49 Geary Street Suite 530 San Francisco, CA 94108-5731 (415) 544-0788 www.rockman.com

July 2009

Rosetta Stone Evaluation Report

Executive Summary

Rosetta Stone[®] is a language learning system that specializes in foreign language acquisition. The Rosetta Stone Spanish language solution includes Rosetta Stone online language software, Rosetta Studio[™] sessions^{*}, and audio CDs for teaching languages. This dynamic immersion system provides the learner with an interactive virtual environment designed to learn naturally, engage interactively, and speak confidently while enjoying the experience.

In May 2009, **Rockman et al** conducted a quantitative study of effectiveness of the Rosetta Stone Spanish language solution on a stratified random sample of 89 adults who took six hours of Rosetta Studio[™] sessions and were exposed on average to 64 ½ hours of work with the software.

Effectiveness of the Rosetta Stone Spanish solution was demonstrated by the results of two different Spanish tests: the American Council on the Teaching of Foreign Languages Oral Proficiency Interview (ACTFL OPI-C.) The researchers found positive results that are briefly described below.

- All of the participants in the study exhibited significant gains as indicated by the ACTFL-OPI-C.
- The differences found in the study most likely reflect differences that would occur in the larger population.
- The majority of the participants successfully advanced 1-2 levels of proficiency in Spanish.
- The Rosetta Stone system received highly positive ratings from participants. Their perceptions of the Spanish solution were overwhelmingly positive.

In sum, based on the overall results of this research study, researchers from Rockman et al conclude that the Rosetta Stone solution is both an effective and efficient means of learning Spanish as a foreign language.

^{*} Live online sessions that allow you to practice your new

language with a native-speaking Coach.

Introduction

Rockman et al (REA) conducted an effectiveness study of the **Rosetta Stone Spanish language solution**. REA is an independent evaluation, research, and consulting firm focusing on education, technology, and media. REA works with preschool, K-12, postsecondary and adult educational institutions undertaking formal education, as well as with broadly educational projects having a wide community or consumer audience. As independent researchers, we often serve as external evaluators for federal grant-funded projects, foundations, and private industry.

REA conducted a quantitative study of effectiveness of the Rosetta Stone[®] Spanish language solution as measured by the *American Council on the Teaching of Foreign Languages Oral Proficiency Interview (ACTFL OPI-C.)* The Rosetta Stone staff recruited a group of adults who expressed interest in taking part in this study. After REA selected a stratified random sample from this group, 89 adults participated in the study. These participants took six hours of Rosetta Studio[™] sessions and were exposed on average to 64 ½ hours of work with the software.

In this report, we describe the demographics of the evaluation participants, examine participant usage, and present our main descriptive statistics relating to effectiveness. We also report on participants' satisfaction with the Rosetta Stone solution, including the software, Rosetta Studio[™] sessions, and audio CDs.

The Rosetta Stone Solution

Rosetta Stone is a language learning system that specializes in foreign language acquisition. The Rosetta Stone Spanish language solution includes Rosetta Stone online language software, Rosetta Studio[™] sessions, and audio CDs for teaching languages. This dynamic immersion system provides the learner with an interactive virtual environment designed to learn naturally, engage interactively, and speak confidently while enjoying the experience. One of the difficulties of learning a new language is the inability to evaluate one's own pronunciation as beginners, something that cannot be adequately satisfied by simplistic systems that do not have practical interactions with the user (Tien-Lok, 2003). Rosetta Stone has an embedded speech-recognition technology that helps learners pronounce syllables, words, and sentences, and provides immediate, ongoing feedback. It presents a native speaker's speech to compare the learner's voice with that of the native speaker. It provides a wide range of tools to work on pronunciation and speaking skills. The Rosetta Stone system also has an Adaptive Recall[®] feature that brings back review exercises to help the learner transfer learning into long-term memory.

The Rosetta Stone Spanish solution (Latin America) Version 3 has three levels. Each level has four units. The units in Level 1 address language basics, greetings and introductions, work and school, and shopping. Level 2 includes units on travel, past and future, friends and social life, dining, and vacation. Level 3 includes home and health, life and world, everyday things, and social life.

Each unit takes about six to eight hours to complete. There are reading, vocabulary, listening, grammar, speaking (with voice recognition), pronunciation, writing, and reading exercises. Users have the choice of taking the lessons sequentially or randomly. The Rosetta Stone solution uses an immersion method with no translations. Pictures are used to guide the associations that the learner makes for different foreign words or phrases.

Research design

Rationale

Technology has been used to teach and learn languages for many years. Powerful and faster tools have been developed to support language learning for both children and adults. There is a growing demand for technology-enhanced or technologydelivered resources for foreign language instruction. Schools, colleges, and universities are considering whether to adopt software or synchronous/asynchronous computer programs to incorporate them into their language curriculum. To inform their decisions, they are looking for research-based evidence of the effectiveness of technology-based tools (Rifkin, 2003). Although research on effectiveness of technology on language learning has been conducted over three decades, scholars are still struggling to determine how effective and efficient technology tools are (Saury, 1998; Graff, 2006). However, recent studies on newer and more powerful technologies have concluded that computer assisted foreign language software could be the "...solution to overcome the language competency problem of school students" (Al Mekhlafi, 2006).

Recently, researchers and scholars have considered the best way to measure effectiveness of technology. Some scholars have pointed out that effective research methods were needed to answer the effectiveness question about technology-based language learning tools (Saury, 1998). More recently, there seems to be an agreement among researchers that the optimal way to evaluate a spoken language learning system is by having a wide spectrum of people use it and measure their progress and satisfaction through some external measures (Tien-Lok, 2003). Classroom research suggests the need for descriptive research documenting the nature of the interaction that learners engage in within various computer-assisted language learning contexts (Chapelle, 1997).

Research Questions

The key research questions for the evaluation included the following:

- How effective is the Rosetta Stone solution at teaching Spanish as a second language?
- How much do the participants learn using the Rosetta Stone solution?

- How many levels of Rosetta Stone do they finish on average? Is there a correlation between the number of levels they finish and the improvement in their language proficiency?
- Are participants satisfied with the software? Would they recommend the software to others? Do they enjoy using it?

Instruments

REA assessed effectiveness of the Rosetta Stone Spanish solution through a quantitative analysis of pre and post-test data, usage data, and post-surveys gathered by the Rosetta Stone staff.

The American Council on the Teaching of Foreign Languages Oral Proficiency Interview-Computerized (ACTFL OPI-C) is a criterion-referenced, standardized test that assesses functional speaking abilities in 37 different languages. It measures language performance and determines proficiency according to seven different levels (novice-low, novice-mid, novice-high, intermediate-low, intermediate-mid, intermediate-high, and advanced). The content of the ACTFL OPI-C is non-curriculum based. The test is conducted as an interactive and structured conversation between a certified interviewer and a test-taker. Participants in this study took the test on the computer using a headset. Their answers were recorded. Two raters listened to their answers and rated them. Expert interviewers certified the test. At the lowest level of the test-Novice-Low, learners have no functional ability. They may exchange greetings, give their identity, name familiar objects from their immediate environment but are not capable of engaging in conversations.

Sampling and Evaluation Participants

Before the evaluation began, we agreed on a sample size of approximately 100 participants based on a *priori* power analyses and communications with Rosetta Stone. REA was then given a list of over 1200 potential participants from Rosetta Stone. From this pool, we created a stratified random sample of approximately 125 persons. The sample was created to ensure a wide distribution across age, race, gender, household income level, and educational level. Anyone with prior experience with Spanish was not included, and the pool was weighted towards individuals with no foreign language experience. Rosetta Stone recruited from this list, and eventually 105 agreed to participate

in the study. Sixteen individuals did not continue with the evaluation after the initial assessment, so findings are largely based on the remaining 89 participants. This group was evenly split by gender, with 44 males and 45 females. A summary of the data related to demographic characteristics is presented in the tables and figures below.

Table 1. Evaluation participants by age

	Frequency	Percent	
Up to 24	8	9.0	
25 to 34	18	20.2	
35 to 44	18	20.2	
45 to 54	21	23.6	
55 and up	24	27.0	

Table 2. Evaluation participants by race

	Frequency	Percent
Caucasian	49	55.1
African American	25	28.1
Asian	6	6.7
Prefer not to answer	4	4.5
Other	5	5.6

Table 3. Evaluation participants by level of education

	Frequency	Percent	
Some college*	15	16.9	
College graduate	43	48.3	
Graduate degree**	31	34.8	

* This signifies a four-year college, with a BA or equivalent earned.

** This signifies a MA, PhD, or equivalent earned.

	,		
	Frequency	Percent	
Less than \$25,000	4	4.5	
\$25,000-\$50,000	14	15.7	
\$50,001-\$75,000	21	23.6	
\$75,001-\$100,000	7	7.9	
\$100,001-\$125,000	9	10.1	
\$125,001-\$150,000	7	7.9	
Above \$150,000	9	10.1	
Prefer not to answer	18	20.2	

Table 4. Evaluation participants by 2008 household income

Figure 1. Evaluation participants by age







Figure 3. Evaluation participants by level of education



Figure 4. Evaluation participants by 2008 household income



While our sampling was able to create a diverse participant group, the subgroups are not evenly distributed. Nor do subgroup sizes mirror the United States adult population. Within the category of educational level attained, it would have been preferable to have participants who did not graduate from high school or graduated high school but did not pursue post-secondary education. Despite this weakness, we are reasonably confident that our participant sample included a wide range of demographic categories.

Results

Rosetta Stone usage

In this section, we examine the numbers of units participants completed and the time generally spent learning Spanish, including study with the Rosetta Stone software and audio CD. Our expectation was that participants would complete two levels and that this amount of instruction would correspond with language proficiency gain on the ACTFL OPI-C. By the end of the evaluation period, participants generally completed at least two levels, and take six Rosetta Studio[™] sessions. This did not vary significantly within any demographic category. College graduates tended to complete slightly more levels than people with graduate degrees or those with some college. Males completed slightly more levels than females. Participants older than 54 years of age completed significantly fewer levels with 59% of the available twelve units.

Another component of the program was study with the Rosetta Stone software. On average, participants spent about 64 ½ hours working with the software. Like level completion, there was not a great deal of variation within demographic categories. Interestingly, the slight differences did not mirror the differences in percent of level completion. Females tended to report slightly higher hours than males, but the difference was only six minutes. Although participants older than 54 years of age completed fewer levels, they tended to have the greatest number of independent study hours. This could reflect that they had slightly more difficulty with the content, but based on post-test scores and survey ratings (described later), a more likely explanation is that they simply went through the content at a slightly slower, and perhaps more deliberate, pace.

Participants also used audio Rosetta Stone compact discs, and submitted weekly reports on their usage in minutes. Interestingly, there was a great deal of variation in the amount of time participants opted to spend with the audio CDs. Because of this great variation, we explored the demographic categories associated with CD use closely. Despite having notably different mean numbers of hours usage, there were few significant differences within demographic categories. This is due to large standard deviations, meaning that there was great variation within each category. Therefore, while the variation was great, it does not appear that this is related to any demographic characteristics.

Table 5. Rosetta Stone audio CD usage, by gender

	Mean	Std. Deviation
Male	12.37	13.36
Female	16.12	10.63
Total	14.27	12.14

Table 6. Rosetta Stone audio CD usage, by age range

	Mean	Std. Deviation
Up to 24	6.66	8.23
25 to 34	6.84	6.63
35 to 44	15.61	14.84
45 to 54	17.81	11.37
55 and up	18.27	11.93
Total	14.27	12.14

Table 7. Rosetta Stone audio CD usage, by highest grade completed

	Mean	Std. Deviation
Some college	15.21	13.66
College graduate	12.29	11.39
Graduate degree	16.56	12.33
Total	14.27	12.14

Table 8. Rosetta Stone audio CD usage, by total
household income

	Mean	Std. Deviation
Less than \$25,000	9.92	12.89
\$25,000-\$50,000	24.30	14.91
\$50,001-\$75,000	9.99	10.50
\$75,001-\$100,000	11.35	13.90
\$100,001-\$125,000	13.50	9.43
\$125,001-\$150,000	16.32	11.25
Above \$150,000	11.53	10.09
Prefer not to answer	14.51	10.57
Total	14.27	12.14

Language Proficiency Gain

Of the 89 participants who took the ACTFL OPI-C pretest, 84 scored in the lowest category, Novice Low. The remaining five scored in the next level up, Novice Mid. At the posttest, 69 participants (77.5%) gained one to two levels of language proficiency. Only 20, or 22.5%, remained in the same level as the pretest. The figure below illustrates the percentage of participants who experienced different levels of proficiency gain.

Figure 5. Language proficiency gain on the ACTFL OPI-C



As the Figure 5 illustrates, nearly half of participants experienced one level of proficiency gain and one-quarter experienced two levels of gain. Converting the pre- and posttest scores to numbers, where 1 represents "Novice Low" and 7 represents "Advanced" also allowed for calculation of means and mean growth. Using this method, the pretest mean was 1.06 and the posttest mean was 2.20. Using the paired-samples test, we determined this difference to be highly significant (p < .001).

Statistical power refers to the ability of a sample effect size to predict actual effect sizes. As the power level approaches 1.0, mean differences in samples are more likely to reflect actual differences in the larger population. A power level of 0.8 is generally accepted as adequate for publication. Based on the sample size of 89 participants and the difference between the pre- and posttests, the ACTFL OPI-C results obtained a power level above .9.

Between the pre- and posttest, most participants (42, or 47.2%) rose by 1 level; in most cases, this is a Novice Mid rating. 22 participants (24.7%) rose by 2 levels (generally, this is to Novice High). Four participants rose three levels and one rose by four levels. This growth is summarized by participant demographic characteristics below.

Table 9. Language proficiency gain on the ACTFL OPI-C, by gender

	Gain by proficiency level					
Gender	0	1	2	3	4	
Male (%)	23	48	23	7	0	
Female (%)	22	47	27	2	2	

Table 9 highlights that males and females experienced similar gain through participation in Rosetta Stone. Analysis of covariance suggests that there were no gender effects. Males on average experienced 1.14 levels of gain, and females experienced 1.16 levels of gain.

		Gain by proficiency level				
	0	1	2	3	4	
24 and under (%)	13	13	38	25	13	
25 to 34 (%)	22	56	17	6	0	
35 to 44 (%)	28	39	28	6	0	
45 to 54 (%)	29	67	5	0	0	
55 and up (%)	17	42	42	0	0	

Table 10. Language proficiency gain on the ACTFL OPI-C, by age

Table 10 suggests that different age groups experienced different amounts of average gain. For example, participants aged 24 and under experienced on average 2.12 levels of gain, while those aged 45-54 only experienced 0.76 levels of gain. The relationship is unclear however, because gain does not simply decrease as age increases; certain groups just experienced more growth than others.

	Gain by proficiency level					
	0	1	2	3	4	
Some college (%)	53	47	0	0	0	
College graduate (%)	19	44	28	7	2	
Graduate degree (%)	13	52	32	3	0	

Table 11. Language proficiency gain on the ACTFL OPI-C, by level of education

Table 11 shows that there were notable differences between the gain experienced by level of education. Participants with some college tended to see 0.47 levels of gain, while college graduates and those with graduate degrees experienced 1.30 and 1.26 levels of growth, respectively.

Table 12. Language proficiency gain on the ACTFL OPI-C, by household income

	Gain by proficiency level					
	0	1	2	3	4	
Less than \$25,000 (%)	25	75	0	0	0	
\$25,000-\$50,000 (%)	29	43	7	14	7	
\$50,001-\$75,000 (%)	33	33	29	5	0	
\$75,001-\$100,000 (%)	0	86	14	0	0	
\$100,001-\$125,000 (%)	33	33	33	0	0	
\$125,001-\$150,000 (%)	29	43	29	0	0	
Above \$150,000 (%)	0	89	11	0	0	

There was some variation in growth across income levels, but the differences between groups were less stark than education level. Again, the relationship between the two variables was unclear in that it was no clear pattern. Participants who did not provide their household income level experienced the most growth, 1.39 levels on average. Those with income levels under \$25,000 experienced the least growth, with 0.75 levels on average. Because that segment is so small, we are reluctant to place great importance on this difference.

Conclusions

In sum, the 89 participants in the evaluation of the Rosetta Stone[®] solution for learning Spanish exhibited significant gains as indicated by both the ACTFL OPI-C. These differences were highly powerful in that the differences obtained here most likely reflect differences that would occur in the larger population. Additionally, the Rosetta Stone solution received highly positive ratings from participants. While there was variation on some ratings, it does not appear that variations were related with the various demographic characteristics reported. The positive ratings reinforce the ACTFL OPI-C results: clearly, participants found the Rosetta Stone solution to be effective in learning Spanish.

In fact, based on the ACTFL OPI-C results, we can state that after completing at least two levels of training and an average of 64 hours of study with the Rosetta Stone software, the majority of the participants successfully advanced 1-2 levels of

proficiency. They went from Novice-Low to Novice-Mid and in 22 cases from Novice-Low to Novice-High (2 levels up), meaning that they went from having non functional language ability to being able to answer short (2-3 words) questions. Those who advanced two levels were able to converse with ease and confidence and handle uncomplicated tasks. Perceptions of the Rosetta Stone solution were overwhelmingly positive. Factoring in participant usage and instructional hours, we conclude that the Rosetta Stone solution is both an effective and efficient means of learning Spanish as a foreign language.

References

Al Mekhlafi, A.G. (2006). The effect of Computer Assisted Language Learning (CALL) on United Arab Emirates English as a foreign Language (EFL) school students' achievement and attitude. *Journal of Interactive Learning Research*, 17(2), 121-142.

Chapelle, C. (1997). CALL in the Year 2000: Still in Search of Research Paradigms? *Language, learning & Technology*, 1(1) 19-43.

Graff, M.M. (2006). A study of Rosetta Stone's effectiveness on improving English pronunciation. Master of Arts Thesis. California State University Dominguez Hills, CA.

Saury, R.E. (1998). Creating a psychological foundation for the evaluation of pre-packaged software in second language learning. *ED-MEDIA/ ED-TELECOM 98 World Conference on Educational Telecommunications*. Proceedings (10th, Freiburg, Germany, June 20-25, 1998).

Tien-Lok, J.L. (2003). SLLS: An online conversational spoken language learning system. Master Thesis of Engineering in Computer Science. Massachusetts Institute of Technology, MA.