of those interactions so that the system is responsive to the core performances targeted. Together, all aspects of the design work together to create an experience that feels genuinely useful and authentic, balanced in terms of difficulty scaling, and easy to use.

References

American Council on the Teaching of Foreign Languages. (2012). *American Council on the Teaching of Foreign Languages proficiency guidelines* [Electronic version]. Retrieved from http://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf

Dangauthier, P. Herbrich, R., Minka, T., & Graepel, T. (2008). TrueSkill through time: Revisiting the history of chess. In *Advances in Neural Information Processing Systems 20*, Cambridge MA: MIT Press.

Gee, J. P. (2010). *An Introduction to Discourse Analysis: Theory and Method (3rd Edition)*. London: Routledge.

Levy, R., & Mislevy, R. J. (2004). Specifying and refining a measurement model for a simulation-based assessment. *International Journal of Testing, 4,* 333–369.

Peterson, M. (2010). Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation & Gaming, 41*, 72–93. doi:10.1177/1046878109355684

Schoettler, S. (2012). *Learning analytics: What could you do with five orders of magnitude more data about learning?* Keynote address at the 2012 Strata Conference, February 29, 2012, Santa Clara, CA.

Shute, V.J. & Becker, B.J. (2010). *Innovative assessment for the 21st century*. NY: Springer.

Supiki (2014). Linguacomm Enterprises Inc. (version 2.6.1) [iOS]. Retrieved from https://itunes.apple.com/us/app/supiki-english-conversation/id443013473?mt=8

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013, February 4). A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology*. Advance online publication. doi: 10.1037/a0031311

Young, M.F., Slota, S., Cutter, A.B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & and Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61–89.