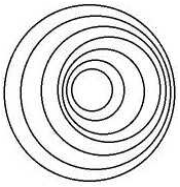


The IALLT Journal

A publication of the International Association for Language Learning Technology

ROSETTA STONE FOR LANGUAGE LEARNING: AN EXPLORATORY STUDY

Gillian Lord
University of Florida

ABSTRACT

The Rosetta Stone program advertises that it can teach language as effectively as, or even more effectively than, a typical classroom-learning environment. Little research has examined this claim, but as institutions are asked to cut costs and simultaneously embrace digital technologies, these programs are often considered as a possible solution to potentially replace teachers or other personnel. This exploratory multiple case study examines the claims and learning outcomes of the Rosetta Stone program among beginning Spanish learners to assess the effects of a semester-long treatment in which participants used Rosetta Stone as their class textbook or alone (instead of any class attendance), as compared to a control group. Data analysis focuses on learner outcomes in terms of linguistic production as well as their attitudes about the materials. Results reveal qualitative differences characterizing learners' speech and strategies, as well as their reactions to the program. While continued investigation is needed, these initial results do not yet provide indication that the Rosetta Stone program, although possibly able to deliver success in some areas, would be capable of replacing the classroom language learning experience.

INTRODUCTION

As the field of Computer Assisted Language Instruction (CALL) continues to grow, its scope has expanded from simple computer programs designed to practice verb conjugations to ever-expanding social networking tools to foster communication and collaboration across learner and speaker populations (e.g., Bax, 2003; Warschauer 1996). Most recently, CALL researchers are contending with a new kind of tool: stand-alone self-paced language learning programs such as Rosetta Stone (Rosetta Stone, http://www.rosettastone.com), Transparent Languages (http://www.transparent.com) or Pimsleur (http://www.pimsleur.com). The claims made by such programs are lofty indeed, asserting that they can be more effective, more fun, and/or more efficient than more traditional forms of learning. For example, Rosetta Stone's advertisements feature claims that it is the "only way to learn a language." More perhaps than any competitors, Rosetta Stone has significantly increased their marketing efforts in recent years, with an extra push to break in to the education arena, both K-12 and higher education. According to Kantar Media, Rosetta Stone "spent $98.5 million on advertising in 2011, up from $70.5 million in 2010" (Newman 2012: n.p.).

In the face of such marketing, educational administrators – who are increasingly needing to cut costs while maximizing student outcomes – have begun to consider tools like Rosetta Stone to supplement or replace their existing language programs, even in the absence of convincing empirical evidence supporting such claims (e.g., Rundquist 2010). The goal of this multiple case study is to attempt an initial assessment of the outcomes that are possible with the Rosetta Stone program, and to compare those to outcomes obtained in a typical university-level language course. By considering a variety of data sources, we begin to assess the feasibility of the claims made by Rosetta Stone.

By way of background, we consider the following:

1) the theoretical underpinning of the Rosetta Stone program;
2) scholarly reviews of the Rosetta Stone program;
3) previous empirical studies exploring the Rosetta Stone program's outcomes.

As will become evident, however, data in these areas are lacking, and further work is needed. The present study adds experimental qualitative and quantitative data from an exploratory investigation into the use of Rosetta Stone for beginning

language learning. These data complement the above information by providing the following additional perspectives:

> 4) attitudes: user attitudes about and reactions to the Rosetta Stone program (including usability and language learning perceptions);
> 5) fluency: analysis of user oral production based on standard measures of fluency.

Taken together, these data sources lay the beginnings of a more solid foundation on which to assess the potential of the Rosetta Stone program than previous works, which have considered only single data sources, and rarely from learner outcomes. Given the fact that this issue has not been fully addressed in scholarly work to date, and the broad nature of the research, the study is exploratory in nature and does not presume to address all potential variables. However, as the goal of any exploratory research is to better understand a given problem and to determine appropriate research methods for future investigations, this study opens the door for those works.

BACKGROUND

Theoretical premise

With any emergent CALL tool, it is worthwhile to explore the theoretical premises that underlie its development, and to assess their validity within current second language acquisition (SLA) frameworks. The cornerstone of Rosetta Stone's program, according to their own promotional materials, is Dynamic Immersion, which is described on their website as follows: "By eliminating translation and grammar explanations from language learning, Dynamic Immersion activates your own natural language-learning ability. You begin to think in your new language from the very beginning—the same way you learned your first language." In essence, the program relies on target language input and visual aids, without translations or explicit instruction. The fundamentals of this approach are not new; they are the same ideas that formed the basis of the Natural Approach to language learning (Terrell 1977; Krashen & Terrell 1983), which operates under the premise that adults learn a second language (L2) in the same way that a child learns a first language (L1). In terms of pedagogical approach, the Natural Approach is characterized by "the use of familiar techniques within the framework of a method that focuses on providing comprehensible input and a classroom environment that cues comprehension of

input, minimizes learner anxiety, and maximizes learner self-confidence" (Richards & Rodgers 2001: 186).

Although the Natural Approach enjoyed popularity in its time (Markee 1997), there is no consensus among SLA researchers that the processes of acquiring the L1 and the L2 are indeed the same (e.g., Ervin-Tripp 1974). In spite of agreement on certain aspects – such as the importance of input, for example (e.g., Krashen 1985) – other areas of L2 acquisition differ from L1 acquisition. Well-known work by Bley-Vroman (1989, 1990, 2009), for example, proposes that there is a 'fundamental difference' between learning first and second languages. His research highlights not just the cognitive differences between L1 and L2 acquisition, but also the key social and affective differences in the typical environments for each process. Others have proposed that interaction and output are equally important to the SLA process (e.g., Long 1996, Swain & Lapkin 1995), and such tenets form the underlying premises of most modern approaches to foreign language pedagogy.

While a more in-depth discussion of SLA theories is beyond the scope of this paper, mention is made here in an effort to point out that Rosetta Stone's assumption that learning a L2 is best accomplished as if one were learning their native language is not universally accepted in today's language classrooms. Further, the Natural Approach is no longer widely employed today, precisely because most language educators have come to realize that adult learners do indeed need some degree of explicit instruction and cannot rely on mere input alone (see, for example, work related to Focus on Form and Focus on Forms, e.g., Long 1988, 1991, Long & Robinson 1989). Thus, the premises underlying the Dynamic Immersion principle do not appear to enjoy unanimous support from researchers or teachers. It is not possible to confirm, therefore, either the theoretical or the pedagogical soundness of the underlying premise of the Rosetta Stone programs.

Scholarly reviews

Only a handful of researchers have assessed Rosetta Stone's potential for fostering successful language acquisition. Rifkin (2003), for example, evaluates a number of online language tools on the basis of certain pedagogical criteria. Rifkin notes that Rosetta Stone (admittedly a much earlier version without many of the enhancements of the current version) falls short in many of these areas, citing artificial dialogues and the program's inability to account for natural and

acceptable variations in language. Later work by Lafford, Lafford and Sykes (2010) evaluates Rosetta Stone and other self-study language programs, based on a number of features that previous SLA research has shown to be important in the acquisition process, such as opportunities for interaction, the relevant contextualization of language, etc. The authors conclude that Rosetta Stone does "…not incorporate a number of the research-based insights … that informed SLA scholars might have given [it]" (516).

A subsequent software review by Santos (2011) assesses the Rosetta Stone Portuguese program, in which he notes that in spite of the advantages to its appealing interface, there is a fairly significant lack of context in the materials and an inability to respond to spontaneous student speech. Santos concludes that what Rosetta Stone calls interaction is "a rather poor and limited version of what one would encounter in a real-life conversation" (187). Again, it should be noted that this review occurred before many of the online interactive functionality was built in to the Rosetta Stone program, but his main critiques, i.e., lack of contextualization and spontaneity, remain valid.

Finally, a recent review essay by DeWaard (2013) explores the possibility of Rosetta Stone replacing classroom instruction. She bases her assessment on personal experience and on her own expertise in language teaching. DeWaard too notes the appealing interface of the Rosetta Stone program, but finds it lacking in a number of areas. Specifically, she notes shaky theoretical foundations, cultural inauthenticity and the overall limitations of a nonhuman system, among other limitations. DeWaard concludes that Rosetta Stone is "not a viable replacement of current instruction at the postsecondary level" (61).

Empirical studies

In light of the above reviews, it is surprising that there are still so few data-driven studies examining Rosetta Stone's effectiveness. (In fact, to date, very few other commercial self-study programs have been the focus of academic study either.) One of the few existing studies looking specifically at Rosetta Stone is one that was commissioned by Rosetta Stone itself (Vesselinov 2009), which claims to "decisively [determine] the effectiveness" (1) of the program. Vesselinov finds that after using Rosetta Stone for 55 hours, students "significantly improve" their language skills, while "enjoying" the program. However, these claims must be taken with caution, as the study population was older than a traditional student population (average age = 41) and were a highly

educated group, with 75% already having earned a bachelor's degree or higher. Additionally, no measures of enjoyment or attitude were reported, and no information is provided regarding the participants' native languages or other languages they previously studied or spoke. What's more, Vesselinov's claims of improved proficiency are only improvements when compared to absolute zero: participants obtained an average posttest WEbCAPE (a well-established test for placement in college-level language courses) score of 238, which, in most universities, represents language skills comparable to those of first-semester courses; while the ACTL Oral Proficiency Interview (OPI) results show that although 50-75% of the Rosetta Stone users improved their proficiency by at least one ACTFL sublevel after 55 program hours, 35.6% showed no change. Even so, 94% of the participants (127/135) remained at the Novice level in the posttest. Therefore, what Vesselinov considers a language learning success may not constitute the evidence that language researchers and educators would require in order to demonstrate considerable language acquisition.

There are virtually no other empirical investigations into Rosetta Stone and its ability to foster language learning, although a few other studies have involved the program. Nielson (2011) explores the use of self-study programs in the workplace to examine how a population of professionals uses and learns from them. Her study tracked 326 U.S. government employees using either Rosetta Stone (n=150) or Auralog's Tell Me More[1] (n=176); the Rosetta Stone users were evenly divided between Spanish (n=50), Chinese (n=50) and Arabic (n=50) and were asked to spend ten hours a week with the program, for a 20-week period, as well as to keep a learner log to track their use and progress, and to engage in various assessments. Data on linguistic outcomes in Nielson's study are scant, due to severe attrition: only 6/150 Rosetta Stone users completed the second assessment, while only 1/150 users completed the third and fourth. Additionally, the learner that did complete the final assessment received only a Novice High rating, in spite of having achieved perfect scores on his in-program assessments, indicating again that mastery of the Rosetta Stone material may not correlate with other measures of effective communication. Nielson concludes that although programs like Rosetta Stone offer attractive options, "they are not yet able to offer an alternative to human support or interaction" (125).

With these three factors as background, it would seem that there is not yet convincing evidence in favor of recommending the Rosetta Stone program as a valid language-learning tool. However, the scarcity of both quantitative and qualitative empirical findings leaves the question open, and it is precisely this lack of information that motivated the present study, which is described below.

As mentioned above, this experiment is an exploratory first step into the process of studying outcomes from self-study programs such as Rosetta Stone.

METHODOLOGY

Participants and treatment

Participants were students at the University of Florida enrolled in a Beginning Spanish 1 class, which was designed for and restricted to those with no prior Spanish. The primary goal of this course is to offer students an introduction to basic communicative skills in Spanish, while developing an awareness and appreciation of Hispanic cultures. Beginning Spanish classes at this university are blended, meeting three hours each week with online asynchronous work equivalent to two additional hours. The course adopts a communicative approach in which class time is reserved for communication, and students are expected to use their out-of-class work to prepare for class meetings. Such preparation consists of online grammar or vocabulary tutorials and a series of mechanical-type practice activities. Subsequent class time is then devoted to small group and paired activities to engage learners in meaningful interaction with their peers using the target structures and language. Instructors may begin the period with a brief review of especially complicated grammar topics, and ask students if they have questions, but otherwise there is little explicit grammar instruction or lecturing during class time. While students are engaged in their group work, the instructor mingles throughout the class, answering questions and assisting when necessary. If s/he notices a particularly common or problematic area, s/he may stop class to go over that point, but then returns to the communicative activities.

A total of 68 participants, whose ages ranged from 18-30 years (average age = 20), took part in the study. Only participants who were native speakers of English and spoke no other second/foreign languages were included in the analysis. Because the course was restricted only to true beginners, and it was originally assumed that everyone had that same starting point of zero, pre-tests were not conducted. However, a post-hoc investigation into language background revealed that only 13 had legitimately placed into the course and had in fact not received any formal high school instruction in Spanish. Therefore, in order to rule out any effect of previous Spanish study and to be able to compare end-of-treatment outcomes, only the data from the true beginners were considered for

this analysis, resulting in a much smaller sample size of 12[2]. This reduction in sample size is highly unfortunate, but in order to maintain the integrity of the results it was necessary to limit the current analysis to only the true beginners.

Participants belonged to one of three groups. The Control group consisted of the true beginners (N=4) from of an intact section of the course. The course instructor had several years of teaching experience in face-to-face, hybrid and online formats, and with a wide range of textbooks and materials. The students in the Control group were informed of the study and the consent protocol at the beginning of the semester, when it was also explained that their curriculum was not impacted. They followed the standard syllabus for the course, and used the materials and assessments that all other 'typical' sections used.

Two experimental groups were formed, in order to assess the Rosetta Stone program both as a truly stand-alone program, as well as instructional materials that could be supplemented by class time. The Rosetta Stone (RS) group was voluntarily self-selected from students enrolled in other sections of the Beginning Spanish course[3], as the Institutional Review Board deemed it unfeasible to randomly assign students to this condition given the exploratory nature of the study and the different learning styles involved in the Rosetta Stone program. Only those participants who had never taken Spanish before (N=4 in RS group) are included in this analysis.

These learners received 16-week licenses to Rosetta Stone Version 4 TOTALe® Spanish (see Appendix A for screenshots), which had been purchased with grant funds secured by the investigator. The syllabus for this group was modeled after a sample Rosetta Stone-created program purported to cover material comparable to a university class. Participants had to complete six units of the Rosetta Stone program during the semester. They were also required to attend a minimum of six synchronous Rosetta Studio™ sessions which, according to Rosetta Stone's website, are tutoring class-like sessions designed to provide learners the opportunity to practice with a native speaker, and can include two or three other learners in one session. Finally, these participants had to spend a minimum of eight hours in Rosetta World™, the program's "interactive social language-learning community," where students can connect with other learners in games and chats designed to offer opportunities for interaction in the target language. These participants did not attend any class meetings or engage in any other pedagogical activities beyond the Rosetta Stone program.

The other experimental group was the Rosetta Stone + Class (RS+C) group, which consisted of another intact section taught by the same instructor as the Control group. Again, only the true beginners who completed all aspects of the study are included in this analysis (N=4). Sixteen-week Rosetta Stone licenses were provided to this group, also at no cost to the students, which served as their "text" for the semester. All of the features and requirements described for the Rosetta Stone group hold for this group as well, except that these students were additionally required to attend the three scheduled class meetings each week. (These students were given the option of switching to a non-experimental section of the course, although none chose to do so.) The instructor developed weekly lesson plans based on the syllabus for the Rosetta Stone materials, incorporating conversation and interaction in to the class time; he also developed additional assessment materials relevant to the content and structures covered in the Rosetta Stone program, which were not always comparable to those covered in the course text.

As can be seen in Table 1, all participants indicated on their background questionnaire (Appendix B) that that they were enrolled in the course to fulfill the language requirement for their majors. About 1/3 mentioned that they chose Spanish specifically because it would be useful in their futures.

Table 1.
Summary of participants' language background

Group	High school Language Background	Why Spanish?	Why volunteer for Rosetta Stone?
Control	No Spanish French 3	I am required to take a language for my major.	
Control	No Spanish Latin 3	Language requirement and for myself since I feel Spanish is a good asset as a physician.	
Control	No Spanish	Fulfill [college requirement].	NA
Control	No Spanish Latin AP	I am going to Panama on a service trip. I believe formal classes would help me gain a better grasp of the language than picking it up on my own.	
Rosetta Stone	No Spanish	Foreign language requirement.	Heard a lot of good things about

			Rosetta Stone so decided to try it.
Rosetta Stone	No Spanish Latin 3	CLAS requirement.	Can better manage my time and schedule and move more at my own pace without dealing with class.
Rosetta Stone	No Spanish French 3	Required for major.	Sounded more beneficial.
Rosetta Stone	No Spanish ASL 3	Spanish is useful in my state/needed FL requirement.	I was going to use my own to supplement education anyway.
Rosetta Stone + Class	No Spanish	I am taking Spanish because I feel like it will be beneficial later on in life.	
Rosetta Stone + Class	No Spanish	As a requirement and to benefit my future jobwise.	NA
Rosetta Stone + Class	No Spanish French 2	I need two semesters of a foreign language to graduate.	
Rosetta Stone + Class	No Spanish French 4	College requirement.	

Although the participants in the study may not have possessed intrinsic motivation to learn Spanish, as we traditionally think of it, there is something to be said for their need to earn a passing grade in order to fill a graduation requirement and for wanting to maintain their GPA, and there was no attrition. When asked their motivation for volunteering specifically for the Rosetta Stone group (the only self-selected group of the three), three of the four said they were curious about Rosetta Stone or had considered buying it to supplement their language classes anyway. The fourth chose this option because his other classes were on a different campus and this would make it easier for him to manage time and logistics. They did not report any differing levels of motivation or interest than the participants in the other groups.

Data sources

Various data sources were employed to assess the participants' experiences as well as their linguistic outcomes. At the first of three required meetings with the researcher, participants filled out the background questionnaire mentioned above and a Likert-type survey to gauge their attitudes towards aspects of language learning (Appendix C), which was repeated at the final meeting. Answers from the attitude survey were analyzed to discern any changes in attitudes between pre- and post-treatment sessions, while a content analysis of the English interviews (at all meetings) was carried out in order to expose common student themes related to their attitudes towards their learning materials and experiences, and their perceived learning over the course of the semester.

In terms of assessing linguistic outcomes, multiple data sources were also employed. At all meetings, participants completed an oral interview in English and another in Spanish, along with a Spanish writing task not discussed here. The English interviews focused primarily on status reports, problem-solving, and discussing the participant's reactions to and thoughts on their learning and materials. The Spanish interviews were structured based on the language content that had been covered in recent units, and were designed to engage the students in conversation to practice that content. Because of differences in curriculum, the same questions/prompts could not always be used with all groups, but the content was held consistent to the extent possible. After the end of the semester, students took two standardized tests to assess proficiency: the Versant Automated Oral Proficiency Test in Spanish (http://www.pearsonhighered.com/versant/), and a portion of the Spanish CLEP test (http://clep.collegeboard.org/exam/spanish-language). However, due to the drastic reduction in viable participants, even nonparametric statistical analysis on these results are unreliable, and those data are not considered here.

Therefore the primary assessment of linguistic outcomes comes from an analysis of all Spanish language interviews (3 interviews for each of 12 participants = 36 interview transcripts). Given the difficulty in operationalizing a complex and multi-faceted concept such as proficiency, (Lantolf & Frawley 1988; Salaberry & Cohen 2006), and its frequent synonymy with fluency, the analysis was based on research on L2 fluency (e.g., Cucchiarini, Strik & Boves 2000; Derwing, Rossiter, Munro & Thomson 2004; García-Amaya 2009; Schmidt 1992; Towell, Hawkins & Bazergui 1996). Following this work, transcripts were assessed for the following measures: total number of words; number of Spanish words; number of English words; number of dysfluencies

(e.g., repetitions, self-corrections, false starts); number of unique Spanish words (i.e., not including repeated words); and number of fillers. Any non-lexical item was considered a filler, since at this level most filler words (e.g., "um" and "uh") do occur in English. Any lexical words, even if used as fillers (e.g., "wait a sec" or *no sé*) were counted as words in the language in which they were spoken. Some of the more standard fluency measures were not appropriate for this dataset and are thus not included here. For example, temporal measures such as rate of speech are frequently used, but were impossible to calculate here given that these recordings come from interviews that involved frequent back-and-forth, overlapping, and long pauses. Additionally, the standard measure of a pause ($\geq$ 0.2 seconds) would be an unrealistic measure for these participants, who tended to speak in isolated words or chunks with excessive pausing and dysfluencies between utterances. Likewise, longest turn or mean length of turn were similarly deemed inappropriate measures at this level. An initial assessment of accuracy, based on the number of error free clauses, was attempted but also discarded given frequent, pervasive errors that would have resulted in overall low accuracy rates and thus rendered the measure meaningless.

RESULTS AND DISCUSSION

Results are discussed here in the order introduced above, beginning with user attitudes about and reactions to the Rosetta Stone program and then moving on to the results of the fluency analysis of participants' oral production.

Attitudes

In order to assess participants' attitudes towards various aspects of the Rosetta Stone program and their experiences, a content analysis of the English interviews was undertaken with the dual goals of determining the primary themes as they emerged, and then attempting to quantify their frequency. Although hundreds of comments were extracted for analysis, the discussion here focuses on the most frequent topics that emerged. Table 2 summarizes these findings, divided into the main themes of usability (including interface, technology, flexibility) and learning (linguistic processes and outcomes). The table also provides information on the frequency of these comments, and examples are taken verbatim from participant interviews.

Table 2.
Participant attitudes: most frequent themes

Theme and Frequency	Topic	Example(s)
Usability [RS: 42/181 comments] [RS+C: 80/124 comments]	**Interface**	• I like Rosetta Stone a lot. It's pretty easy to use… Because it's a lot of visual stuff, and I feel like I'm a visual learner. • I like how it's like uh, more like a game, so I'm more willing to actually do it… • It was just kind of a lot harder [to use] than I expected.
	Technology problems	• I can't get the microphone to work. • I was doing my Studio session and … I had no audio, like, I could hear them, but they couldn't hear me the whole time.
	Flexibility	• Like you're able to kind of do it like on your own time, you know. I'm not like restricted. • It's nice not to go to class. • I have always preferred to learn language, like, on my own. • It's more flexible with my schedule.
Language learning [RS: 31/181 comments] [RS+C: 11/123 comments]	**Comments on effectiveness, successes** [RS: 9/31 positive] [RS+C: 8/11 positive]	• It just didn't show what words I needed to use before it. • I [don't like] the lack of human interaction. • Sometimes it'll show the person speaking, and sometimes it'll say like he or she, and sometimes it'll be I. And I couldn't tell the difference. • The program is really good with like teaching like vocabulary. • With like vocabulary, it's like really good, and you get by. • I feel like it's more like how you naturally learn the language instead of like. "These are … your vocabulary words this week".
	Problems, concerns, lack of learning [RS: 22/31 negative] [RS+C: 3/11 negative]	• It's just like the grammar, and how to like, put it together. • You can't ask questions. • You don't get any writing, and then all of a sudden there's one writing thing. • You have that whole grammar and conjugation issue on Rosetta, because they don't really explain it. • You really need to have communication with a real person. • I would enjoy getting more grammar lessons … just to get a foundation of knowledge, then building up on that. • I'm always just frustrated because I'm like, I don't really understand it. • I feel like it should be more structured. • Rosetta Stone doesn't give you too much, like, actual instruction so you don't learn.

These extracted comments and their frequency highlight some general tendencies. Between one quarter and two thirds of the comments made by participants focused on the usability of the program – the interface, the technology, etc. – as opposed to the actual learning experience. It is difficult to categorize these comments as overall positive or negative toward the program, given that they were often made as mere observations. Generally speaking, participants found the Rosetta Stone program easy to use, visually appealing, and at times fun. These participants appreciated the flexibility of the program and enjoyed being able to work on their own schedule. There were some ongoing technological issues that came up at during the interviews, more often than not related to audio, although overall there were few complaints regarding the technology itself.

With respect to the participants' perceived learning, comments were mixed. In general, the Rosetta Stone + Class group reacted more positively towards the potential of the Rosetta Stone program (73% of comments about learning were positive, 27% were negative), than the Rosetta Stone group, which generally felt that they had not succeeded in learning what or as much as they had hoped (29% positive comments, 71% negative comments). The nature of the comments themselves was more or less consistent across groups, with the only variation being the frequency of each type of comment, indicating a general difference between those who attended class with an instructor regularly versus those who didn't.

The most positively assessed feature of the Rosetta Stone program overall was its presentation and practice of vocabulary, as the vast majority of the positive comments referenced lexical learning. The visual appeal of the Rosetta Stone interface is undeniable, and the early focus on tangible objects that are easily depicted with images is conducive to such a presentation. Additionally, a few comments mentioned that the Rosetta Stone presentation of vocabulary seemed more natural than a textbook, although given the relative lack of any contextualization in the materials, this impression was most likely due to the large quantities of input presented audio-visually, assisting learners in making form-meaning connections.

On the other hand, participants in both experimental groups also noted that they were frequently lost and not sure what they were supposed to be doing or learning, and that they felt that there was not a clear path. Participants did have a syllabus, though, and the Rosetta Stone program clearly indicates a suggested path for progressing through the materials, so this sense of confusion may in fact

be due the individual and somewhat isolated nature of the experience. In the Rosetta Stone group particularly, participants acutely perceived the absence of an instructor or a mentor to whom they could turn for guidance. Participants in the Rosetta Stone + Class group did not have these same concerns, because they did have a teacher with whom they met three times each week; the instructor commented that the vast majority of questions he received in class related to not understanding what an image was supposed to portray, or a specific grammar point. Participants in both groups noted the lack of explicit instruction, mostly with respect to grammar. As noted earlier, research has shown that adult learners of a second language, being aware of the existence of grammatical rules, often benefit from explicit focus on grammatical forms (e.g., Long 1988, 1991). These comments also indicate that students may want this explanation, and that without it, ambiguity can lead to frustration.

Finally, it is also worth noting that the pre- and posttest attitude survey revealed only two items with significant differences at the two testing times, and both relate to the comments highlighted here. Item #11 read: "Interacting via chat or telephone is comparable to interacting face-to-face", and although the Rosetta Stone group agreed strongly at the beginning of the semester, by the end of the term their agreement significantly decreased, as indicated by a Mann-Whitney U test ($z = -2.446$, $p < 0.05$). The other groups did not experience significant changes on this item, so it seems that perhaps the Rosetta Stone group realized over the course of the term that they were not getting the same experience they would have received in a classroom setting.

The only other item to show significant differences pre- and post-semester was item #19: "I would prefer to learn a language on my own time and at my own pace than in a group or classroom setting." Both the Rosetta Stone and the Rosetta Stone + Class group significantly increased their agreement with this statement at the end of the semester (RS: $z = -2.74$, $p < 0.05$; RS+C: $z = -2.88$, $p < 0.05$). This finding corroborates the positive assessments from the participants regarding the flexibility of the Rosetta Stone program, the ability to work on their own schedules, and the freedom of not having to attend a class (in the case of the Rosetta Stone only group).

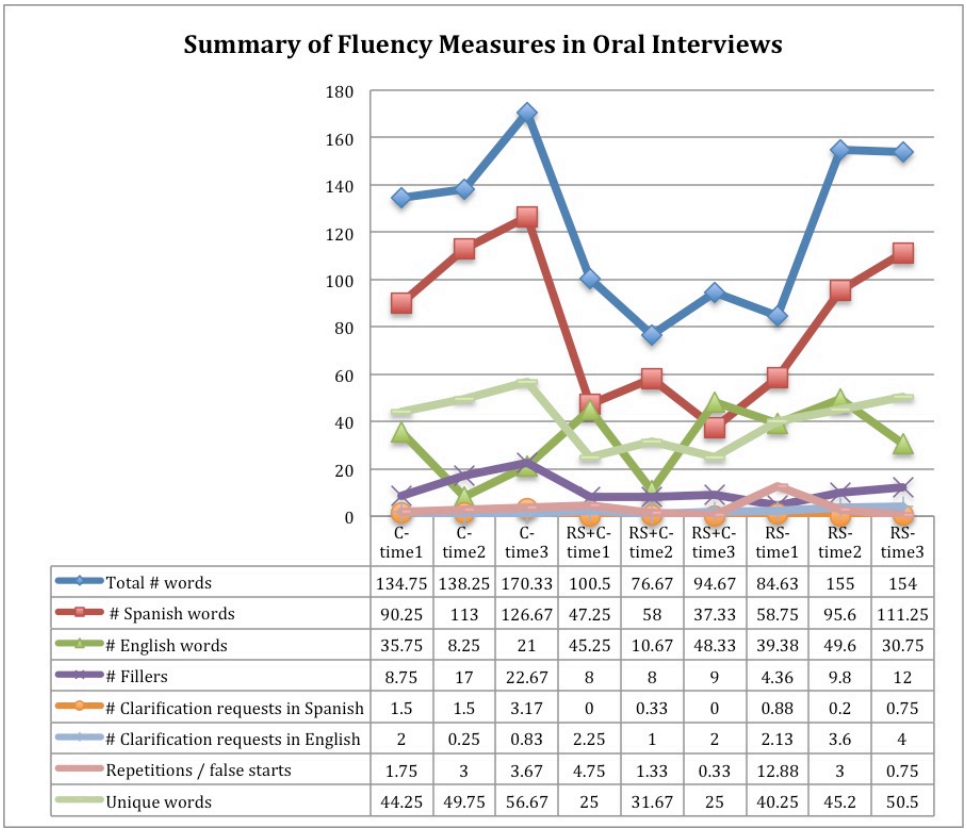
Fluency analysis

Turning to linguistic outcomes, Table 3 presents the summary data from each interview session, by group, for the measures used, while Figure 1 visually displays the trends across groups and interviews.

Table 3.
Fluency measures from oral transcripts, by group

Participant	Total # words	# Spanish words	# English words	# Fillers	# Clarification requests in Spanish	# Clarification requests in English	Repetitions/ false starts	# Unique words
C-time1	134.75	90.25	35.75	8.75	1.5	2	1.75	44.25
C-time2	138.25	113	8.25	17	1.5	0.25	3	49.75
C-time3	170.33	126.67	21	22.67	3.17	0.83	3.67	56.67
RS+C-time1	100.5	47.25	45.25	8	0	2.25	4.75	25
RS+C-time2	76.67	58	10.67	8	0.33	1	1.33	31.67
RS+C-time3	94.67	37.33	48.33	9	0	2	0.33	25
RS-time1	84.63	58.75	39.38	4.36	0.88	2.13	12.88	40.25
RS-time2	155	95.6	49.6	9.8	0.2	3.6	3	45.2
RS-time3	154	111.25	30.75	12	0.75	4	0.75	50.5

Due to the individual nature of the data and the qualitative intention of this analysis, statistical tests have not been run on these numbers.



	C-time1	C-time2	C-time3	RS+C-time1	RS+C-time2	RS+C-time3	RS-time1	RS-time2	RS-time3
Total # words	134.75	138.25	170.33	100.5	76.67	94.67	84.63	155	154
# Spanish words	90.25	113	126.67	47.25	58	37.33	58.75	95.6	111.25
# English words	35.75	8.25	21	45.25	10.67	48.33	39.38	49.6	30.75
# Fillers	8.75	17	22.67	8	8	9	4.36	9.8	12
# Clarification requests in Spanish	1.5	1.5	3.17	0	0.33	0	0.88	0.2	0.75
# Clarification requests in English	2	0.25	0.83	2.25	1	2	2.13	3.6	4
Repetitions / false starts	1.75	3	3.67	4.75	1.33	0.33	12.88	3	0.75
Unique words	44.25	49.75	56.67	25	31.67	25	40.25	45.2	50.5

Figure 1. Summary of fluency measures in oral interviews, by group and time.

There appears to be little evidence of change over time, which may be expected given the relatively short treatment period (16 weeks). An analysis of these data from an overall group perspective, though, reveals interesting trends. For example, the use of English differs notably between groups. The ratio of English-to-Spanish was calculated by dividing the number of English words by the number of Spanish words; a ratio of 0 would indicate that the entire production was exclusively in Spanish, while a ratio of 1 would mean one English word was produced for every Spanish word. Figure 2 displays the overall group averages for this ratio calculation.

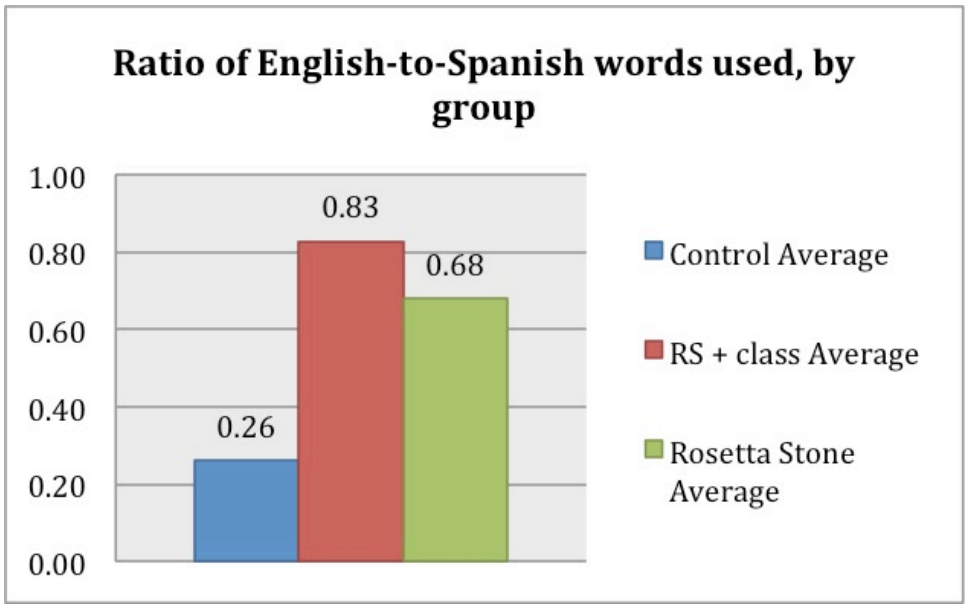


Figure 2. Ratio of English-to-Spanish words, by group.

The Rosetta Stone + Class group produced approximately eight English words to for every 10 Spanish words they produced, while the Rosetta Stone group produced almost seven English words for every Spanish word; in other words, they used almost as much English as Spanish in their interviews. The Control Group produced only two to three English words for every Spanish word, implying a greater ability to remain in the target language while trying to get their point across. Similarly, the groups seem to display different behaviors when confronted with a communication breakdown. All participants struggled to express their meaning and stumbled frequently due to unknown vocabulary items, and there were several instances of requests for help or clarification, but the group behaviors were different.

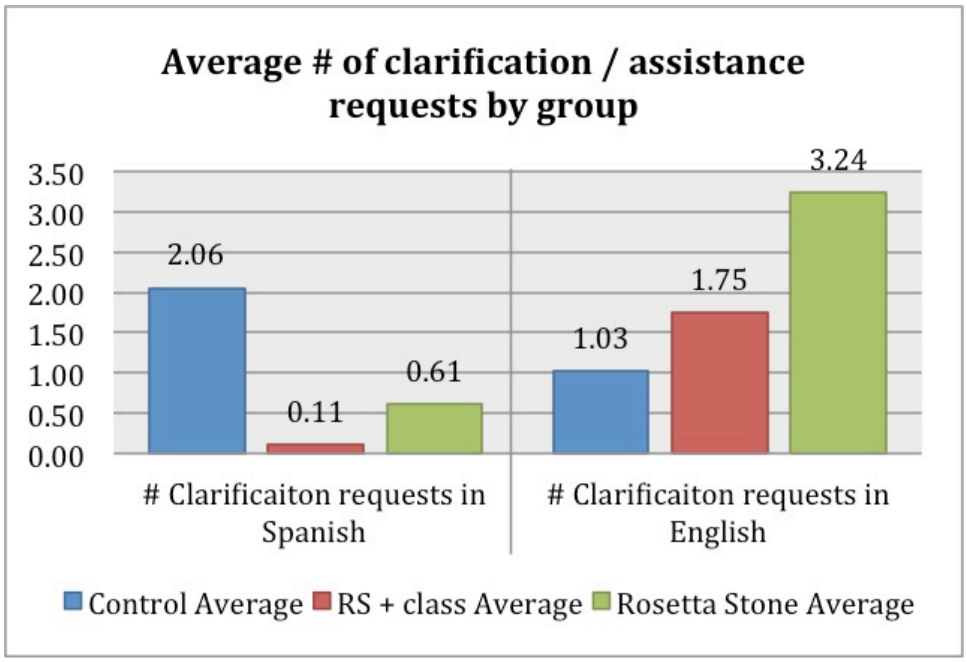


Figure 3. Average number of clarification / assistance requests in English and Spanish, by group.

Figure 3 shows group averages for assistance requests. The Control group produced these requests in Spanish more often than English, while the Rosetta Stone and Rosetta Stone + Class groups produced English requests more than twice as often as Spanish requests. In fact, closer examination of the language produced by learners in each of these groups confirms that the learners approach the second language in different ways. Although all learners at these levels will inevitably struggle to convey meaning, how they handle potential difficulties or failures to communicate can be telling. Transcripts from these groups (see Lord, 2015 for further details and transcript examples) revealed that classroom learners not only use more Spanish but also ask for repetition, clarification, etc., when they are attempting to negotiation meaning. On the other hand, the Rosetta Stone learners often give up and automatically revert to English in order convey their meaning. The Rosetta Stone learners also demonstrate greater levels of frustration at not being able to say what they want to say than either of the classroom learner groups.

Taken together, these measures – and particularly those relating to the use of English – confirm the impressions of those carrying out the interviews: although all learners were clearly novices who struggled to communicate, the Rosetta Stone group seemed to struggle more and frequently resorted to English, while

the classroom groups, but particularly the Control group, were better equipped to request assistance when needed or attempt to convey their message even in spite of linguistic lacunae. The Rosetta Stone + Class group seems to represent an odd mixture, as in some ways they outperformed the Rosetta Stone group, and in others appear to be less proficient than both groups. Anecdotally, the interviewers found that the Rosetta Stone group was frequently unable to respond to anything other than simple naming tasks (*¿Qué es esto?*), while the Control group was able to engage, albeit haltingly, in basic conversations; again, the Rosetta Stone + Class group represented an interesting middle ground, with the same limited conversational tools as the Rosetta Stone group but slightly more disposition to form discourse length utterances.

In sum, the analysis of both English and Spanish interview transcripts reveals different behaviors and trends among the groups. In terms of student satisfaction, the learners enjoyed the Rosetta Stone program and appreciated the flexibility it offered, although those who did not attend any class meetings felt lost at times and desired more explicit instruction. The Control group's satisfaction was mixed, as in any class, and related less to instructional materials and processes than other, external factors (e.g., instructor personality, etc.) In terms of language skills, both in-class groups – regardless of instructional materials – demonstrated a somewhat greater communicative competence, while both Rosetta Stone groups seemed to lack basic vocabulary and conversational strategies. Class time in these environments was spent largely on interactive small-group work, so oral proficiency developed in these classes; but virtually no conversation takes place through the Rosetta Stone materials, which explains why those participants lacked effective communication strategies.

We also must recognize though that time on task differed greatly between the groups, by virtue of the fact that the Rosetta Stone group did not attend classes (a potential of 45 contact hours). Table 4 provides usage data for all participants, and for the Rosetta Stone groups, the number of hours they spent in Rosetta World™ and Rosetta Studio™. After considering absences from class sessions, the number of hours spent in class was recorded for the two in-class groups.

Table 4.
Usage data for all participants, by group

Group	% Complete	Average Score (/100)	Total course usage (in hours)	Total Class Time	Total World hours (9 required)	Total Studio sessions (6 required)
Control	96.72	94.97	83.25	40	--	--
Control	99.64	83.56	68.00	39	--	--
Control	99.27	95.55	42.50	38	--	--
Control	92.34	89.01	86.25	39	--	--
RS+C	76.67	99.00	30.00	40	12.25	6
RS+C	98.00	98.00	36.50	33	1.00	2
RS+C	100.00	99.50	26.25	35	20.00	6
RS+C	100.00	98.00	38.50	41	11.75	5
RS	93.33	93.50	23.50	--	14.00	6
RS	98.67	98.50	28.25	--	9.00	5
RS	100.00	96.50	44.50	--	13.75	7
RS	98.67	95.00	26.50	--	9.50	6

As can be seen in Figure 4, the Control group averaged 109 hours of exposure over the course of the semester, including classroom hours and homework hours online, while the Rosetta Stone group averaged only 48 hours over the semester.

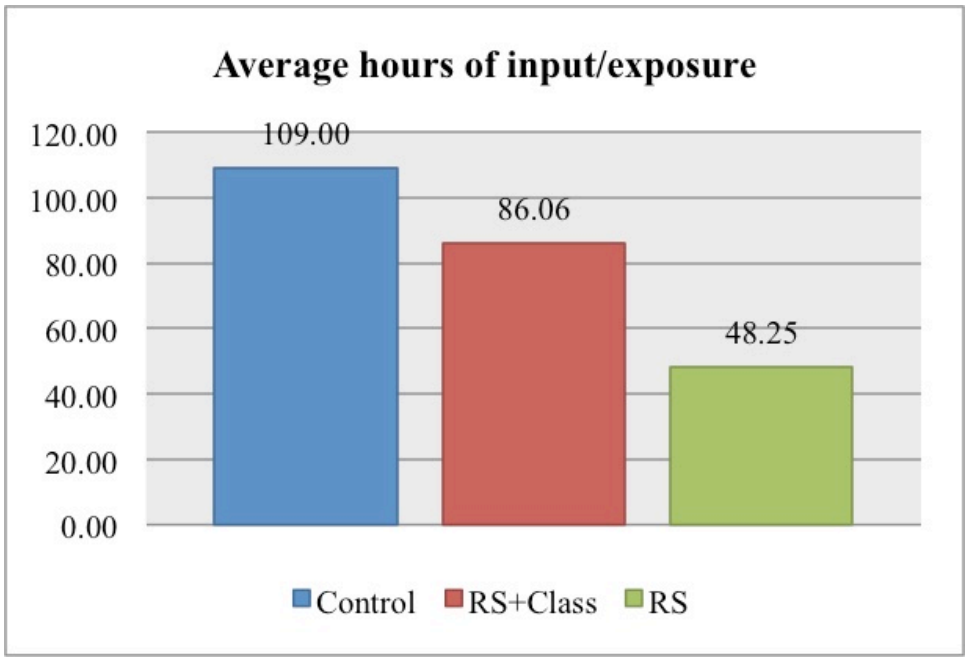


Figure 4. Average time online and in class, by group.

Clearly, the Rosetta Stone units require a great deal less time to complete than the activities assigned in the Control class. The Rosetta Stone + Class group is, logically, between the other groups in terms of usage, since they had the reduced materials but also the class time. Seat time, and subsequently input, are of course essential to the language learning process, so it is more than likely that time on task contributed to differing outcomes. The Control group had twice as much time to learn, practice and use the language than the Rosetta Stone groups, and that additional time may be behind the qualitative differences observed in the oral performance of the groups here. Future work in this area will need to control for exposure.

CONCLUSION

While this study provides much-needed exploratory data regarding the potential of the Rosetta Stone program for learning Spanish in an academic setting, a great deal of additional work is still needed in order to understand the true potential for this program as compared to classroom learning, Likewise, there are some limitations of the present design that future researchers should endeavor to correct. To begin with, a larger sample size is clearly needed, as the findings from this small-scale analysis cannot be generalized across populations.

These students do tend to represent a typical true-beginner population at college, although we should also consider the instruction effects on high beginners as well as different levels of instruction. Further analyses of the current data set are underway, including all learners and analyzing background as a co-variable, and stand to shed more light on the potential of the Rosetta Stone program in various settings. Additional research is also necessary in order to examine outcomes over a longer treatment period that could encompass a potentially greater range of acquisition, ideally several semesters. It would also be beneficial to follow up with participants as they move on to the next levels of language study to determine if differences in their basic language instructional methods lead to variable outcomes in subsequent semesters. Another important consideration not addressed in this analysis is that of cultural awareness and appreciation, a crucial element of most face-to-face language classes. Virtually all major language textbooks on the market today include ample cultural information, though the Rosetta Stone materials do not focus in any way on culture. Subsequent studies should attempt to examine this aspect of language learning, as well as its relation to linguistic development, in an effort to better understand the potential outcomes in each of the learning environments.

In spite of these methodological limitations, the data presented here provide little evidence to confirm Rosetta Stone's marketing claims of being superior to in-class learning, in spite of learners' appreciation for the flexibility and usability of the program. In fact, the analysis of the learner outcomes indicates that the Rosetta Stone program, while capable of presenting isolated, decontextualized language elements, does not seem as adept at helping learners develop crucial communicative strategies in the foreign language. These findings mirror the reservations expressed in previous reviews of the program (e.g., DeWaard 2013; Lafford, Lafford & Sykes 2010).

This project was undertaken with the awareness that most language instructors have an inherent distrust of, if not disdain for, stand-alone programs such as Rosetta Stone. Whether this reaction is based on our recognition of the fact that language teaching is too complex a process to be successfully executed by a computer program, or perhaps the fear that our administrators are seeking to replace language programs with software, is not a question that can be answered here. What does become evident from the present findings is that the program itself is not as inadequate as many language educators would like to believe, and is even capable of teaching isolated elements such as lexical items at the beginning levels; nonetheless, the present data reveal some potentially serious limitations when it comes to fostering communicative competence and oral

proficiency. We must continue to investigate the nuances of learner experiences and outcomes in order to better understand what Rosetta Stone can offer to language study and, importantly, to further investigate what it cannot offer.

ACKNOWLEDGMENTS

This research was supported by a grant from the University of Florida College of Liberal Arts and Sciences and a University of Florida Humanities Scholarship Enhancement Fund award. Special thanks go to Caroline Reist, Brandon Shufelt, Keegan Storrs and Diana Wade for their assistance with data collection and coding, and to Carlos Enrique Ibarra for performing the statistical analyses. The study was not funded or endorsed by Rosetta Stone, but I am grateful to their higher education team for their information and assistance throughout the process. Finally, the comments provided by anonymous reviewers on this and earlier versions of this manuscript were immensely valuable. All errors of course are my own.

NOTES

1. Since data collection, Rosetta Stone has acquired Tell Me More, in addition to Live Mocha, Lexia Learning, and Fit Brains.

2. One of these 13 participants did not complete all assessment tasks, so those data were eliminated, resulting in an even four participants per group condition.

3. To ensure that they received academic credit for the course, the students remained officially enrolled in their original section of the Beginning Spanish 1 class. At the end of the term, the researcher reported their earned grades to the instructors of the classes. This method was pre-approved by the University of Florida's Institutional Review Board, Academic Advising Center, and College of Liberal Arts and Sciences.

ABOUT THE AUTHOR

Gillian Lord is Professor of Spanish and Linguistics at the University of Florida, where she Chairs the Department of Spanish and Portuguese Studies. She has worked for years with the lower division language programs, and has a particular interest in language learning and teacher education. Her primary research focuses on second language acquisition, and particularly on the effects of computer assisted language instruction.

REFERENCES

Bax, S. (2003). CALL – past, present and future. *System 31*(1), 13-28.

Bley-Vroman, R (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (41–68). New York: Cambridge University Press.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis 20*, 3–49.

Bley-Vroman, R. (2009). The evolving context of the Fundamental Difference Hypothesis. *Studies in Second Language Acquisition 31*, 175–198.

Cucchiarini, C., Strik, H. & Boves, L. (2000). Quantitative assessments of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America 107*(2), 989–999.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, J. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning 54*(4), 655–679.

DeWaard, L. (2013). Is Rosetta Stone a viable option for L2 learning? *ADFL Bulletin 42*(2), 61–72.

Ervin-Tripp, S. M. (1974). Is second language learning like the first? *TESOL Quarterly 8*(2), 111–127.

García-Amaya, L. (2009). New findings on fluency measure across three different learning contexts. In J.

Collentine et al. (Eds.), *Selected proceedings of the 11th Hispanic Linguistics Symposium* (68–80). Somerville, MA: Cascadilla Proceedings Project.

Grossman, S. (2013, June 19). Rosetta Stone is no replacement for in-class learning, study finds. *The Chronicle of Higher Education: Wired Campus*. Retrieved from http://chronicle.com/blogs/wiredcampus/rosetta-stone-is-no-replacement-for-in-class-learning-study-finds/44269

Krashen, S. D. (1985). *The Input Hypothesis: Issues and implications.* New York: Longman. Print.

Lafford, B. A., Lafford P., & Sykes, J. (2007). Entre dicho y hecho …: An assessment of the application of research from second language acquisition and related fields to the creation of Spanish CALL materials for lexical acquisition. *CALICO Journal 24*(3), 427–529.

Lantolf, J. P. and Frawley, W. (1985). Oral-proficiency testing: A critical analysis. *The Modern Language Journal 69*(4), 337–345.

Long, M. H. (1988). Instructed Interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (115–141). Rowley, MA: Newbury House.

Long, M. H. (1991). Focus on Form: A design feature in language teaching methodology. In K. DeBot, R. Ginsberg & C. Kramsch (Eds.), *Foreign language research in a crosscultural perspective (*39–52). Amsterdam: John Benjamins.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (413–468). San Diego: Academic Press.

Long, M. H. & Robinson, P. (1998). Focus on Form: Theory, research and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (15–41). Cambridge: Cambridge University Press.

Lord, G. (2015). 'I don't know how to use words in Spanish': Rosetta Stone and learner proficiency outcomes. *The Modern Language Journal 99*(2), 401-405.

Markee, N. (1997). *Managing curricular innovation*. New York: Cambridge University Press. Print.

Newman, A. A. (2012, June 19). An emphasis on fun for language learners. *New York Times*, Media & Advertising Section. Retrieved from http://www.nytimes.com/2012/06/20/business/media/rosetta-stone-ads-emphasize-fun-not-efficiency.html

Nielson, K. B. (2011). Self-study with language learning software in the workplace: What happens? *Language Learning and Technology 15*(3), 110–129.

Richards, J. C. & Rodgers, T. S. (2001). *Approaches and methods in language teaching.* New York: Cambridge University Press.

Rifkin, B. (2003). Criteria for the assessment of foreign language instructional software and web sites. *ADFL Bulletin 34*(2), 53–56.

Rundquist, J. (2010, June 27). Computer Programs Replace Foreign Language Teachers in N. J. Classrooms After Budget Cuts. *New Jersey Real Time News*. Retrieved from http://www.nj.com/news/index.ssf/2010/06/computer_programs_replace_fore.html

Salaberry, R. and Cohen, A. (2006). Testing Spanish. In R. Salaberry & B. A. Lafford (Eds.), *The art of teaching Spanish: Second language acquisition from research to praxis* (149–172). Washington, D.C.: Georgetown University Press.

Santos, V. D. O. (2011). Review of *Rosetta Stone* Portuguese (Brazil), Levels 1, 2, & 3. *CALICO Journal 29*(1), 177–194.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition 14*(4), 357–385.

Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics 16*, 371–391.

Terrell, T. D. (1977). A natural approach to second language acquisition and learning. *The Modern Language Journal 61*(7), 325–337.

Towell, R., Hawkins, R. & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics 17*(1), 84–119.

Vesselinov, R. (2009). Measuring the effectiveness of Rosetta Stone. *Rosetta Stone Resources.* Retrieved from http://resources.rosettastone.com/CDN/us/pdfs/Measuring_the_Effectiveness_RS-5.pdf

Warschauer, M. (1996). Computer assisted language learning: An introduction. In S. Fotos (Ed.), *Multimedia Langauge Teaching* (3–20). Tokyo: Logos International.

APPENDICES

Appendix A: Rosetta Stone Screen Shots

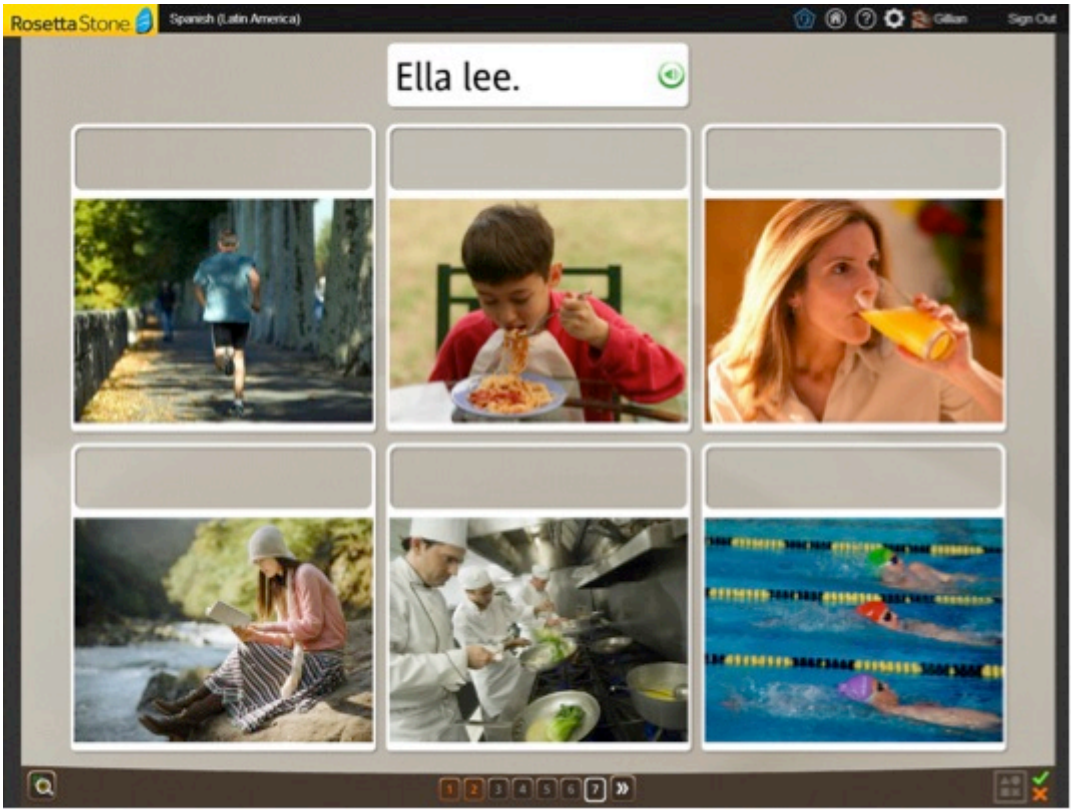


Figure 5. Sample vocabulary lesson.

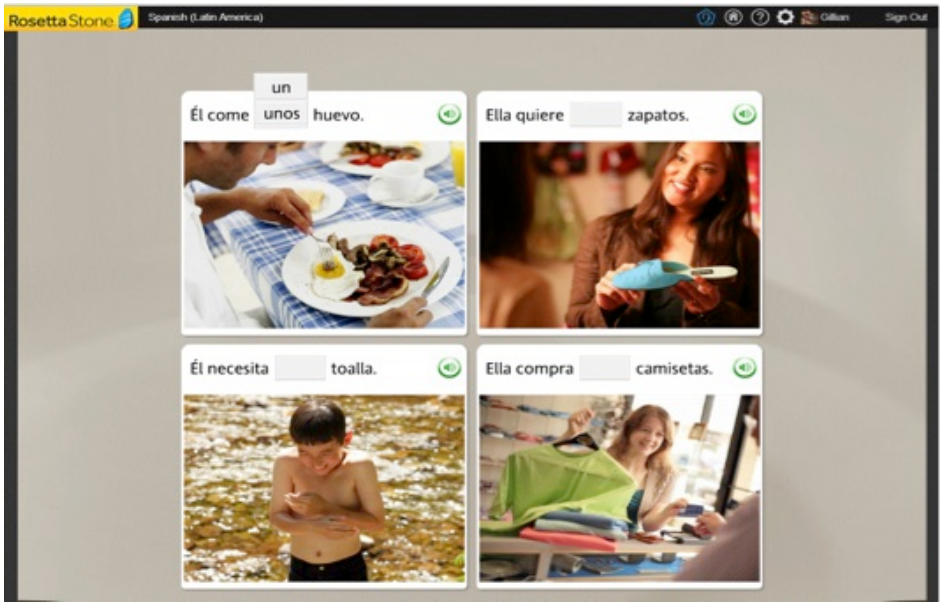


Figure 6. Sample grammar lesson.

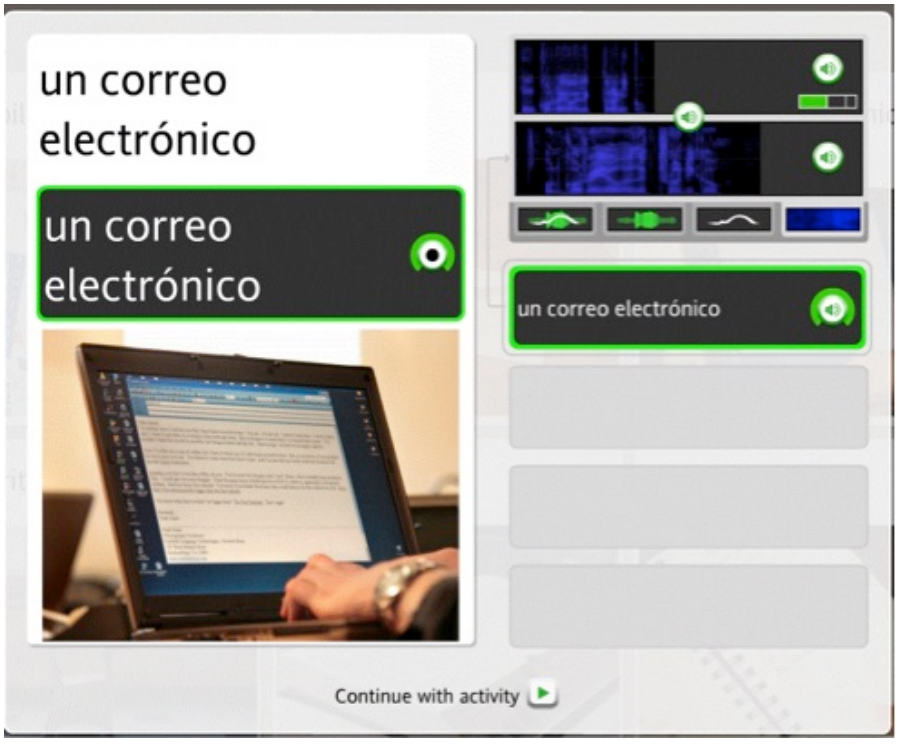


Figure 7. Sample pronunciation lesson.

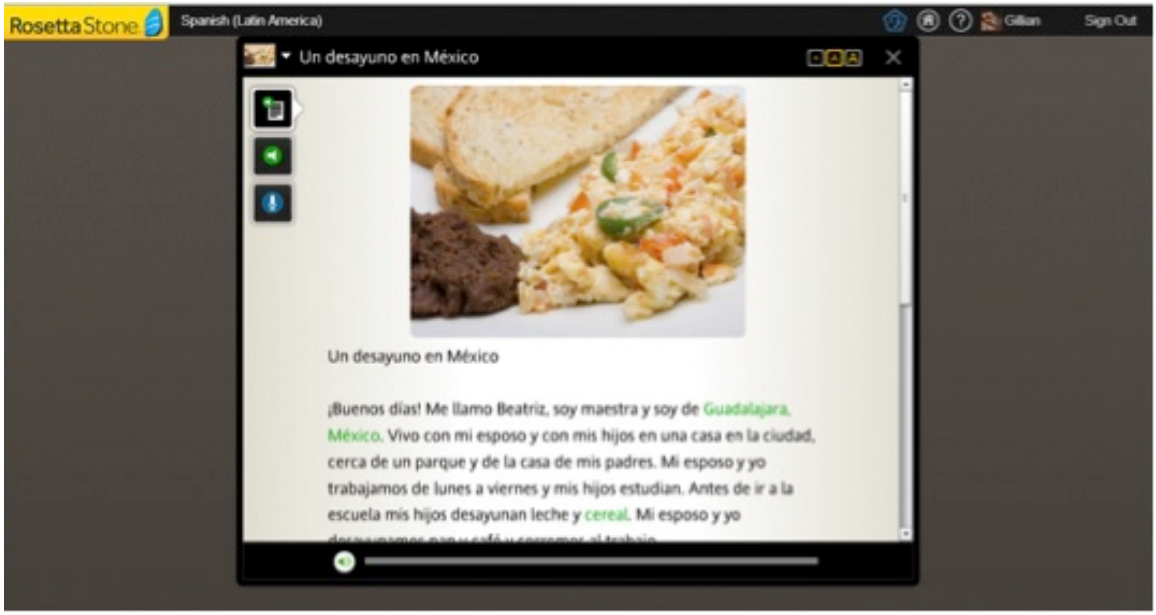


Figure 8. Sample reading from Rosetta World ™.

Appendix B: Language Background Questionnaire

General information

1. Gender: Male Female
2. Age:
3. Country of birth:
4. Standing at UF: Freshman Sophomore Junior Senior Graduate
5. Do you have any known vision or hearing problems? If so, are they corrected (i.e., glasses)?

Language background

1. What is your native language?

2. Do you consider yourself proficient in a language other than your native language? If so, which language? How did you learn this other language?

3. Have you taken college-level language coursework (other than the SPN 1130 course in which you are currently enrolled)? Which?

4. Which skills do you find easiest when learning a foreign language, either based on experience or on your intuitions? Please rank these, 1 = easiest, 5 = most difficult:

   Listening: _____
   Speaking: _____
   Reading: _____
   Writing: _____
   Pronunciation: _____
   Vocabulary: _____

5. Why are you taking Beginning Spanish 1?

6. *For the Rosetta Stone group*: Why did you decide to participate in this study?

Appendix C: Student attitude survey (pre and post)

Group: _____ Interview #1 #3

Participant name: _____

Please read each of following items in terms and think about how well each statement applies to you and your thoughts on learning a foreign language, specifically Spanish. Rate the statements according to a 7-point scale, where 1 = "Never true of me, strongly disagree" and 7 = "Always true of me, strongly agree," using the numbers in between for more nuanced reactions.

1. ____ Acquiring a large and varied vocabulary in Spanish is important to me.

2. ____ Communicating effectively is more important than sounding like a native speaker.

3. ____ I am enjoying my Spanish-learning experience this semester.

4. ____ I am taking this class to fill a language requirement only.

5. ____The interactive technological tools I am using this semester are helping me learn Spanish.

6. ____ I believe emphasis should be given to communication and one-on-one interaction.

7. ____ I plan to continue studying Spanish after this semester.

8. ____ Acquiring proper pronunciation in Spanish is important to me.

9. ____ I put a great deal of effort in to learning Spanish on a regular (daily, weekly) basis.

10. ____ I'd like to sound as native as possible when speaking Spanish.

11. ____ Interacting via chat or telephone is comparable to interacting face-to-face.

12. ____ It is more effective to converse with native speakers of Spanish than with fellow language learners.

13. ____ It is possible to have an effective conversation in Spanish without being able to conjugate verbs.

14. ____ It isn't possible to be successful in language learning without memorizing verb conjugations.

15. ____ I believe more emphasis should be given to proper pronunciation in class.

16. ____ Learning Spanish will be important to my future career plans.

17. ____ My goal is to become fluent in Spanish.

18. ____ Vocabulary and grammar are the most important aspects of learning a language.

19. ____I would prefer to learn a language on my own time and at my own pace than in a classroom setting.

20. ____Learning a language via computer can be as effective as learning in a classroom setting.