The 2019 Rosetta Stone Efficacy Study

FINAL REPORT

RESEARCH TEAM

ROUMEN VESSELINOV[1,2], PhD

Economics Department
Queens College, City University of New York

JOHN GREGO[2], PhD

Statistics Department
University of South Carolina

STEVEN J. SACCO[2], PhD

San Diego State University (Emeritus)
Sacco Global Consulting, Inc.

MILA TASSEVA-KURKTCHIEVA[2], PhD

Linguistics Program
University of South Carolina

January 2019

[1] Corresponding author: roumen.vesselinov@qc.cuny.edu .
[2] This report represents the individual opinion of the authors and not necessarily of their institutions.

THIS PAGE LEFT INTENTIONALLY BLANK

EXECUTIVE SUMMARY

The research team independently completed the study data collection from June 2018 to August 2018 and finalized the analysis by January 2019. The study was based on a random representative sample of 143 Rosetta Stone users.

The participants took one Spanish college placement test, then studied Spanish with Rosetta Stone for two months and took the same college placement test again. We measured the improvement in language abilities as the difference between the final and the initial language test results. We measured the efficacy of Rosetta Stone as language proficiency improvement per one hour of study.

ROSETTA STONE EFFICACY

- The efficacy of Rosetta Stone is a gain of about 21 test points per one hour of study. The 95% confidence interval for the efficacy is between 16 and 26 test points.
- Rosetta Stone users need on average 13 study hours in a two-month period to cover the requirements for the first college semester of Spanish. The lower and upper limits are between 10 and 17 study hours.

USER SATISFACTION

- The majority of users thought that Rosetta Stone was easy to use (97%), helpful (94%), enjoyable (94%), satisfying (89%).
- Rosetta Stone received a positive Net Promoter Score of +51 from the users.
- Rosetta Stone efficacy was not affected by gender, age, education, native language, device used, etc.
- Participants' motivation was very high and remained high after two-month study with average level of 75% of the maximum motivation score.

THIS PAGE LEFT INTENTIONALLY BLANK

CONTENTS

INTRODUCTION

This is the tenth study of the Research Team testing the efficacy and motivation of language learning apps (Vesselinov et al., 2009-2018). Our previous studies evaluated Rosetta Stone, Duolingo, busuu, Babbel, Hello English, italki and a new language app. The statistical design and methodology are practically the same for all ten studies. The only deviation is the measurement of study time in the first Rosetta Stone study in 2009 when the time was self-reported versus recorded because participants were given CDs with the software and no objective measure for study time was technologically available. In addition, there was a requirement for 55 study hours and thus the results from this study cannot be scientifically compared to the other studies.

This is a new study designed to evaluate the efficacy of Rosetta Stone (RS)[3].
According to the company, Rosetta Stone Inc. (NYSE: RST) is dedicated to changing people's lives through the power of language and literacy education. The company's innovative digital solutions intend to drive positive learning outcomes for the inspired learner at home or in schools and workplaces around the world. Founded in 1992, Rosetta Stone's language division uses cloud-based solutions to help all types of learners read, write, and speak more than 30 languages. Rosetta Stone states that the key features of the application include:

- TruAccent Technology: patented speech recognition technology – TruAccent - checks users' speech 100 times a second to make sure their pronunciation is on par with the native pronunciation and helps them correct it, allowing users to speak more authentically.
- Bite-sized lessons: Everyone learns at their own speed, so users can break up lessons into convenient, short chunks.
- Phrasebook: User's guide to perfectly pronouncing key phrases.
- Stories: Practice speaking by reading aloud while listening to native speakers.
- Audio Companion: Take a break from the screen by listening to lessons wherever you go.
- Cross-platform access: Lessons update in real time across mobile, desktop, and tablet.

[3] www.RosettaStone.com

This study was funded by Rosetta Stone, but the data collection and the analysis were carried out independently by the Research Team. The language test used in the study was designed and developed by an external independent testing company.

THE ROLE AND IMPORTANCE OF ONLINE LANGUAGE LEARNING SYSTEMS

By Steven J. Sacco, PhD[4]

Online language learning systems, which often include apps as one of their learning platforms, are playing an increasing role in providing language instruction. The market is huge and growing exponentially; and it is worldwide. The primary providers are Rosetta Stone, Transparent Language, Babble, busuu, and Duolingo. The contenders include Berlitz, Fluenz, Yabla, Rocket Languages, Mango Languages, WeSpeke, Linguistica360, Fluent Forever, and many others.

Online language learning comprises digital content and products that facilitate the learning of languages through Information Communication Technology (ICT) tools. These tools include mobile apps, activities, e-books, games, videos, audio clips, and in some cases, evaluation using the Common European Framework for Reference of Languages (CEFR). Online platforms are interactive, allow real-time feedback, and enhance learning processes, as it involves different formats of learning. With a laptop and a cell phone, a user can study any one of numerous languages anywhere, anytime 24/7/365. This type of versatility, flexibility, and accessibility overmatch traditional language instruction provided by high schools, colleges, and universities. As a result, online language learning systems are syphoning off enrollments from these traditional sources (MLA, 2017), but online systems face cutthroat competition from each other.

Traditional foreign language instruction at U.S. colleges and universities has changed little over the last few decades. Teaching methodology remains basically the same since the 1970s. Courses still take place in physical classrooms, 35 or more students to a classroom; scheduling remains the standard four to five times a week at the elementary level. Instructors, the primary language providers, continue to possess limited knowledge of foreign language acquisition and teaching methodology. Language offerings are still limited to the commonly taught languages: Spanish, French, and German. Less commonly taught languages are usually offered at large universities. In other words, traditional foreign language instruction remains a one-size-fits-all model. The model neither features versatility, convenience or customization of instruction. The model is designed to benefit the provider versus the end user.

For global professionals and dedicated language learners seeking to acquire foreign language fluency, the traditional model seldom works in terms of convenience (time and place), cost, language offerings, the pace of instruction, formal testing or customized language learning. This anachronistic business model would have doomed most businesses, but language instruction is fueled by its college or university.

Rosetta Stone is the patriarch of all online language learning systems. Founded in 1992, Rosetta Stone offers instruction in 30 languages. Its unique system simulates immersion and provides the learner with the most comprehensible input in the industry. Rosetta Stone's ingenuous use of visuals enables it

[4] Professor Emeritus of French and Italian, San Diego State University
Director Emeritus, The Center for International Business Education and Research (CIBER)
President, Sacco Global Consulting, Inc., stevenjsacco52@gmail.com

to avoid using English in teaching a second language. To the best of our knowledge no other system can match this teaching technique. Rosetta Stone also provides access to native-speaking tutors to provide guidance and speaking practice. It assesses student weaknesses and provides remedial instruction to eliminate them; it allows learners to move freely through each unit, which are created by language experts.

Rosetta Stone is a trusted language learning system that is used by the U.S. military, the medical community, corporations, and many colleges and universities in the U.S. As corporate online language learning grows over the next decade, more adaptive providers, like Rosetta Stone and Transparent Language are likely to benefit most from this new market, one that requires customization.

Weaknesses of online language systems

Online language systems pride themselves on being stand-alone learning systems[5]. And that is their biggest weakness. "Blended or hybrid learning is a combination of online and face-to-face learning opportunities as part of the same course" (McGarell, 2013)[6]. Blended learning uses and takes advantages of the strengths of each. The elite teacher needs technology and technology needs the elite teacher. After spending close to a thousand hours working on Rosetta Stone, Transparent Language, Living Language, Babbel, busuu, and Duolingo, here's what elite teachers can do and what those systems can't do:

- Simulations involving problem-solving communication like applying for a job
- Discussion of complex cultural misunderstandings
- Discussion of current events via social media and traditional media
- Use of case studies to analyze complex communicative situations
- Teaching of language through commands like with Total Physical Response
- Story telling

An elite teacher could add to the list indefinitely.

In addition, online language learning systems lag far behind traditional classroom learning in the teaching of both big and little C culture. They have designed their platforms as if language and culture are NOT intertwined. It's one thing to know grammar and vocabulary, it's another thing to solve communication problems dealing with culture. Traditional foreign language textbooks introduce cultural concepts in each chapter and discuss the impact of cultural values, norms, and taboos on cross-cultural communication.

In any future courses designed for multinational corporations or the military, online language providers must incorporate culture throughout the curriculum or in a blended format.

[5] Rosetta Stone and transparent Language are exceptions to the stand-alone claim; both are used at West Point and other military installations along with traditional language classes.

[6] A representative of Fluent Forever admitted that their claims to fast and forever fluency actually come from blended learning in addition to immersion.

Research on the efficacy of online language learning systems

Online language learning systems face fierce competition to control market share despite an ever-expanding market. Most are funded with investment capital adding even more pressure to meet the expectations of investors.

Claiming product superiority is the premier strategy in convincing an unsophisticated public, desperate to learn another language. Number of subscribers and rate of customer satisfaction are two valid measures to cite, but unsubstantiated claims of fluency in record time top the list of marketing strategies. Unfortunately for potential subscribers, some online providers neither define "fluency" nor "in record time."

Babbel claims that "you will learn a language in 3 weeks." Given its recommended rate of 20 minutes of study a day, that means you can "learn" a language in seven hours. Seven hours! However, when I asked Babbel how long it would take to reach B1, it responded: "we don't provide that information to the general public." Babbel's efficacy study (Vesselinov & Grego, 2016) finds that one needs to average 21 hours of study with Babbel during a two-month period to cover the placement requirements of the first college semester of Spanish.

Fluent Forever features "a proven 5-star method." It claims that learners will attain B1 Intermediate "in a short time." The could not provide me with a clear definition of "in a short time." Unsubstantiated claims like that make research critical to assessing the efficacy of online language systems.

So far very little research exists on the efficacy of online language learning systems (Vesselinov & Grego, 2009-2018, nine studies). The research, for the most part, has been funded by the providers themselves although executed by independent researchers. The focus so far has been on the number of hours it takes to match a traditional elementary language course. Most of the studies tested only Spanish skills using systems like Duolingo, busuu, and Babbel, etc. There is one efficacy study of English as a second language (ESL) using HelloEnglish (Vesselinov & Grego, 2017).

So, why did the online providers stop at one study? Because they received the positive results they were hoping for. And they used the single study with a single language to extrapolate across all languages they offer. For example, Duolingo claims that "it takes only 34 hours to match the content of an elementary college language course." By not specifically mentioning Spanish, Duolingo is deceptively claiming that it only takes 34 hours to match an elementary college course in any of the 28 languages it offers. Spanish takes many fewer hours for native speakers of English to attain fluency than German, Turkish, Swahili or Japanese and Chinese. According to the Foreign Service Institute, to reach the superior level in speaking, it takes . . .

- 600 hours for Category I languages (Spanish, French, Italian, Dutch, etc.)
- 750 hours for Category II languages (German)
- 900 hours for Category III languages (Swahili)
- 1100 hours for Category IV languages (Russian, Turkish, Vietnamese, etc.) and
- 2200 hours for Category V languages (Mandarin Chinese, Korean, Japanese)

If 34 hours of Duolingo or 22 hours of busuu equals an elementary college language course, successfully completing that elementary language course's final exam should come naturally. That's another line of research to pursue. In my personal experience, I took UCLA's Swedish 101/102 final exam after completing over 200 hours of Swedish exclusively on Duolingo but received a grade of F. Duolingo teaches grammar and vocabulary via single sentences or phrases using the grammar-translation method, a discontinued method that dates back to the 1950s and 1960s. The UCLA final focused on free expression in both writing and speaking, which reflect its communication focus. Most elementary language final exams focus on communication, not only grammar knowledge.

Duolingo was the subject of a third study. In a Master's thesis, Ratzlaff (2015) compared German language performance levels in two groups at Fresno State University. The first group enrolled in an elementary German class while the second group used Duolingo. The Goethe Institute's A1 level test was the measure used for grammar in addition to listening, reading, speaking and writing skills. The elementary German class group outperformed the Duolingo German group in nearly every skill. Ratzlaff concludes that 34 hours of Duolingo is not the equivalent of an elementary German course. Unfortunately, the sample size was too small with insufficient statistical power.

This line of research is promising, but future comparative studies must (1) feature a larger sample size than the Ratzlaff study, (2) include other languages, and (3) other online systems like busuu, Babbel, and Rosetta Stone that participated in the Vesselinov & Grego studies. Unlike Duolingo, busuu and Babbel focus heavily on free expression at the end of each unit. Users of those two online language providers might excel at elementary language final exams given their focus on communication. This hypothesis is worth testing in future studies.

Multinational corporations and the military (among others) may not care about comparing performance between Duolingo and Spanish 101. They may be more interested in knowing how long it takes to reach workplace or battlefield proficiency. Cambridge English in its English at Work survey (2016) studied the language needs of over 3,000 multinational corporations worldwide. Almost 50% listed advanced or near-native proficiency as their benchmark. They may want to know the timeframe needed to reach B1, B2 or C1. A comparison test pitting the leaders, Rosetta Stone, Transparent Language, busuu, Babbel, Duolingo and Living Language would be the subject of future research studies.

The future of online language learning systems

The past: the key to the future

Online language learning systems have mastered technology as brilliantly as the field of aviation that leaped from Kitty Hawk to the moon in only three generations. Our Kitty Hawk is arguably PLATO, developed in the 1970s at the University of Illinois. Designed for computer-based education, PLATO nurtured an online community through groundbreaking communication and interface capabilities. In addition to pioneering chat rooms (Talkomatic) and instant messaging (Term-Talk), PLATO excelled at online games which would later evolve into the learning management systems (LMS) that make online language providers like Rosetta Stone famous.

Within a decade Ohio State would introduce Telephone-Assisted Self-Paced Master-Based Instruction to teach Eastern European and Central Asian languages through grants from the Department

of Defense and the National Endowment for the Humanities (Twarog & Perezlenyi-Pinter, 2011). The program would combine telephone-based instruction using language tutors in addition to the completion of assignments found on the website. Rosetta Stone and Transparent Language took it further to introduce language education via carefully crafted instructional CD-ROMs. Today, we have learning management systems at which to marvel.

So, what does the future hold? "I hesitate to make projections about the contribution of technology because it changes so rapidly," states Diane Larsen-Freeman (2018), one of the most distinguished applied linguists in the world. Larsen-Freeman and Kern (2014) point to "augmented reality, game-based learning, and other innovative uses of mobile devices will be exploited for the teaching and learning of languages."

Let's look at gamification as an example. PLATO started out teaching language via gaming. I developed a version of "hangman" to teach language vocabulary on a Commodore 64 computer in my language technology class in 1984 at Ohio State. The technological evolution to CD-ROMs and language management systems got away from gaming to create complete courses.

Online language learning systems of the future may add an interactive videogame dimension to its current courses. Rosen (2014) in her article *eLearning Future: What Will eLearning Look Like in 2075* reveals that . . .

"Candy Crush and World of Warcraft have taught us a lot about the cognitive psychology behind engagement. Learners like games. They like challenges, interactive elements, and opportunities to develop strategies. They also like mastering concepts (leveling up), immediate feedback, and characters with distinct personalities. Great courses of the future will likely include many of these elements which will make the learning experience so exciting, interactive, and fun that learners can't wait to participate and reap the benefits by mastering the content."

Simone Bregni, Associate Professor of Italian at St. Louis University, has already pioneered the use of videogames for teaching Italian. His numerous teaching awards attest to his success. Bregni's most recent article "Assassin's Creed Taught Me Italian: Video Games and the Quest for Lifelong, Ubiquitous Learning" is a description of use of gamification. It is a must-read for online language providers.

Accessibility

Accessibility will take center stage as access to high-quality education in foreign languages is not available to all students in the U.S. (Larsen-Freeman, 2018). Larsen-Freeman cites African-American students in inner-city schools and rural children in small high schools. These two groups of students have access to few languages, usually Spanish and French. Working with school districts to offer 20 to 30 languages offered through sophisticated learning systems will fill the gap due to budget cuts and the unavailability of language teachers. Making close connections to international baccalaureat (IB) programs in the U.S. will enhance their language programs, which are usually limited to five languages. The addition of an online language program within IB Programs will make them superior to European or Asian high schools in foreign language study. Other online providers will mimic Rosetta Stone's close relationship with colleges and universities.

In Africa, Worldreader, a global NGO, provides children in the developing world with free access to a library of digital books via e-readers and mobile phones. Since 2010, 6.5 million people across 50 countries have read from Worldreader's digital library of over 40,000 e-books. Worldreader works with

device manufacturers, local and international publishers, government agencies, education officials, and local communities to support readers everywhere. An unintended consequence of this extraordinary program is the possibility of teaching languages online to students in the developing world. Online language providers should explore this possibility with Worldreader. Sub-Saharan Africa, by itself, boasts a population of over 595 million, a sizeable potential market of future subscribers of online language systems.

Strategic partnerships between online providers and foreign language book publishers

In 2015, McGraw-Hill Education became a strategic minority investor in busuu. McGraw-Hill Education holds the exclusive distribution rights to sell busuu to organizations and institutions worldwide while busuu now has an entry into the lucrative market of high schools, colleges, and universities. Busuu course completion carries with it certification from McGraw-Hill Education along with a CEFR language skills rating. McGraw-Hill Education and busuu now have their sights on Rosetta Stone and Transparent Language's monopoly on the corporate language learning market. In the future other online providers can adopt their system to the parameters of a publisher's foreign language textbook which often sells for over $100. Students continue to use their textbooks AND their e-workbook/lab designed by Babbel, Duolingo or Transparent Language. The strategic partnership eliminates the either-or proposition, benefitting both traditional foreign language textbook providers and their online counterparts.

The Corporate Online Language Learning Market

The corporate online language learning market is largely untapped except for industry veterans Rosetta Stone and Transparent Language. In two recent reports on the future market potential of the corporate online language learning market for the period 2017-2021, I examined two markets: China and Europe. Technavios analysts forecast the corporate online language learning market in China to grow at a CAGR of 21.58% during the period 2017-2021. CAGR refers to the compound annual growth rate. That's 107.9% growth over that five-year period. In Europe, the CAGR is forecasted at 9.31%. Rosetta Stone, Transparent Language, and Berlitz are cited among the key vendors. Among the many languages learned in China, English and Japanese dominate accompanied by a strong interest in French, Spanish, and Portuguese. In this future scenario, I envision business clients in China, Europe or on any continent completing the traditional Rosetta Stone or Transparent Language course and then proceed to a customized course designed for a corporation or an industry. For example, my team and I designed a learning management system to teach SAFETY ENGLISH to immigrant workers employed by U.S. agribusinesses (Sacco et al., 2015).

Disappearing languages and dialects

"Every two weeks a language dies," reports *National Geographic* reporter Nina Strochlic. (Strochlic, 2018). There is a growing movement to preserve disappearing languages like those spoken among First Nation peoples in North America. New online providers are offering instruction, but their technology skills don't match the skills honed through decades of experience the Rosetta Stone and Transparent Language.

Recently, I approached an Italian Diaspora organization in New York City about creating online courses for disappearing Italian dialects (Sacco, 2018b). There are over 150 dialects in Italy, most of which are threatened by the hegemony of Italian, Italy's official language since 1861. Second, third and fourth generation Diaspora Italians in the U.S., Australia, Canada, Argentina, Brazil, Egypt, Germany and in other countries are starving for formal instruction in the dialects spoken by their grandparents and great grandparents, many of whom are in their 80s and 90s. Assimilation has slowly strangled these dialects. In addition, languages and dialects are disappearing on every continent from Inuit in northern Canada to Maori in New Zealand. This is yet another potential market for the online language leaders. Funding opportunities abound for preserving or resurrecting disappearing languages and dialects, which would reduce costs for online course creation.

Cited Literature

Bregni, S. (2018). Assassin's Creed Taught Me Italian: Video Games and the Quest for Lifelong, Ubiquitous Learning." Profession – The Journal of the Modern Language Association.

Cambridge English Language Assessment. (2016). *English at Work: Global Analysis of Language Skills in the Workplace*.

Foreign Service Institute. Foreign Language Difficulty Rankings. https://www.state.gov/m/fsi/ .

Kern, R. (2014). Technology as Pharmakon: The promise and perils of the Internet for foreign language education. *Modern Language Journal*, 98, 340–357.

Larsen-Freeman, D. (2018). Looking ahead: Future directions in, and future research into, second language acquisition. *Foreign Language Annals*. https://doi.org/10.1111/flan.12314

Ratzlaff, N. (2015). A Cross Comparison and Efficacy Study of Duolingo and an Entry-Level German 1A Course. Master's thesis, Fresno State University. https://csufresno-dspace.calstate.edu/bitstream/handle/10211.3/147680/RATZLAFF_Niklas.pdf?sequence=1

Rosen, D. (2014). eLearning Future: What Will eLearning Look Like in 2075. *https://elearningindustry.com/elearning-future-what-will-elearning-look-like-2075* .

Sacco, S.J. et al. (2015). Toward a Learning Management System to Teach *Safety English* to Immigrant Workers Employed by U.S. Agribusinesses: A Proposal to GAI, December 17, 2015.

Sacco, S.J. (2018). A Proposal to Design an Online Program for Southern Italian Dialects. A paper to be presented at Diaspore Italiane, Italy in Movement: A Symposium on Three Continents, Queen's College, NY: November 1-3, 2018.

Strochlic, N. (2018). The Race to Save the World's Disappearing Languages. https://news.nationalgeographic.com/2018/04/saving-dying-disappearing-languages-wikitongues-culture/

Corporate Online Language Learning Market in the US to Grow at a CAGR of Over 16% Through 2021 Reports Technavios. https://www.businesswire.com/news/home/20170704005286/en/Corporate-Online-Language-Learning-Market-Grow-CAGR

Twarog, L. & Perezlenyi-Pinter, M. (2011). Telephone-Assisted Language Study at Ohio State University: A Report. *Modern Language Journal* 72(4):426 – 434.

Vesselinov, R. & Grego, J. 2009-2018, nine studies, http://comparelanguageapps.com/ .

THIS PAGE LEFT INTENTIONALLY BLANK

MOTIVATION AND LANGUAGE PROFILE

By Mila Tasseva-Kurkchieva, PhD[7]

Motivation

By some estimations 80% of world population speaks 2 or more languages regularly. A good portion of those 80% have learned the language later in life, in an instructed environment. Starting with Krashen (1976) and continuing with Bley-Vroman (1990, 2009), a lot of research has been devoted to the difference between language acquisition (the typical native language apprehension) and language learning (the typical case of an instructed second language attainment), the most crucial difference between the two being that language acquisition is thought to be unconscious, effortless, and uninfluenced by learner-internal characteristics while language learning is conscious, effortful, and highly dependent on learner-internal characteristics such as age, sex, attitude towards the language and associated culture, integrative abilities, and last but not least motivation.

Learner motivation has been in the focus of second language (L2) acquisition research for more than 3 decades now. The early research (c.f., Gardner, 1985) was focused on the big divide between learners' integrative and instrumental orientation, the former being associated with desire to integrate with the L2 community, positive attitudes towards the learning situation, and active approach to learning, the latter having to do with association of the L2 with potential pragmatic gains. This big divide soon proved to be too strict and unaccommodating for the multitudes of issues associated with motivation. Later developments in our understanding of the issue involved orientation towards the source of motivation (learner vs. teacher/instructor), the time and context in which motivation is studied (beginning, middle, or end of the learning process), the social factors interacting with motivation (attitudes of the learner and the target language community towards each other), temporal characteristics of the motivation (for how long and how persistent). With the rise of post-structuralism and the developments of the socio-cultural and sociolinguistic theories, the views on motivation also shifted from motivation as a constant, once and for all set learner-internal characteristic, to a fluid, ever changing characteristic, affected by a number of other variables.

Our current understanding of this complex and fluid concept, influenced primarily by work by Dörnyei and colleagues (Dörnyei, 2001, 2005, 2009, Csizér & Dörnyei, 2005 to name a few) takes motivation to be a cognitive process, a psychological predisposition towards learning an L2, influenced by/projected as a number of factors. The earliest, and possibly most obvious, components of motivation are *integrativeness* and *instrumentality* (Gardner, 1985). The former reflects willingness to become an integral part of the target language society and culture, the latter is associated with perceived pragmatic benefits from learning and using the L2. Another concept associated with the motivational milieu is *attitudes towards the L2 speakers/community* which takes different forms but is generally associated with desire, or lack thereof, of meeting and communicating with speakers of the target language. A fourth component is the *cultural interest* associated with interest in deeper understanding of the culture through music, movies or direct contact with speakers. Another component is the *ethnolinguistic vitality of the target language community* which further breaks down into social, political and economic stability of the community, size and distribution of the community, representation of the language and culture in the media, educational system and government. Following the contribution of this last factor, recent studies compare motivation to learn languages of high international prestige such as English to

[7] Research Associate Professor, Program Director, Linguistics Program, University of South Carolina, tassevak@mailbox.sc.edu

motivation to learn less commonly taught languages. The final commonly included learner-external component is what Csizér & Dörnyei (2005) call *milieu*, or the influences stemming from the social networks of the learner and the attitudes which other (speakers or non-speakers) of the L2 have and impose on the learner. Among the learner-specific factors, *linguistics self-confidence* and *ideal self* are the two that stand out as most strongly contributing to motivation to learn an L2. Note that both are based on the learner's perceived ability to learn, stay on task and achieve goals. These components are the key factors comprising Dörnyei's L2 Motivational Self System, a questionnaire and testing tool which is currently among the most widely used systems in motivational research.

In Dörnyei's own words, motivation is "only indirectly related to learning outcomes/achievement because it is, by definition, an antecedent of behavior rather than of achievement" (Csizér & Dörnyei, 2005), yet, motivation has been studied widely as one of the most crucial contributing factors to a successful L2 acquisition in the classroom. Far less is known about the effects of motivation on L2 acquisition in an exclusively Computer Assisted Language Learning (CALL) or mixed classroom-CALL environment (Vesselinov & Grego, 2009-2018). While it is true that technological advances have changed the way, learners interact with new languages: information is readily available, new learning systems provide immediate feedback, and, notably, there is an immediate reward for achievements, language learning in any form is predicted to still be highly influenced by learner-internal and leaner-external factors such as age, sex, attitudes, aptitude, and motivation.

Language Profile

Language shares a number of the general characteristics of the cognitive system, the most important of which is that a necessary but not sufficient condition for its development is exposure to the language that is being acquired. This exposure can come in different forms. In native language acquisition it is naturalistic, unplanned, and unmodulated. In second language acquisition, it is, in the typical case, classroom-based (broadly speaking), planned, and consciously modified to support the development of the language skills. The influence of the quantity and quality of exposure to the language becomes apparent not only in the final outcomes of the learning experience but also through the internal factors affecting language learning.

A complicating factor in second (L2), and any subsequent, language acquisition, whether it starts in the pre-school years or later in life, is that the subsequent language comes after the native language system has been built at least partially. As such, the two languages influence each-other throughout the lifespan of the speaker with the correlation being stronger when the initial exposure came earlier in life and continued for most of the lifespan, and weaker when the subsequent language was acquired later in life and used for a shorter period of time.

Once a third (L3) or subsequent language is added, we are facing an even more complex relationship. In both theory and practice, both the first and the second language can and do influence the acquisition of the third one. This influence is never unilateral and its strength depends on, among other factors, time, length, and type (classroom vs. naturalistic) of exposure to the three languages, typological closeness between the pairs of languages (L1-L3 and L2-L3), attained skills level in L2 and L3. Relevant to the current study is knowledge of, achieved proficiency in, and recency of acquisition of any of the Romance languages, e.g., Portuguese, Italian, French, etc., which share both grammatical and lexical features with the target language and might influence the outcomes of learning.

Given that multilingualism is the worldwide norm, rather than exception, a well-controlled study of subsequent language acquisition needs to account for the differences between prior monolingualism and bilingualism, and the types of bilingualism with which the learners come to the new language. For the purposes of establishing these connections, Birdsong and colleagues (Birdsong et al, 2012) have created the Bilingual Language Profile (BLP), a survey-based assessment of language abilities and associations with each language and culture for bilingual speakers. What BLP does is that it asks exactly the same 19 questions once in one and then in the other language that the bilingual speaks. The responses in both languages are then scored and combined in the overall score for the speaker. This allows researchers to measure the hidden attitudes of the speakers towards the languages, as well as their perceived association with their cultures. The BLP has become a standard measure of language dominance for bilingual speakers in the field of second language acquisition.

The current study uses BLP not as a measure of language dominance but rather as a measure of knowledge of and exposure to other languages which may ultimately affect the outcomes of learning, namely the Romance languages. Participants completed only the English portion of survey and were asked about prior language-learning experiences and attitudes towards the languages they have been exposed to.

Cited Literature

Birdsong, D., Gertken, L.M., & 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/ .

Bley-Vroman, R. 1990. The Logical Problem of Foreign Language Learning. Linguistic Analysis, 20(1–2), 3–49.

Bley-Vroman, R. 2009. The Evolving Context of the Fundamental Difference Hypothesis. Studies in Second Language Acquisition, 31(Special Issue 02), 175–198.

Csizér, K. and Z. Dörnyei. 2005. The internal structure of language learning motivation and its relationship with language choice and learning effort. *Modern Language Journal*, 89(i), 19-36.

Dornyei, Z. 2001. *Teaching and researching motivation*. London: Longman.

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*, 9-42. Bristol, UK: Multilingual Matters.

Gardner, R. C. 1985. *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.

Krashen, S. 1976. Formal and informal linguistic environments in language learning and language acquisition. *TESOL Quarterly* 157-168.

Vesselinov, R. & Grego, J. 2009-2018, nine studies, http://comparelanguageapps.com/ .

THIS PAGE LEFT INTENTIONALLY BLANK

RESEARCH DESIGN

The random sample for this study was drawn from existing or new Rosetta Stone users residing in the U.S. There were some additional requirements for the potential participants, who had to:

- Be willing to study Spanish using only Rosetta Stone for two months with at least two hours of study;
- Take two sets of language placement tests;
- Be at least 18 years of age;
- Be novice or beginner learners of Spanish.

The last requirement was in place because the written language placement test used in the study has limited abilities for assessing very advanced users. Its highest evaluation group is equivalent to college Semester 4+. The age limitation was in place because the language test was designed as college placement test.

Sample size and power analysis.

We based our power and sample calculations on the typical results from our previous nine studies. We designed the study to test the hypothesis that the language improvement measured as test points gain per hour of study is statistically significant. We hypothesized an average efficacy improvement of 10 points per one hour of study and at most 30 points standard deviation (std). This is a conservative estimation given the last study results with italki in 2018. The estimated[8] sample size was at least 97 participants. With this sample size we can test the null hypothesis with 5% level of significance (Alpha) and 90% statistical power. We also incorporated an expectation of about 30% drop-out rate. This way our initial sample size was designed to be at least 150 people.

The Spanish language was selected as one of the most popular languages and because of the existence of previous research on Spanish for other language learning apps (Vesselinov & Grego, 2009-2018). The study lasted approximately 8 weeks and it was conducted between June 2018 and August 2018. Participants who successfully completed the study were given two free subscriptions for any language on the Rosetta Stone platform for two years. No monetary or other incentives were offered to the participants.

[8] SAS 9.4 PROC POWER

The main instrument for evaluating the level of knowledge of Spanish was the Web based Computer Adaptive Placement Exam[9] (WebCAPE) test. This 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 on their website[10].

The 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 a 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:

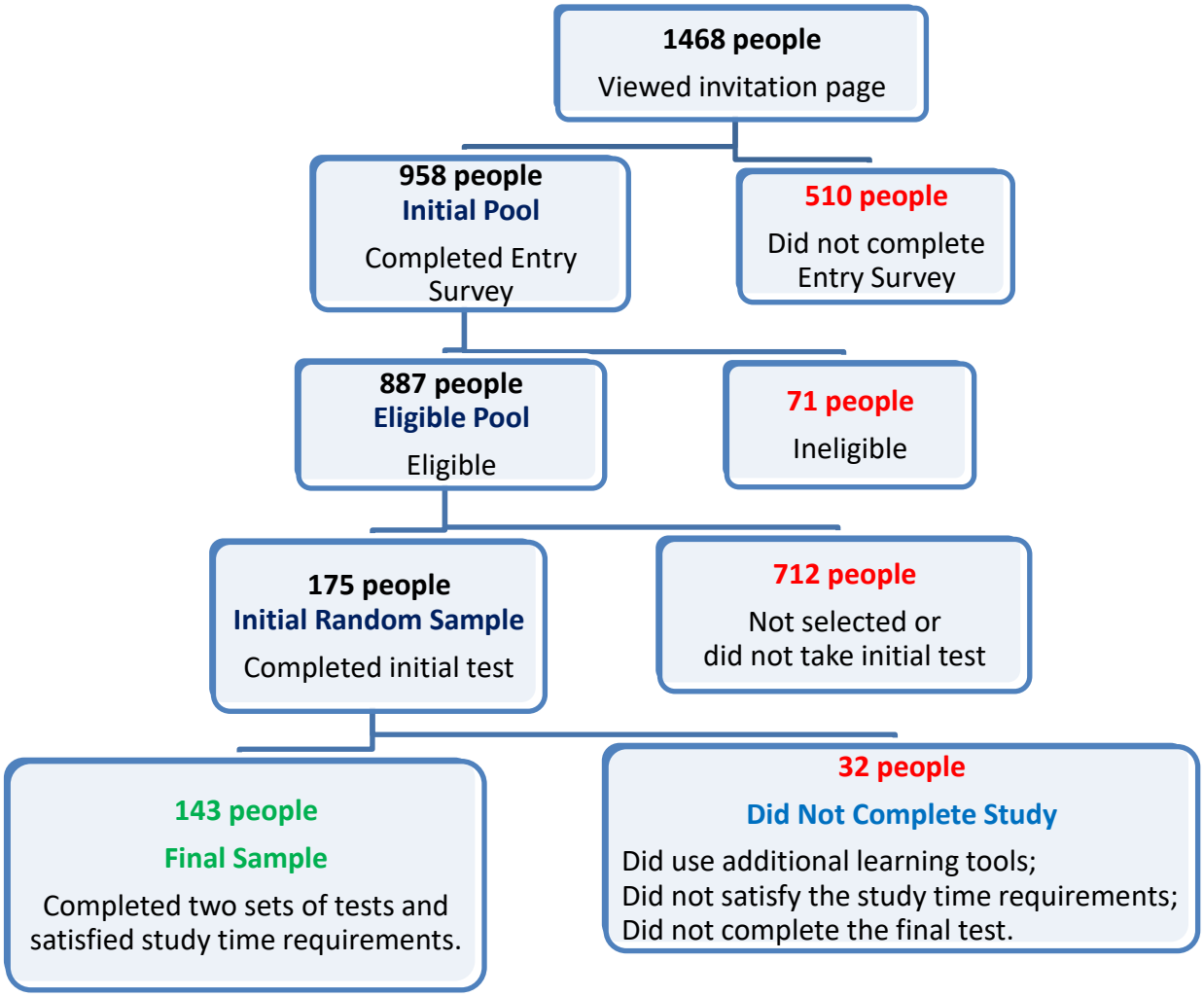
$$Efficacy = \frac{\text{Effect}}{\text{Effort}} = \frac{\text{Improvement of language skills}}{\text{Study time}} = \frac{\text{Final-Initial WebCAPE test score}}{\text{Hours of study}}$$

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. This is a direct and objective measure 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 time recorded on the computer servers (instead of self-report).

[9] 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.
[10] https://perpetualworks.com/

Figure 1. Sample Selection Tree



FINAL STUDY SAMPLE

We randomly selected 200 people of the eligible pool of participants and 175 of them completed the initial language test. They constituted our initial random sample (N=175). Rosetta Stone study continued for approximately two months (8 weeks), starting in June 2018 and ending in August 2018. During the study the Research Team sent weekly e-mail reminders to the participants with information detailing the amount of time they had used Rosetta Stone each week. 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 test.
- People who used additional learning tools during the study.

All participants were instructed at the beginning of the study that they could use only Rosetta Stone to study Spanish for the duration of the study. In the exit survey two people stated that they had regularly used other language apps and they were excluded from the study. Other people had occasionally used internet dictionaries, YouTube and translation websites and they were allowed to stay in the study.

The final study sample consisted of 143[11] people who had used only Rosetta Stone, with at least two hours of study and valid initial and final WebCAPE tests.

Final Study Sample Versus Not Completed

From the initial random sample (N=175) only 32 people (18.3%) did not complete the study for different reasons: participants who did not satisfy the study time requirements; participants who did not take the final test; and participants who used additional learning tools during the study. The 18.3% dropout rate is relatively small in this line of research.

We compared the two groups, the final sample of 143 people and the 32 people who did not complete the study by gender, age, education, employment status, initial knowledge of Spanish (initial WebCAPE score) and reason for studying Spanish. There were no statistically significant differences (at 5% error), which means that participants who did not complete the study were not very different from the ones that did, and they did not introduce a bias.

[11] Some people declined to answer specific survey questions, so the number of answers can be less than 143.

Sample Description

In the final study sample about 56% were female. The age varied from 20 to 71 years of age, with a mean age of 40 years. The pool was very well educated with majority participants having some college, undergraduate or graduate degree.

Figure 2. Age Distribution

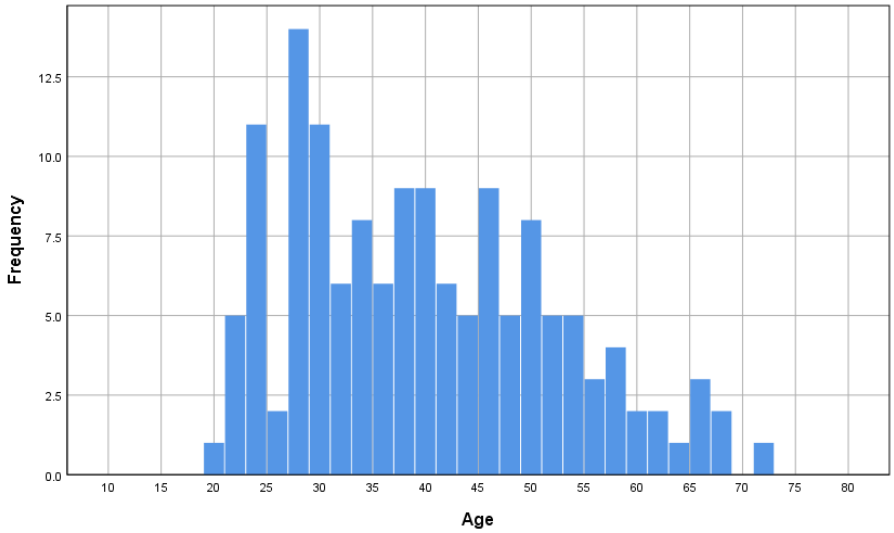


Table 2. Age and Gender Distribution

Age	Female (N)	Male (N)	Total (N)	Total (%)
18-20 years old	0	1	1	0.7
21-30 years old	23	20	43	30.1
31-40 years old	27	11	38	26.6
Over 40 years old	29	31	60	42.7
Total	**79**	**63**	**142**	**100**

Table 3. Education

Category	N	Percent
2. HS diploma or equivalent	11	7.9
3. Some college but no degree	48	34.3
4. College graduate, BA or equivalent	45	32.1
5. Some graduate school but no degree	7	5.0
6. Master's degree (MA, MS)	21	15.0
7. PhD/MD/JD	8	5.7
Total	**140**	**100**

The majority of the people were employed full time (73%), followed by part time employment (11%), etc.

Table 4. Employment Status

Category	N	Percent
1. Employed full time	103	72.5
2. Employed part time	16	11.3
3. Homemaker	6	4.2
4. Student	7	4.9
5. Retired	5	3.5
6. Unemployed	2	1.4
7. Other employment	3	2.1
Total	**142**	**100**

Ninety-two percent of the participants were English native speakers and the rest included the native speakers of the following languages: Farsi, German, Hindi, Korean, Mandarin, Navajo, Polish, Russian, Tagalog, and Vietnamese.

All participants in the final sample described themselves as beginner users or never studied Spanish. About 11% of the respondents' spouse, partner, or close friends spoke Spanish. About 9% of their parents, grandparents, or great-grandparents spoke Spanish.
About 43% of the final sample had formally studied a foreign language before (mostly at high school or college).

About 15% have lived outside U.S. in a non-English speaking country for more than 6 months. About 16% of the respondents were raised in a multilingual or non-English speaking household. The primary reason for studying Spanish was personal interest (44%), followed by business or work (41%), travel (11%), etc.

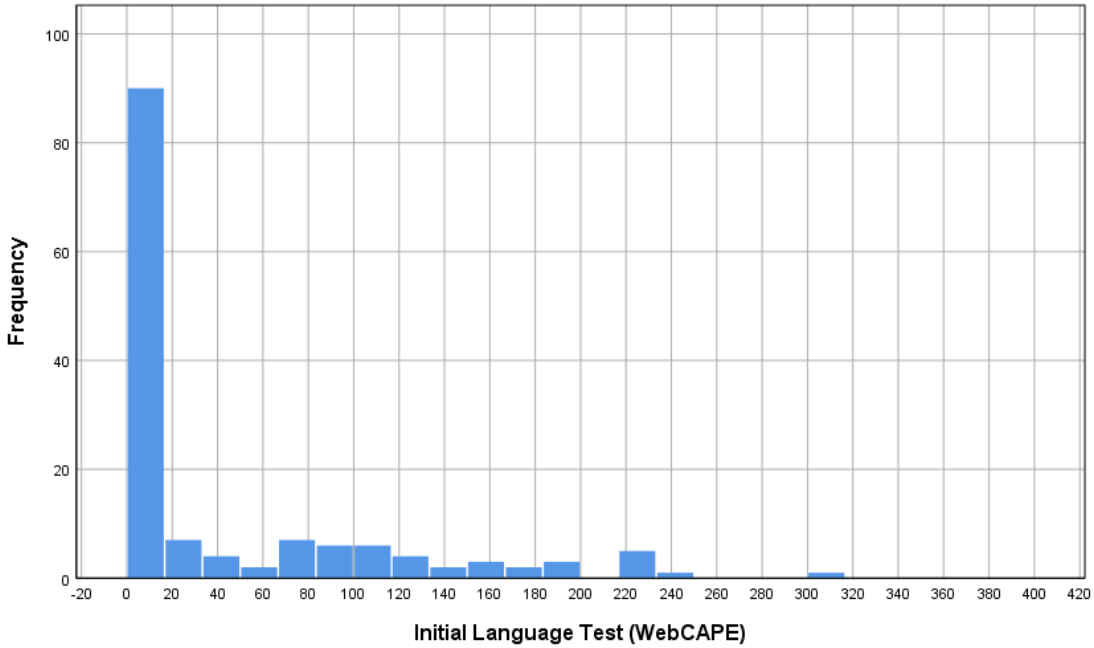
Table 5. Reason for Studying Spanish

Category	N	Percent
1. Personal Interest	62	43.7
2. Business/Work	58	40.8
3. Travel	16	11.3
4. School	1	0.7
5. Other	5	3.5
Total	**142**	**100**

Initial Language Test

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

Figure 3. Initial Language Test Distribution (WebCAPE points).



As expected, the majority of the participants (58%) scored zero on the initial language test. One participant was placed in semester 2 and one in semester 3 by the WebCAPE test. The overall mean WebCAPE score was 42.0 (std=68.5) corresponding to first college semester of Spanish.

Motivation

All participants completed a motivation scale in the beginning of the study. The idea was to evaluate the role of motivation on efficacy. In the previous Rosetta Stone motivation study (Vesselinov et al., 2009a and 2009b) we used a set of classical motivation scales: Academic Motivation Scale (Vallerand et al., 1992), Intrinsic Motivation Inventory (Deci et al., 1994) and the classical Attitude Motivation Test Battery (AMTB), (Gardner, 1985).

For this study we adopted a more modern motivation scale approach largely based on the second language (L2) motivational self-system (Dörnyei, 2005, 2009) which stems largely 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 pushes them from societal obligation or a fear of failure (ought-to L2 self).

We adopted a specific 33 questions, 6 factors version of L2 Motivational Self System (see Appendix, Table A2) 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 are users determined to make and how hard they are ready to study.

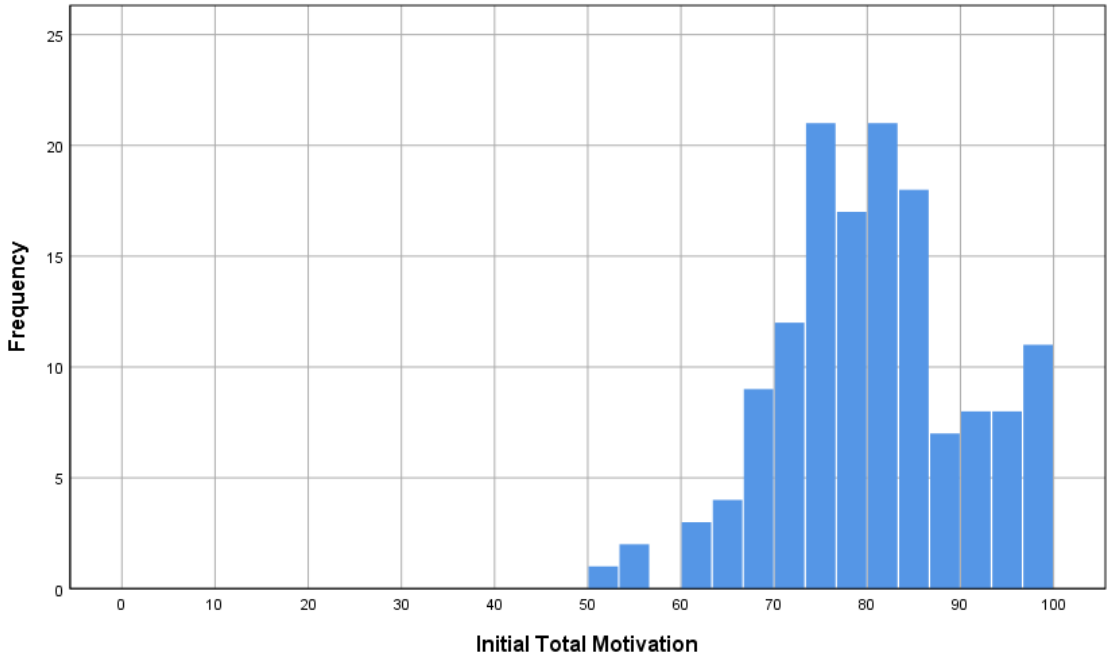
As we can see from Figures 4 to 10 below, the participants vary a lot on initial motivation. The scale dimensions were recoded, so the maximum motivation is equal to 100%. It is obvious that very few of the participants were 100% motivated.

Table 6. Initial Motivation Levels (%) Max=100

Motivation Dimensions	1st Quartile[12]	Median[13]	3rd Quartile[14]
1. Ideal Self	70.0	82.5	95.0
2. Ought-to-Self	51.4	62.9	72.1
3. International Posture	70.0	80.0	86.7
4. Competitiveness	76.7	83.3	90.8
5. Learning Attitude	80.0	90.0	100.0
6. Intended Effort	80.0	86.7	96.7
Total Motivation	**73.8**	**80.4**	**86.5**

The initial average level of total motivation is very high (Me=80%). From the motivation elements the highest level (90%) 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 (63%) which suggests that the participants were not very afraid of failure or they were not that susceptible to pressure from societal obligation.

Figure 4. Initial Total Motivation Level (%) **Max=100**



As noted above the average initial level of motivation is very high (Me=80%) and most people are overall highly motivated. Only a handful of people have motivation level less than 60%.

[12] First 25% of the sample.
[13] 50% middle point.
[14] First 75% of the sample.

INITIAL MOTIVATION LEVELS:

Figure 5. L2 Motivation (Max=100)
M1 "Ideal Self"

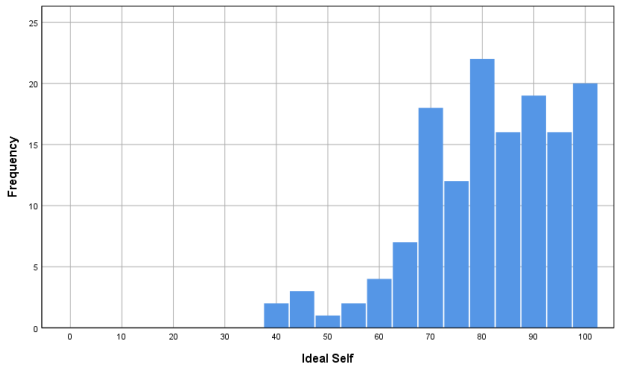


Figure 6. L2 Motivation (Max=100)
M2. "Ought-to-Self"

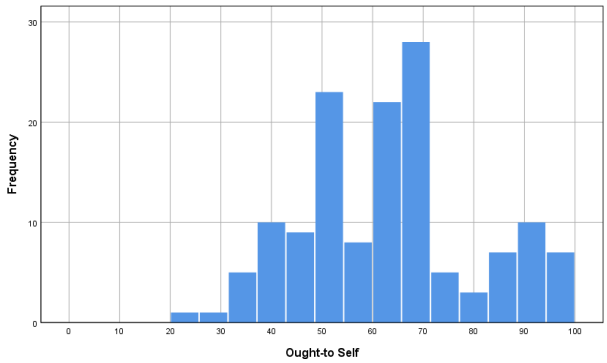


Figure 7. L2 Motivation (Max=100)
M3. "International Posture"

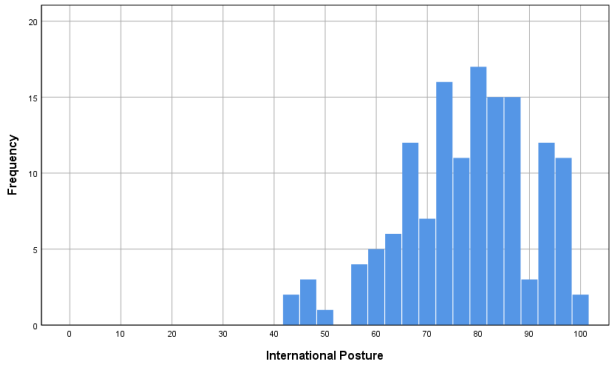


Figure 8. L2 Motivation (Max=100)
M4. "Competitiveness"

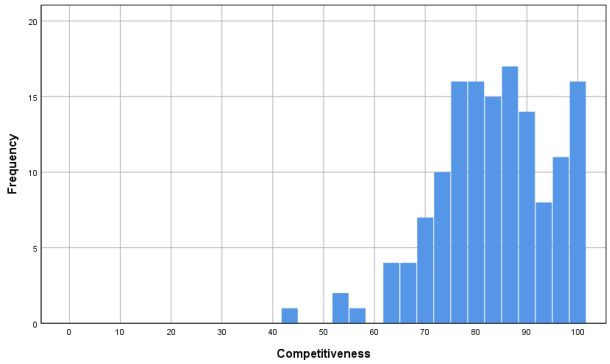


Figure 9. L2 Motivation (Max=100)
M5. "Learning Attitude"

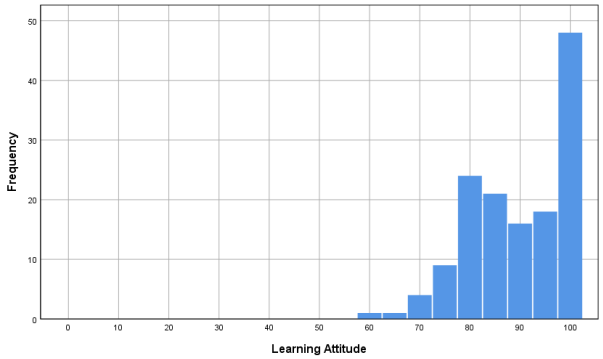
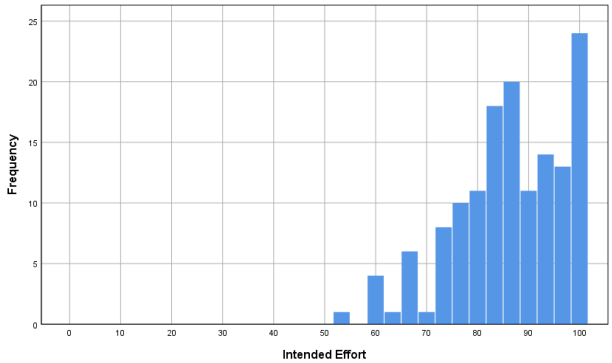


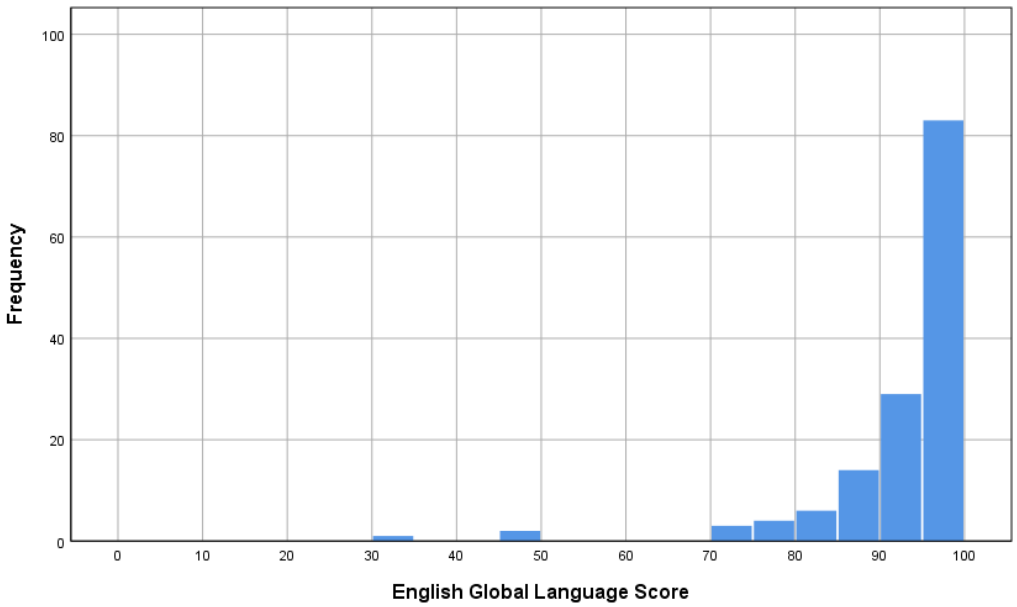
Figure 10. L2 Motivation (Max=100)
M6. "Intended Effort"



Language Profile

We asked participants to complete an adapted version of Bilingual Language Profile (Birdsong et al., 2012). A Global Language Score (GLS) for English and for a second language for participants with a second language in use. 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 recoded it so the maximum is equal to 100.

Figure 11. GLS score for English Language (N=142) Max=100



The median GLS percent was 96.2 (IQR[15]=6.3) which corresponds to initial sample of strong English native speakers. Twenty-eight participants or 20% of the sample, felt comfortable enough to complete GLS for their second non-English language.

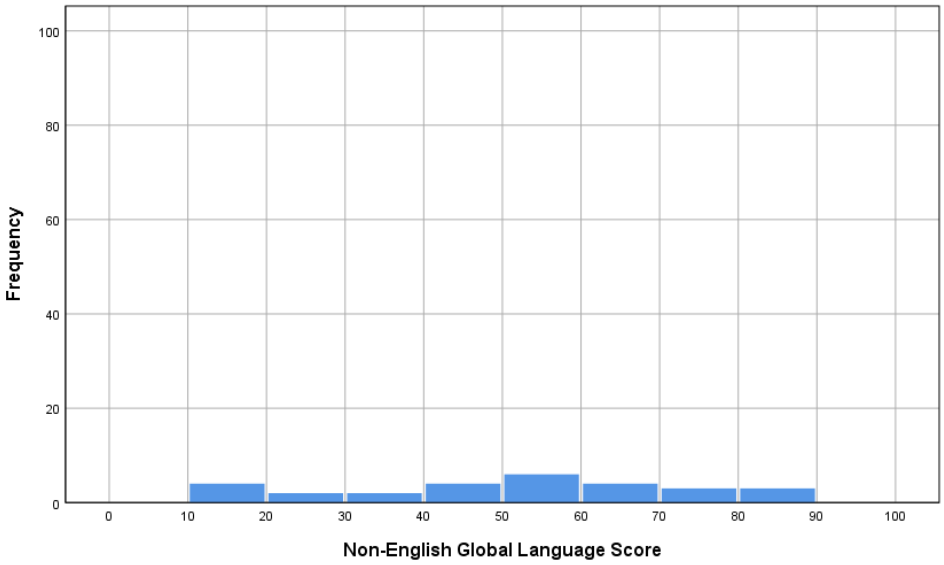
For example, a GLS score of 218 (100%) would be appropriate for participants born in English speaking family, in an English-speaking country, who started studying English immediately, for whom all classes at school were in English, who speak only English all the time with family, friends, and at work.

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.

[15] Interquartile Range = 3rd Quartile – 1st Quartile

GLS score will lose some points if the participants started learning English at older age; or some of the school teaching is in another language; or they use another language to speak with some of their friends and family, etc.

Figure 12. GLS score for Language other than English (N=28) Max=100

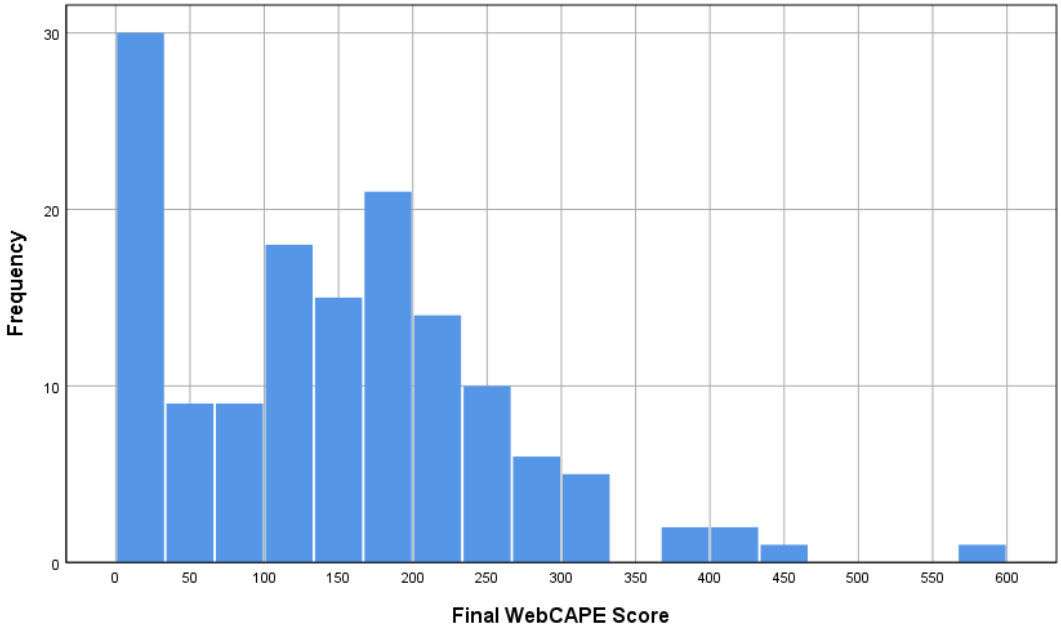


The median GLS percent for language other than English (n=28) was 53.4% (IQR=35.5%).

Final Language Test

The participants in the study completed the same language test at the end of the study.

Figure 13. Final Test Distribution (WebCAPE points).



The overall average WebCAPE score from the second test was 145.2 (std=109.9) but several people reached levels of 400 points and above.

Final Motivation Levels

Participants completed the same motivation scale (Kong et al., 2018) at the end of the study.

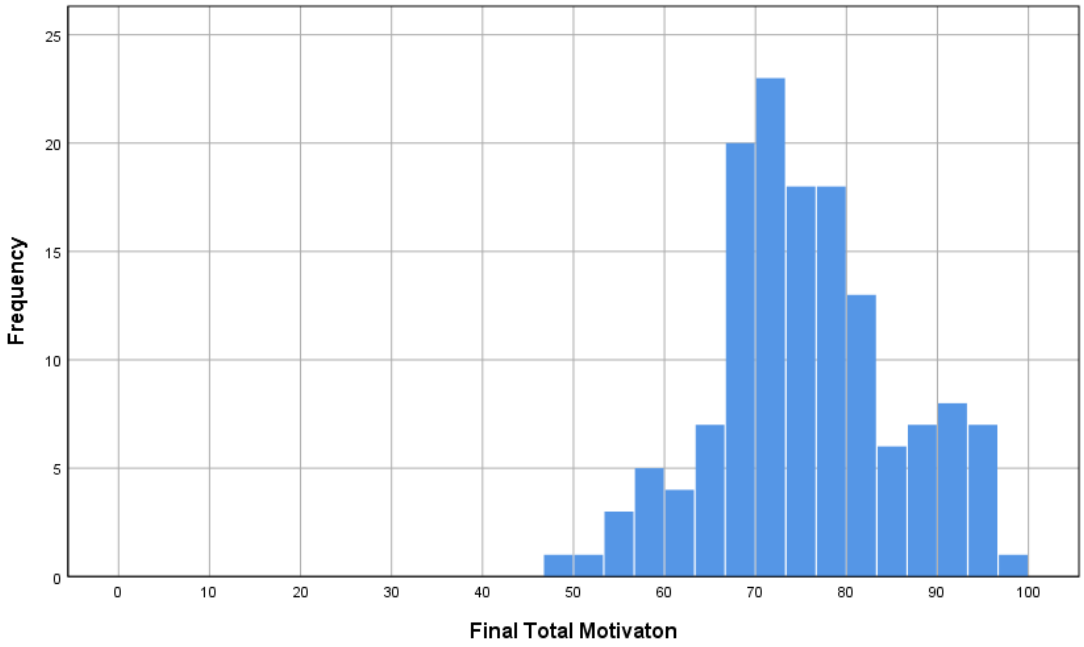
Table 7. Final Motivation Levels (Percent) Max=100

Motivation Dimensions	1st Quartile	Median	3rd Quartile
1. Ideal Self	70.0	80.0	90.0
2. Ought-to-Self	45.7	60.0	68.6
3. International Posture	63.3	73.3	83.3
4. Competitiveness	73.3	83.3	90.0
5. Learning Attitude	75.0	80.0	95.0
6. Intended Effort	73.3	80.0	86.7
Total Motivation	**69.3**	**75.1**	**81.8**

The total motivation level after two-months of studying the language remains remarkably high (Me=75%), although it is slightly below the initial level of motivation (Me=80%). The champion is “Competitiveness” (83%) with “Learning Attitude”, “Ideal Self”, and “Intended

Effort" as close second with 80% motivation level. Still, at the end of the study the participants were not very afraid of failure or societal obligations with "Ought-to-Self" 60% level of motivation.

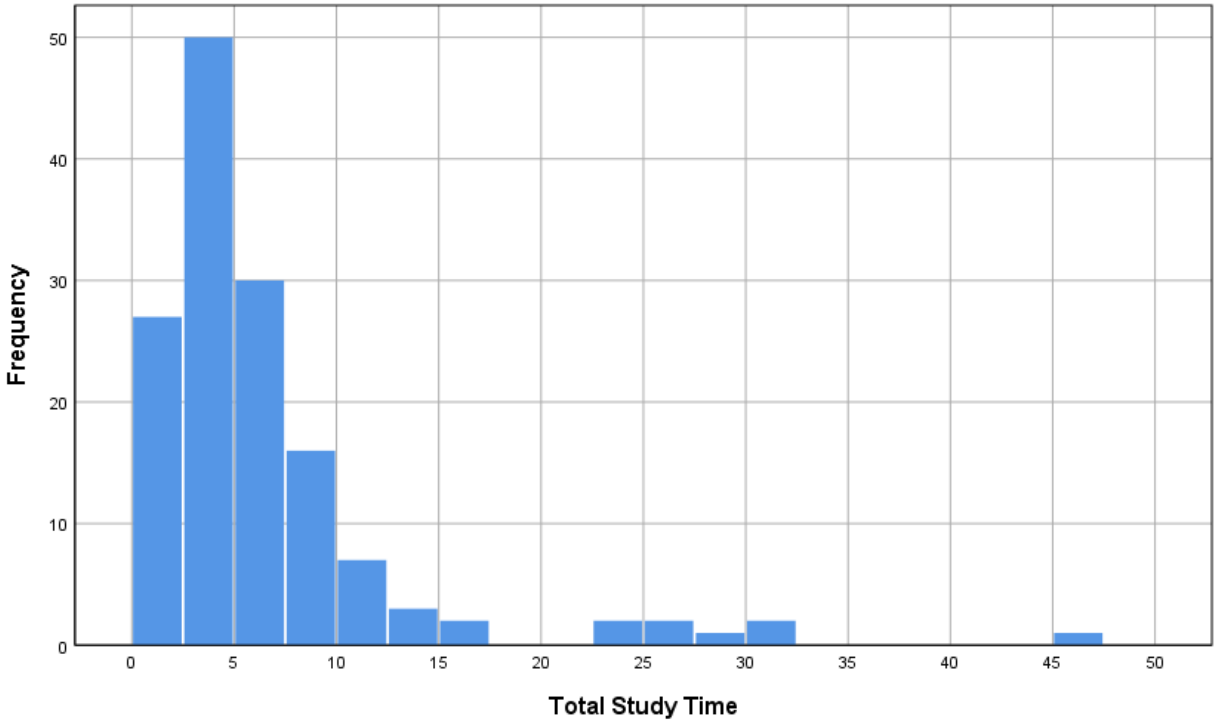
Figure 14. Final Total Motivation Levels (%) Max=100



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 average study time for the final study sample (N=143) was about seven hours, or a little less than one hour a week. The study time varied from about two hours to 46 hours.

Figure 15. Study Time Distribution Hours



LANGUAGE IMPROVEMENT

WebCAPE Test Results

All participants took 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 one.

Table 8. Language Improvement

WebCAPE Test Points

Statistics	**Initial WebCAPE**	**Final WebCAPE**	**Improvement (Final-Initial)**
Mean (std)	42.0 (68.5)	145.2 (110.0)	103.3 (121.7)
Median (IQR)	0 (80)	144 (158)	99 (174)
95% Confidence Interval[16]	31.2 – 53.5	127.4 – 163.7	83.6 – 123.1

The average overall improvement of 103 WebCAPE test points was statistically significant with a 95% Confidence Interval from 84 to 123 points. This means that the improvement in the language proficiency for the final sample was statistically significant (at 5% error). Overall 88% of all participants retained or improved their language proficiency with 95% Confidence Interval[17] of 81.7% to 92.5%.

Only 17 participants out of 143, or 12% decreased their score. There are two plausible explanations for this fact. First, some participants were more advanced learners of Spanish and gaining points at this higher level is generally more difficult and requires more time (see Fig. 16). Second, some participants studied irregularly with more efforts and more study time in the beginning of the study and less time towards the end of the study.

[16] Bootstrapped (N=10,000) confidence intervals.
[17] 95% CI with Agresti-Coull correction (Agresti & Coull, 1998).

College Semester Placement

We can measure progress by movement from one semester level to a higher semester level and the results are presented below.

Table 9. WebCAPE Semester Placement

College Semester	Initial Test		Final Test	
	People (N)	%	People (N)	%
First	142	99.3	127	88.8
Second	1	0.7	10	7.0
Third			3	2.1
Fourth+			3	2.1
Total	143	100	143	100

Participants at First Semester level decreased from 99.3% to 88.8% and the proportion in Second to Fourth+ Semester level increased notably.

Table 10. Semester Improvement

Level (Semester Change)	Improved		Study Time
	People (N)	%	Mean (Hours)
-1 Negative change	1	0.7	5.1
0 Same/No Change	126	88.1	6.4
1 One Semester Up	10	7.0	4.9
2 Two Semesters Up	3	2.1	8.2
3 Three Semesters Up	3	2.1	20.6
Total	**143**	**100**	**6.6**

The problem with the semester improvement 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 one-point progress is needed to move to Second semester. Another person can start with 10 points (First semester), then gain 200 points and the new level (210 points) is still First Semester. This measure is not very consistent for interpretation and it is presented here only for completeness.

EFFICACY

The main efficacy measures are presented below.

Table 11. Efficacy of Rosetta Stone

Statistics	**Efficacy = Improvement per one hour of study** WebCAPE Test Points	**Time to cover the placement requirements for first semester of college Spanish** Hours
Mean (std)	21.0 (32.5)	12.9[18]
95% Confidence Interval	15.7 – 26.4[19]	10.2 – 17.2[20]

On average Rosetta Stone users will gain 21 WebCAPE test points per one hour of study with 95% Confidence Interval of 16 to 26 test points per one hour of study.

The main measure of the Rosetta Stone 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 the first 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.

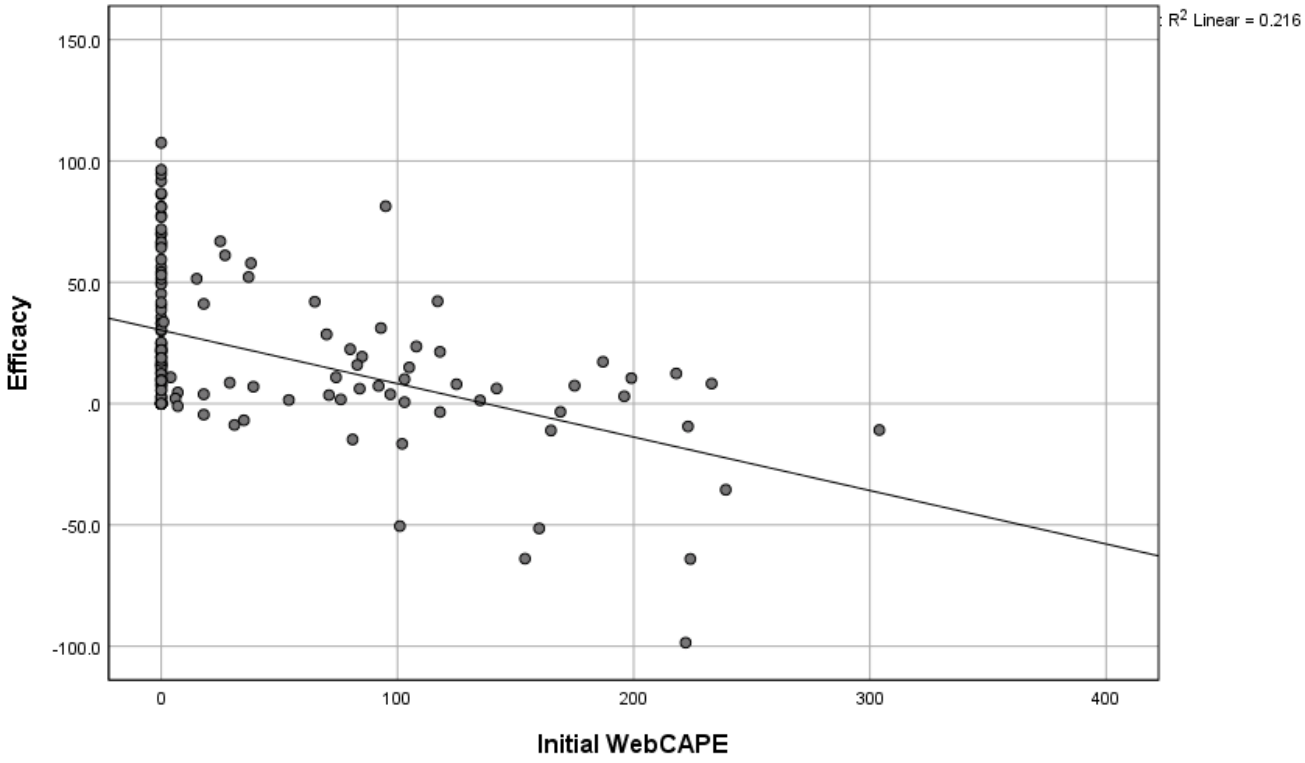
In other words, Rosetta Stone users will need on average about 13 hours of study during a two-month period to cover the requirements for the first college semester of Spanish.
The transformed lower and upper limits are from 10 to 17 hours of study during a two-month period.

[18] The threshold of 270 points divided by the mean efficacy (21 points).
[19] Bootstrapped (N=10,000) confidence interval.
[20] The threshold of 270 points divided by the lower limit (15.7) and the upper limit (26.4) of the 95% CI.

FACTORS FOR EFFICACY

Initial Level of Knowledge of Spanish

Figure 16. Effect of initial Level of Knowledge on Efficacy



The efficacy is the highest for participants with no or very little knowledge of Spanish and it decreases for more advanced learners. The biggest study effect is typical for people with very low starting point.

Demographic Factors

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

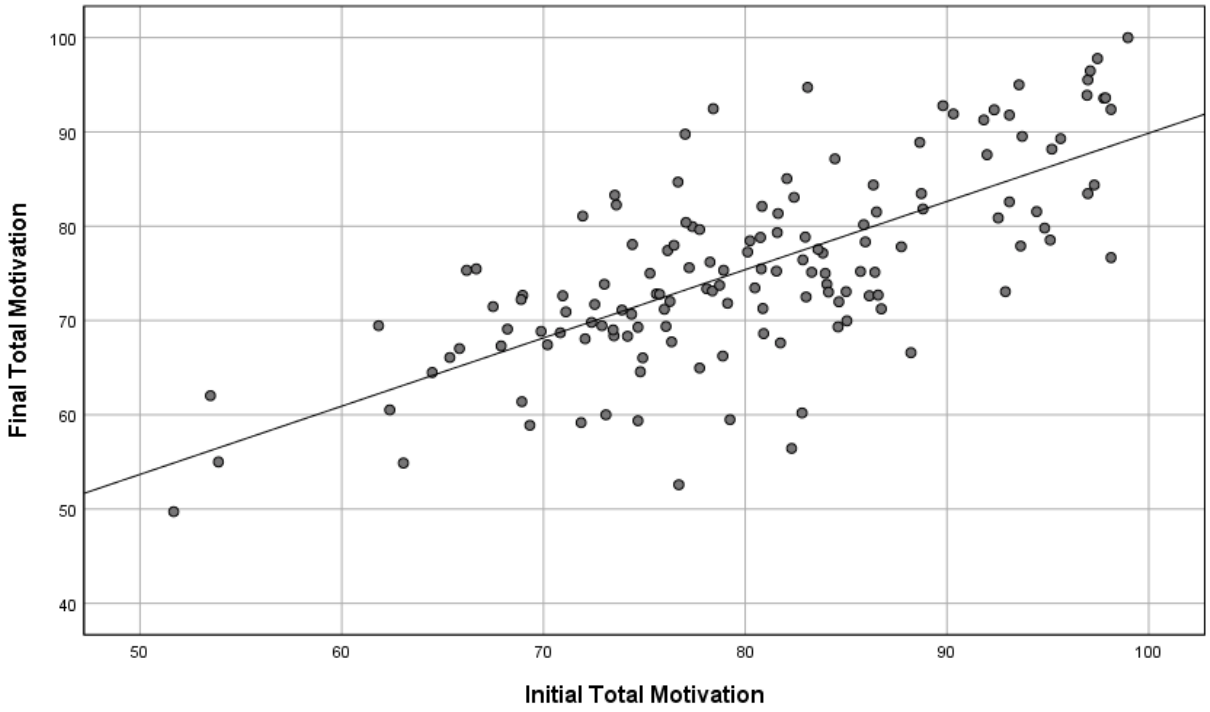
None of these potential factors had a statistically significant effect on the efficacy (p=.05). In some instances, the number of cases by subgroups was too low to expect enough statistical power for the test of hypotheses.

This means that the Rosetta Stone app works similarly well for people with different gender, age, native language, education, employment status, etc.

Motivation Effect

Initial and final motivation were highly correlated as expected and as factor only the initial motivation level will be considered.

Figure 17. Initial and Final Motivation



Our working hypothesis was that the more motivated participants would exhibit higher efficacy than less motivated. It turned out that this relationship is complex and not linear. The standard correlation and regression coefficients were not statistically significant. In order to discover the nature of this relationship we applied Classification and Regression Tree (CART) models (Breiman et al, 1984; Zhang & Singer, 2010). CART models are nonparametric recursive partitioning models and they are very effective in discovering interaction terms and partitioning variables in groups according to a specified outcome. The CART model results for the effect of the initial total motivation on efficacy are presented below.

Table 12. Motivation Effect on Efficacy

Factor/Outcome	**Group M1 N=37**	**Group M2 N=91**	**Group M3 N=15**
Total Motivation	**Below Average** (≤ 74%)	**Average** (>74 – 95%)	**High** (>95%)
Efficacy Mean (std)	**High** 32 (35)	**Low** 15 (30)	**High** 32 (31)
Initial WebCAPE Mean (std)	53 (74)	43 (70)	10 (27)
Study Time (hours) Mean (std)	6 (6)	6.6 (5.7)	8 (11.4)

The lowest efficacy was exhibited by participants who had about average motivation. They were neither highly motivated nor less motivated. Please note that for this sample the lowest level of total motivation was 52%. This could be interpreted as uncertainty in their level of motivation given the overall level of motivation for these participants.

The lower motivated with high efficacy Group M1 had relatively high initial level of knowledge in Spanish (mean WebCAPE=53) and they studied about the average time while the high motivated Group M3 had a very low starting point (mean WebCAPE=10) but studied the most (8 hours).

The results for the six different elements of motivation (Ideal Self, Ought-to-Self, etc.) were similar to the total motivation (data not shown).

Language Profile Effect

Similar to the motivation, the effect of English Global Language Score (GLS) on efficacy is not linear in nature. The CART model revealed the following relationship.

Table 13. English Language Profile Score Effect on Efficacy

Factor/Outcome	**Group P1 N=31**	**Group P2 N=28**	**Group P3 N=71**	**Group P4 N=13**
GLS	**Below Average** (≤ 91)	**Average** (>91 – 95)	**High** (>95-99)	**Extremely High** (>99)
Efficacy Mean (std)	**Low** 12 (23)	**High** 42 (35)	**Average** 20 (32)	**Low** 2 (26)
Initial WebCAPE Mean (std)	63 (82)	23 (44)	40 (68)	45 (78)
Study Time (hours) Mean (std)	7.7 (8)	4.9 (3)	6.7 (7)	7.4 (8)

The lowest efficacy (only 2 test points per study hour) exhibited people with almost perfect (>99%) English language score. This can be explained by the fact that these people had almost no exposure to a foreign language in their life so learning a new language may be not easy. People with the highest efficacy (42 points per study hour) had English GLS between 91% and 95%. These people had 5%-10% exposure to foreign language and that makes learning a new one easier.

On the other hand, lower English GLS (≤ 91) seems to be an impediment to learning a new language with relatively low efficacy (12 points per study hour).

The results for the components of the language profile score (language history, use, proficiency, and attitudes) were similar to the total score (data not shown).

USER SATISFACTION

After the study the participants were asked for their opinion about Rosetta Stone, specifically how easy it was to use, how helpful, enjoyable, and satisfactory. The 5-point Likert scale was recoded into two categories: Strongly Agree/Agree vs Strongly Disagree/ Disagree/Neutral.

Table 14. User Satisfaction (N=143)

	Percent	
Do you agree with the following statement?	**Strongly Disagree/ Disagree/Neutral**	**Agree/ Strongly Agree**
"Rosetta Stone was easy to use"	3.5	96.5
"Rosetta Stone was helpful in studying Spanish"	6.3	93.7
"I enjoyed learning Spanish with Rosetta Stone"	6.3	93.7
"I am satisfied with Rosetta Stone"	11.2	88.8

After two months of study, the overwhelming majority of users (89% to 97%) agreed with the positive statements that: Rosetta Stone was easy to use, helpful, they enjoyed learning with Rosetta Stone and were satisfied with it.

Almost all (99.3%) of the respondents in the exit survey declared that they will continue to use Rosetta Stone after the study ends.

In the exit survey a special question was included: "How likely are you to recommend Rosetta Stone 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 indicator for the company.

From our exit survey the "Promoters" were 58.7%, the "Detractors" were 7.7% and "Passives" were 33.6%. The Rosetta Stone NPS was +51.

LIMITATIONS OF THE STUDY

The population of people who are seeking to study foreign language with language app is highly educated with majority of them having college level education and above. This is true not only for the U.S.[21], but also Europe[22] and the rest of the world[23]. This was confirmed by all our previous studies[24]. This population has higher education level than the general population. Our current sample for the 2019 Rosetta Stone study is representative of this population and it should not be compared to the general population.

The WebCAPE test used in this study is not tailored to any specific learning tool, including Rosetta Stone. On the one hand, some participants in the study complained that the test sometimes contained words or expressions that were not part of their regular course with Rosetta Stone. 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, their progress evaluation of language proficiency is generally conservative.

For future studies we highly recommend in addition of WebCAPE, to include other language tests like the Oral Proficiency Interview by Computer® (OPIc)[25] created by Language Testing International (LTI). LTI is the exclusive licensee of the American Council on the Teaching of Foreign languages (ACTFL).

There are some limitations of the study, mostly related to the instruments and technological limitations. The online WebCAPE written test measures the progress of beginner/novice users of Spanish well, but it is not very 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 as true beginners (WebCAPE score of zero) gained much faster per study hour than people who started at higher level.

---

[21] Rosetta Stone (2009, 2019), Duolingo (2012), italki (2018)
[22] Babbel (Germany & US), Busuu (UK and US).
[23] New Language App, 2015 report, (world sample).
[24] Except Hello English (2017) where the participants were of high school age.
[25] http://www.languagetesting.com/oral-proficiency-interview-by-computer-opic

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. The average study time was a little less than one hour 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 one hour a week for two months.

The study results could be generalized for studying Spanish with Rosetta Stone. For other languages the results could be markedly different. But the results could be generalized for non-English native speakers studying Spanish residing in the US or other countries.

The results of this study or the nine previous studies, cannot be compared to a standard college semester of Spanish for at least two reasons. 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. The WebCAPE test is only used as initial placement test and not for determining course grades. The progress is measured very differently in a college setting compared to our ten studies so far. Second, the study time at college is difficult to measure scientifically and it is not the same for everybody.
If there is a need to compare to the first college semester of Spanish, a new study can be designed to satisfy the efficacy definition. To the best of our knowledge such a study has not been done yet and until then the two sets of results cannot be compared scientifically.

There are not enough 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.
Probably in near future new language apps will be required to present independent efficacy test to their customers.

CONCLUSION

The Rosetta Stone efficacy study is based on a random sample of 143 people, 18 years of age or older, residing in the U.S. All participants were self-reported novice/beginner users of Spanish.

The main goal of measuring the efficacy of Rosetta Stone was achieved with this study. The results show that, on average, one hour of study with Rosetta Stone alone leads to an improvement of 21 points on the college placement test WebCAPE. There is a lot of variability of the efficacy and the 95% confidence interval is between 16 and 26 points per hour.

In other words, Rosetta Stone users would need on average 13 study hours in a two-month period to complete the requirements for the first college semester of Spanish. The transformed upper and lower limits are between 10 and 17 hours of study. These results are not generally valid in different context: they cannot imply, for example, 13 study hours for a shorter than two-month period. Also, the results are not necessarily valid for languages other than Spanish.

The main factor for the efficacy is the initial level of language knowledge of the participants. The novice/beginner users gain points faster than the more advanced users.

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/ .

Breiman, L., Friedman, J., Stone, C., Olshen, R., 1984. Classification and Regression Trees, Wadsworth Statistics/Probability.

Deci, E., Eghrari, H., Patrick, B., Leone, D. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, 62, 119-142.

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.

Gardner, R., 1985. Social psychology and second language learning: The role of attitudes and motivation. London: Edward Arnold.

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.

Vallerand, R., Pelletier, L., Blais, M. Brière, N., Senécal, C., Vallières, E., 1992. The academic motivation scale: a measure of intrinsic, extrinsic, and amotivation in education, *Educational and Psychological Measurement*, 52, 1003-1017.

Vesselinov, R. and Grego, J., 2018, italki Efficacy Study.
http://blog.italki.com/wp-content/uploads/2017/12/italki2018FinalReport.pdf or
http://comparelanguageapps.com/documentation/italki2018FinalReport.pdf

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

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

Vesselinov, R. and Grego, J., 2016, 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

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

Vesselinov, R. and Grego, J., 2012, Duolingo Effectiveness Study. http://comparelanguageapps.com/documentation/DuolingoReport_Final.pdf, or http://static.duolingo.com/s3/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/MeasuringTheAttitudeandMotivationofRSUsers.pdf, or http://resources.rosettastone.com/CDN/us/pdfs/Measuring_the_Effectiveness_RS-5.pdf.

Zhang, H., Singer, B., 2010. Recursive Partitioning and Applications, Springer Series in Statistics.

APPENDIX

Table A1. Study Participants' Geographic Distribution: US States

Number of people

	State	ST	Initial Pool	Eligible Pool	Initial Sample	Final Sample
1	Alabama	AL	8	8		
2	Alaska	AK				
3	Arizona	AZ	22	21	8	7
4	Arkansas	AR	6	6	1	1
5	California	CA	123	122	27	21
6	Colorado	CO	28	28	5	4
7	Connecticut	CT	15	14	4	4
8	Delaware	DE	1	1		
9	Florida	FL	72	68	12	9
10	Georgia	GA	40	39	2	2
11	Hawaii	HI	4	4		
12	Idaho	ID	3	3		
13	Illinois	IL	28	27	4	3
14	Indiana	IN	10	10		
15	Iowa	IA	11	10		
16	Kansas	KS	3	3		
17	Kentucky	KY	8	8		
18	Louisiana	LA	8	8	3	1
19	Maine	ME	1	1		
20	Maryland	MD	27	25	5	5
21	Massachusetts	MA	15	13	1	1
22	Michigan	MI	27	27	3	2
23	Minnesota	MN	7	6	3	1
24	Mississippi	MS	4	4		
25	Missouri	MO	12	12	5	3
26	Montana	MT	3	3	1	1
27	Nebraska	NE	10	9	3	3
28	Nevada	NV	7	7	1	1
29	New Hampshire	NH	2	2	1	1
30	New Jersey	NJ	17	17	9	7
31	New Mexico	NM	6	6	1	
32	New York	NY	39	39	9	6
33	North Carolina	NC	37	37	5	4
34	North Dakota	ND				
35	Ohio	OH	17	17	2	2

Table A1. Continued

	State	ST	Initial Pool	Eligible Pool	Initial Sample	Final Sample
36	Oklahoma	OK	7	7	1	1
37	Oregon	OR	14	14	2	1
38	Pennsylvania	PA	26	26	4	4
39	Rhode Island	RI	1	1	1	1
40	South Carolina	SC	10	9	2	1
41	South Dakota	SD	2	2		
42	Tennessee	TN	13	13	2	2
43	Texas	TX	86	83	18	16
44	Utah	UT	16	16	7	7
46	Virginia	VA	25	25	5	4
45	Vermont	VT	1	1	1	1
47	Washington	WA	14	14	3	3
49	Wisconsin	WI	15	15	3	3
48	West Virginia	WV	5	4		
50	Wyoming	WY	2	2		
	District of Columbia	DC	3	3		
	Unknown state (but US)		47	47	11	10
	Outside US		50			
Total			**958**	**887**	**175**	**143**

Table A2. Motivation Scale

Developed by Kong et al., 2018.

A. Ideal L2 self (4 items)

1. I can imagine myself living abroad and having a discussion in Spanish.
2. I can imagine myself speaking Spanish with international friends or colleagues.
3. I can imagine myself speaking Spanish as if I were a native speaker of Spanish.
4. Whenever I think of my future career/life, I imagine myself using Spanish.

B. Ought-to L2 self (7 items)

1. I study Spanish because close friends of mine think it is important.
2. Learning Spanish is necessary because people surrounding me expect me to do so.
3. I consider learning Spanish important because the people I respect think that I should do it.
4. Studying Spanish is important to me in order to gain the approval of my peers/teachers/family/boss.
5. It will have a negative impact on my life if I don't' learn Spanish.
6. Studying Spanish is important to me because an educated person is supposed to be able to speak it.
7. Studying Spanish is important to me because other people will respect me more if I have knowledge of it.

C. International posture (6 items)

1. I want to make friends with foreigners visiting U.S.
2. I would feel somewhat uncomfortable if a foreigner moved in next door. (reverse-coded)
3. I want to participate in a volunteer activity to help foreigners living in the surrounding community.
4. I am interested in an international career/living abroad.
5. I often read and watch news about foreign countries.
6. I have thoughts that I want to share with people from other parts of the world.

D. Competitiveness (6 items)

1. I want to survive in the future.
2. I don't want to be an illiterate person.
3. I want to succeed in life.
4. Other people will consider me an elite if I have a good command of Spanish.
5. I don't want to place behind any of my friends.
6. I want to have a head start on other people.

E. L2 learning Experience or Attitudes (4 items)

1. I like the atmosphere of my Spanish classes with Rosetta Stone.
2. I find learning Spanish with Rosetta Stone really interesting.
3. I always look forward to Spanish classes with Rosetta Stone.
4. I really enjoy learning Spanish with Rosetta Stone.

F. Learners' Intended Effort or Motivated Behavior in L2 Learning (6 items)

1. If Spanish course were offered in the future, I would like to take it.
2. I expend a lot of efforts in learning Spanish.
3. I do my best to learn Spanish.
4. I spend lots of time studying Spanish.
5. I concentrate on studying Spanish more than any other topic.
6. Compared to other people I know, I think I study Spanish relatively hard.

Table A3. Language Profile

Developed by Birdsong et al., 2012.

I. Biographical Information

II. Language history

In this section, we would like you to answer some factual questions about your language history.

1. At what age did you **start learning** English?

   Since birth 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

2. At what age did you **start to feel comfortable** using English?

   As early as I can remember 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+ not yet

3. How many years of **classes (grammar, history, math, etc.)** have you had in English (primary school through university)?

   0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

4. How many years have you spent in a **country/region** where English is spoken?

   0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

5. How many years have you spent in a **family** where English is spoken?

   0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

6. How many years have you spent in a **work environment** where English is spoken?

   0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20+

III. Language use

In this section, we would like you to answer some questions about your language use.

7. In an average week, what percentage of the time do you use English **with friends**?

   0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

8. In an average week, what percentage of the time do you use English **with family**?

   0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

9. In an average week, what percentage of the time do you use English **at school/work**?

   0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

10. When you talk to yourself, how often do you **talk to yourself** in English?

    0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

11. When you count, how often do you **count** in English?

    0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

IV. Language proficiency

In this section, we would like you to rate your language proficiency.

	0=not well at all						*6=very well*
12. How well do you speak **English**?	0	1	2	3	4	5	6
13. How well do you understand **English**?	0	1	2	3	4	5	6
14. How well do you read **English**	0	1	2	3	4	5	6
15. How well do you write **English**?	0	1	2	3	4	5	6

V. Language attitudes

In this section, we would like you to respond to statements about language attitudes.

	0=disagree						*6=agree*
16. I feel like myself when I speak **English**.	0	1	2	3	4	5	6
17. I identify with an **English-speaking** culture.	0	1	2	3	4	5	6
18. It is important to me to use (or eventually use) **English** like a native speaker.	0	1	2	3	4	5	6
19. I want others to think I am a native speaker of **English**.	0	1	2	3	4	5	6

THIS IS THE LAST PAGE

OF THE 2019 ROSSETA STONE EFFICACY STUDY

********** No text bellow this line **********