# Duolingo Effectiveness Study

## FINAL REPORT

**RESEARCH TEAM**

ROUMEN VESSELINOV, PhD

Visiting Assistant Professor
Queens College
City University of New York
roumen.vesselinov@qc.cuny.edu

JOHN GREGO, PhD

Professor and Chair
Statistics Department
University of South Carolina
grego@stat.sc.edu

**December 2012**

# EXECUTIVE SUMMARY

The research study of Duolingo effectiveness was independently conducted in September-November of 2012. The study lasted for approximately eight weeks. A random representative sample was selected from Duolingo users who studied Spanish. The participants were at least 18 years of age, native speakers of English, not from Hispanic origin and not advanced users of Spanish, and all of the participants resided in the United States

The participants took one college placement Spanish language test in the beginning of the study and one test at the end of the study. The test results were measured in points (the higher the better). The improvement of language abilities was measured as the difference between the final and the initial language test results. The effectiveness of Duolingo was measured as language improvement per one hour of study.

**MAIN RESULTS**

- Overall the average improvement in language abilities was 91.4 points and the improvement was statistically significant.
- The effectiveness measure showed that on average participants gained 8.1 points per one hour of study with Duolingo.
- The 95% Confidence Interval for the effectiveness is from 5.6 points to 10.7 points gained per one hour of study.
- The study estimated that a person with no knowledge of Spanish would need between 26 and 49 hours (or 34 hours on average) to cover the material for the first college semester of Spanish. This result is based on the language test's cut-off point for the second college semester and the 95% Confidence Interval of the effectiveness measure.
- The main factor for higher effectiveness was the motivation of the participants, with people studying for travel gaining the most and people studying for personal interest gaining the least.
- Another factor for higher effectiveness was the initial level of knowledge of Spanish with beginners gaining the most and more advanced learners gaining the least.

## Introduction

Learning a foreign language has become a very appealing and important ability in the contemporary world. In many cases learning a foreign language involves using language learning software or computer assisted self-study. There are many language learning software tools available, some more popular than others. But there is very little research specifically dealing with these tools. Our research team (Vesselinov et al, 2008, 2009a, 2009b) has conducted three studies related to effectiveness, attitude and motivation of language learning software packages (Rosetta Stone ®, Auralog ® and Berlitz ®).

The goal of the current research study was to evaluate the effectiveness of Duolingo, a newly developed free language-learning website which became publicly available in 2012.

This study was funded by Duolingo but the data collection and the analysis were done independently by the Research team.

## 1. Research Design

Duolingo has distinct advantages from the research point of view compared to other language learning software packages. Duolingo users have to register online and provide a working e-mail address. Duolingo also allows extracting the exact time of use/study by date and time and by different activities: time used for lessons, time used for translation and time used for other activities.

Our research design included selection of a random representative sample of Duolingo users who were:

- Willing to participate in the study;
- Studying Spanish as a foreign language;
- At least 18 years of age;
- Native speakers of English;
- Residing in the U.S.;

- Not of Hispanic origin;
- Not advanced users of Spanish.

The last requirement was due to the fact that the language placement test used in the study has placement in college Semester 4+ as its highest evaluation group.

The recommended goal for the participants in the study was to use Duolingo for at least 30 hours during the two month study. We knew in advance that this recommendation would not be feasible for some participants. For this study we imposed a threshold of two hours of Duolingo use. The notion was that if a participant is studying foreign language for two months and they end up studying a total of two hours or less (15 minutes or less a week) this is not sufficient effort for measurable progress.

Spanish language was selected as one of the more popular languages and also because of the existence of previous research on Spanish for other language learning software packages. The length of the study was 8 weeks and was conducted between the months of September and November of 2012. A $20 gift certificate from Amazon.com was given to the people who successfully completed the study.

The main instrument for evaluating the level of knowledge of Spanish was the Web Based Computer Adaptive Placement Exam[1] (WebCAPE test). It is an established university placement test and it is offered in ESL, Spanish, French, German, Russian and Chinese. It was created by Brigham Young University and maintained by the Perpetual Technology Group. More detailed description of the test can be found at their website: http://www.perpetualworks.com/webcape/overview.

The Spanish WebCAPE test has a very high validity correlation coefficient (0.91) and very high reliability (test-retest) value of 0.81. The test is adaptive so the time for taking the test varies with an average time of 20-25 minutes. The WebCAPE test gives a score (in points) and based on that score places the students in different level groups.

[1] Spanish WebCAPE Computer-Adaptive Placement Exam by Jerry Larson and Kim Smith, WWWeb version Charles Bush. ©1998, 2004 Humanities Technology and Research Support Center, Brigham Young Univ.

**Table 1. Spanish WebCAPE Test Cut-off Points**

| Points | College Semester Placement |
|---|---|
| Below 270 | Semester 1 |
| 270-345 | Semester 2 |
| 346-428 | Semester 3 |
| Above 428 | Semester 4+ |

The measure of Effectiveness for this study was defined as follows:

$$Effectiveness = \frac{\text{Effect}}{\text{Efforts}} = \frac{\text{Improvement of language skills}}{\text{Study time}} = \frac{\text{Final-Initial test score}}{\text{Hours of study}}$$

This measure includes both the amount of progress made by each study participant and the amount of their efforts and it is a fair measure of effectiveness.

## 2. Sample Description

The entire sample selection process is graphically represented on Figure A1 in the Appendix A.

The Duolingo study on effectiveness was announced on the web. Duolingo included a link advertising the new study in Spanish on its website and put out a Google ad. In Duolingo the link was only visible to users who were logged in and were studying Spanish. People who were interested in participating in the study were asked to click on the link and go to the invitation page. On this page the study plan and requirements were explained and a short entry survey was included. The link was available for a week and 727 people viewed the invitation page and of them 556 successfully completed the entry survey. This was the initial pool of respondents in the study.

### Initial Pool

The initial pool (N=556) of potential participants had an average age of 30.4 years with 46.2% females and 98.4% of them were novice to intermediate (self-report) users of Spanish. A small portion of them (7.6%) were of Hispanic origin. The majority of respondents were White

(74.6%), followed by Asian (11.2%), Black/African American (5.4%), Native American, Alaskan or Pacific Islander (0.9%) and of other race (7.9%), including multiracial categories.

The primary reason for studying Spanish was personal interest (61.8%), followed by business/work (14.4%), travel (10.5%), school (11.4%), and other reasons (2.0%). For other reasons the respondents mentioned: “all of the above”, “boredom”, “family”, “fun”, “to help my son learn Spanish”, “to talk with my Spanish family members”, etc.

A small portion (13.6%) of the respondents’ spouse, partner, or close friends spoke Spanish. Similar proportion (10.1%) of their parents, grandparents, or great grandparents spoke Spanish. The majority (92.6%) of the respondents had English as their native language. Other native languages included: Arabic, Armenian, Bambara, Bengali, Bulgarian, Chinese, French, German, Hebrew, Hindi, Polish, Portuguese, Romanian, Russian, Tagalog, Tajik, Tamil, Tulu, Urdu. A third (32.7%) of the respondents knew at least one other foreign language.

Educational composition was as follows: 0-11 grade (6.5%), High school diploma/GED (7.4%), some college (31.3%), college graduate, BA or equivalent (37.4%), graduate degree - MA, PhD or higher degree (17.4%).

The majority of the respondents were employed either full time (45.6%) or part time (11.0%). Almost a third (29.2%) of the respondents were students and the rest were unemployed (9.7%) and with other employment (4.5%). For other employment the respondents listed: “homemaker”, “retired”, “disabled”, “self-employed”, “stay at home mom”, etc.

The majority (97.7%) of the initial pool stated that they resided in the US. The rest of them were from Brazil, Bulgaria, Canada, China, Costa Rica, Germany, and United Kingdom.

## Pool of Eligible Participants

From the Initial Pool (N=556) we excluded the following ineligible participants:

1. People who were younger than 18 years of age.
2. People whose native language was not English.

3. People from Hispanic origin.
4. People who did not live in the US (self- report and by IP address).
5. People whose IP address was not identifiable (blank).

Altogether 170 people were ineligible for this study and the final pool of eligible participants for sample selection was N=386.

The average age of the pool of eligible participants was 32.0 years with 48.4% females, and 99.0% of them were novice to intermediate (self-report) users of Spanish. The racial composition was as follows: Black/African American (5.7%), Asian (9.3%), White/Caucasian (80.1%), Native American, Alaskan or Pacific Islander (1.3%) and other race, including multiracial categories (3.6%).

The primary reason for studying Spanish was as follows: business/work (14.5%), travel (8.8%), school (8.3%), personal interest (66.5%) and other reasons (1.8%). For other reasons the respondents mentioned: “all of the above”, “boredom”, “family”, “fun”, “to help my son learn Spanish”.

A small proportion (11.0%) of the respondents’ spouse, partner, or close friends spoke Spanish. An even smaller proportion (4.5%) of their parents, grandparents, or great grandparents spoke Spanish. All participants were native speakers of English and they were not of Hispanic origin.

More than a quarter (28.5%) of the respondents knew at least one other foreign language. Educational composition was as follows: 0-11 grade (0.8%), High school diploma/GED (8.0%), some college (33.4%), college graduate, BA or equivalent (39.9%), graduate degree - MA, PhD or higher degree (17.9%).

Employment composition was as follows: unemployed (10.8%), student (23.2%), full time employed (49.9%), part time employed (11.6%) and other employment (4.6%). For other employment the participants listed: “homemaker”, “retired”, “disabled”, “self-employed”, “stay at home mom”, etc.

The respondents were geographically from 46 states (see Table A1 in Appendix A).

## Initial Random Sample

The people in the initial sample were randomly selected from the pool of eligible participants. They are people of 18 years of age and older, native speakers of English, not of Hispanic origin, not advanced users of Spanish and residing in the US. The country of residence was based on the IP address identification.

Originally 211 people were selected and they completed the baseline WebCAPE placement test in Spanish. But 8 people scored above 428 points which put them in the highest group the test can place them (Semester 4+). They were too advanced to be tested with the WebCAPE test and they were dropped from the study. An additional 7 people refused to participate in the study. The initial random sample consisted of 196 people.

The average age of the initial sample participants was 31.3 years with 45.4% females, and 99.5% of them being novice to intermediate users of Spanish (self-report). The racial composition was: Black/African American (4.1%), Asian (9.2%), White/Caucasian (81.1%), Native American, Alaskan or Pacific Islander (2.0%) and other race, including multiracial categories (3.6%).

The primary reason for studying Spanish was as follows: business/work (15.8%), travel (9.2%), school (6.1%), personal interest (67.9%) and other reasons (1.0%). Other reasons included "fun" and "family".

**Table 2. Initial Random Sample: Age and Gender Distribution**

| Age | Female (N) | Male (N) | Total (N) | Percent |
|---|---|---|---|---|
| Up to 20 years old | 3 | 14 | 17 | 8.7 |
| 21-30 years old | 53 | 45 | 98 | 50.0 |
| 31-40 years old | 13 | 32 | 45 | 23.0 |
| Over 40 years old | 20 | 16 | 36 | 18.4 |
| **Total** | **89** | **107** | **196** | **100.0** |

A small proportion (10.3%) of the respondents' spouse, partner, or close friends spoke Spanish. An even smaller proportion (4.1%) of their parents, grandparents, or great grandparents spoke

Spanish. More than a quarter (29.7%) of the respondents knew at least one other foreign language.

Educational composition was as follows: 0-11 grade (0.5%), High school diploma/GED (8.2%), some college (29.6%), college graduate, BA or equivalent (43.9%), graduate degree - MA, PhD or higher degree (17.9%).

Employment composition was as follows: unemployed (10.5%), student (18.4%), full time employed (53.2%), part time employed (13.7%), and other employment (4.2%). For other employment the participants listed: “homemaker”, “retired”, “disabled”, “self-employed”, “stay at home mom”, etc.

The participants in the initial random sample were geographically from 42 states (see Table A1 in Appendix A).

After the selection the study participants were asked to go online and complete the first WebCAPE placement test in Spanish. The results are as follows:

**Table 3. Initial WebCAPE Semester Placement**

| College Semester | Number | Percent |
|---|---|---|
| First | 157 | 80.1 |
| Second | 25 | 12.8 |
| Third | 14 | 7.1 |
| Fourth+* | | |
| **Total** | **196** | **100.0** |

* Eight people who scored above 428 (Fourth+) on the initial test were excluded from the study.

The majority of people in the sample (80.1%) were evaluated to be in the first semester group. About 13% were in the second semester group and about 7% in the third semester group. About 18% (N=35) of the participants scored the lowest score of 0 on the placement test which makes them absolute novice/beginner learners of Spanish. The mean WebCAPE score was 158 (Median=154) which is well below the cut-off point for second college semester of Spanish.

**Table 4. Initial WebCAPE Placement Test Statistics**

| Statistics | WebCAPE points |
|---|---|
| Mean (std) | 158.0 (115.4) |
| Median | 154.0 |
| Min | 0 |
| Max | 414 |
| N | 196 |

## Final Study Sample

The study continued for 8 weeks, starting in September 2012 and ending on November 9, 2012. During the study the research team sent e-mail reminders two times a week to the participants with information about how much time they have used Duolingo each week.

At the end of the study we reviewed the time use of the participants. The initial target for this study was at least 30 hours of use for the two months of study. A quarter of the final sample (N=22) did have 30 hours or more of use. The lower threshold for inclusion in the conclusion of the study was defined as 2 hours. People who had studied Spanish for 2 hours or less for the whole period of two months were considered as not seriously studying and they did not complete the study. At the end 90 people completed the study and took the final WebCAPE test. Two of them were eventually excluded from the study. One of them, in addition to Duolingo had participated in another Spanish course and this was the reason for exclusion. The second person's final test was actually automatically closed after 4 hours inactivity and was considered invalid.

The question about additional help during the study was asked in the exit survey as a way to confirm that Duolingo was the only tool for studying Spanish. In addition to the above excluded person, a couple of participants said that they used occasionally some web tools for additional information and translation, watched some Spanish videos etc.

The final study sample consisted of 88 people with more than 2 hours use of Duolingo and valid initial and final WebCAPE tests. They are people of 18 years of age and older, native speakers of English, not from Hispanic origin, not advanced users of Spanish and residing in the US.

The average age of the final sample was 34.9 years, ranging from 18 to 66 years of age. There were exactly 50% females, and 85.2% of the final sample were novice to intermediate users of Spanish (self-report). The racial composition was: Black/African American (3.4%), Asian (8.0%), White/Caucasian (81.8%), Native American, Alaskan or Pacific Islander (2.3%) and other race, including multiracial categories (4.5%).

**Table 5. Final Study Sample: Age and Gender Decomposition**

| Age | Female (N) | Male (N) | Total (N) | Percent |
|---|---|---|---|---|
| 18 to 20 years old | 0 | 5 | 5 | 5.7 |
| 21-30 years old | 23 | 13 | 36 | 40.9 |
| 31-40 years old | 7 | 14 | 21 | 23.9 |
| Over 40 years old | 14 | 12 | 26 | 29.5 |
| **Total** | **44** | **44** | **88** | **100.0** |

The primary reason for studying Spanish was as follows: business/work (18.2%), travel (11.4%), school (2.3%), and personal interest (68.2%).

A small proportion (10.3%) of the respondents' spouse, partner, or close friends spoke Spanish. An even smaller proportion (2.3%) of their parents, grandparents, or great grandparents spoke Spanish. More than a quarter (28.4%) of the respondents knew at least one other foreign language.

Educational composition was as follows: High school diploma/GED (3.4%), some college (27.3%), college graduate, BA or equivalent (42.0%), graduate degree - MA, PhD or higher degree (27.3%).

Employment composition was as follows: unemployed (15.3%), student (9.4%), full time employed (55.3%), part time employed (15.3%) and other employment (4.7%). For other employment the participants listed: "retired", "disabled", "self-employed", "stay at home mom", etc.

The participants in the final sample were from 31 states (see Table A1 in Appendix A).

Duolingo offered the possibility to estimate the time used for different activities while studying Spanish with Duolingo. Three of the people in the final sample had anti-tracking browser add-ons and their total time of use was computed based partially on server reports and partially on self-report. For the majority (N=85) of the final sample only the objective measure of time use was applied. On average the participants spent 70.0% of the time using the lessons provided by Duolingo. The rest of the time was devoted to translation (9.0%) and other activities (21.0%). Other activities included browsing around the website, questions, vocabulary, home, etc.

**Table 6. Structure of the Study Time**

| | Percent of Study Time Spent for: | | |
|---|---|---|---|
| **Statistics** | **Lessons** | **Translation** | **Other Activities** |
| Mean (std) | 70.0 (11.9) | 9.0 (11.2) | 21.0 (9.8) |
| Median | 70.4 | 5.0 | 20.1 |
| Min | 26 | 0 | 1 |
| Max | 95 | 52 | 46 |
| N | 88 | 88 | 88 |

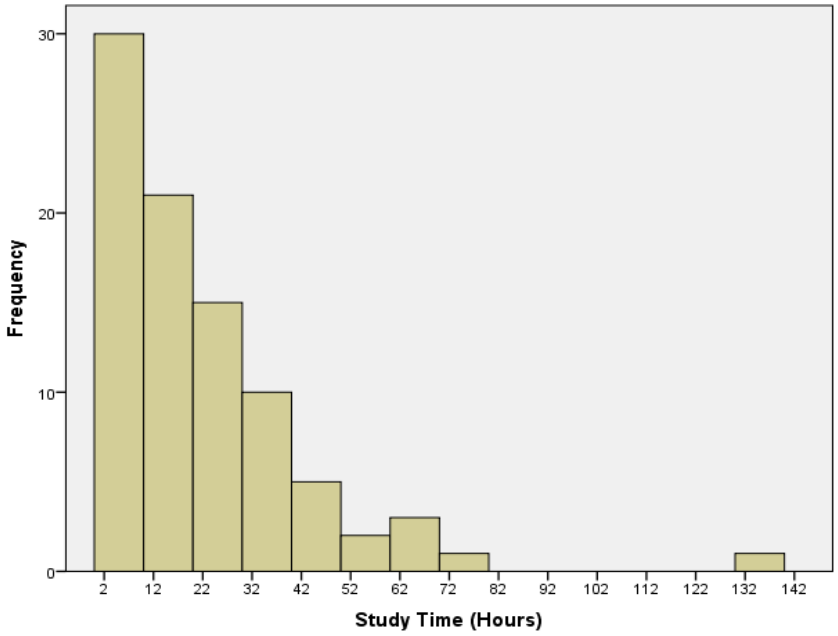
The average study time was about 22 hours with 2 hours as the lowest and 133 hours as the highest.

**Tale 7. Study Time**

| **Statistics** | **Hours of Study** |
|---|---|
| Mean (std) | 22.0 (20.4) |
| Median | 16.9 |
| Min | 2 |
| Max | 133 |
| N | 88 |

A quarter of the participants studied between 2 and 8 hours, and a quarter of them have 30 hours or more, including 7 people with 50 hours or more.

**Figure 1. Study Time Distribution**




# Main Results

## Effectiveness

Usually in these kind of studies with two measures (initial and final test) a simple measure of the progress is computed and analyzed. This is the difference score which is the difference between the final and initial score. The problem with the progress measured as difference score is that it does not take into account how much time each participant actually studied. For example, the highest improvement in this study was a change score of 341 points and it was achieved after more than 33 hours of study while a quarter of the participants have studied between 2 and 8 hours. It is reasonable to expect that the time of study does matter in most cases. The study time varied a lot; from 2 hours to 133 hours for the two month study. That is why a more reliable and analytical measure of progress was constructed. We created a new indicator of the effectiveness of the Spanish language study. The new indicator is the ratio of the amount of progress (the difference score) divided by the time of study (in hours). For example a participant with 20 hours of study and improvement of 40 WebCAPE points from the

initial to the final test will have an effectiveness measure of 40/20=2. The resulting number 2 means that this person gained 2 points per one hour of study. This is a more fair and objective measure of progress because it takes into account the two major elements: time for study and improvement.

**Table 8. Language Improvement**

| Statistics | Initial WebCAPE | Final WebCAPE | Improvement (Final-Initial) |
|---|---|---|---|
| Mean (std) | 162.5 (116.5) | 253.9 (110.5) | 91.4 (88.0) |
| Median | 161 | 262.5 | 81.5 |
| Min | 0 | 0 | -57 |
| Max | 405 | 539 | 341 |
| N | 88 | 88 | 88 |

On average there was an improvement of 91.4 WebCAPE points and this difference was statistically significant (paired samples t-test=9.74, p<.001). The 95% confidence interval of this difference was (72.7-110.0). A little over 84% (N=74) of the participants did improve their WebCAPE score at the end of the study. Only 16% (N=14) of the participants had the same or lower WebCAPE score at the end of the study. Some of these cases are the result of people studying a couple of hours in the beginning of the study and then they stopped studying for the next 4-6 weeks and then they took the test. Another group of people started the study with a high WebCAPE score (300 or above) and it was difficult for them to add more points at this high level. Language improvement has limited importance because it does not account for the length of study.

**Table 9. WebCAPE Semester Placement**

| College Semester | Initial Test | Final Test |
|---|---|---|
| | Percent (Number) | Percent (Number) |
| First | 77.3 (68) | 52.3 (46) |
| Second | 14.8 (13) | 29.5 (26) |
| Third | 8.0 (7) | 14.8 (13) |
| Fourth+ | | 3.4 (3) |
| **Total** | **100 (88)** | **100 (88)** |

Almost 23% of the participants moved-up one semester and more than 9% moved-up two semesters in the placement test. Semester placement dynamics is not as precise as the effectiveness because it does not account for the time of study. In addition it could be the case that a person gains 100 points and still be within the range of the same semester while another person can gain only 10 points and this might be enough to move up to the next semester.

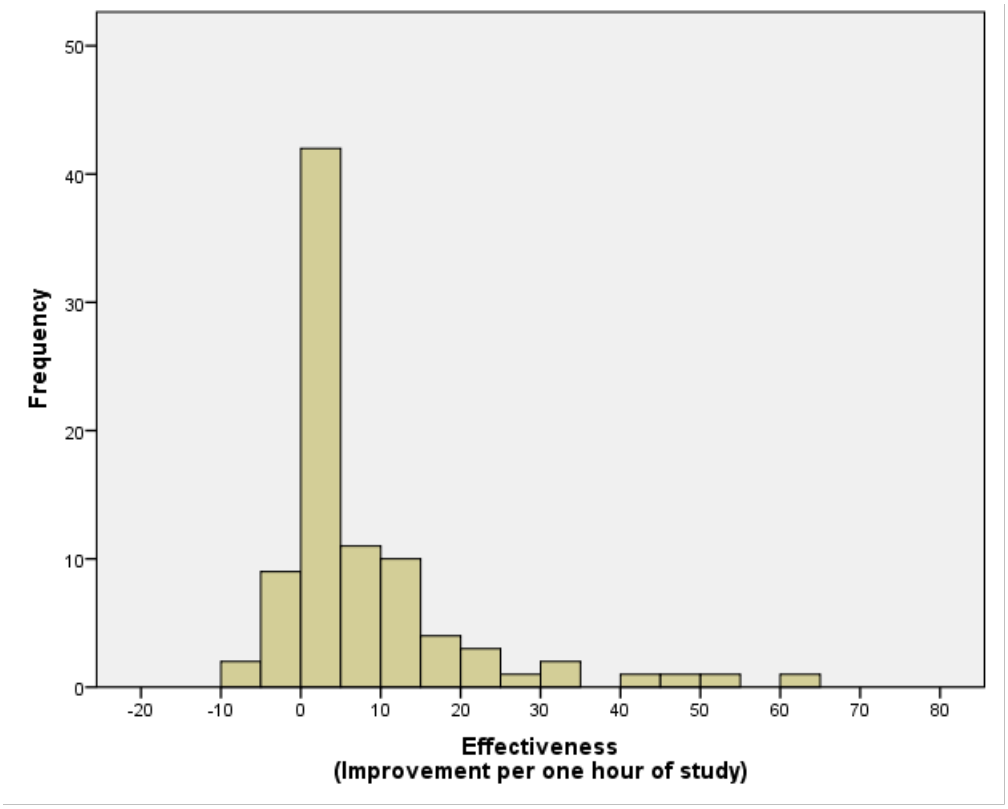
**Table 10. Effectiveness**

| Statistics | Improvement per one hour of study |
|---|---|
| Mean (std) | 8.1 (12.1) |
| Median | 3.9 |
| Min | -5.6 |
| Max | 60.4 |
| 95% Confidence Interval (CI) | 5.6 - 10.7 |
| N | 88 |

The average effectiveness (gain) was 8.1 WebCAPE points per one hour of study. The 95% Confidence Interval was between 5.6 points and 10.7 points. The distribution of effectiveness is slightly asymmetric and with higher than normal excess. That is why for the confidence interval we performed also a bootstrapping procedure based on 1000 samples. The new bootstrapped confidence interval (5.8-10.7) was almost identical with the classical one. We chose the classical because it was slightly more conservative (wider).

The threshold for WebCAPE for Semester 2 is 270 points (i.e. people with at least 270 points are placed at least in Second college semester or more). So if a person starts with no knowledge of Spanish (WebCAPE=0) they will need on average 34 hours (=270/8.1) of study with Duolingo in order to cover the material for the First college semester and move to Second college semester. The transformed 95% confidence interval is between 26 hours (=270/10.7) and 49 hours (=270/5.6) of study with Duolingo (results rounded upward).

**Figure 2. Effectiveness Distribution**




The lower 25% of the participants gained up to 1.1 points per one hour of study, while the upper 25% of them gained over 10.4 points per one hour of study. The largest gain of the study was 60.4 points per one hour of study.

## Factors for Effectiveness

From an analytical point of view it is interesting to know what factors if any, determine improvements in language skills. We started with demographic factors. Gender results were unexpected; men had better progress than women. Under further investigation we found that 11 out of 44 women decreased slightly their results. Most of them started the study with a high initial WebCAPE score and could not improve their scores, while none of the men had such negative “progress”. Three men had the same score (zero progress) but none decreased.

There were no statistically significant differences by racial groups. There were some differences in progress by age groups with the 31-40 year old group having the highest results but the differences were not statistically significant. There were expected differences between the education groups with the MA/PhD group having the highest progress report but these differences were not statistically significant.

There was no effect on progress of the presence or absence of relatives or friends who spoke Spanish. The same lack of influence was true for the type of employment.

Although people who knew another foreign language did a little bit better than those who did not, the difference between the two was not statistically significant.

The only significant factor (p=.006) was the reason for studying Spanish. The best progress was achieved by people studying for travel with an average progress of 17.6 points improvement per one hour of study. The majority of the people studied for personal interest and school (60 personal interest and 2 school) but this group had the least progress with 5.7 points improvement per one hour of study. The difference between the Travel and Personal Interest/School groups was statistically significant (p=.009 with Tukey HSD correction for multiple comparisons). The difference between Travel and Business/Work groups and the difference between Business/Work and Personal Interest/School groups were not statistically significant.

**Table 11. Reason for Studying Spanish as Factor for Effectiveness**

| Reason | N | Effectiveness |
|---|---|---|
| | | Mean (std) |
| Travel | 10 | 17.6 (22.7)* |
| Business/Work | 16 | 11.4 (15.7) |
| Personal Interest or School | 62 | 5.7 (7.0)* |
| **Total** | **88** | **8.1 (12.1)** |

* p=.009 with Tukey HSD correction for multiple comparisons

The structure of time using Duolingo (percent of time for lessons, translation and other activities) did not have statistically significant effect of progress and language improvement.

Another interesting aspect of the analysis is the question about the initial level of knowledge of Spanish and the effectiveness of the study. The study participants were placed in 4 groups based on their initial WebCAPE test scores and the effectiveness by these 4 groups is presented in the next table.

**Table 12. Effectiveness by Initial Level of Language Ability**

| **Initial Level College Semester** | **N** | **Effectiveness*** Mean (std) |
|---|---|---|
| First | 68 | 9.2 (12,2) |
| Second | 13 | 6.4 (13.4) |
| Third | 7 | 0.6 (2.6) |
| Fourth+ | | |
| **Total** | **88** | **8.1 (12.1)** |

* The group differences are not statistically significant

As expected participants in the beginner/novice group (First semester) had the biggest improvement in their language skills. As the same time the more advanced participants (Third semester) showed more modest improvement in their language skills. This is an expected result. Still there are no statistically significant differences so we cannot partition the effectiveness results by the initial level of knowledge of Spanish. But the result is noteworthy because of the large effect sizes (9.2 vs 0.6 points progress) and with larger sample the result would have been significant.

## 3. User Satisfaction

At the end of the study the participants were asked to complete an exit survey with questions mostly related to their experience with Duolingo and their recommendations. Overall 66 people completed the exit survey.

**Table 13. Users Satisfaction**

| Do you agree with the following statement? | Strongly Disagree | Disagree | Neither Disagree nor Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| | Percent | | | | |
| "Duolingo was easy to use" | | | 4.5 | 37.9 | 57.6 |
| "Duolingo was helpful in studying Spanish" | | | 7.6 | 40.9 | 51.5 |
| "I enjoyed learning Spanish with Duolingo" | | 3.0 | 9.1 | 45.5 | 42.4 |
| "I am satisfied with Duolingo" | | 4.5 | 16.7 | 40.9 | 37.9 |

If we combine the "Agree" and "Strongly Agree" answers we can say that 95.5% of Duolingo users consider it easy to use, 92.4% think that Duolingo helps them study Spanish, 87.9% enjoy learning Spanish with Duolingo and 78.8% are satisfied with Duolingo.

In the exit survey a special question was included: "How likely are you to recommend Duolingo to a colleague or friend?" with 11 possible answers, from 0 "Very unlikely" to 10 "Very likely". The answers to this question were used to compute the so called Net Promoter Score (NPS). This is "a management tool that can be used to gauge the loyalty of a firm's customer relationships" (Wikipedia). It was developed by Reichheld (2003) and it categorizes users in three categories: "Promoters" (answers 9, 10), "Passives" (answers 7, 8), and "Detractors" (answers 0-6). The Net Promoter Score (NPS) is equal to the difference between "Promoters" and "Detractors" and in general it can vary from -100 (all detractors) to + 100 (all promoters). As a rule positive NPS is good news for the company and any score of +50 or more is considered an excellent indicator for the company.

From our exit survey (N=66) the "Promoters" were 63.6% and the "Detractors" were 12.1% and "Passives" were 24.2%. That way the Duolingo NPS=+51.5. This is an excellent result.

Also 93.8% of the participants in the exit survey declared that they will continue to use Duolingo after the study ends.

# Conclusion

This study on effectiveness of Duolingo answered some very important questions. The vast majority of the participants in the study liked the product and most of them succeeded in improving their knowledge of Spanish. The improvement was statistically significant and on average Duolingo users gained a little over 8 points of WebCAPE placement test per one hour of study. Based on these findings we can say that for a completely novice user of Spanish it would take on average 26 to 49 hours of study with Duolingo to cover the material for the first college semester of Spanish. This result is based on the transformed 95% Confidence Interval of the effectiveness measure and the language test's cut-off points for 4 semester Spanish college sequence.

Only two factors were shown to matter for the progress of the participants per one hour of study. First, their motivation; with people who studied Spanish to travel having the biggest improvement. People who studied mainly for personal interest and school had more modest improvement. Second, the initial level of knowledge of Spanish was another contributing factor. People who were beginners (Semester 1) had the biggest improvement and more advanced people (Semester 2 and 3) had the smallest improvement.

## Limitations of the Study

We cannot generalize the results for this study for languages other than Spanish. It is fair to expect similar results with other languages but there are not enough empirical studies in the literature to prove this expectation. More studies with other languages are necessary to expend the current results.

As expected many people had difficulties keeping up with the study and their use of Duolingo for the two months of study was very uneven. Many people dropped out of the study or spent less than two hours studying Spanish. Although we informed the participants twice a week of their study time for the week this was not enough for many of them. It is highly recommended that Duolingo develops some kind of individual online clock which shows how much time each

user spends by date or week. This recommendation is not specifically for future research studies but for the everyday use of Duolingo. Some people do not have an accurate notion about how much time they have spent studying a foreign language. And if they spend less than two hours studying for two months their expectations for improvement cannot be very high.

Some of the participants considered the incentive of $20 gift certificate to be very low and probably with a higher incentive we could have kept some of the people we lost.

Some of the participants (N=11) decreased slightly in their WebCAPE scores. Some of them studied very irregularly, e.g. a few hours in the beginning of the study followed by long periods of inactivity. So it could be expected that they forgot some of the new knowledge they acquired in the beginning of the study. But the majority of them came with more advanced knowledge of Spanish (Second and Third semester placement). And when starting at high level it was not easy to improve their score. So it should be acknowledged that more advanced users cannot expect the same rapid success as the beginner/novice users although this difference was not statistically significant.

Similarly as it was explained above we excluded 8 very advanced users of Spanish (Semester 4+) of the study. The reason was that the WebCAPE test has its limitations as language placement test and it cannot effectively handle very advanced users of Spanish.

As many participants mentioned in their exit survey recommendations the WebCAPE test did not evaluate a lot of language skills they acquired for the two months of study. In that regard for future studies we would recommend in addition to the written placement test to include some test of spoken proficiency (e.g. ACTFL®).

This research study draws upon some previous empirical studies (Vesselinov et al, 2008, 2009a, 2009b) and it is a fair question whether the results could be compared. In very general terms it is possible to compare the results but in a narrow statistical sense such comparison is not easy to accomplish. There are some major differences between the two sets of studies.

- The previous studies were based on the 2008 version of the respective language software packages. Four years in the digital world can make a big difference. A more fair comparison would be if there is a newer study using the latest version of other language learning software packages and web applications.
- The definition of effectiveness was different for the two sets of studies. In the current study the improvement in the language skills was related to the study time because there was an objective measure of this time. Such measure was not available in the 2008 study because half of the participants worked at home without online record for time of use.
- In the 2008 study a test for spoken proficiency was performed while no such test was included in the current study.

## Cited Literature

Vesselinov, R., 2008, Measuring the Effectiveness of Rosetta Stone®, Final Report, manuscript available through Rosetta Stone®.

Vesselinov, R., J. Grego, B. Habing, A. Lutz, 2009a, Measuring the Attitude and Motivation of Rosetta Stone ® Users, Final Report, manuscript available through Rosetta Stone®.

Vesselinov, R., J. Grego, B. Habing, A. Lutz, 2009b, Comparative Analysis of Motivation of Different Language Learning Software, Final Report, manuscript available through Rosetta Stone®.

Reichheld, Frederick F. (December 2003). "One Number You Need to Grow". Harvard Business Review.

# Appendix A

**Figure A1. Sample Selection Tree**

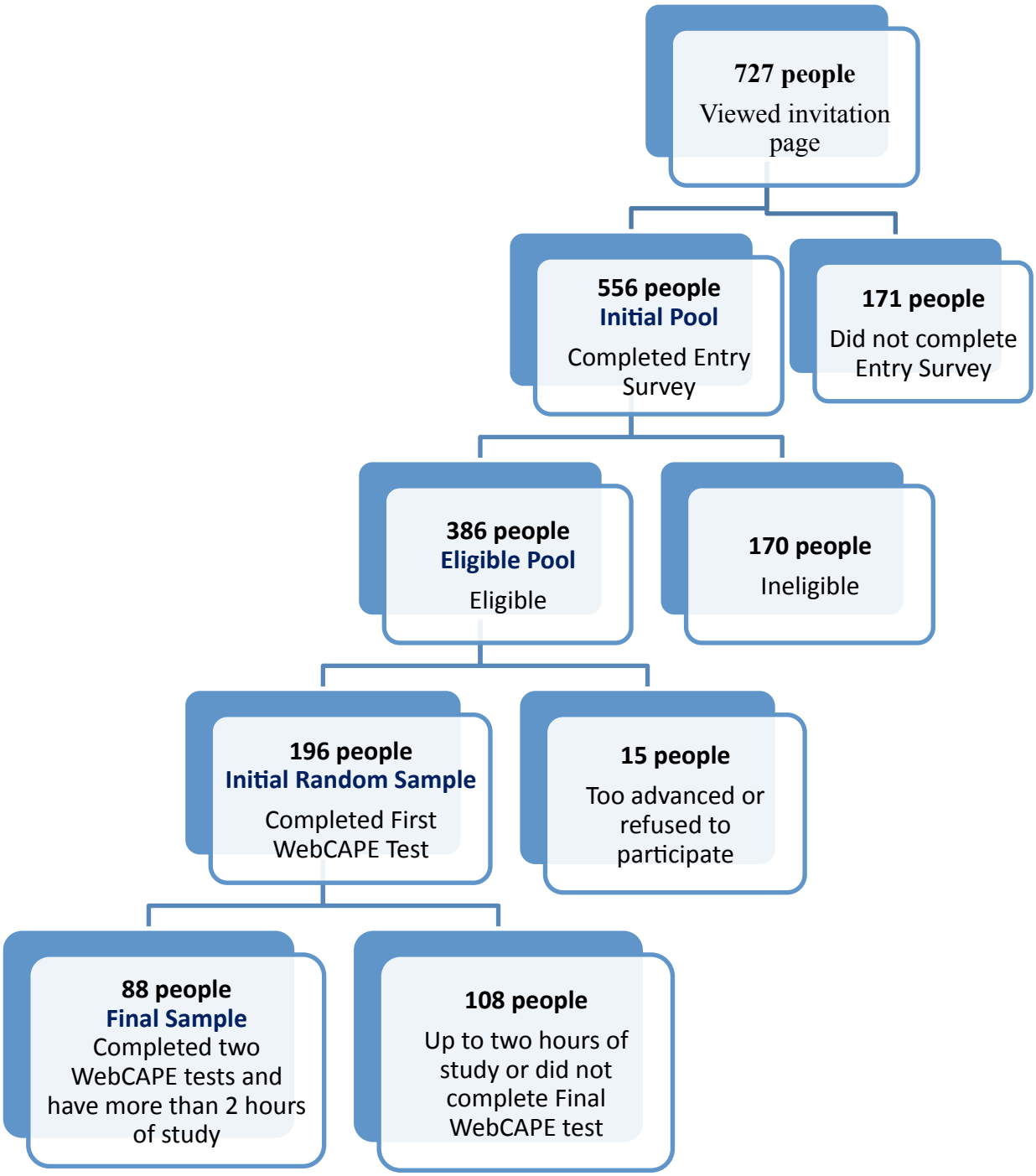


**Table A1. Geographic Distribution (Number of people)**

| | State | ST | Eligible Pool | Initial Sample | Final Sample |
|---|---|---|---|---|---|
| 1 | Alaska | AK | 1 | 1 | |
| 2 | Alabama | AL | 2 | 1 | 1 |
| 3 | Arkansas | AR | 1 | 1 | |
| 4 | Arizona | AZ | 4 | 4 | 2 |
| 5 | California | CA | 54 | 28 | 12 |
| 6 | Colorado | CO | 8 | 4 | 2 |
| 7 | Connecticut | CT | 6 | 3 | 2 |
| 8 | District of Columbia | DC | 2 | 2 | 2 |
| 9 | Delaware | DE | 2 | 1 | 1 |
| 10 | Florida | FL | 15 | 6 | 2 |
| 11 | Georgia | GA | 12 | 7 | 4 |
| 12 | Iowa | IA | 5 | 3 | 2 |
| 13 | Idaho | ID | 3 | 2 | |
| 14 | Illinois | IL | 16 | 8 | 4 |
| 15 | Indiana | IN | 5 | 2 | 1 |
| 16 | Kansas | KS | 2 | 1 | |
| 17 | Kentucky | KY | 5 | 3 | 2 |
| 18 | Massachusetts | MA | 11 | 4 | 2 |
| 19 | Maryland | MD | 7 | 3 | 1 |
| 20 | Maine | ME | 1 | 1 | |
| 21 | Michigan | MI | 12 | 4 | 2 |
| 22 | Minnesota | MN | 7 | 4 | 3 |
| 23 | Missouri | MO | 4 | 3 | 2 |
| 24 | North Carolina | NC | 10 | 4 | 2 |
| 25 | North Dakota | ND | 2 | | |
| 26 | Nebraska | NE | 6 | 4 | 2 |
| 27 | New Hampshire | NH | 4 | 3 | 1 |
| 28 | New Jersey | NJ | 6 | 4 | |
| 29 | New Mexico | NM | 1 | | |
| 30 | Nevada | NV | 5 | 3 | 1 |

**Table A1 Continued**

| | State | ST | Eligible Pool | Initial Sample | Final Sample |
|---|---|---|---|---|---|
| 31 | New York | NY | 24 | 10 | 3 |
| 32 | Ohio | OH | 5 | 2 | |
| 33 | Oregon | OR | 7 | 4 | 2 |
| 34 | Pennsylvania | PA | 17 | 8 | 5 |
| 35 | Rhode Island | RI | 2 | | |
| 36 | South Carolina | SC | 5 | 4 | 2 |
| 37 | South Dakota | SD | 1 | 1 | |
| 38 | Tennessee | TN | 5 | 2 | |
| 39 | Texas | TX | 18 | 10 | 4 |
| 40 | Utah | UT | 4 | 1 | |
| 41 | Virginia | VA | 10 | 5 | 3 |
| 42 | Vermont | VT | 2 | 1 | |
| 43 | Washington | WA | 19 | 11 | 5 |
| 44 | Wisconsin | WI | 6 | 4 | 1 |
| 45 | West Virginia | WV | 1 | | |
| 46 | Wyoming | WY | 1 | 1 | 1 |
| | Unknown state (but US) | | 40 | 18 | 9 |
| **Total** | | **US** | **386** | **196** | **88** |