The italki Efficacy Study

FINAL REPORT

RESEARCH TEAM

ROUMEN VESSELINOV[1,2], PhD

Economics Department
Queens College,
City University of New York
roumen.vesselinov@qc.cuny.edu

JOHN GREGO[2], PhD

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

JANUARY 2018

---

[1] Corresponding author.
[2] This report represents the individual opinion of the authors and not necessarily of the two institutions.

THIS PAGE LEFT INTENTIONALLY BLANK

EXECUTIVE SUMMARY

This study was independently conducted by the Research Team from August 2017 to November 2017. A random representative sample of 125 italki users was drawn. The participants took one college placement Spanish language test and one oral proficiency test, then studied Spanish with italki for two months and took the same tests again. The improvement in language abilities was measured as the difference between the final and the initial language test results. The efficacy of italki was measured as written proficiency improvement per one hour of study and the proportion of users who improved their oral proficiency. The final sample size was 102.

MAIN RESULTS

Written Proficiency Gain:

• Overall 78% of the participants improved their written proficiency.

• italki users need on average 19 hours of study in a two-month period

 to cover the requirements for one college semester of Spanish.

Oral Proficiency Gain:

• About 72% of the participants increased their oral proficiency by at least one level.

DETAILED RESULTS

Written Proficiency:

• The efficacy of italki is a gain of about 14 test points per one hour of study.

 For beginners, the gain is 21 points and for advanced users it is 6-7 points.

• About 51% of the participants moved up at least one college semester level.

 Of those, 32% moved up one semester, 16% - two semesters, and 3% - three semesters.

Oral Proficiency:

• About 72% of the participants increased their oral proficiency by at least one level.

  Of those, about 49% moved up one level, 19% - two levels, and 4% - more than two levels.

User Satisfaction:

• The clear majority of users thought that italki was easy to use (89%),

  helpful (98%), enjoyable (98%), and they were satisfied with it (92%).

• italki received a positive Net Promoter Score of 52.8 from users.

• italki efficacy was not affected by gender, age, education, native language, etc.

CONTENTS

1. Introduction

There is a growing interest in the world in using language software packages or language applications (apps) by individuals, organizations and school districts. On the other hand, there are many language apps and their number is growing. There is an urgent need for users, investors, and educators to have better and comparable scientific evaluation about the different language apps' abilities and efficacy. Users cannot rely on the self-serving marketing campaigns by the respective companies. There are many claims like "Learn a foreign language in 10 days!" It is not clear how much you can learn for 10 days, or how many hours you have to study during these 10 days. It is very unrealistic and unfounded expectation but more importantly usually this is just a claim without any scientific evidence.

Since 2008 our research team has been working in the area of evaluating language learning apps. We have already conducted eight studies attempting to directly evaluate the efficacy, attitude and motivation of some popular language learning apps like, Rosetta Stone, Duolingo, Busuu, Babbel, and Hello English (Vesselinov 2009, Vesselinov et al. 2009a, 2009b, Vesselinov & Grego, 2012, 2015, 2016a, 2016b, 2017).

In this study the Research Team is trying to evaluate the efficacy of a well-known language learning product: italki[3] which helps language learners connect with language teachers from all over the world. Whether users are learning Chinese for business, Spanish for travel, or Arabic for personal interest, italki connects people with professional teachers or tutors to help them become fluent.

Many people study a foreign language in a traditional way for years, and are still unable to speak the language. Most of these people have very few opportunities to communicate with a person from that country. By connecting learners with online teachers, learners can become fluent faster and have an authentic cultural experience.

---

[3] www.italki.com Note. The name of the company is spelled with lower case "i", italki.

italki is one of the well-known leaders in online language education, and according to the company, it has over 3 million students and 5000+ teachers of over 100 world languages. italki connects students with teachers around the world for 1-on-1 language lessons.

italki was founded by Kevin Chen (US) and Yongyue Jiang (China) in 2007. The company is registered in Hong Kong, with a development office in Shanghai, China.

This study was funded by italki but the data collection and the analysis were carried out independently by the Research Team. The two language tests used in the study were designed, developed and managed by two external independent testing companies.

2. Research Design

A random sample for this study was selected from existing italki users. There were some requirements for the potential participants.

- Willing to study Spanish using only italki for two months, and take two sets of language tests;
- At least 18 years of age;
- Not advanced learners of Spanish.

The last requirement was due to the fact that the written language placement test used in the study has placement in college Semester 4+ as its highest evaluation group and it has limited capabilities to test very advanced users. The oral proficiency test has no limitations because the evaluation is done by independent raters.

The recommended goal for the participants in the study was to use italki for at least 9 hours during the two-month study. Based on our experience with previous studies we imposed a threshold of at least two hours of study for the written test. People with less than two hours of study were not allowed to complete the study because there was not a sufficient effort for measurable progress. For the oral proficiency test, the requirement was at least 9 hours of study. The testing company requires 2-3 months of study for test-retest but does not specify how many hours of study is required.

Spanish language was selected as one of the more popular languages and because of the existence of previous research on Spanish for other language learning apps. The length of the study was approximately 8 weeks and it was conducted between the months of August 2017 and November 2017. People who successfully completed the study were given a 30% refund (up to $100) of the amount they paid for lessons during the study.

The main instrument for evaluating the level of knowledge of Spanish was the Web Based Computer Adaptive Placement Exam[4] (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 is maintained by the Perpetual Technology Group. A more detailed description of the test can be found at their website[5].

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 (college semesters).

Table 1. Spanish WebCAPE Test Cut-off Points

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

The WebCAPE results alone cannot give a clear picture about the efficacy of the language learning app because they do not account for the time spent studying. That is why we are relying on a **direct and objective** measure of efficacy which is defined as follows:

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

Efficacy=Improvement per one hour of study

This measure includes both the amount of progress made by each study participant and the amount of their effort. It is a fair measure of efficacy and also a direct and objective measure

---

[4] Spanish WebCAPE Computer-Adaptive Placement Exam by Jerry Larson and Kim Smith, online version Charles Bush. ©1998, 2004 Humanities Technology and Research Support Center, Brigham Young University.
[5] http://www.perpetualworks.com/webcape/overview

of efficacy. Direct, because it includes directly the effect and the effort. Objective, because the effect is measured by an independent college placement test (instead of our own test) and the effort is measured by the actual time for italki lessons.

The second test used in the study was the Oral Proficiency Interview by Computer® (OPIc)[6] created by Language Testing International (LTI). LTI is the exclusive licensee of the American Council on the Teaching of Foreign languages (ACTFL). The online test is recorded and the recording of the test is reviewed and evaluated by independent raters.

Table 2. OPIc Ratings

UR	Un-Ratable	AL	Advanced Low
NL	Novice Low	AM	Advanced Mid
NM	Novice Mid	AH	Advanced High
NH	Novice High	S	Superior
IL	Intermediate Low		
IM	Intermediate Mid		
IH	Intermediate High		

The specific definition of the levels is presented on the company's webpage[7].

---

3. Sample Description

The entire sample selection process is graphically represented in the Appendix, Figure A1. E-mail messages were sent out to italki clients with an invitation to participate in the research study. If they accepted the invitation they were asked to complete the online Entry Survey with some demographic questions and questions about their knowledge of Spanish. In all 1631 people viewed the invitation page and of those 388 successfully completed the Entry Survey. This was the initial pool of respondents in the study.

Initial Pool (N=388)

The initial pool of potential participants consisted of people from the US (N=222) and outside the US (N=166) and almost half of them (48.8%) were female. The mean age was 36.3 years and they were well educated: 37% had a graduate (MA, PhD) degree and 45% had a BA or equivalent college degree. Only 8.5% had just High School or less and about 9% has some college but no degree. Most of the people were employed, either full time or part-time (72%), 4% were unemployed, 11% were students, 6% retired and 8% had other employment.

For 75% of the initial pool English was their native language and the remainder (25%) included 29 other languages: Arabic, Bengali, Chinese, Croatian, Dutch, Filipino, Finnish, French, German, Greek, Hebrew, Hindi, Hungarian, Indonesian, Italian, Japanese, Korean, Macedonian, Mandarin, Mongolian, Polish, Portuguese, Russian, Slovak, Swedish, Telugu, Turkish, Urdu, Vietnamese.
Almost 77% described themselves as Beginner users or Never Studied Spanish. About 14% of the respondents' spouse, partner, or close friends spoke Spanish. A very small proportion (5%) of their parents, grandparents, or great-grandparents spoke Spanish.

The primary reason for studying Spanish was personal interest (63%), followed by travel (18%), business or work (15%), etc.
More than 60% of the initial pool had studied a foreign language before (mostly at school or college).

Pool of Eligible Participants (N=284)

From the Initial Pool (N=388) we excluded the following ineligible participants:

- People who were younger than 18 years of age.

- People with advanced or fluent Spanish.

- People who lived in a Spanish speaking country.

Altogether 104 people were ineligible for this study and the final pool of eligible participants for sample selection was N=284. Of them 163 were from the US and the rest (N=121) were from outside the US.

The pool of eligible potential participants had a mean age of 36 years, from 18 years old to 73 years old, with 49.6% female users. The pool of eligible users was very well educated with only about 5% with just a High School diploma or less and 38% had a masters or PhD. About 46% had a BA or equivalent. About 74% were employed full time or part time, 9% were students, and 5% were unemployed. For 78% of them, English was their native language and 56% of the pool knew at least one foreign language.

Initial Random Sample (N=125)

The research design suggested an initial sample size of N=150 based on our previous studies' effect size results and drop-out rate. The people in the initial sample were randomly selected from the pool of eligible participants. They completed the baseline WebCAPE placement test in Spanish and the oral proficiency OPIc test. Some people (n=25) could not complete the initial tests due to technical problems with OPIc and other reasons.

The initial random sample included 125 people with a mean age of 38 years, from 19 years old to 67 years old, with 52.8% female users. The initial sample participants were from the US (N=78) and outside the US (N=47).

More than half (52.5%) of the initial sample were female, the mean age was 38 years and they were well educated: 43% had a graduate (MA, PhD) degree and 43% had a BA or equivalent college degree. Only 2% had just High School or less and 12% had some college but no degree. Most of the people were employed either full time or part-time (78%), 8% were unemployed, 4% were students, 6% were retired and 8% had other employment.

For 78% of the initial sample, English was their native language and the remainder (22%) included 14 other languages: Chinese, Dutch, Filipino, French, German, Greek, Hungarian, Portuguese, Russian, Swedish, Telugu, Turkish, Urdu, and Vietnamese.

All of them described themselves as Beginner users or Never Studied Spanish. About 9% of the respondents' spouse, partner, or close friends spoke Spanish. A very small proportion (3%) of their parents, grandparents, or great-grandparents spoke Spanish.

The primary reason for studying Spanish was personal interest (63%), followed by travel (18%), business or work (16%), and 3% other reasons.

More than 63% of the initial sample had studied a foreign language before (mostly at school or in college).

Table 3. Initial Random Sample: Age and Gender Distribution (N=125)

Age	Female (N)	Male (N)	Total (N)	Percent
18-20 years old	1	0	1	0.8
21-30 years old	24	22	46	36.8
31-40 years old	20	15	35	28.0
Over 40 years old	21	22	43	34.4
Total	**66**	**59**	**125**	**100.0**

The written proficiency of the initial study sample was as follows:

Table 4. Initial WebCAPE Semester Placement (N=125)

College Semester	People (N)	Percent
First	66	52.8
Second	34	27.2
Third	25	20.0
Fourth+	0	0
Total	**125**	**100.0**

The majority (53%) of the participants were evaluated as novice/beginner users of Spanish and they were placed in First Semester of Spanish. About 27% of the participants were placed in

Second Semester and 19% in Third Semester of Spanish. The mean WebCAPE score was 237 (std[8]=121), corresponding to First college semester of Spanish.

The oral proficiency of the initial sample was as follows:

Table 5. Initial Oral Proficiency (OPIc) (N=125)

Proficiency level	People (N)	Percent
0 Un-Ratable	9	7.2
1 Novice Low	29	23.2
2 Novice Mid	34	27.2
3 Novice High	33	26.4
4 Intermediate Low	12	9.6
5 Intermediate Mid	7	5.6
6. Intermediate High	1	0.8
Total	**125**	**100.0**

The majority of the participants (84%) scored at the Novice level of oral proficiency. A handful of people (N=20) scored at the intermediate level. The unrated cases were basically people who cannot speak any Spanish and kept silent or spoke English on their oral test.

---

[8] Standard Deviation.

Final Study Sample (N=102)

The work on the study started in August 2017 and ended in November 2017 with about 8 weeks dedicated to study with italki with time for tests in the beginning and at the end of the study. During the study, the Research Team sent weekly e-mail reminders to the participants with information detailing the amount of time they had used italki each week.

At the end of the study, we reviewed the time use of the participants. For the written proficiency test (WebCAPE) based on our previous studies (Vesselinov & Grego, 2012, 2016) the threshold was established at about two hours of study. For the oral proficiency test the threshold was established at about 9 hours of study. In other words, participants with at least two hours of study could take the written test and complete the study. If they had less than 9 hours of study, they were not eligible to take the oral test. People with 9 or more hours of study could take both written and oral tests.

Based on these criteria from the initial sample the following people were excluded:
- People who did not satisfy the study time requirements.
- People who did not take the final tests.

All participants were instructed at the beginning of the study that they were allowed to use only italki services to study Spanish for the duration of the study. Italki users study with a tutor or teacher and in many cases their study plan included the use of other tools. The restriction for this study was that the participants cannot use other language tools independently of their tutors or teachers' recommendations.

Nine people (7%) had less than two hours of study and they were excluded from the study. Altogether 102 people took the final written test and 81 people took the final oral test. The mean study time for the final study sample was about 13 hours.

The final study sample for written proficiency consisted of 102 people with at least two hours or more of italki use and valid initial and final WebCAPE tests. The final subsample for oral proficiency was part of the 102-person sample and consisted of 81 people with about 9 hours of study or more and valid initial and final OPIc tests. The mean study time for the oral test sample was about 16 hours.

The final study sample (N=102) had a mean age of 39.5 years, from 19 years old to 67 years old, with 53% female users. The users were very well educated with 44% holding a graduate degree and 44% with a BA or equivalent, one person with less than HS, and 11% with some college but no degree. About 78% of them were employed full time or part time, 4% were students, 4% unemployed, 6% retired and 8% had other employment.

For 80% of the participants, English was their native language and the rest included: Chinese, Dutch, Filipino, French, German, Greek, Hungarian, Nepalese, Russian, Swedish, Turkish, Urdu, and Vietnamese. About 58% of the sample knew at least one other foreign language (not Spanish).

About 8% of the respondents' spouse, partner, or close friends spoke Spanish. A small portion (2%) of their parents, grandparents, or great-grandparents spoke Spanish.

The primary reason for studying Spanish was personal interest (65%), followed by travel (18%), business or work (14%), and other reasons (3%).

Table 6. Final Study Sample: Age and Gender Distribution (N=102)

Age	Female (N)	Male (N)	Total (N)	Total (%)
18 to 20 years old	1	0	1	1.0
21-30 years old	18	14	32	31.4
31-40 years old	17	12	29	28.4
Over 40 years old	18	22	40	39.2
Total	**54**	**48**	**102**	**100.0**

People from the final sample used different devices to study Spanish with italki. The majority of them (92%) used a desktop/laptop computer. Other devices were also used (categories are not mutually exclusive): tablet (11%), Android smartphone (10%), Apple iPhone (18%) and other devices (9%).

Final Study Sample vs Not Completed

From the initial random sample (N=125) 23 people (18.4%) did not complete the study for different reasons: people who did not satisfy the study time requirements or people who did not take the final tests.

We compared the two groups, the final sample of 102 people and the 23 people who did not complete the study by gender, age group, education, employment status, initial knowledge of Spanish (initial WebCAPE score and OPIc) and reason for studying Spanish. There were no statistically significant differences (p=.01) which means that people who did not complete the study were not very different from the ones that did, and their exclusion did not introduce an evident bias.

We also compared the sample composition for the US versus all other countries. There were no statistically significant differences (p=.01) between the two subsamples on gender, age groups, education, employment status, initial knowledge of Spanish (initial WebCAPE score and OPIc) and reason for studying Spanish.

4. Language Improvement and Study Time

Study Time

The study time was measured objectively by the actual purchased lesson time with an italki tutor or teacher. The average study time for the final study sample (N=102) was about 13 hours, or one hour and 40 minutes a week. The average study time for the oral test eligible participants (N=81) was about 16 hours, or two hours a week.

WebCAPE Test Results

All participants took an initial WebCAPE test before the start of the study and then again at the end of the study. The progress or improvement was measured as the difference between the final test score and the initial one.

Table 7. Language Improvement Written Proficiency (N=102)

WebCAPE Test Points

Statistics	Initial WebCAPE	Final WebCAPE	Improvement (Final-Initial)
Mean (std)	236.9 (123.8)	336.8 (163.7)	99.9 (149.5)
Median	268.5	334.5	74.5
95% Confidence Interval[9]	212.6 – 261.2	304.6 – 368.9	70.5 – 129.2

The overall average improvement of about 100 points WebCAPE test points was statistically significant with a 95% confidence interval from 71 to 129 points. This also means that the improvement in the written proficiency for the final sample was statistically significant (at 5% error). Overall 78% of all participants improved their written proficiency (increased their WebCAPE score) with a 95% confidence interval[10] of 68% to 85%.

There were 23 cases (22%) where study participants did not improve their WebCAPE result or had a lower result at the end of the study compared to their initial level.

There are two plausible explanations for this fact. First, some of them were more advanced learners of Spanish and gaining points at this higher level is generally more difficult and requires more time. Second, some of them studied irregularly with more effort and study time in the beginning of the study and less towards the end of the study. These users were not excluded from the sample, so the results can be generalized for all types of users.

---

[9] We also bootstrapped (N=10,000) the confidence intervals but the results remained practically the same.
[10] 95% CI with Agresti-Coull correction (Agresti & Coull, 1998).

College Semester Placement

Progress can be measured by movement from one semester level to a higher semester level and the results are presented below.

Table 8A. WebCAPE Semester Placement (N=102)

College Semester	Initial Test		Final Test	
	People (N)	%	People (N)	%
First	54	52.9	30	29.4
Second	28	27.5	29	28.4
Third	20	19.6	25	24.5
Fourth+			18	17.6
Total	102	100	102	100

People at First Semester level decreased from 53% to 29% and the proportion of people in Second and Third Semester level increased as well. There was nobody in 4th semester level in the beginning compared to about 18% at the end of the study.

Another way of presenting the results is direct comparison in the following table.

Table 8B. Direct Comparison of WebCAPE Semester Placement (N=102)

Semester		FINAL				
		1	2	3	4	Total
INITIAL	1	27	15	9	3	54
	2	2	9	10	7	28
	3	1	5	6	8	20
	Total	30	29	25	18	102

Oral Proficiency

The oral proficiency results are presented below.

Table 9. Oral Proficiency Placement (N=81)

Level	Initial Test		Final Test	
	People (N)	%	People (N)	%
0 Un-Ratable	6	5.9	0	0
1 Novice Low	26	25.5	3	3.7
2 Novice Mid	27	26.5	10	12.3
3 Novice High	25	24.5	31	38.3
4 Intermediate Low	11	10.8	21	25.9
5 Intermediate Mid	6	5.9	15	17.3
6 Intermediate High	1	1.0	2	2.5
Total	**81**	**100**	**81**	**100**

In the oral proficiency area, the results are strong. At the beginning of the study the truly novice users (Un-Ratable and Novice Low) were almost 32% of the sample while at the end their numbers decreased to about 4%. This is a completely different level of oral proficiency.

5. Main Results

Written Proficiency

Table 10. Written Proficiency Improvement (N=102)

Level (Semester Change)	Improved	
	People (N)	%
-1 Negative change/Down	8	7.9
0 Same/No Change	42	41.2
1 One Semester Up	33	32.4
2 Two Semesters Up	16	15.7
3 Three Semesters Up	3	2.9
Total	**102**	**100**

Overall about 51% of the participants moved up at least one semester. About 32% moved up one semester, 16% moved up two semesters and 3% moved up three semesters. About 41% stayed in the same semester they started in and 8 people moved down a semester. The average study time for the people who decreased their score (n=8) was about 5 hours while the average time for the people who improved at least by one semester was more than 17 hours. People who stayed in the same semester on average studied about 8 hours.

As the results indicate, people who had invested the lowest amount of effort and study time were unsurprisingly the ones who did not improve or worsened their written proficiency measured by semester level.

The problem with this measure is that first, it does not account for the effort (study time) and second, moving up a semester is dependent on the exact initial level. For example, if a person has initially 269 test points (First semester), only a 1-point gain is needed to move to Second semester. Another person can start with 10 points as initial level (First semester), then gain 200 points and the new level (210 points) is still First Semester.

The main efficacy measures are presented below.

Table 11. Main Result. Efficacy of italki (N=102)

Statistics	Efficacy Improvement per one hour of study WebCAPE Test Points	Time to cover the requirements for one semester of college Spanish Hours
Mean	14	19[11]
95% Confidence Interval	7 − 21[12]	13 − 38[13]

On average italki users will gain 14 WebCAPE test points per one hour of study with a 95% confidence interval of 7 to 21 test points per hour.

The main measure of italki efficacy is the improvement per one hour of study. In addition, if we divide the required cut-off point (270) for WebCAPE Second Semester placement by the efficacy mean we can construct a new measure representing the time needed to cover the requirements for one college semester of Spanish. This is the one measure of efficacy that is easy to understand and given the nature of the WebCAPE placement test, can be used for comparison with other language apps.

In other words, on average, italki users will need 19 study hours during a period of two months to cover the requirements for one college semester of Spanish with transformed lower and upper limits of 13 hours to 38 hours of study.

---

[11] The threshold of 270 points divided by the mean efficacy (14 points).
[12] We also bootstrapped (N=10,000) the confidence interval but the result remained practically the same.
[13] The threshold of 270 points divided by the lower limit (7) and the upper limit (21) of the 95% CI.

Efficacy and the Initial Level of Knowledge of Spanish

Table 12. Efficacy by Initial Level of Language Ability (N=102)

Initial Level College Semester	People N	Efficacy Mean (std)
First	54	20.9 (48.1)
Second	28	7.6 (18.7)
Third	20	6.0 (15.4)
Total	**102**	**14.3 (37.5)**

The overall efficacy is 14 WebCAPE points per one hour of study, but novice users of Spanish managed a bigger gain of 21 points per hour of study. For the second and third semester levels the improvement was more modest at about 6-7 points per hour.

Factors for Written Proficiency

We investigated the impact of some factors on the efficacy measure, namely age, gender, education, employment, reason for studying Spanish, presence of people around the participant who spoke Spanish (spouse, friend, parents, grandparents, etc.), native language, and knowing another foreign language.

None of the available factors had a statistically significant effect on the efficacy. In some instances, the number of cases by subgroups was too low to expect enough statistical power for the test of hypotheses.

Italki users could chose freely a teacher or a tutor for their study. About 15% of the participants chose only tutors for their study, and about 58% chose only teachers, with the rest using a mix of teachers and tutors. People who used only teachers for their study had slightly higher improvement rate (80% vs 74%), but the difference was not statistically significant.

This result can be interpreted as a positive finding because it means that italki works similarly well for people with different gender, age, employment status, native language, etc.

Oral Proficiency

Table 13. Oral Proficiency Improvement (Initial to Final Level)

Progress in Levels	Progress	
	People (N)	%
0 Same Level	23	28.4
1 One Level Up	40	49.4
2 Two Levels Up	15	18.5
3 Three Levels Up	2	2.5
4 Four Levels Up	1	1.2
Total	**81**	**100**

Overall 71.6% of the participants improved their oral proficiency by at least one level. The 95% Confidence Interval[14] is 61% to 80%. Almost half (49.4%) improved by one level, a fifth (18.5%) improved by two levels and 3 people (3.7%) improved by more than two levels.

The average study time for people who did not improve (same level) was 13.7 hours; the same time for people who improved one level was 14.5 hours, and people who improved more than one level studied on average more than 20 hours.

Not surprisingly people who did not improve their oral proficiency were the people with the lowest amount of study time.

Given the study time requirements for the oral test subsample it looks as though 16 hours within two months of study, or at least two hours a week, was sufficient time for most people to achieve significant progress in their oral proficiency. The majority of people improved by one or two levels which seems like a reasonable expectation. The ACTFL requirement to have 2-3 months between test-retest looks appropriate. This study simply confirmed the additional requirement that in practice this means two months of study with at least two hours of study a week.

---

[14] 95% CI with Agresti-Coull correction (Agresti & Coull, 1998).

Table 14. Oral Proficiency Improvement by Initial Level

Initial Level	People	Improved (One or more levels)
	N	% Improved
0 Un-Ratable	2	100
1 Novice Low	17	88.2
2 Novice Mid	23	87.0
3 Novice High	23	65.2
4 Intermediate Low	11	54.5
5 Intermediate Mid	4	0
6 Intermediate High	1	0
Total	**81**	**71.6**

People who started at rock bottom as Un-Ratable and people at Novice level improved the most (65% to 100%). People at higher intermediate levels improved less than novice. The more advanced levels had very few cases to make meaningful conclusions.

Factors for Oral Proficiency

We investigated the impact of some quantifiable factors on the oral proficiency measure, namely age, gender, education, employment, reason for studying Spanish, presence of people around the participant who spoke Spanish (spouse, friend, parents, grandparents, etc.), native language, using teachers/tutors and knowing another foreign language. None of the available factors had a statistically significant effect on the oral proficiency.

As with the factors for written proficiency, this result can be interpreted as a positive finding because it means that italki works the same way for people with different gender, age, employment status, native language, device used etc. regarding their oral proficiency.

The Best Results: People Who Gained in Both Written and Oral Proficiency

Overall 78% of the participants improved their written proficiency (gain in WebCAPE points) and 72% improved their oral proficiency by at least one level. The relationship between the two is presented below.

Table 17. Oral and Written Proficiency Improvement (N=81)

Improved Written Proficiency (Gain in WebCAPE score)	Improved Oral Proficiency (One or more levels)	
	No	Yes
No	5 (6.2%)	12 (14.8%)
Yes	18 (22.2%)	46 (56.8%)

The interpretation of this relationship is as follows: almost all people (93.8%) who took both the written and oral tests twice gained either in oral or written proficiency. Almost 57% of the people improved both their written and oral proficiency. About 22% improved only their written proficiency and about 15% improved only their oral proficiency.

6. User Satisfaction

After the study the participants were asked for their opinion about italki, specifically how easy it was to use, how helpful, enjoyable, and satisfactory.

Table 18. User Satisfaction (N=89)

			Percent
Do you agree with the following statement?	**Strongly Disagree/ Disagree**	**Neither Disagree nor Agree**	**Agree/ Strongly Agree**
"italki was easy to use"	6.7	4.5	88.8
"italki was helpful in studying Spanish"	1.1	1.1	97.8
"I enjoyed learning Spanish with italki"	0	2.2	97.8
"I am satisfied with italki"	2.2	5.6	92.1

After two months of study, the vast majority of users (89% to 98%) agreed with the positive statements that: italki was easy to use, helpful, they enjoyed learning with italki and they were satisfied with it.

In the exit survey a special question was included: "How likely are you to recommend italki 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). 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 the higher the score the better the indicator for the company. From our exit survey the "Promoters" were 55.6% and the "Detractors" were 2.8% and "Passives" were 41.7%. The italki NPS was +52.8 which is a quite substantive result.

Almost all respondents (97.8%) in the exit survey declared that they will continue to use italki after the study ends.

7. Limitations of the Study

On the positive side, neither of the two language tests was tailored to any specific learning tool, including italki. On the one hand, some participants in the study complained that the tests sometimes contained words or expressions that were not part of their regular course with italki. On the other hand, people insisted that they had learned a lot more than the tests asked for. The tests are valuable as an independent tool for evaluation which allows us to compare efficacy across different apps, however they do not provide a complete measure of the exact progress of users.

There are some limitations of the study, mostly related to the instruments and technological limitations. The WebCAPE written test measures the progress of beginner/novice users of Spanish well, but it is not suitable to measure the progress of very advanced users. Also, more study time is required for advanced users because it takes longer to achieve mastery of higher language levels. Participants who started at rock bottom as true beginners (WebCAPE score close to 0) gained much faster per study hour than people who started at the level of a second or third college semester of Spanish.

The Research team sent e-mail messages every week with individualized information about the study time for the previous week. This seemed to stimulate the study process. In normal settings when people work individually on their studies, this stimulation is not available. Many participants suggested adding a clock and time tracker to the software, so they can be aware of how much time they spend studying. The average study time was a little less than two hours of study a week but for some of the participants this was too much. The results of the study should be valid in a setting where the users study regularly for about two hours a week for two months.

The study results could be generalized for studying Spanish with italki. For other languages, more studies are necessary to confirm these findings, although there is no obvious reason in the literature that the results should be markedly different.

There are few other studies with a direct objective measure of efficacy available to compare with the results of this study. More help is needed from users, investors, and analysts to require the creators of language learning apps to provide independent efficacy measures.

8. Conclusion

The italki efficacy study is based on a final random sample of 102 people, 18 years of age or older.

Overall 78% of the participants improved their written proficiency (gained WebCAPE points). The main goal of measuring the efficacy of italki was achieved with this study. The results show that, on average, one hour of study with italki alone leads to an improvement of 14 points on the college placement test WebCAPE. There is a lot of variability in the efficacy and the 95% confidence interval is between 7 and 21 points per study hour.

In other words, an italki user would need on average 19 hours to complete the requirements for one college semester of Spanish. The transformed upper and lower limits are between 13 and 38 hours of study.

The main factor for progress is the initial level of language knowledge of the participants. The novice/beginner users (First semester) gain faster, with an average of 21 points per one hour of study and the more advanced users (Second and Third semester) gain on average 6-7 points per one hour of study.

Using italki for two months (about two hours of study a week) improved the oral proficiency of 72% of the users. The 95% Confidence Interval is 61% to 80%.

There are only a handful of known studies with direct objective measures of efficacy of language learning apps. Among them, the efficacy of the written proficiency of italki is the best so far. The creators of other language apps should be encouraged to provide efficacy measures so users and investors can make more educated choices.

Cited Literature

1. Agresti, A., B. Coull, 1998, Approximation is better than "exact" for interval estimation of binomial proportions, *American Statistician*, 52, pp. 119–126.

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

3. Vesselinov, R. and J. Grego, 2017, **Hello English Efficacy Study**. http://centralsquarefoundation.org/grant/hello-english-efficacy-study/ , or http://comparelanguageapps.com/documentation/HelloEnglish_2017Study.pdf

4. Vesselinov, R. and J. Grego, 2016b, **The Babbel Efficacy Study**. http://comparelanguageapps.com/documentation/Babbel2016study.pdf , or http://press.babbel.com/en/releases/2016-09-29-Spanish-Study.html

5. Vesselinov, R. and J. Grego, 2016a, **The Busuu Efficacy Study**. http://comparelanguageapps.com/documentation/The_busuu_Study2016.pdf , or https://blog.busuu.com/wp-content/uploads/2016/05/The_busuu_Study2016.pdf

6. Vesselinov, R. and J. Grego, 2015, **Efficacy of New Language App**, http://comparelanguageapps.com/documentation/LA_Final_Report.pdf .

7. Vesselinov, R. and J. Grego, 2012, **Duolingo Effectiveness Study**. http://comparelanguageapps.com/documentation/DuolingoReport_Final.pdf , or http://static.duolingo.com/s3/DuolingoReport_Final.pdf

8. Vesselinov, R., J. Grego, B. Habing, A. Lutz, 2009b, **Measuring the Attitude and Motivation of Rosetta Stone® Users.** http://comparelanguageapps.com/documentation/MeasuringTheAttitudeandMotivationofRSUsers.pdf

9. Vesselinov, R., J. Grego, B. Habing, A. Lutz, 2009a, **Comparative Analysis of Motivation of Different Language Learning Software.** http://comparelanguageapps.com/documentation/ComparativeMotivationAnalysisofDifferentLanguageSoftware.pdf

10. Vesselinov, R., 2009, **Measuring the Effectiveness of Rosetta Stone®.** http://comparelanguageapps.com/documentation/MeasuringTheAttitudeandMotivationofRSUsers.pdf , or http://resources.rosettastone.com/CDN/us/pdfs/Measuring_the_Effectiveness_RS-5.pdf .

9. Appendix

Figure A1. Sample Selection Tree
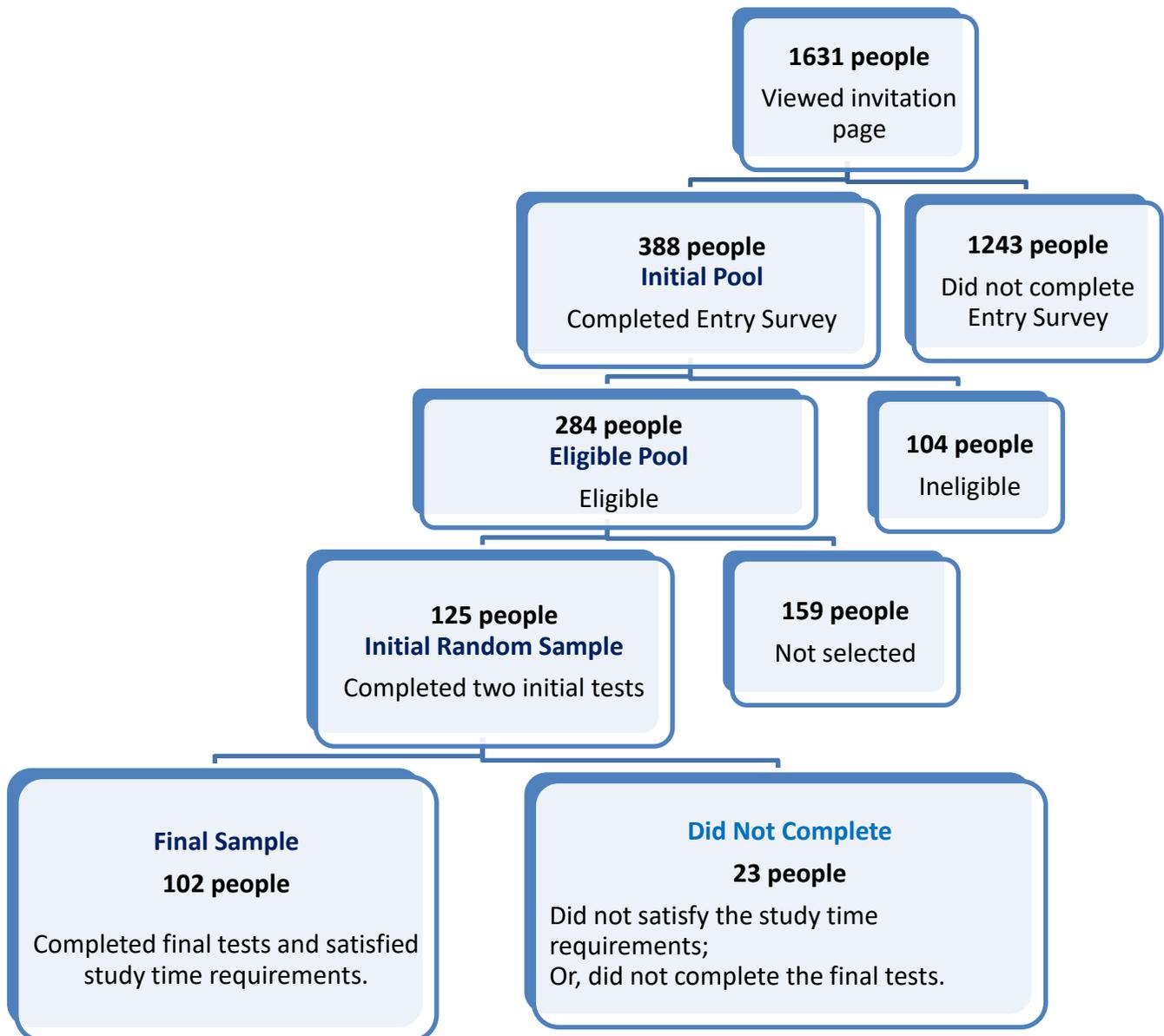
Table A1. Geographic Distribution Countries (Number of people)

Code	State	Country Code	Initial Pool	Eligible Pool	Initial Sample	Final Sample
38	United Arab Emirates	AE	1	1		
2	Austria	AT	2	2		
1	Australia	AU	19	12	4	4
3	Belgium	BE	1	1	1	
5	Brazil	BR	7	3	1	
4	Belize	BZ	1	1		
6	Canada	CA	10	7	5	5
34	Switzerland	CH	2	1		
7	China	CN	5	5		
11	Germany	DE	7	6	2	2
8	Estonia	EE	1	1	1	1
9	Finland	FI	1			
10	France	FR	8	6	2	1
39	United Kingdom	GB	42	33	17	13
12	Greece	GR	2	2		
13	Hungary	HU	1	1	1	1
15	Indonesia	ID	2	2		
16	Ireland	IE	4	3	1	
14	India	IN	4	4		
17	Italy	IT	4	4		
18	Jamaica	JM	2	2	1	1
19	Japan	JP	2	2	1	1
20	Kazakhstan	KZ	2	1	1	1
21	Latvia	LV	1	1		
23	Mali	ML	1	1	1	1
22	Malaysia	MY	1	1		
24	Netherlands	NL	1	1	1	1
26	Norway	NO	1	1	1	
25	New Zealand	NZ	1			
27	Philippines	PH	1	1		
28	Poland	PL	2	1	1	
29	Portugal	PT	1	1		
30	Russia	RU	6	5	3	3
33	Sweden	SE	1	1	1	1
31	Singapore	SG	3	3	1	1

Table A1. Continued and end.Geographic

Code	State	Country Code	Initial Pool	Eligible Pool	Initial Sample	Final Sample
32	Slovakia	SK	1	1		
35	Thailand	TH	1	1		
36	Turkey	TR	1	1		
37	Ukraine	UA	1	1		
40	USA	US	222	163	78	65
	Spanish-Speaking Countries					
41	Chile	CL	1			
42	Colombia	CO	2			
43	Costa Rica	CR	1			
44	Dominican Republic	DO	1			
45	Ecuador	EC	1			
46	Honduras	HN	1			
47	Mexico	MX	2			
48	Paraguay	PY	1			
49	Spain	ES	2			
Total			**388**	**284**	**125**	**102**

Table A2. Geographic Distribution: US States (Number of people)

Code	State	ST	Initial Pool	Eligible Pool	Initial Sample	Final Sample
2	Alaska	AK				
1	Alabama	AL	1	1		
4	Arkansas	AR				
3	Arizona	AZ	1	1	1	1
5	California	CA	38	31	18	15
6	Colorado	CO	8	4	2	1
7	Connecticut	CT	1	1	1	1
9	Delaware	DE	3	2	1	1
10	Florida	FL	12	11	1	1
11	Georgia	GA	4	3	1	1
15	Iowa	IA	2	2	2	2
12	Idaho	ID	1	1		
13	Illinois	IL	12	9	8	6
14	Indiana	IN	1			
16	Kansas	KS				
17	Kentucky	KY	3	3	1	1
18	Louisiana	LA	1			
21	Massachusetts	MA	8	5	2	2
20	Maryland	MD	6	4	3	2
19	Maine	ME	1	1	1	1
22	Michigan	MI	3	2	1	1
23	Minnesota	MN	3	1		
25	Missouri	MO	1	1		
24	Mississippi	MS				
26	Montana	MT				
33	North Carolina	NC	5	2	2	2
34	North Dakota	ND				
27	Nebraska	NE				
29	New Hampshire	NH	1			
30	New Jersey	NJ	10	9	2	1
31	New Mexico	NM				
28	Nevada	NV				
32	New York	NY	25	18	9	9
35	Ohio	OH	2	1		

Table A2 Continued and end.

Code	State	ST	Initial Pool	Eligible Pool	Initial Sample	Final Sample
36	Oklahoma	OK				
37	Oregon	OR	4	2	1	1
38	Pennsylvania	PA	5	4	3	3
39	Rhode Island	RI				
40	South Carolina	SC	1	1		
41	South Dakota	SD				
42	Tennessee	TN	3	1	1	
43	Texas	TX	13	11	3	2
44	Utah	UT	2	2	1	1
46	Virginia	VA	9	5	3	3
45	Vermont	VT				
47	Washington	WA	4	4	3	3
49	Wisconsin	WI	2	2	1	
48	West Virginia	WV				
50	Wyoming	WY				
	District of Columbia	DC	1			
	Unknown state (but US)		25	18	6	4
Total		**USA**	**222**	**163**	**78**	**65**

****************** **LAST PAGE OF THE italki FINAL REPORT** ******************