

LingQ Efficacy Study

2023

RESEARCH TEAM

- Roumen Vesselinov, PhD¹
- John Grego, PhD
- Mila Tasseva-Kurktchieva, PhD

¹Email for correspondence: study@comparlanguageapps.com

EXECUTIVE SUMMARY

The Research Team independently conducted this study from January to April 2023. A random representative sample of 101 novice Spanish learners participated in the study. The participants took one set of Spanish language vocabulary/grammar and oral proficiency tests in the beginning of the study, then studied Spanish with LingQ for two months and took the same tests again.

MAIN RESULTS

Overall Vocabulary/Grammar Proficiency Gain:

- 86.1% of the participants improved their vocabulary/grammar proficiency².
- Novice users need on average 13 hours of study in a two-month period to cover the requirements for first college semester of Spanish placement.

Overall Oral Proficiency Gain:

- 89.1% of all study participants improved their oral proficiency³.

Oral Proficiency Gains for Participants with at least 8 Hours of Study:

- 87.5% of the participants improved their oral proficiency³.
- 77.1% of the participants improved their TNT-ACTFL estimation⁴ level.
- 77.1% of the participants improved their TNT-CEFR estimation⁵ level.

Efficacy:

- On average participants gained 23 vocabulary/grammar proficiency WebCAPE test points per one hour of study.
- On average participants gained 0.06 points of the TNT oral proficiency test per one hour of study.

User Satisfaction:

- The majority of participants thought that LingQ was easy to use (70%), helpful (86%), enjoyable (71%), and satisfying (74%).
- LingQ received a positive Net Promoter Score of +4.2 from the participants.
- Participants' motivation was very high with average level of 71%.

2 Based on the college placement test WebCAPE.

3 Based on TrueNorth (TNT) oral proficiency test.

4 TNT estimation of American Council for Teaching Foreign Languages (ACTFL) levels.

5 TNT estimation of Common European Framework of Reference for Languages (CEFR) levels.

Contents

EXECUTIVE SUMMARY	1
MAIN RESULTS	1
INTRODUCTION	3
RESEARCH DESIGN	3
Study Instruments	3
STUDY SAMPLE	4
Sample Description	6
Initial Language Tests	6
Motivation	7
Second Language Profile	8
Study Time	9
LANGUAGE IMPROVEMENT	9
Vocabulary/Grammar Proficiency Results	9
Oral Proficiency Results	11
EFFICACY	12
Vocabulary/Grammar Proficiency Efficacy	12
Oral Proficiency Efficacy	12
FACTORS FOR EFFICACY	13
USER SATISFACTION	14
LIMITATIONS OF THE STUDY	15
CONCLUSION	15
CITED LITERATURE	16
APPENDIX	17

INTRODUCTION

This is the 15th study of the Research Team testing the efficacy of different language learning apps (Vesselinov et al., 2009-2021). Our previous studies⁶ evaluated Rosetta Stone, Duolingo, Busuu, Babbel, Mango Languages, Pimsleur, Hello English, italki, and Language Zen. The statistical design and methodology are comparable for all studies⁷.

This study was designed to evaluate the efficacy of LingQ.⁸ According to their website, LingQ was launched in late 2007 by Steve Kaufmann and his son Mark to develop better tools for language learners.

The cost for this study was covered by LingQ, but the data collection and the analysis were carried out independently by the Research Team. The language tests used in the study were designed and developed by an external independent testing company⁹.

RESEARCH DESIGN

The random sample for this study was drawn from existing LingQ users from around the world.

There were some requirements for the potential participants, who had to:

- be willing to study Spanish using only LingQ for two months.
- take two sets of vocabulary/grammar and oral proficiency language tests.
- be at least 18 years of age.
- be novice or beginner learners of Spanish.

Sample Size and Power Analysis

We based our power and sample calculations on the typical results from our previous studies. We designed the study to test the hypothesis of proportions of 0.6 or more for statistical significance, corresponding to the lowest expected proportion of participants that improve their oral proficiency. We considered a 5% statistical significance level ($\alpha=0.05$)

⁶ <http://comparelanguageapps.com/>

⁷ Except 2009 Rosetta Stone study when the study time was self-reported due to technological limitations.

⁸ www.lingq.com

⁹ <https://emmersion.ai>

and at least 80% statistical power. Under these assumptions we would need a sample of 40 people or more to test for statistical significance of improvement. Our initial sample size was set to 200 in order to reflect possible dropouts from the study and reduce the impact of outliers.

Spanish language was selected for this study because it is one of the most popular languages studied by users of language learning apps, and for comparability with previous research on Spanish for other language learning apps (Vesselinov et al., 2009-2021)¹⁰. The study lasted approximately 8 weeks and was conducted between January and April 2023. Participants who successfully completed the study were given free 6 months Premium LingQ subscription for two. No other incentives were offered to the participants.

Study Instruments

Test 1. WebCAPE: Vocabulary/Grammar Proficiency

We used a college placement test called the Web based Computer Adaptive Placement Exam (WebCAPE). This is an established university placement test and is offered in English, Spanish, French, German, Russian and Chinese. It was created by Brigham Young University and is hosted by Emmersion Learning¹¹.

The WebCAPE test has a very high validity correlation coefficient (0.91) and a very high reliability (test-retest) value of 0.81¹². 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 group levels (college semesters; see Table 1).

Table 1. Spanish WebCAPE Test Cut-off Points

WebCAPE Test Points	College Semester Placement
Below 270 ¹³	Semester 1
270-345	Semester 2
346-428	Semester 3
Above 428	Semester 4+

¹⁰ Except Hello English study in 2017.

¹¹ Currently at <https://emmersion.ai/>.

¹² Personal correspondence with Dr. Jerry Larson, Professor of Spanish Pedagogy at Brigham Young University.

¹³ same threshold of 270 points was used for all our previous language studies (Vesselinov et al., 2009-2021).

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

$$\text{Efficacy} = \frac{\text{Effect}}{\text{Effort}} = \frac{\text{Improvement of language skills}}{\text{Study time}} =$$

$$\frac{\text{Final-initial WebCAPE test score}}{\text{Hours of study}}$$

Efficacy=Improvement per one hour of study.

A similar efficacy measure will be computed for oral proficiency, using the True North Test score described below.

The efficacy measure includes both the amount of progress made and the amount of effort invested by each study participant. This is a direct and objective measure of efficacy: direct, because it directly includes 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 time recorded on computer servers (instead of self-report).

Test 2. True North Proficiency Test (TNT)¹⁴

This is a newly developed oral proficiency test based on Elicited Imitation (EI) as a testing method in which participants hear an utterance in the target language and are prompted to repeat the utterance as accurately as possible.

a. TNT oral proficiency score.

The TNT test gives an incremental score from 0.0 to 10.0 with 0 being the lowest level and 10 being the highest. TNT also provides estimation of ACTFL and CEFR levels.

b. TNT-ACTFL estimation of oral proficiency.

The American Council for Teaching Foreign Languages (ACTFL) developed a proficiency scale to assess foreign language abilities. This scale includes four main groups¹⁵ (Novice, Intermediate, Advanced, and Superior), with the first three divided into levels. The levels are shown below:

Novice:	1. Low	2. Mid	3. High
Intermediate:	4. Low	5. Mid	6. High
Advanced:	7. Low	8. Mid	9. High
10. Superior (no levels)			

c. TNT-CEFR estimation of oral proficiency.

The Common European Framework of Reference for Languages (CEFR)¹⁶ is designed as a global standard for describing language proficiency. It has six levels, A1-A2 for beginner, B1-B2 for intermediate, and C1-C2 for proficient.

STUDY SAMPLE

This study started in January of 2023 when emails were sent to existing LingQ users with an invitation to participate in a Spanish language study for two months. They were directed to an online survey designed by the Research Team. This survey collected demographic information, and self-evaluation of their language proficiency level. We received complete responses from 1119 people which constituted the initial pool for the study (see Figure 1). From this pool we excluded people who lived in Spanish-speaking countries, were younger than 18 years of age, or had an intermediate or high level of knowledge of Spanish; the remainder formed the eligible pool (N=499) for this study.

We randomly selected 200 people from the eligible pool of participants and 192 of them completed the initial language tests; they constituted our initial random sample (N=192).

The LingQ study continued for approximately two months (8 weeks), starting in January 2023. During the study, the Research Team sent weekly e-mail reminders to the participants with information detailing the amount of time they had used LingQ each week. The final study sample consisted of 101 people who had at least two hours of study and at least one final test. All participants were instructed at the beginning of the study that they could use only LingQ to study Spanish for the duration of the study. In the exit survey no participant reported regularly using additional language apps, tutors, or language courses. Occasional use of translation websites and internet dictionaries was allowed in this study as part of the usual learning process.

Final Study Sample versus Not Completed

¹⁴ Currently at <https://emersion.ai/>

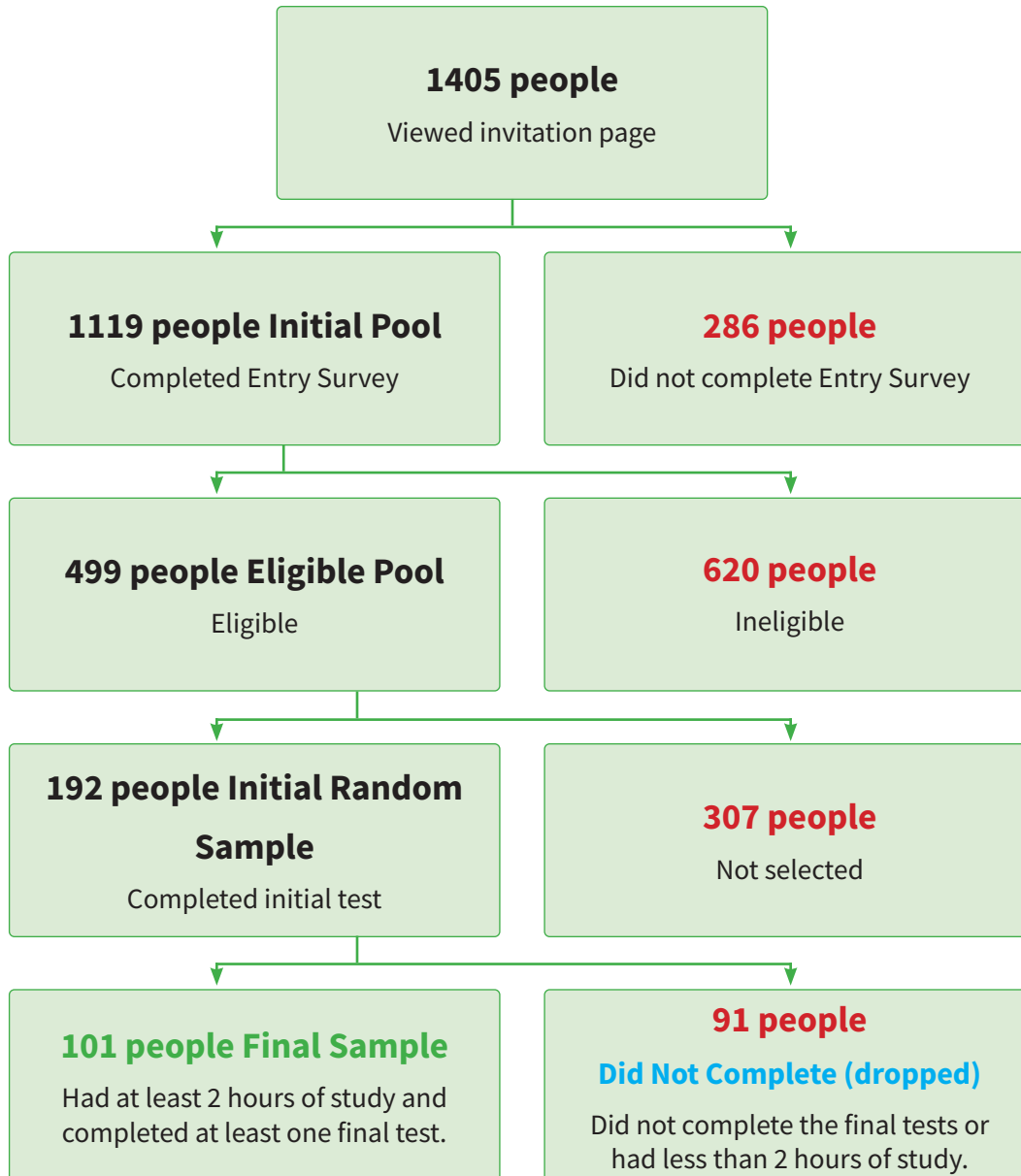
¹⁵ <https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012>

¹⁶ <https://www.coe.int/en/web/common-european-framework-reference-languages>

From the initial random sample (N=192), 91 people (47.4%) did not complete the study for two different reasons: 1) they did not take the final tests, or 2) they studied for less than 2 hours during the two-month study. This dropout rate is about average in this line of research (Vesselinov et al., 2009-2021).

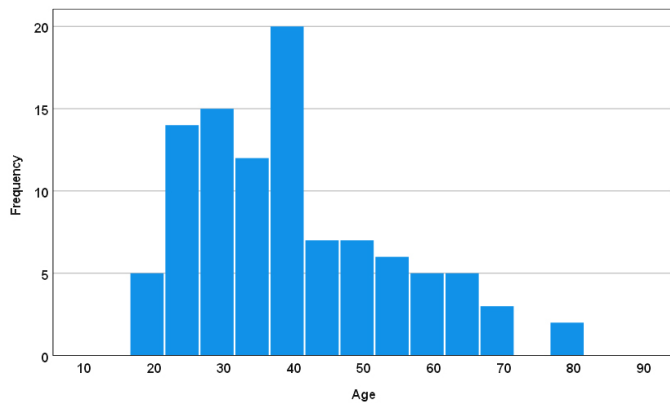
We compared the two groups, the final sample of 101 people and the 91 people who did not complete the study by gender, age, education, employment status, initial knowledge of Spanish and reason for studying Spanish. There were no statistically significant differences (at $p < 0.05$), which means that participants who did not complete the study were not very different from the ones that did.

Figure 1. LingQ Sample Selection Tree



Sample Description

In the final study sample (N=101¹⁷), 62.9% were female (see Table A2 in the Appendix). The age of participants varied from 19 to 79 years of age, with a mean age of 39.8 years. About 20% of the participants had at least some college experience but no degree, about 35% had an undergraduate degree (BA/BS), and 34% had graduate degree. Only 11% of the participants had only a High School diploma or less education.



The initial pool of participants included people from 83 countries (see Appendix, Table A1). The final random sample included representatives from 28 countries with the biggest group from the U.S. (n=43, 42.6%). In the final sample, 57.4% (n=58) listed English as their native language, while other users listed more than 20 other native languages, including Belorussian, Bulgarian, Chinese, Croatian, Flemish, French, German, Greek, Hungarian, Hindi, Italian, Korean, Lithuanian, Malay, Nepali, Norwegian, Polish, Portuguese, Punjabi, Russian, Tamil, Turkish, and Ukrainian.

All participants in the final sample described themselves as beginner or novice Spanish learners. Only 13% of the respondents had a spouse, partner, or close friend who spoke Spanish and one person had parents, grandparents, or great grandparents who spoke Spanish.

About 86% of the final sample had studied a foreign language before (mostly in high school or college). About 29% of the participants had lived in a foreign language country for more than 6 months. The primary reason for studying Spanish among participants was personal interest (63%), followed by travel (22%), business/work (12%), and other reason (3%).

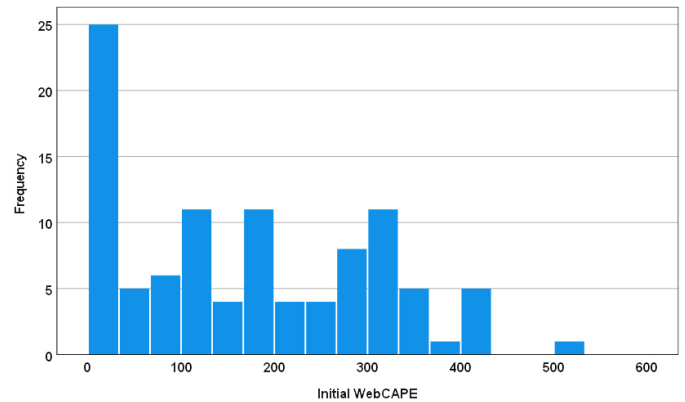
¹⁷ Occasionally participants declined to answer some survey questions, so the number of observations can vary.

Initial Language Tests

Test 1. Vocabulary/Grammar proficiency: WebCAPE

All participants took an initial vocabulary/grammar proficiency test (WebCAPE) and the results are as follows.

Figure 3. Initial WebCAPE Score (N=101)



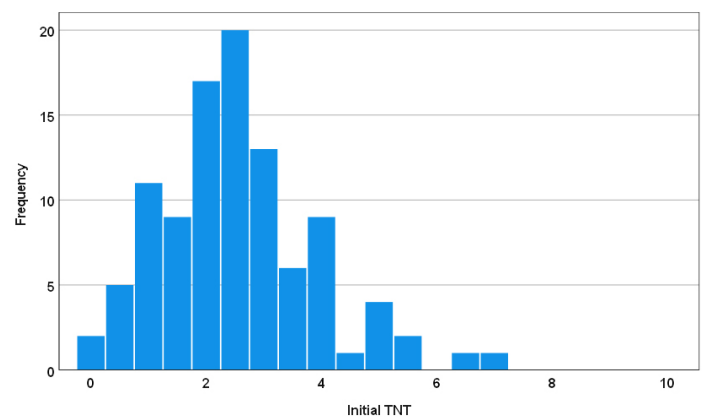
As expected, a large portion of the participants had a very low WebCAPE score. The overall median WebCAPE score was 160 (IQR¹⁸ =275) corresponding to first college semester of Spanish proficiency.

Test 2. Oral Proficiency:

a. TNTScore.

The TNT score can vary from 0.0 to 10.0, and the initial test scores ranged from 0 to 7.1. The overall median value was 2.5 (IQR=1.5) with 25% scoring very close to zero.

Figure 4. Oral Proficiency: Initial TNTScore (N=101)



¹⁸ Interquartile Range= Third Quartile – First Quartile

Motivation

All participants completed a motivation survey in the beginning of the study to evaluate the effect of motivation on efficacy.

We adopted a motivation scale approach based on the second language (L2) motivational self-system (Dörnyei, 2005, 2009), which stems from the concepts of possible selves and self-discrepancy theory. The model proposes that language learners are guided by visions of ‘second language selves’, one which attracts them toward becoming an idealized L2 user (ideal L2 self) and one which motivates them from societal obligation or a fear of failure (ought-to L2 self).

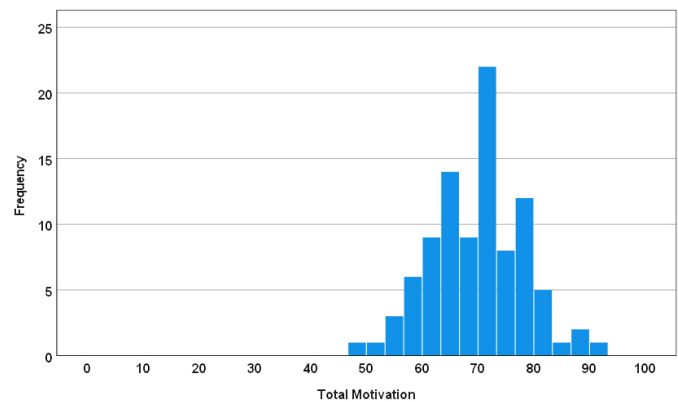
In our study we used 33 question/6 factor version of the L2 Motivational Self System created by Kong et al. (2018). Kong et al. (2018) offer the following descriptions of the motivation scale elements:

1. **Ideal L2 self:** “The ideal L2 self refers to a positive future image of the L2 self. For example, learners who have developed a vivid ideal L2 self are likely to endeavor to learn an L2 by imagining themselves communicating fluently using the L2 in the future.”
2. **Ought-to L2 self:** “(This element) pushes people from societal obligation or a fear of failure.”
3. **International posture:** “It captures a tendency to relate oneself to the international community rather than any specific L2 group. The key characteristics of international posture are described as an interest in global issues or international affairs, a willingness to travel, stay, or work abroad, and a readiness to interact with foreigners or foreign cultures.”
4. **Competitiveness:** “Competitiveness can be described as the desire to excel in comparison to others and contends that a learner constantly compares oneself with one's idealized self-image or with other learners, feels pressured to out-do other students.”
5. **L2 learning Experience or Attitudes:** “L2 learning experience is related to the learners' environment including teachers, peer groups, curriculum, and their attitudes toward L2 learning.”
6. **Learners' Intended Effort or Motivated Behavior in L2 Learning:** This motivation element evaluates how much effort users are determined to make and how hard they are ready to study.

Ninety-five participants completed the motivation survey. As we can see from Table 2, participants' motivation was very high. The scale dimensions were recoded, so the maximum motivation is equal to 100.

The average level of the overall motivation was high (Median=71%). From the motivation elements, the highest average level (80%) belongs to “Learning Attitude” which indicates that the participants were extremely eager to learn a new language. The element “Ought-to-Self” has the lowest level of all (43%) which suggests that the participants were not very afraid of failure, or they were not that susceptible to pressure from societal obligation.

Figure 5. Total Motivation Level (N=95)



As noted above, the average initial level of motivation was very high, and most people were highly motivated overall. The lowest individual motivation level was 47%.

Table 2. Motivation Levels (N=95) (%)

Motivation Dimensions	1st Quartile ¹⁹	Median ²⁰	3rd Quartile ²¹
Ideal Self	65	75	85
Ought-to-Self	31	43	57
International Posture	67	77	83
Competitiveness	63	73	83
Learning Attitude	75	80	85
Intended Effort	67	77	80
Total Motivation	64	71	75

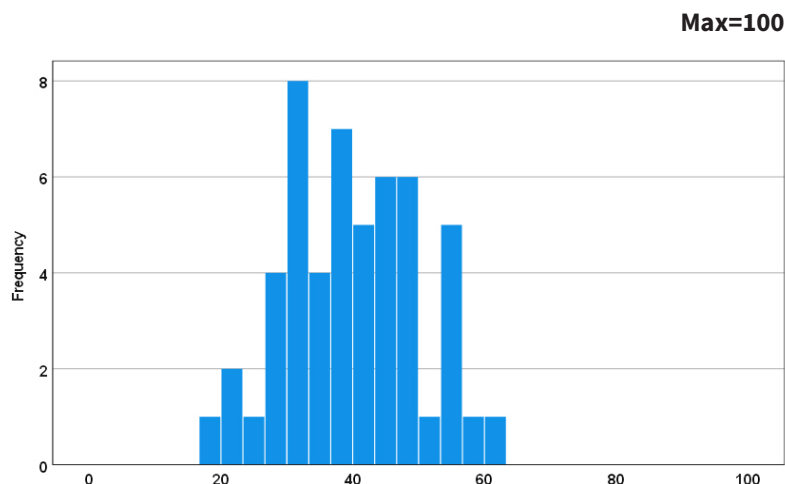
Second Language Profile

We asked participants to complete an adapted version of the Bilingual Language Profile (Birdsong et al., 2012) for their second (foreign) language. The profile provides a Global Language Score (GLS) for foreign language spoken by the participants. GLS is based on separate modules for evaluating language history, language use, language proficiency and language attitudes. GLS can vary from 0 to 218, and we rescaled it, so the maximum GLS is equal to 100.

A GLS score of 218 (or rescaled as 100) in English would be appropriate for participants born into an Englishspeaking family, in an English-speaking country, who started studying English immediately, for whom all classes at school were in English, who speak only English all the time with family, friends, and at work. Their language history and language use are entirely English- based. They feel totally proficient in English, and they identify themselves with an English-speaking culture.

Fifty-two participants felt confident of their knowledge of a second language and completed the GLS questionnaire; the results are shown below.

Figure 6. Second Language Profile (N=52)



The median GLS score for second language was 38.2 (IQR=15) which corresponds to the intermediate level proficiency.

19 First 25% of the sample.

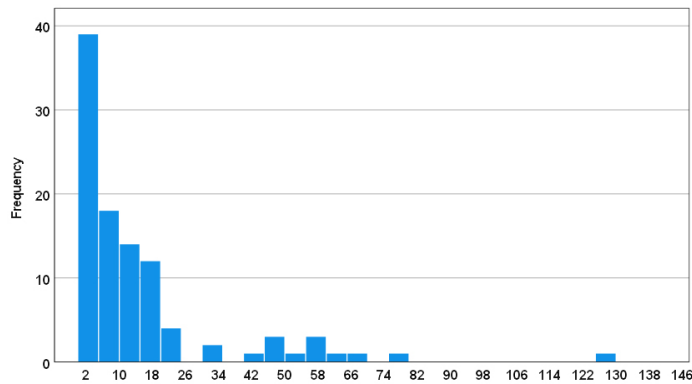
20 50% middle point.

21 First 75% of the sample.

Study Time

We measured the study time objectively by the actual server time on a weekly basis and we reported the time to the participants regularly via e-mail in order to encourage them to keep studying. The median total study time for the final study sample was about 8 hours, or about one hour per week. The total study time ranged from two hours to 129 hours.

Figure 7. Study Time Distribution in Hours (N=101)



LANGUAGE IMPROVEMENT

Vocabulary/Grammar Proficiency Results

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

Figure 8. WebCAPE Gain (n=101)

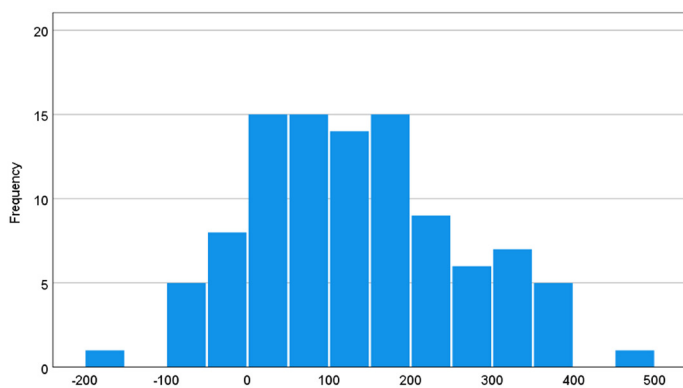


Table 3. Language Improvement (Vocabulary/ Grammar)

WebCAPE Test Points (n=101)

Statistics	Initial WebCAPE	Final WebCAPE	Improvement (Final-Initial)
Mean (std)	167.7 (137.1)	298.2 (139.4)	130.5 (126.8)
95% Confidence Interval	140.7 – 194.8	270.7 – 325.7	105.4 – 155.5

The average overall improvement of 130.5 WebCAPE test points was statistically significant with a 95% confidence interval from 105.4 to 155.5 points.

Overall, 86.1% of all participants improved their vocabulary/ grammar language proficiency with a 95% confidence interval²² of 77.9% to 91.6%.

Only 14 participants out of 101, or 13.9% did not increase their WebCAPE score. Participants who did not improve their

WebCAPE score studied on average much less (12.5 hours) than participants who improved their score (15.5 hours).

College Semester Placement

We can measure progress by movement from an initial semester level to a final semester level with the results presented below.

²² All 95% CI for proportions use Agresti-Coull correction (Agresti&Coull, 1998).

Table 4. WebCAPE Semester Placement

College Semester	Initial Test		Final Test	
	People (N)	%	People (N)	%
First	70	69.3	31	30.7
Second	21	20.8	37	36.6
Third	9	8.9	20	19.8
Fourth+	1	1.0	13	12.9
Total	101	100	101	100

Participants at first semester level decreased from 69.3% to 30.7%.

Table 5 below shows the semester-level change as counts of people who moved from one to another level. Forty-four of our participants (43.5%) did not improve their semester-level while 57 participants, or 56.5%, improved by at least one semester-level.

Participants who did not increase their semester placement studied on average much less (8.8 hours) compared to the participants who did increase their semester level (20 hours).

Table 5. Semester Improvement

Level (Semester Change)	Improved		Study Time
	People (N)	%	Mean (Hours)
-1 Negative change	6	5.9	Small n
0 Same/No Change	38	37.6	8.8
1 One Semester Increase	38	37.6	19.5
2 Two Semester Increase	15	14.9	21.0
3 Three Semester Increase	4	4.0	Small n
Total	101	100	10.0

The problem with using the semester improvement as a sole measure of proficiency gains 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 initially has 269 test points (first semester), only a one-point gain is needed to move to the second semester level. Another person can start with 10 points (first semester), then gain 200 points and the new level (210 points) is still first semester.

Oral Proficiency Results

Ninety-two people successfully took the final TNT oral proficiency test and the results are presented below. As with the grammar and vocabulary test, we measured their proficiency gains as the difference between their scores at the initial and final testing. The average overall oral proficiency improvement of 1.3 TNT test points was statistically significant with a 95% confidence interval from 1.0 to 1.6 points.

Figure 9. Oral Proficiency: TNT Score Gain in Points (n=92)

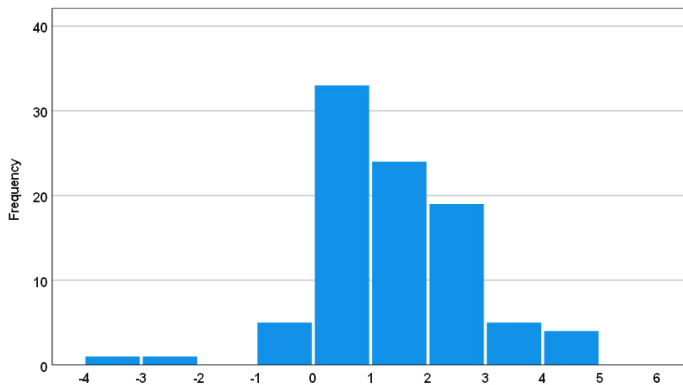


Table 6. Language Improvement. Oral Proficiency by TNT score (n=92)

Statistics	Initial TNT score	Final TNT score	Improvement (Final-score)
Mean (std)	2.5 (1.3)	3.8 (1.2)	1.3 (1.4)
95% Confidence Interval	2.3 – 2.8	3.5 – 4.0	1.0 – 1.6

Overall, 89.1% of all participants improved their oral proficiency with 95% confidence interval of 80.9% to 94.1%.

Oral Proficiency for Participants with at least 8 Hours of Study

Language Testing International, the exclusive licensee of ACTFL recommends²³ 8 weeks as the minimum time between test and retest for Group I languages, which includes Spanish. In our previous studies, we used the ACTFL computerized oral proficiency test (OPIc) and we required at least 8 hours for 8 weeks of study for oral evaluation. For completeness, and in order to make the results of this study comparable to the previous studies, we present similar results here.

Table 7. Oral Proficiency for Participants with 8 Hours of Study (n=48)

Statistics	Initial TNT score	Final TNT score	Improvement (Final-score)
Mean (std)	2.8 (1.5)	4.1 (1.3)	1.3 (1.5)
95% Confidence Interval	2.4 – 3.2	3.7 – 4.4	0.9 – 1.7

The average overall oral proficiency improvement of 1.3 TNT test points was statistically significant with a 95% confidence interval from 0.9 to 1.7 points.

Overall, 87.5% of all participants improved their oral proficiency with a 95% confidence interval of 74.8% to 94.4%.

TNT-ACTFL and TNT-CEFR Progress Estimation

Similar to our previous studies (Vesselinov et al., 2016-2021), we evaluated oral proficiency by TNT-ACTFL and TNT-CEFR estimation only for participants with 8 hours of study or more (n=48 participants).

The majority of the participants (77.1%) with at least 8 hours of study increased their TNT-ACTFL and TNT-CEFR estimation level by at least one level. The 95% confidence interval is from 63.2% to 86.8%.

²³ <https://www.languagetesting.com/how-long-does-it-take>

EFFICACY

Vocabulary/Grammar Proficiency Efficacy

As mentioned above, our study investigated the efficacy of the LingQ app as a function of total proficiency gains per one hour spent studying. Table 8 shows the overall results.

Table 8. Vocabulary/Grammar Efficacy of LingQ (n=101)

Statistics	Efficacy = Improvement per one hour of study (WebCAPE Test Points)	Time to cover the placement requirements for first semester of college Spanish proficiency. (Hours)
Mean (std)	23.3 (36.2)	12.6 ²⁴
95% Confidence Interval	16.1 – 30.4	8.9 – 16.8 ²⁵

On average LingQ users gained 23.3 WebCAPE test points per one hour of study with a 95% confidence interval of 16.1 to 30.4 test points per one hour of study.

The LingQ vocabulary/grammar efficacy measures 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 mean efficacy, we can construct a new measure representing the time needed to cover the requirements for the first college semester of Spanish. This is one measure of efficacy that is easy to understand, given the nature of the WebCAPE placement test.

Based on this measure, LingQ users will need on average about 12.6 hours of study during a two-month period to cover the requirements for the first college semester of Spanish.

The transformed lower and upper limits are from 8.9 to 16.8 hours of study during a two-month period.

Oral Proficiency Efficacy

Similar to the vocabulary/grammar efficacy, we estimated the oral proficiency efficacy as function of the improvement (final TNT score minus initial TNT score) and the effort (number of study hours) and the results are presented below.

Table 9. Oral Proficiency Efficacy for LingQ for Users with 8 Hours of Study (n=48)

Statistics	Efficacy = Improvement per one hour of study (TNT Test Points)	Time to reach max TNT=10.0 (Hours)
Mean (std)	0.0593 (0.1)	168.6 ²⁶
95% Confidence Interval	0.0294 – 0.0893	112 – 340 ²⁷

Regarding oral proficiency efficacy, it will take LingQ users on average 168.6 hours of study in a two-month period to reach the upper limit of the TNT test. The 95% confidence interval of this estimate is between 112 to 340 hours. This estimate assumes linear improvement trajectory.

²⁴ The threshold of 270 points divided by the mean efficacy (23.3 points).

²⁵ The threshold of 270 points divided by the lower limit (8.9) and the upper limit (16.8) of the 95% CI.

²⁶ TNT=10 divided by gain per hour (0.0593).

²⁷ TNT=10 divided by the upper (0.0893) and lower (0.0294) limits of the 95% confidence interval

Table 10 below presents the improvement in both vocabulary/grammar and oral proficiency.

Table 10. Improvement in both Vocabulary/Grammar and Oral Proficiency

Language Proficiency Improvement Vocabulary/Grammar & Oral	Improved		Final Test
	N	%	Mean (Hours)
No improvement	1	1.1	Small n
Improved Only Vocabulary/Grammar Proficiency	9	9.8	5.6
Improved Only Oral Proficiency	11	12.0	9.2
Improved Both Vocabulary/Grammar & Oral Proficiency	71	77.2	22.5
Total	92	100	20.4

The above results show that 99% of the participants improved either their vocabulary/grammar or oral proficiency, or both. Only one person was unable to improve their language proficiency. About 77% of the participants improved both their vocabulary/grammar and oral proficiency.

FACTORS FOR EFFICACY

Demographic Factors

We investigated the impact of several factors on language improvement (WebCAPE and TNT), namely age, gender, education, employment, and reason for studying Spanish. None of these potential factors had a statistically significant effect ($p < 0.05$). This means that the LingQ app works similarly for people regardless of gender, age, education, employment status, etc.

Motivation Effect

We evaluated the effect of motivation on oral and vocabulary/grammar language improvement. There was no statistically significant effect of the motivation score.

One possible explanation is that the people in this study are highly motivated with an average level of motivation of 71%.

At this high average level, additional higher motivation does not have an effect on the results. This result is consistent with our previous studies (Vesselinov et al., 2009-2021).

Language Profile Effect

GLS of a second language does not have a significant effect on language improvement.

Study Time

As expected, the more time participants studied, the better results they achieved. Table 11 below shows the effect of study time on the language improvement. Participants who improved their vocabulary/grammar proficiency on average studied more (15.5 hours) than the participants that did not improve (12.5 hours). Similarly, participants who improved their oral proficiency on average studied more (16.6 hours) than the participants that did not improve (8.6 hours).

Table 11. Study Time as Factor for Language Proficiency Improvement

Language Proficiency Improvement	Improved		Study Time
	N	%	Mean (Hours)
No improvement <input type="checkbox"/> Yes	14	13.9	12.5 (14.7)
	87	86.1	15.5 (20.9)
No improvement <input type="checkbox"/> Yes	10	9.9	8.6 (5.4)
	82	89.1	16.6 (21.4)

Initial Vocabulary/Grammar and Oral Proficiency

As in our previous studies (Vesselinov et al., 2009-2021), this study confirmed the inverse relationship between the initial level of language proficiency and the gain in new knowledge (Figures 10 and 11). The biggest gain in new knowledge can be attributed to the novice/beginner users, while more advanced learners gain test points slowly, both in vocabulary/grammar and oral proficiency. This is especially noticeable for oral proficiency: the regression line for oral proficiency is steeper than the one for vocabulary/grammar proficiency (regression coefficient -0.63 vs -0.41).

Figure 10. Initial Vocabulary/Grammar Proficiency and Improvement

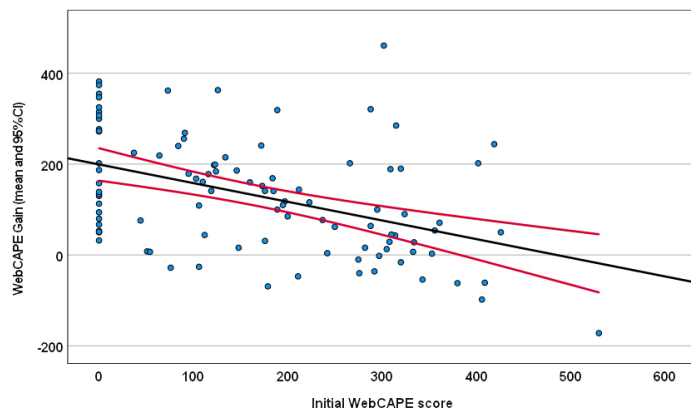
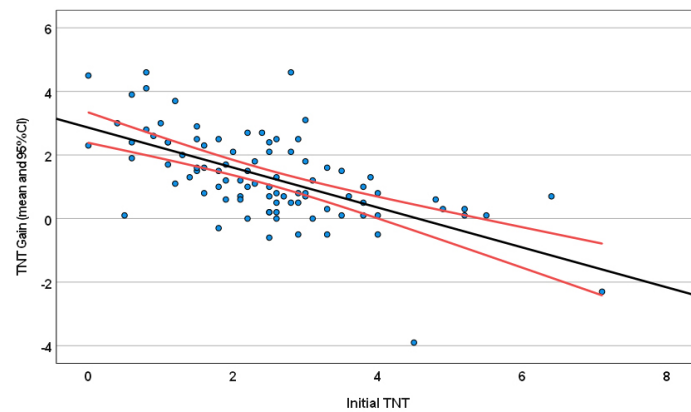


Figure 11. Initial Oral Proficiency and Improvement



USER SATISFACTION

After the study the participants were asked for their opinion about the LingQ app; specifically how easy it was to use, how helpful it was for learning Spanish, how enjoyable it was, and how much they were satisfied with it. The 5-point Likert scale was recoded into two categories: Strongly Agree/Agree vs Strongly Disagree/ Disagree/Neutral.

Table 12. User Satisfaction(N=93)

Do you agree with the following statement?	Agree/Strongly Agree
“LingQ was easy to use”	69.9
“LingQ was helpful in studying Spanish”	86.0
“I enjoyed learning Spanish with LingQ”	71.0
“I am satisfied with LingQ”	74.2

After two months of study, the majority of participants (70% and above) agreed with the positive statements that: LingQ was easy to use, it was helpful, they enjoyed learning with LingQ, and they were satisfied with it.

In the exit survey, a special question was included: “How likely are you to recommend LingQ 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, a positive NPS is desirable news for the company and the higher the score, the better the indicator for the company.

From our exit survey the “Promoters” were 35.5%, the “Detractors” were 29.0% and “Passives” were 35.5%. The LingQ NPS was positive, +4.2.

LIMITATIONS OF THE STUDY

The population of adult people who are seeking to study foreign language with language apps is highly educated with the majority of them having at least a college-level education. This is true not only for the U.S.²⁸, but also Europe²⁹ and the rest of the world³⁰. This was confirmed by all our previous studies³¹. This population has a higher education level than the general population. Our current sample for the 2023 LingQ efficacy study is representative of this particular population, but it may not be comparable to the general population.

This study measures the progress of novice/beginner users of Spanish. The study results cannot be generalized to intermediate or advanced users of Spanish. We conducted one language efficacy study of intermediate users for Busuu (Vesselinov et al., 2021).

The independently developed tests used in this study were not tailored to any specific learning tool, including LingQ. On the one hand, some participants in the study complained that the test contained words or expressions that were not part of their regular course with LingQ. On the other hand, people insisted that they had learned a lot more than the test asked for. The test is valuable as an independent tool for evaluation which allows us to compare efficacy across different apps, however it does not provide a complete measure of the full progress of users, so the evaluation of their progress in language proficiency is generally conservative.

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. The results of the study should be valid in a setting where users study regularly for two months and receive weekly reminders.

The study results could be generalized for studying Spanish with LingQ. For other languages the results could be markedly different. Also, this study's results may not be generalizable to study periods shorter than two-months.

28 Rosetta Stone (2009, 2019), Duolingo (2012), italki (2018)

29 Babbel (Germany & US), Busuu (UK and US).

30 Language Zen Efficacy Study, 2015 report, (world sample).

31 Except Hello English (2017) where the participants were of high school age.

This study results cannot be compared directly to a regular language study in a college. There are at least two limitations related to comparing the results of this study to a standard college semester of Spanish. First, progress or success in college is determined usually with one midterm exam and one final exam, plus some form of testing for oral proficiency and homework. These are typically done by the instructor and are not subject to a standardized testing and both semester-long perceptions of the student progress and deep knowledge of the material that has been presented in the classroom affect the construction and evaluation of the tests. Progress is measured very differently in a college setting compared to this study. Second, study time in college is difficult to measure scientifically.

CONCLUSION

The LingQ efficacy study is based on a random sample of 101 people, 18 years of age or older, residing all over the world. All participants were self-reported novice users of Spanish. We applied a set of instruments for vocabulary/grammar and oral proficiency.

The use of the AI-driven TNT gave us the opportunity to detect any change in oral proficiency in both directions: increase and decrease. In some of our previous studies, we used ACTFL OPIc, which is a level-based test and cannot detect incremental change.

The main goal of measuring the efficacy of LingQ was achieved with this study.

We found that 86.1% of the participants improved their vocabulary/grammar proficiency during the study. Novice users need on average 13 hours of study in a two-month period to cover the requirements for first college semester of Spanish.

Comparably, 89.1% of all study participants increased their oral proficiency (TNT) during the study.

The vocabulary/grammar efficacy was an average gain of 23 WebCAPE test points per one hour of study and the oral proficiency was a gain of 0.06 TNT points per one hour of study.

CITED LITERATURE

- Agresti, A., Coull, B., 1998, Approximation is better than “exact” for interval estimation of binomial proportions, *American Statistician*, 52, pp. 119–126.
- Birdsong, D., Gertken, L., & Amengual, M. Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. COERLL, University of Texas at Austin. Web. 20 Jan. 2012. <https://sites.la.utexas.edu/bilingual/>.
- Dörnyei, Z., 2005. *The psychology of the language learners*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z., 2009. The L2 motivational self-system. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9e42). Bristol, UK: Multilingual Matters.
- Habing, B., Grego, J., Vesselinov, R. “Predictive and Psychometric Properties of the TrueNorth Test (TNT)”, 2020. http://comparelanguageapps.com/documentation/TNT_Final_Report.pdf
- Kong, J., Han, J., Kim, S., Park, H., Kim, Y., Park, Hy. L2 Motivational Self System, international posture and competitiveness of Korean CTL and LCTL college learners: A structural equation modeling approach, *System*, Volume 72, February 2018, Pages 178-189
- Reichheld, F., 2003, “One Number You Need to Grow”, *Harvard Business Review*, 2003 December.
- Vesselinov, R., Grego, J., Tasseva-Kurktchieva, M., Sedaghatgoftar, N. “The Busuu Efficacy Study 2021”, 2021. http://comparelanguageapps.com/documentation/Busuu_Efficacy_Study_2021.pdf
- Vesselinov, R., Grego, J. “Mango Languages Efficacy Study”, 2019. http://comparelanguageapps.com/documentation/Mango_Languages_FinalReport2019.pdf
- Vesselinov, R., Grego, J. “Pimsleur Efficacy Study”, 2019. http://comparelanguageapps.com/documentation/Pimsleur_EfficacyStudy2019.pdf
- Vesselinov, R., Grego, J., Sacco, S., Tasseva-Kurktchieva, M., 2019, The 2019 Rosetta Stone Efficacy Study. http://comparelanguageapps.com/documentation/The2019_RS_FinalReport.pdf
- Vesselinov, R. and Grego, J., 2018, italki Efficacy Study. <http://comparelanguageapps.com/documentation/italki2018FinalReport.pdf>
- Vesselinov, R. and Grego, J., 2017, Hello English Efficacy Study. http://comparelanguageapps.com/documentation/HelloEnglish_2017Study.pdf
- Vesselinov, R. and Grego, J., 2016b, The Babbel Efficacy Study. <http://comparelanguageapps.com/documentation/Babbel2016study.pdf>
- Vesselinov, R. and Grego, J., 2016, The Busuu Efficacy Study. http://comparelanguageapps.com/documentation/The_busuu_Study2016.pdf
- Vesselinov, R. and Grego, J., 2015, Language Zen Efficacy Study, http://comparelanguageapps.com/documentation/Language_Zen_Efficacy_Study.pdf.
- Vesselinov, R. and Grego, J., 2012, Duolingo Effectiveness Study. http://comparelanguageapps.com/documentation/DuolingoReport_Final.pdf
- Vesselinov, R., Grego, J., Habing, B., Lutz, A., 2009a, Measuring the Attitude and Motivation of Rosetta Stone Users. <http://comparelanguageapps.com/documentation/MeasuringTheAttitudeandMotivationofRSUsers.pdf>
- Vesselinov, R., Grego, J., Habing, B., Lutz, A., 2009b, Comparative Analysis of Motivation of Different Language Learning Software. <http://comparelanguageapps.com/documentation/ComparativeMotivationAnalysisofDifferentLanguageSoftware.pdf>
- Vesselinov, R., 2009, Measuring the Effectiveness of Rosetta Stone. http://comparelanguageapps.com/documentation/EffectivenessStudy2008_RS.pdf.

APPENDIX

Table A1. Study Participants' Geographic Distribution

Number of people

	Country	Initial Pool N=1119	Eligible Pool N=499	Initial Sample N=192	Final Sample N=101
1	Algeria	3	1	1	
2	Argentina	1			
3	Armenia	1	1		
4	Australia	25	6	5	4
5	Austria	4	1		
6	Azerbaijan	2	1	1	
7	Bahrain	1	1		
8	Belarus	2	1		
9	Belgium	2	2	2	1
10	Brazil	33	15	7	3
11	Bulgaria	1	1	1	1
12	Cambodia	1	1		
13	Canada	70	37	16	9
14	Cayman Islands	2	2	2	1
15	China	3	2	1	1
16	Congo-Kinshasa	1			
17	Costa Rica	1			
18	Croatia	6	6	6	3
19	Czech Republic	5	3		
20	Denmark	2	1		
21	Egypt	9	4		
22	Estonia	3	2		
23	Finland	2	1		
24	France	17	5	5	4
25	Georgia	1	1		
26	Germany	45	12	4	3
27	Gibraltar	1			
28	Greece	6	4	2	1
29	Guyana	1			
30	Haiti	1			
31	Hungary	10	5	2	1
32	India	17	9	5	3
33	Indonesia	1			
34	Ireland	9	3	2	1
35	Israel	5	4	1	1
36	Italy	9	2	2	1
37	Jamaica	1	1		
38	Japan	3	2		
39	Jordan	1	1		
40	Kenya	2	1		

	Country	Initial Pool N=1119	Eligible Pool N=499	Initial Sample N=192	Final Sample N=101
41	Latvia	1			
42	Lebanon	1	1		
43	Lithuania	5	3	2	2
44	Malaysia	2	1	1	1
45	Malta	1	1		
46	Montenegro	2	1		
47	Morocco	4	2		
48	Nepal	1	1	1	1
49	New Zealand	3			
50	Nigeria	2	2	1	
51	Norway	3	1	1	1
52	Pakistan	2	2	1	
53	Philippines	16	9	2	
54	Poland	16	7	4	2
55	Portugal	3	2		
56	Qatar	1			
57	Romania	3	2	2	
58	Russia	15	6	3	1
59	Saudi Arabia	5	3	3	
60	Scotland	3	1		
61	Serbia	3			
62	Singapore	2	1	1	1
63	Slovakia	2			
64	Slovenia	3	2		
65	South Africa	8	5	2	1
66	South Korea	1	1		
67	Spain	1			
68	Sudan	1			
69	Sweden	7	2		
70	Switzerland	6	2		
71	Taiwan	5	1		
72	The Bahamas	1			
73	The Netherlands	9	4		
74	Trinidad and Tobago	1	1		
75	Tunisia	1	1		
76	Turkey	17	5	4	2
77	UAE	2	1		
78	UK	100	41	10	5
79	Ukraine	14	9	3	3
80	USA	532	238	86	43
81	Vietnam	2			
82	Wales	1			
83	Yemen	1	1		

Table A2. Background Information on the Participants

Categories	N	Percent	Total N
Age: mean (std)	39.8 (14.3)		101
Female	61	62.9	101
Education			96
Less than High School	4	4.2	
High School diploma or equivalent	6	6.3	
Started college but did not graduate	19	19.8	
College graduate, BA degree or equivalent	28	29.2	
Started graduate school but did not graduate	6	6.3	
Master's degree (MA, MS)	26	27.1	
PhD/MD/JD	7	7.3	
Employment			98
Employed Full Time	44	44.9	
Student	22	22.4	
Homemaker	10	10.2	
Retired	7	7.1	
Employed Part Time	6	6.1	
Unemployed	2	2.0	
Other Employment	7	7.1	
Second Language	58	58	100
Reason for Studying Spanish			101
Personal Interest	64	63.4	
Travel	22	21.8	
Business/Work	12	11.9	
School	1	1.0	
Other	2	2.0	
Have close friend or spouse who speaks Spanish	13	13.0	100
Have parents of grandparents who speak Spanish	1	1.0	100
Lived 6+ months in foreign language country	29	29.0	100

Figure A1. L2 Motivation (Max=100)
M1 “Ideal Self”

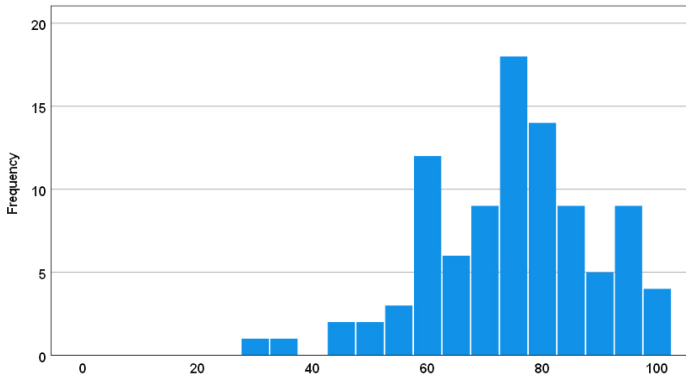


Figure A4. L2 Motivation (Max=100)
M2. “Ought-to Self”

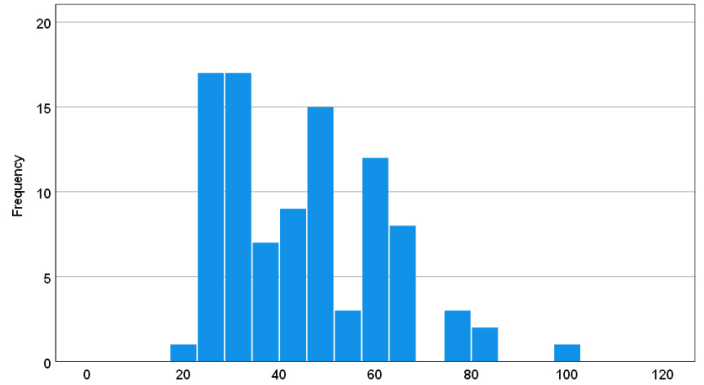


Figure A2. L2 Motivation (Max=100)
M3. “International Posture”

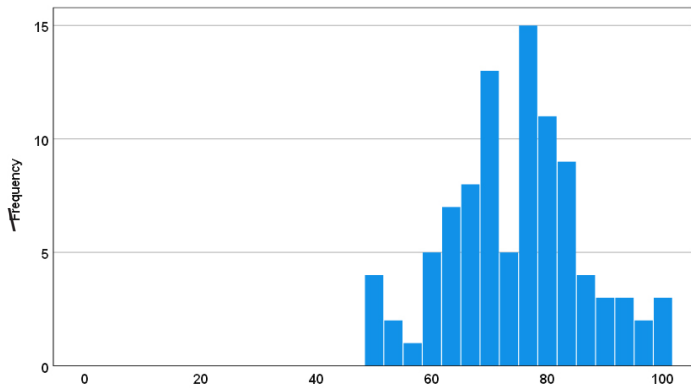


Figure A5. L2 Motivation (Max=100)
M4. “Competitiveness”

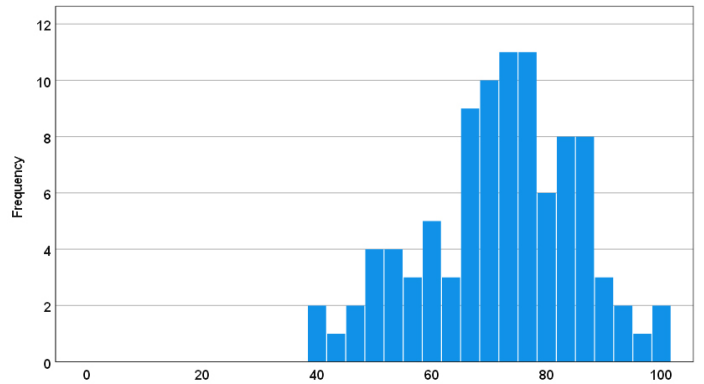


Figure A3. L2 Motivation (Max=100)
M5. “Learning Attitude”

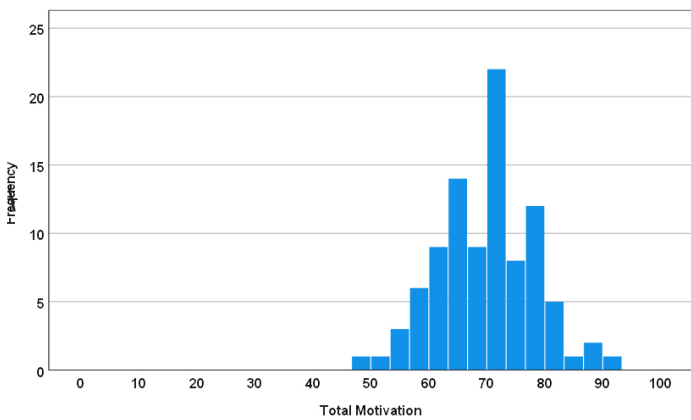


Figure A6. L2 Motivation (Max=100)
M6. “Intended Effort”

