A TEST OF SUBJECTIVE VERSUS OBJECTIVE LANGUAGE PROFICIENCY:
AN INVESTIGATION OF AGE OF ACQUISITION AND EXPOSURE IN
MULTILINGUALS

By

Anastina B. Steiber

A Thesis Presented to
The Faculty of California Polytechnic State University Humboldt
In Partial Fulfillment of the Requirements for the Degree
Master of Arts in Psychology: Academic Research

Committee Membership
Dr. Kauyumari Sanchez, Committee Chair
Dr. Christopher Walmsley, Committee Member
Dr. Brandilynn Villarreal, Committee Member
Dr. Amber Gaffney, Program Graduate Coordinator

December 2023
Abstract

A TEST OF SUBJECTIVE VERSUS OBJECTIVE LANGUAGE PROFICIENCY:
AN INVESTIGATION OF AGE OF ACQUISITION AND EXPOSURE IN
MULTILINGUALS

Anastina B. Steiber

This research used the BEST dataset gathered in 2017 by de Bruin and colleagues, which contains information gathered from 650 multilingual participants (435 female, $M_{\text{age}} = 25.02$) from the Basque Country at the Basque Center on Cognition, Brain and Language (BCBL). Multilingual individuals may have different levels of proficiency in each language they speak, and there are various factors that can influence proficiency. Such factors include the age at which they begin learning a language and how much exposure they have to the language. Proficiency can be assessed both subjectively (personal perception of proficiency) and objectively (observed and/or measured performance on proficiency tests). While most research on these topics have focused on bilingualism and relied heavily on subjective measures of proficiency, this study expands the literature by investigating the role of age of acquisition and exposure on multilingualism using both subjective and objective measures of proficiency. This thesis demonstrated the importance of taking into consideration factors beyond age of acquisition that may impact language proficiency, such as exposure, and supported the need to use a wide range of objective tests when measuring language proficiency.
Acknowledgements

I would like to express my deepest gratitude to my thesis advisor, Dr. Mari Sanchez, without whom this project, and my completion of Cal Poly Humboldt’s Master’s program would not be possible. Additionally, I am indebted to Dr. Amber Gaffney, who has aided and supported me through my entire graduate degree. I would also like to thank my committee members, Dr. Christopher Walmsley and Dr. Brandilynn Villarreal for their time and support. Finally, thank you to Karisa Guettler and Tyler Erickson for all the emotional and IT support.
# Table of Contents

Abstract ....................................................................................................................................... ii

Acknowledgements .................................................................................................................. iii

List of Tables ........................................................................................................................ vi

Introduction .............................................................................................................................. 1

Proficiency ................................................................................................................................ 2

Subjective Measures .......................................................................................................... 3

Objective Measures ............................................................................................................. 3

Results ....................................................................................................................................... 5

Age of Acquisition ................................................................................................................. 13

Exposure ................................................................................................................................. 16

Current Study ......................................................................................................................... 19

Methods .................................................................................................................................... 20

Participants .............................................................................................................................. 20

Measures ................................................................................................................................ 20

Age of Acquisition ................................................................................................................. 20

Exposure ................................................................................................................................. 20

Subjective Proficiency ........................................................................................................ 20

Objective Proficiency ........................................................................................................ 21

Procedure ................................................................................................................................ 21

Tables .................................................................................................................................... 22

Results .................................................................................................................................... 24
Hypothesis 1 ................................................................................................................. 24
Hypothesis 2 ................................................................................................................. 25
Hypothesis 3 ................................................................................................................. 28
Discussion ......................................................................................................................... 31
Limitations ...................................................................................................................... 33
Conclusion ....................................................................................................................... 36
References ....................................................................................................................... 37
List of Tables

Table 1 .............................................................................................................................................. 22
Table 2 .............................................................................................................................................. 23
Introduction

Research in the psychological field regarding multilingualism has picked up within just the past 40 years. There are many explanations for this increased need for research on multilingualism, one of particular prominence being globalization. Globalization is comprehensively described as an increased connectedness of people around the world “driven by technological innovations mainly in the field of […] communication technology, and resulting in new patterns of global activity, community organization and culture” (Blommaert, 2010, p. 13). The growing ease of communication between people around the world has led many to learn new languages and continue to pass the languages down to their family. Being multilingual is different for everyone; one person may study for years learning new languages while another may grow up speaking multiple languages. Each multilingual person may have different levels of proficiency in each language they speak, and there are various factors that can influence proficiency. Such factors include the age at which they begin learning a language and how much exposure they have to the language. Proficiency can be assessed both subjectively (personal perception of proficiency) and objectively (observed and/or measured performance on proficiency tests). While most research on these topics have focused on bilingualism and relied heavily on subjective measures of proficiency, this paper will expand the literature by investigating the role of these factors on multilingualism using both subjective and objective measures of proficiency. This thesis investigates the relationship between subjective and objective language proficiency as impacted by age of acquisition and exposure.
Proficiency

Language proficiency, for the purpose of this paper, refers to the ability to speak with ease in any context “with a native-like fluency and pronunciation” (Angela de Bruin, Manuel Carreiras, & Jon Andoni Duñabeitia, 2017). Most studies use subjective measures (Hulstijn, 2012), while only a few (de Bruin et al., 2017; Kałamala, Senderecka, & Wodniecka, 2022; Macbeth, Bruni, de La Cruz, Erens, Atagi, Robbins, Chiarello & Montag, 2022) have incorporated the use of objective measures. Language proficiency is difficult to measure since there is no set standard for what exactly it means to be proficient in a language. While using subjective measures, such as self-report, to determine proficiency is convenient, it leaves a lot of room for personal interpretation and possibly inconsistent data. This does not mean that self-report is completely unreliable, but the literature on bilingual and multilingual individuals may be expanded by using more objective measures to determine language proficiency.

de Bruin et al. (2017) compiled a multilingual dataset that incorporates both subjective and objective measures, and will be discussed thoroughly due to its relevance with the current thesis. de Bruin and colleagues collected data from 650 multilinguals using several objective tests along with a self-report questionnaire. All participants were gathered from the Basque Country in Spain and are trilingual in Spanish, Basque and English, with Spanish and Basque being learned before age six and English being learned at or before age 12. Participants’ data was gathered over the course of 18 months using both online and in-person collection techniques. Online data included a subjective test of language proficiency (i.e., questionnaire) and an objective test of language proficiency.
LEXICAL DECISION TESTS. In-person data collection emphasized objective language measures and included both picture naming tests and an interview with a multilingual linguist.

**Subjective Measures**

**Self-report Questionnaires.** The self-report questionnaire asked participants to rate their proficiency in each language (Spanish, Basque and English) “from 0 (‘lowest level’) to 10 (‘native or native-like level’)” (de Bruin et al., 2017). Participants were also asked to provide an estimate for their level of exposure to each language “on a scale from 0 (‘never’) to 100 (‘always’)” (de Bruin et al., 2017). Finally, the participants reported the age at which they first began learning/using each language. The self-report questionnaire aimed to address the participant's age of acquisition, subjective proficiency, and subjective exposure for Spanish, Basque and English, and was adapted from Marian et al.’s (2007) “Proficiency Questionnaire”.

**Objective Measures**

**LexTALE.** Next, the participants completed a LexTALE test (Lemhöfer & Broersma, 2012) which is a lexical decision test that has been demonstrated to be able to give a valuable assessment of one’s knowledge of a language. A lexical decision test presents a target item that is either a word (e.g. slain) or nonword (e.g. skave) and participants are asked to categorize whether the target item is a word or nonword (Meyer and Schvaneveldt, 1971). Proficiency scores were determined by the percentage of correct responses. Each participant took three versions of the test- the original English version, a Spanish version adapted by Izura et al. in 2014, and a Basque version
developed for the BEST dataset collection. Each version had between 60 and 90 items, with an equal number of real words and nonwords.

**Picture Naming Test.** To test expressive vocabulary, participants engaged in a picture-naming test modified from Gollan et al.’s Multilingual Naming Test (2012). Modifications allowed the test to be specific for Spanish, Basque and English, such that none of the images shown had names that are the same/similar between languages (cognates). The pictures were grouped into “categories, such as animals (24 items) or body parts (8 items)” (de Bruin et al., 2017). Participants were presented with all 65 images three times, beginning by naming all images in Spanish, then Basque, and finally English. A score was given for each language, ranging between 0 (all incorrect) and 65 (all correct). The researchers had a list of acceptable names in all three languages for each image; in order to receive 1 point per image the participants had to name each image in line with one of the pre-assessed acceptable names. For example, to receive a point for a picture of pants, the participants could name the image in Spanish: pantalón or pantalones, in Basque: praka, prakak, galtza, or galtzak and in English: trousers or pants. Higher scores represent greater proficiency in the language.

**Interview.** Finally, short interviews in each of the three languages were conducted for each participant by linguists who were highly proficient in English, and native speakers of Spanish and Basque. Each interview in each language was designed to ask a range of questions for the participant to demonstrate grammatical skill and proficiency within five minutes, for a total of 15 minutes. After the interviews, the
linguists rated proficiency “on a Likert-like scale from 1 (‘lowest level’) to 5 (‘native or native-like’)” (de Bruin et al., 2017).

**Results**

For the perceived proficiency questionnaire, nearly all participants rated themselves as being most proficient in Spanish, with a wider variety of proficiency in Basque and English. All participants estimated themselves as having at least 10% or more exposure to Spanish. Basque exposure varied from 0% to 80%, with the majority of participants rating themselves to have about 30% exposure. The majority of participants estimated they had 30% or less exposure to English. Interview scores for Spanish were perfect (5/5) for all participants, but Basque and English scores varied widely. All participants scored a 60 or higher (out of 65) on the Picture Naming test in Spanish, with more variation in scores for the other languages, but there were a larger proportion of high scores for Basque than English. The LexTALE test had the largest range in scores for all three languages, with Spanish and Basque having larger proportions of scores above 70%, and English having a larger proportion of scores below 70%. Across participants overall, the results showed the greatest rates of both subjective and objective proficiency for Spanish, proficiency for Basque and English ranged widely, but more participants demonstrated higher proficiency rates for Basque than English.

de Bruin et al. (2017) ran correlations between all subjective (i.e., age of acquisition, exposure, and self-rated proficiency) measures. For the subjective measure of age of acquisition, age of acquisition was significantly negatively correlated with self-rated proficiency for Spanish and Basque, and moderately significantly negatively
correlated for English. This means that for all three languages, the younger the participants were when they acquired the language, the higher they rated their proficiency for that language. Age of acquisition was significantly negatively correlated with exposure for Spanish and Basque, but not correlated for English. This means that the younger the participants were when they began using Spanish and Basque, the greater their estimate was of exposure to the language. For the subjective measure of exposure, exposure was significantly positively correlated with self-rated proficiency for all three languages. This means that the more exposure each participant had to each language, the higher they rated their proficiency for that language.

The correlations between subjective measures are mostly in agreement with one another. It is expected that the younger one is when they begin using a language, the more proficient they would consider themselves to be in that language. A younger age of acquisition was related to a greater amount of exposure for Spanish and Basque, but not for English. Although this may seem to put the correlations in disagreement, it is not a surprising finding for this particular sample of participants. All of the participants were gathered from the Basque country in Spain, where Spanish and Basque are the main languages spoken, so it is expected that their level of exposure would be lower for English than Spanish and Basque. It is also unsurprising that greater levels of exposure were related to greater self-rated proficiency for all three languages. Being exposed to and using a language more frequently allows the speaker to practice the language.

de Bruin et al., (2017) ran correlations between all subjective (i.e., age of acquisition, exposure, and self-rated proficiency) and objective (i.e., picture naming test,
LexTALE, and interview) measures. For the subjective measure of age of acquisition, age of acquisition was significantly negatively correlated with the picture naming test for Spanish and Basque, but not correlated for English. This means that the younger the participants were when they began using Spanish and Basque, the higher they scored on the picture naming test in each respective language. Age of acquisition was not significantly correlated with LexTALE scores for Spanish, significantly negatively correlated for Basque, and moderately positively correlated for English. This means that the age at which the participants started using Spanish did not change the score they received on the Spanish LexTALE test, the younger they began using Basque, the higher their score was for the Basque LexTALE test, and the younger they began using English, the lower their score was for the English LexTALE test. Age of acquisition was significantly negatively correlated with interview scores for Basque, and not significantly correlated for English (correlations for Spanish were not provided). This means that the younger the participants were when they began using Basque, the greater the proficiency they demonstrated in the Basque interview, but the age at which they began using English did not change their level of proficiency demonstrated in the English interview. For the subjective measure of exposure, exposure was significantly positively correlated with picture naming test scores for all three languages. This means that the more exposure the participants had to each language, the higher their score was on the picture naming test for each language. Exposure was not significantly correlated with LexTALE scores for Spanish, but was significantly positively correlated for Basque and English. This means that exposure level to Spanish did not change the participants’ score on the Spanish
LexTALE test, but the more exposure they had to Basque and English, the higher their score was for the Basque and English LexTALE tests. Exposure was significantly positively correlated with interview scores for Basque and English (correlations for Spanish were not provided). This means that the more exposure the participants had to Basque and English, the greater the proficiency they demonstrated in the Basque and English interviews. For the subjective measure of self-rated proficiency, self-rated proficiency was significantly positively correlated with picture naming test scores for all three languages. This means that the higher the participants rated their proficiency in each language, the higher their score was in each language’s picture naming test. Self-rated proficiency was significantly positively correlated with LexTALE scores for all three languages. This means that the higher the participants rated their proficiency in each language, the higher their score was in each language’s LexTALE test. Self-rated proficiency was significantly positively correlated with interview scores for Basque and English (correlations for Spanish were not provided). This means that the higher the participants rated their proficiency in Basque and English, the greater the proficiency they demonstrated in the Basque and English interviews.

The correlations between objective and subjective measures were not entirely in agreement with each other across languages. For example, each subjective measure correlated with picture naming scores differently across language both within and between measures, which suggests that subjective measures alone may not be reliable predictors of language proficiency.
de Bruin et al., (2017) ran correlations between all objective (i.e., picture naming test, LexTALE, and interview) measures. For the objective measure of the picture naming test, picture naming test scores were significantly positively correlated with LexTALE scores for all three languages. This means that the higher the participants’ scores were for each language’s picture naming test, the higher their scores were for each language’s LexTALE test as well. Picture naming test scores were significantly positively correlated with interview scores for Basque and English (correlations for Spanish were not provided). This means that the higher the participants’ scores were for the Basque and English picture naming test, the greater the proficiency they demonstrated in the Basque and English interviews. For the objective measure of the interview, interview scores were significantly positively correlated with LexTALE scores for Basque and English (correlations for Spanish were not provided). This means that the greater the proficiency they demonstrated in the Basque and English interviews, the higher their scores were on the Basque and English LexTALE tests.

The correlations between objective measures were in agreement across all three languages. All objective measures were positively correlated with one another, which is to say that the higher the participants scored on any given objective measure of proficiency, the higher they also scored on the others. This is to be expected because the more proficient one becomes in a language overall, the better they should be able to demonstrate their proficiency in multiple domains of that language. However, it should be noted that correlations for the Spanish interview were not provided, though this is
assumed due to the likelihood of being at ceiling levels of proficiency for this language, given that the study was conducted in Spain.

Overall, the correlations within a type of measure of proficiency, either within subjective or within objective measures were found to be in agreement. However, the correlations between subjective and objective measures were found to be less in agreement. This suggests that studies that have used subjective measures exclusively may be in high agreement to other studies that have used subjective measures, but that these studies may potentially differ from those that may rely exclusively on objective measures. Subjective measures are beneficial because they are efficient, but they are not ideal due to the variation in personal interpretation on part of the participant. Objective measures allow for a more robust and unbiased assessment of proficiency, but may not be ideal because they require more resources and time. It is likely that a combination of the two types of measures may provide a more comprehensive view of a person’s language proficiency.

Following the BEST study, de Bruin (2019) wrote a review of literature relevant to the discussion of bilingualism, executive functioning, and the multiple measures available for assessing language use and proficiency. In the review, it is acknowledged that proficiency is multifaceted, and can be difficult to examine as separate from other factors such as age of acquisition (de Bruin, 2019). In addition, proficiency may refer to several different aspects of language competency such as vocabulary, pronunciation, grammar, and so on, depending on how it is being measured. Due to the variability in defining proficiency, using a self-report measure leaves the interpretation up to the
participant. The review discusses how participant interpretation could be influenced by differences in language learning history and personal backgrounds. For example, one’s confidence in a second language could be influenced by who they compare themselves to when evaluating their proficiency. Those who learned their second language in an environment where that language is native, they may be more likely to rate themselves as being less proficient than those who learned their second language in an environment where the other speakers of that language do not speak it natively (de Bruin, 2019). This suggests that self-report measures alone could be unreliable due to differences in participants’ backgrounds even if they all speak the same native language and second language.

Just as the BEST data found relatively weak positive correlations between subjective and objective language proficiency, recent studies (Kalamala et al., 2022, Macbeth et al., 2022) have discussed the necessity of incorporating more than one measure of proficiency in language research. To assess proficiency and daily language use of 60 bilingual University of California, Riverside students, Macbeth and colleagues (2022) used subjective questionnaires, a battery of objective language knowledge tests, and an Electronically Activated Recorder (EAR). The participants spoke English (age of acquisition: 0-12 years) and a heritage language (language varying from participant to participant; age of acquisition: 0). While both subjective and objective heritage language proficiency were predictive of daily heritage language use, none of the correlations between measures were strong (Macbeth et al., 2022). This lends support to the argument for using a diverse range of both subjective and objective proficiency measures to be able
to capture the most authentic data about the personal experiences of those who speak
more than one language.

Using multiple measures, however, is often easier said than done. Kalamala et al.
(2022) demonstrate the challenge of trying to gather uniform data on such an individually
unique topic. Their team examined data from 171 Polish-English bilinguals who all live
in an environment that primarily uses their first language. The data reflected relationships
between measures of diversity in daily language use, self-confidence in language use,
vocabulary knowledge, age of acquisition, and subjective proficiency. Analyses showed
subjective proficiency to be positively correlated to vocabulary knowledge and language
use diversity, however, the correlations were only of moderate strength (Kalamala et al.,
2022). The authors discuss how this reflects the notion that self-rated proficiency should
be used in part with other objective proficiency measures. By combining subjective and
objective measures, researchers may be able to capture a more comprehensive
understanding of bilingual/multilinguals’ language proficiency.

In addition, Cilibrasi and Marková (2022) used the BEST dataset to investigate
how exposure and age of acquisition may predict proficiency in English as an additional
language. They use the term ‘additional language’ to refer to English since it is not
necessarily every participant’s second language. Their study specifically examined the
data related to the LexTALE test and interview as measures of language proficiency in
English. Their findings suggest that exposure is a better predictor of English proficiency
than age of acquisition for the participants in the dataset (Cilibrasi & Marková, 2022).
These results confer that level of exposure to an additional language may be more
important than, or at least mediate the importance of age of acquisition when gaining language proficiency. This thesis will discuss the relevance of age of acquisition and exposure, as well as examine the relationship between subjective and objective measurements of proficiency in the same dataset.

**Age of Acquisition**

Age of acquisition is a term used to describe the age at which one learns or acquires something; for the purpose of this paper, age of acquisition will refer to the age at which one considers themselves to have begun using a language (de Bruin et al., 2017). Originally proposed in 1959, Penfield and Roberts introduced the concept of the Critical Period Hypothesis, which claims that there is a specific age range that is ideal for learning new languages. Since this proposal, there have been a multitude of studies examining the effects of age of acquisition on proficiency. While many studies (Hartshorne, Tenenbaum, & Pinker 2018; Johnson & Newport, 1989; 1991) support the hypothesis, it is not completely cut-and-dry as to which aspects of proficiency age of acquisition impacts since there is no set way for all studies to measure proficiency. This thesis will discuss how age of acquisition has been observed in relationship to different types of language proficiency measures.

Although the BEST dataset’s (de Bruin et al., 2017) picture naming and LexTALE tests primarily assess vocabulary knowledge, the interview provides a broader assessment of language proficiency. The interview required the participants to demonstrate other aspects of language proficiency such as grammar and pronunciation, which have been explored in relation to age of acquisition by a number of other studies.
Age of acquisition has been found to affect the aspects of proficiency that relate to one’s ability to process and produce accented speech in a non-native language (Larraza et al., 2016, Saito, 2015). A 2016 study of bilingualism comparing simultaneous, early, and late age of acquisitions examined perception of accented second language (L2) speech and found that simultaneous and early age of acquisition bilinguals perform better when distinguishing between phonetic similarities in L2 speech than late age of acquisition bilinguals (Larraza et al., 2016). This demonstrates evidence that the younger someone begins to learn an additional language, the more likely they will be able to reach a level of proficiency that allows for processing speech in their L2 just as well as a native speaker. Furthermore, accentedness and speech accuracy in L2 oral production has been found to decrease as age of acquisition increases, suggesting that age of acquisition plays a role in the degree to which advanced bilinguals can reach native-like pronunciation as well (Saito, 2015).

Johnson and Newport (1989) investigated the relationship between age of acquisition and grammar performance. They found that participants who began using English as a second language at a young age performed better on grammatical tests of English than those who began when they were older. This difference, however, was only observed for ages up to puberty; after puberty, age no longer held a direct relationship with grammatical proficiency. These results were further supported in a follow up study done by the researchers in 1991 that suggests a wide range language learning
mechanisms undergo a decline in function after puberty (Johnson & Newport, 1991). These findings lend support for the possibility that the Critical Period Hypothesis may apply beyond the learning of a first language.

To refine the research around the effect of age of acquisition on grammar learning ability, Hartshorne and colleagues (2018) conducted analyses on a dataset collected from 669,498 participants. Their study had a uniquely large sample size and examined participants’ current age, age of acquisition, and years of use to reduce the amount of unexplained variability in their data. Their findings expanded on Johnson and Newport’s (1989) report of puberty being a cut off for advanced proficiency attainment in an additional language. Their study also found a steady decline in grammar learning ability with age, however, their findings suggest a later ceiling age of about 17.4 years (Hartshorne et al., 2018). This challenges the idea that the Critical Period Hypothesis is tethered to biological changes that occur with age, but may rather be a function of cultural changes that limit one’s ability to invest in learning an additional language.

Hartshorne and colleagues were not, however, the first to suggest the Critical Period Hypothesis may be flawed in its claim that age of acquisition dictates one’s success potential for learning languages. In 2015, researchers Janet Werker and Takao Hensch explored the literature around critical periods and speech perception development. Their discussion acknowledges the processes that operate during critical periods from infancy to allow us to become specialized in our first language, and in turn narrow our sensitivity to other languages. Rather than leaving it at that, though, they offer the theory that under the right circumstances, critical periods may be able to reopen later
in life. Rather than age itself, an individual’s life circumstances, motivations, and other factors may play a much larger role in language learning than previously posited.

Age of acquisition has played an essential role in our understanding of language learning for decades. Age has been linked to aspects of additional language proficiency such as speech processing (Larraza et al., 2016), pronunciation (Saito, 2015), and grammar (Johnson & Newport 1989, 1991, Hartshorne et al., 2018). Although age of acquisition is important to consider for one’s ability to reach proficiency in an additional language, it is not the only factor that impacts language learning. For example, all of the participants in the BEST dataset had an age of acquisition of younger than 12 years for all languages, but did not demonstrate identical levels of proficiency in each language (de Bruin et al., 2017). These inconsistencies may be explained by differences in individual life conditions such as language exposure. This thesis will continue to discuss and explore the role that language exposure has on proficiency.

**Exposure**

While age of acquisition has long been an area of interest in language research, exposure has become a highly considered variable when studying language proficiency. One’s level of exposure to a language is not easy to measure; in order to objectively determine exposure, one would need to constantly keep record of each time they used or came in contact with their target language. Considering the difficulty of keeping such a record during a participant’s everyday life, many studies allow participants to determine for themselves (often via self-report surveys) how exposed they are to the language of interest. For example, the BEST study asked participants to rate how frequently they
were exposed to each of their spoken languages “on a scale from 0 (never) to 100 (always)” to measure exposure (de Bruin et al., 2017). As previously discussed in the proficiency section, de Bruin’s (2019) review talks about how exposure is involved in language proficiency. She brings up the issue of defining exposure, and the weaknesses of using a single question measure, as was used in the BEST study (de Bruin et al., 2017). Exposure may also be discussed in terms of “language use” which may clarify that what is being measured is not just hearing and/ or seeing the language, but engaging with it and utilizing it as an individual. Depending on the type (Gullifer & Titone, 2019), context (Anderson, Mak, Chahi & Bialystok, 2018), quality and duration/ frequency (Paradis & Jia, 2017), exposure may have varying impacts on individuals’ proficiency outcomes.

To look more deeply at how exposure can influence one’s proficiency, Gullifer and Titone (2019) proposed a new way of examining exposure that is composed of measuring how much each language is used compared to others, and how diverse the daily use of each language is. This measure of exposure is called Language Entropy, wherein higher levels of language entropy are associated with “more balanced language usage and greater language diversity” (Gullifer & Titone, 2019). While language entropy provides a detailed description of the type of exposure one has with their languages, another domain of exposure is the context in which a language is used. In 2018, Anderson and colleagues conducted a study examining young adults’ bilingual experiences across several contexts. Language use was found to differ significantly between non-English use at home and non-English language use socially (Anderson et al., 2018). The distinction that was found between at home and social language use
suggests that language proficiency may present itself differently in different contexts, and therefore a single self-report measure may not be a detailed representation of proficiency.

Language use may also be expanded to consider how frequently one uses their languages, and the quality of the use. A 2017 longitudinal study done by Paradis and Jia, looked into which aspects of exposure are related to accelerated acquisition of language proficiency. The researchers followed 23 school aged Chinese-English bilingual children over the course of 4.5 to 6.5 years to measure their English proficiency in relationship to their English exposure. The aspects of exposure that had the largest influence on proficiency were duration of English exposure, English use at home, and quality of English exposure outside of school. Students with more exposure, more use at home and richer quality of use reached an English proficiency comparable to native speakers faster than those with less exposure, less use at home and a lower quality of use regardless of age of acquisition (Pardis & Jia, 2017). Although the study measured proficiency by comparing the students’ performance to their monolingual peers, the researchers conclude by suggesting that future measurements of proficiency instead be determined in the context of bilingual, rather than monolingual, development.

The current and forthcoming research around language exposure is actively challenging the use of a single self-report measure for obtaining information on language use, however the convenience of self-report remains a leading justification for its use. Although more detailed measures of exposure are important in the development of research around language proficiency, self-report measures still provide valuable
information. In an analysis of the BEST dataset (de Bruin et al., 2017), Cilibrasi and Marková (2022) were able to use the self-report data to determine exposure as a better predictor of language proficiency than age of acquisition. The current thesis will continue to examine the BEST dataset to further assess the impact of exposure and age of acquisition on both subjective and objective measures of proficiency.

**Current Study**

Based upon the previous research, this thesis investigated the relationship between subjective and objective language proficiency as impacted by age of acquisition and exposure. This thesis utilized the BEST dataset gathered in 2017 by de Bruin and colleagues to investigate the following hypotheses:

- **H1:** A younger age of acquisition, more exposure, and higher scores on all three objective proficiency measures will be predictive of higher self-rated proficiency scores for each respective language.

- **H2:** The interview will be the most predictive of age of acquisition, exposure, and self-rated proficiency compared to the picture naming test or LexTALE test for each respective language.

- **H3:** English proficiency (both subjective & objective) will be more impacted by Exposure and age of acquisition than Spanish and Basque.
Methods

Participants

The dataset contains information gathered from 650 participants (435 female, $M_{age} = 25.02$) from the Basque Country at the Basque Center on Cognition, Brain and Language (BCBL) (de Bruin et al., 2017). All participants are trilingual, speaking Spanish, Basque, and English. All participants report acquiring Spanish and Basque before age six ($M_{Spanish} = 0.67, SD = 1.55; M_{Basque} = 1.68, SD = 1.81$), and English by age 12 ($M = 6.37, SD = 2.49$).

Measures

Age of Acquisition

Participants reported the age (in years) that they began learning/using each language.

Exposure

Participants provided an estimate for their level of exposure to each language “on a scale from 0 (‘never’) to 100 (‘always’)” (de Bruin et al., 2017).

Subjective Proficiency

Participants rated their proficiency in each language (Spanish, Basque and English) “from 0 (‘lowest level’) to 10 (‘native or native-like level’)” (de Bruin et al., 2017).
**Objective Proficiency**

Participants completed 3 LexTALE tests: the original English version (Lemhöfer and Broersma, 2012), a Spanish version adapted by Izura et al. in 2014, and a Basque version developed for the BEST dataset collection.

Participants engaged in 3 picture-naming tests: Gollan et al.’s Multilingual Naming Test (2012) modified to allow each test to be specific for Spanish, Basque and English.

Participants completed short interviews in each of the three languages with linguists who were highly proficient in English, and native speakers of Spanish and Basque. The linguists rated each participant’s proficiency for each language “on a Likert-like scale from 1 (‘lowest level’) to 5 (‘native or native-like’)” (de Bruin et al., 2017).

**Procedure**

Data collection began by having participants provide their informed consent and complete an online questionnaire for the subjective measures along with the LexTALE tests. The participants later completed the picture naming tests and interviews in person at the BCBL (de Bruin et al., 2017). This project was deemed exempt from IRB approval due to the use of secondary data (IRB 22-125, 11 May, 2023). The data was analyzed in R Studio, using a series of multiple and simple regressions, along with observing descriptive statistics.
### Table 1

**Predictivity of the Interview Compared to the Picture Naming Test and Interview**

<table>
<thead>
<tr>
<th>Measures</th>
<th>Spanish</th>
<th>Basque</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b^*$</td>
<td>$R^2$</td>
<td>AIC</td>
</tr>
<tr>
<td>LexTALE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age of Acquisition</td>
<td>.041</td>
<td>.002</td>
<td>2422.407</td>
</tr>
<tr>
<td>Exposure</td>
<td>.008</td>
<td>&lt;.001</td>
<td>5709.898</td>
</tr>
<tr>
<td>Self-report Proficiency</td>
<td>.124</td>
<td>.015</td>
<td>1669.438</td>
</tr>
<tr>
<td>Picture Naming Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age of Acquisition</td>
<td>-.231***</td>
<td>.053</td>
<td>2387.949</td>
</tr>
<tr>
<td>Exposure</td>
<td>.201***</td>
<td>.040</td>
<td>5683.214</td>
</tr>
<tr>
<td>Interview</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age of Acquisition</td>
<td>NA</td>
<td>NA</td>
<td>2421.497</td>
</tr>
<tr>
<td>Exposure</td>
<td>NA</td>
<td>NA</td>
<td>5707.941</td>
</tr>
<tr>
<td>Self-report Proficiency</td>
<td>NA</td>
<td>NA</td>
<td>1677.483</td>
</tr>
</tbody>
</table>
Table 2

Proficiency Predicted by Age of Acquisition and Exposure

<table>
<thead>
<tr>
<th>Measures</th>
<th>Spanish</th>
<th>Basque</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b^*$</td>
<td>$R^2$</td>
<td>AIC</td>
</tr>
<tr>
<td>Age of Acquisition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-report Proficiency</td>
<td>-.243***</td>
<td>.1059</td>
<td>2383.880</td>
</tr>
<tr>
<td>LexTALE</td>
<td>.041</td>
<td>.002</td>
<td>2422.407</td>
</tr>
<tr>
<td>Picture Naming Test</td>
<td>-.231***</td>
<td>.053</td>
<td>2387.949</td>
</tr>
<tr>
<td>Interview</td>
<td>NA</td>
<td>NA</td>
<td>2421.497</td>
</tr>
<tr>
<td>Exposure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-report Proficiency</td>
<td>.262***</td>
<td>.069</td>
<td>5663.787</td>
</tr>
<tr>
<td>LexTALE</td>
<td>.008</td>
<td>&lt;.001</td>
<td>5709.898</td>
</tr>
<tr>
<td>Picture Naming Test</td>
<td>.201***</td>
<td>.040</td>
<td>5683.214</td>
</tr>
<tr>
<td>Interview</td>
<td>NA</td>
<td>NA</td>
<td>5707.941</td>
</tr>
</tbody>
</table>
**Results**

**Hypothesis 1**

The first hypothesis proposed that a younger age of acquisition, more exposure, and higher scores on all three objective proficiency measures would be predictive of higher self-rated proficiency scores for each respective language. For this hypothesis I ran 3 multiple regressions in R Studio, and used a Bonferroni adjustment of .016, to determine the predictivity of all measures on self-rated proficiency for each of the three languages:

Regression with age of acquisition, exposure, LexTALE scores, picture naming test scores, and interview scores as the IVs and self-rated proficiency as the DV.

The first hypothesis was supported in full for Spanish, and in part for Basque and English. For Spanish, taken together, the variables significantly predicted self-rated proficiency, $R^2 = .12$, $F(4, 645) = 22.08$, $p < .001$. A younger age of acquisition ($b^* = -.137$, $p = .001$), greater exposure ($b^* = .169$, $p < .001$), and higher scores on the objective measures (LexTALE: $b^* = -.114$, $p = .002$, picture naming test: $b^* = .135$, $p < .001$, interview: $b^* = NA$, $p = NA$) were related to higher self-rated Spanish proficiency scores.

For Basque, taken together, the variables significantly predicted subjective proficiency, $R^2 = .51$, $F(5, 644) = 131.9$, $p < .001$. Although the set of variables significantly predicted subjective Basque proficiency, not all variables contributed to the
prediction. A younger age of acquisition ($b^* = -0.100, p = 0.006$), greater exposure ($b^* = 0.128, p < 0.001$), higher scores on the picture naming test ($b^* = 0.246, p < 0.001$), and higher scores on the interview ($b^* = 0.268, p < 0.001$) were related to a higher self-rated Basque proficiency score. However, there was not a significant relationship between the LexTALE test and self-rated Basque proficiency scores ($b^* = 0.048, p = 0.257$).

For English, taken together, the variables significantly predicted subjective proficiency, $R^2 = 0.41$, $F(5, 644) = 89.4, p < 0.001$. Although the set of variables significantly predicted self-rated English proficiency, not all variables contributed to the prediction. Greater exposure ($b^* = 0.226, p < 0.001$), higher scores on the picture naming test ($b^* = 0.203, p < 0.001$), and higher scores on the interview ($b^* = 0.354, p < 0.001$) were related to higher self-rated English proficiency scores. However, age of acquisition ($b^* = -0.067, p = 0.027$) and the LexTALE test ($b^* = -0.012, p = 0.735$) were not related to self-rated English proficiency scores.

**Hypothesis 2**

The second hypothesis proposed that the interview would be the most predictive of age of acquisition, exposure, and self-rated proficiency compared to the picture naming test or LexTALE test. For my second hypothesis, I ran 18 simple regressions in R Studio, and used a Bonferroni adjustment of 0.003, to determine the predictivity of each objective proficiency variable on the three subjective proficiency variables for each of the three languages:

Regression with LexTALE scores as the IV and age of acquisition as the DV.
Regression with LexTALE scores as the IV and exposure scores as the DV.
Regression with LexTALE scores as the IV and self-rated proficiency scores as the DV.
Regression with picture naming test scores as the IV and age of acquisition as the DV.
Regression with picture naming test scores as the IV and exposure scores as the DV.
Regression with picture naming test scores as the IV and self-rated proficiency scores as the DV.
Regression with interview scores as the IV and age of acquisition as the DV.
Regression with interview scores as the IV and exposure scores as the DV.
Regression with interview scores as the IV and self-rated proficiency scores as the DV.

The second hypothesis was partially supported for Basque and English, but not Spanish. For Spanish, picture naming test scores significantly predicted age of acquisition, exposure level, and self-rated proficiency scores. Higher scores on the picture naming test were related to a younger age of acquisition ($b^* = -0.231, p < .001$), greater exposure ($b^* = 0.201, p < .001$), and higher self-rated proficiency scores ($b^* = 0.212, p < .001$). LexTALE scores were not significantly related to any of the three subjective measures (age of acquisition: $b^* = 0.041, p = .297$, exposure: $b^* = 0.008, p = .836$, self-rated proficiency: $b^* = 0.124, p = .014$). The interview scores for Spanish were at ceiling.
for all participants, thus variance between the measures could not be determined, making it not possible to evaluate how well the interview predicted the subjective measures in comparison to the other objective measures.

In order to assess the strength of the 3 objective measures’ predictability, I looked at the models' $R^2$ and Akaike Information Criterion (AIC) scores. $R^2$ shows the amount of variance in the dependent variable that can be explained by the independent variable, so a larger $R^2$ is indicative of a better model fit. AIC scores measure how likely it is to observe the current data provided the model. The lower the AIC score is, the more likely it is that the current data would exist given the model, meaning that the model fits the data better. Ultimately, the model with the best fit should have the highest $R^2$ scores and the lowest AIC scores.

For Basque, scores on all three objective tests significantly predicted age of acquisition, exposure level, and self-rated proficiency scores. Higher scores on all three objectives tests were related to a younger age of acquisition (LexTALE: $b^* = -.478, p < .001$, picture naming test: $b^* = -.588, p < .001$, interview: $b^* = -.582, p < .001$), greater exposure (LexTALE: $b^* = .470, p < .001$, picture naming test: $b^* = .503, p < .001$, interview: $b^* = .523, p < .001$), and higher self-rated proficiency scores (LexTALE: $b^* = .546, p < .001$, picture naming test: $b^* = .648, p < .001$, interview: $b^* = .656, p < .001$). Although scores on all three objective tests significantly predicted all three subjective measures, as shown in Table 1, the interview was a slightly better predictor than the LexTALE test and the picture naming test for most of the subjective measures.
For English, scores on all three objective tests significantly predicted exposure level and self-rated proficiency scores. Higher scores on all three objective tests were related to greater exposure (LexTALE: $b^* = .289, p < .001$, picture naming test: $b^* = .340, p < .001$, interview: $b^* = .361, p < .001$) and higher self-rated proficiency scores (LexTALE: $b^* = .296, p < .001$, picture naming test: $b^* = .534, p < .001$, interview: $b^* = .580, p < .001$). The objective test scores did not, however, significantly predict age of acquisition (LexTALE: $b^* = .080, p = .041$, picture naming test: $b^* = -.001, p = .979$, interview: $b^* = -.010, p = .800$). Although English exposure level and self-rated English proficiency were significantly predicted by scores on all three objective tests, as shown in Table 1, the interview was a slightly better predictor than the LexTALE test and the picture naming test.

**Hypothesis 3**

The third hypothesis proposed that, in comparing each language (Spanish, Basque, and English), exposure and age of acquisition would show the greatest impact for English proficiency scores (both subjective and objective), due to this language being the least familiar. Finally, for my third hypothesis, I ran 24 simple regressions in R Studio, and used a Bonferroni adjustment of .002, to determine how predictive age of acquisition and exposure are of the four proficiency measures compared across the three languages:

Regression with age of acquisition as the IV and self-rated proficiency scores as the DV.
Regression with age of acquisition as the IV and LexTALE scores as the DV.
Regression with age of acquisition as the IV and picture naming test scores as the DV.
Regression with age of acquisition as the IV and interview scores as the DV.
Regression with exposure scores as the IV and self-rated proficiency scores as the DV.
Regression with exposure scores as the IV and LexTALE scores as the DV.
Regression with exposure scores as the IV and picture naming test scores as the DV.
Regression with exposure scores as the IV and interview scores as the DV.

The third hypothesis was not supported. As shown in Table 2, Spanish age of acquisition and level of exposure significantly predicted self-rated Spanish proficiency (age of acquisition: $b^* = -0.234, p < .001$, exposure: $b^* = 0.262, p < .001$) and Spanish picture naming test scores (age of acquisition: $b^* = -0.231, p < .001$, exposure: $b^* = 0.201, p < .001$), but not LexTALE test scores (age of acquisition: $b^* = 0.041, p = .297$, exposure: $b^* = 0.008, p = .836$) or interview scores (age of acquisition: $b^* = \text{NA}, p = \text{NA}$, exposure: $b^* = \text{NA}, p = \text{NA}$). Basque age of acquisition (self-reported proficiency: $b^* = -0.511, p < .001$, LexTALE: $b^* = -0.478, p < .001$, picture naming test: $b^* = -0.588, p < .001$, interview: $b^* = -0.582, p < .001$) and level of exposure (self-reported proficiency: $b^* = 0.516, p < .001$, LexTALE: $b^* = 0.470, p < .001$, picture naming test: $b^* = 0.503, p < .001$, interview: $b^* = 0.523, p < .001$) significantly predicted all four proficiency measures.
English age of acquisition (self-reported proficiency: $b^* = -.072, p = .065$, LexTALE: $b^* = .080, p = .041$, picture naming test: $b^* = -.001, p = .979$, interview: $b^* = -.010, p = .800$) did not significantly predict any of the proficiency measures and can therefore not be evaluated in comparison to Spanish and Basque. English level of exposure (self-reported proficiency: $b^* = .419, p < .001$, LexTALE: $b^* = .289, p < .001$, picture naming test: $b^* = .340, p < .001$, interview: $b^* = .361, p < .001$), on the other hand, did significantly predict all four proficiency measures. A greater level of English exposure was related to higher scores on all four proficiency measures.

Although the relationships between level of exposure and the proficiency measures were, for the most part, significant across languages, only Spanish demonstrated a clear difference in how strong of a predictor exposure was for proficiency. Compared to English and Basque, Spanish had consistently lower $R^2$ scores and higher AIC scores across proficiency measures, suggesting that exposure was not as strong of a predictor of proficiency for Spanish as it was for the other languages. In comparing English and Basque, the AIC scores were lower for English, but the $R^2$ scores were larger for Basque, so it is unclear if exposure was a stronger predictor of proficiency for either language.
Discussion

This study investigated the relationship between subjective and objective language proficiency as impacted by age of acquisition and exposure. Overall, there was mixed support for the majority of the hypotheses. The first hypothesis proposed that higher self-rated proficiency scores would be related to a younger age of acquisition and higher scores on all other measures. This was found to be true for Spanish, which is unsurprising considering that the large majority of participants rated themselves to be highly proficient in Spanish. For Basque and English, however, there was not a significant relationship between the LexTALE test and self-rated proficiency. This could suggest that the LexTALE test is not as valid of a measure for assessing overall language proficiency in this sample as the picture naming test and interview. Additionally, for English, age of acquisition and self-rated proficiency did not reach significance after the Bonferroni adjustment, however, without the Bonferroni adjustment, the relationship would have been considered significant. This was unexpected since the participants’ ages of acquisition for English varied the most between languages, however, the lack of relationship in this model may be attributed to less frequent use of English in Basque Country. It is possible that a participant who acquired English at a younger age may use the language less than another who acquired English at an older age, thus leading them to rate themselves as less proficient. This is in line with past research that has demonstrated the importance of exposure and language use when gaining proficiency in a language (Pardis & Jia, 2017).
The second hypothesis proposed that the interview would be the strongest of the three objective measures at predicting the three subjective measures. For Spanish, specifically, all participants scored five out of five on the interview, so it was not possible to evaluate how well the interview predicted the subjective measures in comparison to the other objective measures. It is worth noting, that even though all the participants were rated at the highest level of proficiency for the Spanish interview, not all participants received perfect scores on the other Spanish proficiency tests. For Basque and English, the interview was a stronger predictor than the other objective tests for most of the subjective measures. This was expected because the interview allows the participants to demonstrate more aspects of their proficiency than the other tests, which primarily measure vocabulary knowledge. These findings continue to support the call for tests with more diverse measures of proficiency.

The third hypothesis proposed that, of the three languages, English’s proficiency scores, both subjective and objective, would be most strongly predicted by age of acquisition and exposure compared to Spanish and Basque. This hypothesis was developed because the participants reported older age of acquisitions for English compared to Spanish and Basque, and they live in a location where Spanish and Basque are regularly spoken, so there is more variability in exposure scores for English across participants. English age of acquisition was not significantly related to any of the proficiency measures, so it could not be compared to the other languages. Although a comparison could not be made, this finding once again speaks to the importance of
considering factors other than just age of acquisition when examining language proficiency.

English exposure was significantly related to all four proficiency measures, but in comparison to Spanish and Basque, exposure was only a stronger predictor of proficiency for English over Spanish. English and Basque seemed to be similar in exposure’s predictive power. Spanish having the weakest relationship between exposure and proficiency may be explained by the greater amount of exposure that all participants reported having to Spanish compared to both Basque and English. Even though the hypothesis was not fully supported, Spanish being the least impacted is still in line with the reasoning behind the hypothesis and lends further support to the value of investigating exposure in language research.

Limitations

Given my limited time and resources, this project would not have been possible without the work done by de Bruin and colleagues (2017) in the creation of the BEST dataset. Although the utilization of the BEST dataset was essential for the completion of this project, working with secondary data comes with its own limitations, such as being confined to the information provided by the original researchers. For example, it is unknown if reaction time data was collected for the LexTALE and picture naming tests. Reaction time data would have been useful information because it could provide insight into proficiency level beyond just technical knowledge. For instance, a participant could provide the same number of correct answers on the LexTALE test for both Spanish and
Basque, but if they are more proficient in Spanish, they may have been able to provide the correct answers faster for the Spanish test than the Basque test.

It would also be beneficial to know the degree of inter-rater reliability that was obtained for the interviews. While each interview was conducted by a single linguist, 4 total linguists were involved. It is not clear if all 4 linguists gave scores for each interview or if they worked together to decide on one score for each interview. In regard to the interviews, the dataset lacked access to the full transcripts and complete lists of questions that were used. The researchers gave liberty to the linguists to decide what questions would be asked beyond an initial self-introduction, so I do not know what other types of information were discussed or how consistent the interviews were across participants. Considering the findings of Anderson and colleagues’ 2018 study that the context in which one interacts with a language can influence their overall language use, it would be useful to know more about the topics that were discussed and the environment that the interviews were conducted in. It could be possible that some participants failed to demonstrate accurate proficiency depending on how comfortable they felt or the type of conversation they had during the interview.

In using this dataset, the objective measures were limited to the picture naming test, LexTALE test, and interviews. Although these tests provide valuable proficiency information, they touch on a relatively narrow range of linguistic domains. Both the picture naming test and the LexTALE test primarily capture aspects of semantics (vocabulary knowledge), whereas the interview captures a broader view of language knowledge and use. It would be beneficial to use a combination of tests that are able to
measure a wider variety of linguistic domains such as syntax (the use of phrases and sentences) and pragmatics (understanding the meaning of language within a variety of social contexts).

Another limitation of this study is that all of the participants were immersed in their primary languages. It would be interesting to compare the data to that of participants who are immersed in a secondary language. For example, a 2014 study comparing study-abroad English learners’ and home-country classroom English learners’ test scores on the Test of English as a Foreign Language (TOEFL) exam found that both groups of students received similar exam scores (Gu, 2014). Although these findings oppose the general assumption that there are advantages in language learning for immersion students due to increased exposure, studies such as this exemplify the importance of effort and motivation over sheer exposure.
Conclusion

Although the specific hypotheses of this thesis had mixed support, the results remained in line with the two major investigations of the study. First, this study’s findings continued to demonstrate the importance of including a wide range of objective tests for measuring language proficiency. Second, this study supports the use of considering factors beyond age of acquisition, such as exposure, when evaluating individuals’ language proficiency. Continued research is needed to expand the options for objective language tests, alongside investigating the role of exposure in bi/multilingual research. I hope that future language research continues to expand into a variety of linguistic domains and push for more studies on multilingualism. I would be interested to see the use of an objective measure similar to the interview, that takes into account a variety of contexts and types of language use. I am also interested to see how the use of more objective proficiency tests can inform research about language learning, and potentially aid in the development of teaching techniques. Aside from just measuring proficiency, I believe a variety of objective tests can help researchers to better understand the multilingual experience and take this research into other domains such as how multilingualism can affect other aspects of cognitive performance.
References


de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences, 9*, no. 3: 33.


(EAR) to capture the day-to-day linguistic experiences of young adults. *Collabra: Psychology, 8*(1), 36310.


