# 

# STAT 5310 Final

### Nathan Diffey

Dr. Mhoon

# Literature Review

### Bilingualism: Learning Language in the Modern Age

##### A large majority of the world population is bilingual, with the rates only rising due to the interconnectedness of our world in the modern age. Knowing more than one language is becoming a crucial skill for many, whether that be in the workplace or in an educational setting, enabling opportunities never seen before such as studying abroad or working for an international company.

#### Social media has also enabled us to immerse ourselves in a language other than our mother-tongues more easily than ever. Communicating with people worldwide has never been easier and more instantaneous, but are we making use of this wonderful opportunity? Can everyone take advantage of this opportunity, or is it true that some just can’t learn a new language? This literature review intends to explore the history of language in the United States. research on language acquisition in different life stages, and how technology, and more recently AI, has shaped a new way of learning languages.

### History of Bilingualism in the United States

##### Bilingual education in the United States has a long and complex history, shaped by waves of immigration and shifting political attitudes toward language diversity. In the early years of the nation bilingualism thrived, as large numbers of immigrants from Germany, France, Spain, and other countries brought their languages and cultures with them. Many states adopted bilingual education laws during the 19th century to support this linguistic diversity. However, as the U.S. faced social and economic challenges, such as the Civil War and World War I, the nation moved towards English-only policies, marginalizing non-English speakers. Over time, the debate over bilingual education has been influenced by movements for civil rights, assimilation, and cultural identity, leading to significant shifts in language education policies.

#### The Colonies

##### During the colonial and early years of the United States, bilingualism flourished across the country as waves of immigrants from diverse linguistic backgrounds arrived and settled in various regions. German, French, and Spanish were particularly prominent, as immigrants from these countries formed large, distinct communities that continued to use their native languages in both daily life and education. Throughout the 1800s, several states passed laws to support bilingual education in public schools. For example, Ohio became one of the first states to authorize German-English education in 1839, Louisiana followed suit in 1847, authorizing French-English bilingual education and New Mexico adopted Spanish-English bilingual education in 1850. (Gándara & Escamilla, 2016)

##### These early bilingual education policies were designed to support immigrants in keeping their native languages while also learning English, fostering a sense of inclusion and cultural preservation. Gándara and Escamilla (2016) emphasize that these policies reflected a time when the United States viewed linguistic diversity as a strength, embracing the multicultural makeup of its population. However, this period of flourishing bilingualism was relatively short-lived, as shifting political, social, and economic pressures later led to a push for English-only education.

#### Civil War

##### The end of the nineteenth century marked a significant and tragic turning point for the United States, particularly in terms of its attitudes toward immigration and language. The aftermath of the Civil War brought about a severe economic recession that not only reshaped the nation’s economy but also led to a surge of nativist sentiment. As economic struggles deepened, policies that had previously favored immigrants or promoted multiculturalism were increasingly viewed with suspicion and hostility. This shift in public opinion, coupled with economic anxiety, resulted in the dismantling of many bilingual education programs that had once supported the idea of linguistic diversity.

#### Early 1900s

##### One of the most defining legislative actions that reflected this change was the Naturalization Law of 1906. This law required all immigrants seeking U.S. citizenship to demonstrate proficiency in English, signaling a dramatic shift in the government’s stance on language. By mandating English proficiency, the Naturalization Law institutionalized the idea that English was not just the dominant language, but the only language recognized for civic inclusion. This marked the beginning of a broader movement toward English-only policies that would shape the nation’s education system and immigration laws for decades to come, as the idea of linguistic diversity came to be seen as a barrier to national unity rather than a cultural asset.

##### In the early 1910s, the outbreak of World War I intensified the already growing anti-immigrant sentiment in the United States, leading to even more stringent policies aimed at linguistic and cultural assimilation. By the end of the war, 34 states had adopted laws mandating English-only education in both public and private primary schools. These policies were not just educational in nature; they were tools of cultural assimilation, essentially forcing immigrant communities to abandon their native languages and, in many cases, significant parts of their cultural identities in order to fit into the American mainstream.

##### This movement towards English-only education was particularly harsh on Spanish-speaking communities, especially along the U.S.-Mexico border. In the 1930s, despite a significant decline in overall immigration, the border regions continued to see a high concentration of Spanish-only speakers. Rather than providing bilingual or culturally inclusive education, these students were often segregated from their English-speaking peers and placed into separate schools or classrooms. The explicit purpose of these “Americanization” programs was not to educate Spanish-speaking students on equal terms but to strip them of their linguistic and cultural heritage. This segregation not only deepened the divide between English and non-English speakers but also underscored the prevailing attitudes of the time, which viewed foreign languages and cultures with suspicion.

#### Civil Rights Movement

##### As the Civil Rights movement gave way in the 60s, bilingual education once again became a topic of conversation. Research being done at the time enforcing the idea that forcing language acquisition on people may cause shame in both their cultural identity and language knowledge. The 1968 Bilingual Education Act gave way to major funding of English acquisition programs around the country and encouraged states with high populations of immigrants to once again adopt bilingual education laws which helped provide many bilingual programs.

##### While these legislations came into effect, he real action came when a group of Chinese immigrants arguing to the Supreme Court that they were being denied equal education by their school district, and the court ruled in favor of the students, making them take action to provide the same education with affirmative steps. Unfortunately, these affirmative steps were not specified, which only helped the idea of ‘Americanization’ of immigrants and English-only education officially became the norm; bilingual education laws officially overturned in most states starting with California in 1998 (Cummins, 2003, Baker et al., 2016). Since then, English only education is the norm, with English Second Language programs’ first goal being to integrate the student in with their peers, rather than teaching them in their first language. (Baker et al., 2016, Gándara & Escamilla, 2016)

##### In summary, the history of bilingual education in the United States reflects the nation’s complex and evolving attitudes toward language and immigration. During the colonial and early years, bilingualism was embraced as many states adopted bilingual education laws to accommodate the diverse immigrant populations. However, as political and economic circumstances shifted, the country moved towards English-only policies. The rise of the Civil Rights movement in the 1960s revived the discussion on bilingual education, leading to the Bilingual Education Act of 1968 and renewed support for language diversity in schools. However, the push for assimilation through English-only education prevailed, and by the late 20th century, most states had rolled back bilingual education laws in favor of integration into English-speaking classrooms. Despite some efforts to promote linguistic diversity, the dominant approach in the U.S. education system remains focused on English proficiency, often at the expense of students’ native languages and cultural identities.

### Learning a Second Language Through the Ages

##### Learning a new language is a complex process that changes throughout various stages of life. During childhood, the brain’s heightened capacity for absorbing information makes language acquisition easier. As we transition into adulthood, however, cognitive flexibility decreases, making the learning of a second language (L2) more challenging, especially if the first language (L1) differs significantly. Despite this, motivation plays a crucial role in language proficiency among adults, as those with a personal or social incentive tend to achieve better results. In later life, learning a new language can still be beneficial, as studies show that engaging in this mental exercise improves cognitive health and overall well-being in seniors. While the pace of language learning may slow, the benefits to mental health make it a worthwhile pursuit at any age.

#### Children

##### During childhood, the brain is highly adaptable and capable of absorbing and processing vast amounts of information, making language acquisition relatively effortless, especially when supported by proper guidance from parents or caregivers. When parents and other caregivers (e.g. grandparents) are able to spend time with their children and tell stories or discuss issues with them in a way that develops their mother tongue vocabulary and concepts, children come to school well prepared to learn the school language and succeed educationally. (Cummins, 2003) Studies like those by Rowe et al. (2016) emphasize the importance of early and consistent language practice, as it plays a crucial role in helping children internalize new vocabulary, syntax, and grammar structures.

#### Adults

##### As we transition into adulthood, our ability to learn a new language diminishes, particularly when the first language (L1) differs significantly from the second language (L2). These differences may be related to grammar, phoneme pronunciation, or syntax, making it more challenging for adult learners to grasp the nuances of a new language. However, one of the key factors influencing successful language acquisition in adults is motivation. According to Baker et al. (2008), adults who have a strong desire to communicate with friends or family members in the new language are generally more successful in becoming proficient than those learning for more formal reasons, such as career advancement or fulfilling academic requirements. This suggests that personal motivation and emotional investment can greatly enhance language learning outcomes.

##### Immersion remains a highly effective method for adult learners, but even in immersive environments, adults face certain challenges. A study conducted by Knoch et al. (2015) tracked international university students over three years, examining the role of immersion in language learning. The findings indicated that while fluency and conversational skills improved significantly, the students’ grammatical accuracy and complexity still lagged behind that of their native-speaking peers. This highlights that although adults are capable of learning and using a new language, they may not reach the same level of grammatical precision as younger learners, even in immersive settings.

##### Sarret et al. (2021) further explain that adult learners of an L2 tend to rely on translation strategies, actively translating thoughts and phrases from their L1 to understand and produce sentences in the L2. This contrasts with younger learners, who often acquire language more naturally and intuitively, without the need for conscious translation. This cognitive burden in adults can slow down the learning process and contribute to the difficulty in achieving native-like proficiency in certain aspects of the language, particularly in areas that are structurally different from their L1. Therefore, while language learning in adulthood is entirely possible, it demands greater cognitive effort, motivation, and consistent exposure to the target language.

#### Seniors

##### Later in life, learning a new language offers significant cognitive and mental health benefits, even though the process tends to be slower. As the brain ages, cognitive functions such as memory and processing speed decline, which can make language acquisition more difficult. However, engaging in the process of learning a new language stimulates the brain, enhancing cognitive function and mental resilience. Research by Pikhart and Klimova (2020) in the Czech Republic demonstrates that while older adults may face challenges in learning a language due to slower absorption and retention, the very act of learning has a positive impact on their overall mental well-being. The study found that participants experienced improved mood, greater self-esteem, and a sense of accomplishment as they worked to master new linguistic skills.

##### Language learning in older adults can also serve as a form of mental exercise, akin to activities like puzzles or memory games, which are known to help delay cognitive decline. The process of acquiring vocabulary, practicing pronunciation, and forming sentences requires mental flexibility, which can help maintain brain health and reduce the risk of conditions such as dementia or Alzheimer’s disease. (Blackmore, 2024) Beyond the cognitive benefits, language learning provides older individuals with a way to connect with others, engage with different cultures, and continue growing intellectually, contributing to an overall sense of purpose and fulfillment.

##### In conclusion, language acquisition is a dynamic process that evolves throughout our lives, with distinct challenges and advantages at different stages. During childhood, the brain’s remarkable capacity for learning allows for the seamless acquisition of one or more languages, especially when supported by immersive environments and consistent practice. As we enter adulthood, language learning becomes more complex, particularly when the first and second languages differ significantly. Motivation and immersion play crucial roles in overcoming these challenges, although adults often face difficulties with grammar and fluency compared to younger learners. In later life, while language learning may slow down, it offers profound cognitive benefits, enhancing mental well-being and promoting brain health. Regardless of age, the journey of learning a new language is not only possible but can also be deeply enriching, providing personal, social, and cognitive rewards that continue to shape and enrich our lives.

### Technology and Language

##### New tools are now available to help English as a Second Language (ESL) learners improve their writing skills, such as sequence word-combination checkers. These tools assess whether a sentence structure adheres to the grammatical rules of the English language, providing learners with valuable feedback on their sentence construction. Used worldwide, these tools aim to enhance English language proficiency by giving students a clearer understanding of correct grammar usage. A study conducted in Saudi Arabia examined the effectiveness of such tools by tracking the progress of 74 undergraduate students in their English language courses. The study revealed that students who used the sequence word-combination checker significantly outperformed their peers in the control group, with approximately 20% higher scores across all five categories identified by the researchers. This suggests that integrating technology into language learning can be a highly effective way to enhance language acquisition and overall proficiency.

##### Google+ and WhatsApp have also emerged as valuable platforms for English language learning, particularly in countries where English is not the primary spoken language. In Malaysia, for instance, Google+ has been adopted as an educational tool for students in both primary and secondary schools. The program’s primary objective is to facilitate communication in English only, creating an immersive and interactive learning environment. By engaging with peers who share similar interests, students find language acquisition more enjoyable and less formal. This conversational approach allows them to practice English in a context that feels natural and relevant to their everyday lives, making it an effective way to enhance language skills while fostering connections with others. Similarly, WhatsApp has been used in various educational settings to encourage students to practice English through group chats, voice messages, and real-time conversations, further promoting language learning in a way that is both accessible and engaging.

##### Meanwhile, in Spain, a study involving undergraduate students enrolled in B1 English courses utilized WhatsApp groups as a platform for language practice, where participants were instructed to communicate exclusively in English. This approach aimed to create a casual yet effective environment for enhancing their language skills. Initially, the students exhibited a range of difficulties, including numerous grammatical mistakes and frequent switching of word structures, indicating their limited proficiency at the outset. However, as the students engaged in ongoing conversations within the WhatsApp groups, a noticeable improvement in their English skills was observed over time. The frequency of mistakes began to decrease significantly, demonstrating that consistent practice in a supportive and informal setting can effectively boost language acquisition. This evolution not only reflects the positive impact of peer interaction on language learning but also highlights the potential of digital platforms to facilitate meaningful communication and enhance language competence among learners This shows that technology has had a positive effect on ESL learning, as well as learning in general.

### References

##### Aktan-Erciyes, A., Ünlütabak, B., & Zengin, B. F. (2021). Effects of second-language acquisition on character introductions in 5- and 7-year-old bilingual and monolingual children’s Frog story narratives. First Language, 41(5), 605-622. <https://doi.org/10.1177/01427237211024219>

##### Baker, D.L., Basaraba, D.L. and Polanco, P. (2016) ‘Connecting the Present to the Past: Furthering the Research on Bilingual Education and Bilingualism’, Review of Research in Education, 40(1), pp. 821–883. <doi:10.3102/0091732x16660691>.

##### Baker, W. et al. (2008) ‘Child—Adult Differences in Second-Language Phonological Learning: The Role of Cross-Language Similarity’, Language and Speech, 51(4), pp. 317–342. <doi:10.1177/0023830908099068>.

##### Blackmore, D. G. (2024). Long-Term Improvement in Hippocampal-Dependent Learning Ability in Healthy, Aged Individuals Following High Intensity Interval Training. Aging and Disease. <https://doi.org/10.14336/ad.2024.0642>

##### Cummins, J. (2003). 4. Bilingual Education: Basic Principles. In J. Dewaele, A. Housen & W. Li (Ed.), Bilingualism: Beyond Basic Principles (pp. 56-66). Bristol, Blue Ridge Summit: Multilingual Matters. <https://doi.org/10.21832/9781853596315-007>

##### Gándara, P. and Escamilla, K. (2016) ‘Bilingual education in the United States’, Bilingual and Multilingual Education, pp. 1–14. <doi:10.1007/978-3-319-02324-3_33-1>.

##### Knoch, U. et al. (2015) ‘What happens to ESL students’ writing after three years of study at an English Medium University?’, Journal of Second Language Writing, 28, pp. 39–52. <doi:10.1016/j.jslw.2015.02.005>.

##### Pikhart, M. and Klimova, B. (2020) ‘Maintaining and supporting seniors’ wellbeing through Foreign Language Learning: Psycholinguistics of Second Language Acquisition in older age’, International Journal of Environmental Research and Public Health, 17(21), p. 8038. <doi:10.3390/ijerph17218038>.

##### Rowe, M.L., Leech, K.A. and Cabrera, N. (2016) ‘Going Beyond Input Quantity: Wh-Questions Matter for Toddlers’ Language and Cognitive Development’, Cognitive Science, 41(S1), pp. 162–179. <doi:10.1111/cogs.12349>.

##### Sarrett, M. E., Shea, C., & McMurray, B. (2021). Within- and between-language competition in adult second language learners: implications for language proficiency. Language, Cognition and Neuroscience, 37(2), 165–181. <https://doi-org.uhd.idm.oclc.org/10.1080/23273798.2021.1952283>

# Data Analysis:

#### The literature review highlights the difficulties in learning a new language, but exactly what factors can predict language proficiency in a second language?

### About the data:

##### This dataset is from a study called BEST (Basque, English, and Spanish tests) Dataset of Language Proficiency, in which persons speaking Basque, Spanish and English were evaluated on their language proficiency in all three languages and determining what factors may influence test results.

##### The dataset first highlights the same variables in all three languages tested, in which I only kept English due to the focus of this report being on the United States. The four factors tested were Education level, level of exposure to a language in everyday life, age at which language was first acquired, and the participants’ self-perception of proficiency. The education predictor has been factored into four, high school, technical school, bachelor’s degree and post-graduate degree. While the study had three dependent variables, LEXTALE test score, Picture naming score and interview score, only the LEXTALE score was used to simplify the model. The LEXTALE test is a test in which participants whose first language is not English answer 60 multiple choice vocabulary questions. This test is aimed at moderate to high proficiency speakers.

##### Multiple regression analysis, as well as subset selection and testing of assumptions will be used to find the best model for analytical purposes.

### Preliminary Data Analysis:

#### Initial Model Summary (all predictors present)

##   
## Call:  
## lm(formula = LEXTALE.Test.Score ~ ., data = best.eng)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -25.129 -5.193 -0.596 4.716 31.958   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 49.58861 2.14128 23.158 < 2e-16 \*\*\*  
## Age 0.39647 0.07023 5.645 2.47e-08 \*\*\*  
## Education -0.38461 0.27305 -1.409 0.159   
## Exposed.To.English 0.21885 0.03770 5.805 1.01e-08 \*\*\*  
## Self.Perceived.English 0.10774 0.01856 5.805 1.01e-08 \*\*\*  
## Age.of.Acquisition -0.06573 0.15790 -0.416 0.677   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.425 on 644 degrees of freedom  
## Multiple R-squared: 0.1732, Adjusted R-squared: 0.1668   
## F-statistic: 26.99 on 5 and 644 DF, p-value: < 2.2e-16

#### Pairs Plot

A pairs plot with all variables is used to see if there are any clear relationships between the predictor variables and LEXTALE score.

A screenshot of a graph

Description automatically generated

#### Model Residual Plots

A group of graphs showing different values

Description automatically generated

#### VIF Test

## Age Education Exposed.To.English   
## 1.405832 1.004530 1.138074   
## Self.Perceived.English Age.of.Acquisition   
## 1.159999 1.410636

#### Initial Model Analyzation:

Looking at the pairs plot and residual plots, the data seems to show that there is a positive relationship between the response variable (LEXTALE Test Score) and Age, Exposure to English and Self Perceived English knowledge. The beta coefficients of Age (p = 0.0000000247), Exposure to English (p = 0.0000000101), and self Percieved English (p = 0.0000000101) are statistically very significant and hold a strong relationship to the dependent variable. The data seems to be linear, moderately normal, homoskedastic and multicollinearity is met.

The initial regression formula is as follows:

LEXTALE.Test.Score = 49.58861 + (0.39647 \* Age) - (0.38461 \* Education) + (0.21885 \* Exposed.To.English) + (0.10774 \* Self.Perceived.English) - (0.06573 \* Age.of.Acquisition)

### Model Fitting:

#### Subset Selection:

After analyzing the initial model with all variables, a subset is found which has all statistically significant variables, while leaving non-significant ones out.

## # A tibble: 5 × 5  
## set.name AIC AICc BIC AdjR2  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 LEXTALE.Test.Score, Exposed.To.English 4676. 4682. 4696. 0.0818  
## 2 LEXTALE.Test.Score, Age + Self.Perceived.English 4645. 4653. 4671. 0.124   
## 3 LEXTALE.Test.Score, Age + Exposed.To.English + Self.… 4611. 4621. 4644. 0.167   
## 4 LEXTALE.Test.Score, Age + Education + Exposed.To.Eng… 4609. 4621. 4648. 0.168   
## 5 LEXTALE.Test.Score, Age + Education + Exposed.To.Eng… 4609. 4623. 4654. 0.167

Through checking the AIC, AICc, BIC and Adjusted R2 scores of each subset possible, a subset of Age, Exposure to English and Self perceived English knowledge is the best subset to use for our model, according to both a manual and an automatic (R) examination.

#### Summary of Subset Model:

##   
## Call:  
## lm(formula = LEXTALE.Test.Score ~ Age + Exposed.To.English +   
## Self.Perceived.English, data = best.eng)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -25.4856 -5.1933 -0.5235 4.7413 31.2879   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 48.29155 1.94838 24.786 < 2e-16 \*\*\*  
## Age 0.37955 0.05971 6.357 3.89e-10 \*\*\*  
## Exposed.To.English 0.21930 0.03762 5.829 8.79e-09 \*\*\*  
## Self.Perceived.English 0.10959 0.01847 5.935 4.79e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.426 on 646 degrees of freedom  
## Multiple R-squared: 0.1705, Adjusted R-squared: 0.1666   
## F-statistic: 44.25 on 3 and 646 DF, p-value: < 2.2e-16

The best subset’s three variables are all significant with a p-value well below 0.001. The initial and best subset model have similar Adjusted R-Squared results, 0.1668 and 0.1666, which hints that the difference between the two models is very small, but the smaller model is chosen for simplicity.

#### Testing Assumptions of Best Subset Model:

##### Normality:

##   
## Shapiro-Wilk normality test  
##   
## data: residuals(subset.best)  
## W = 0.99049, p-value = 0.0003329

A graph with numbers and lines

Description automatically generatedA graph of a number of individuals

Description automatically generated

While the distribution looks normal or semi-normal on the plots, the Shapiro-Wilk test failed, so a Box-Cox transformation is attempted.

#### Summary of Transformed Subset Model:

##   
## Call:  
## lm(formula = trans.score ~ Age + Exposed.To.English + Self.Perceived.English,   
## data = best.eng)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.81786 -0.13711 -0.00195 0.13591 0.73435   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.2364355 0.0525702 99.608 < 2e-16 \*\*\*  
## Age 0.0099346 0.0016110 6.167 1.23e-09 \*\*\*  
## Exposed.To.English 0.0055241 0.0010150 5.442 7.48e-08 \*\*\*  
## Self.Perceived.English 0.0029642 0.0004982 5.950 4.41e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2273 on 646 degrees of freedom  
## Multiple R-squared: 0.1616, Adjusted R-squared: 0.1577   
## F-statistic: 41.51 on 3 and 646 DF, p-value: < 2.2e-16

The transformed subset regression formula is as follows:

LEXTALE.Test.Score = 5.2364 + (0.0099 \* Age) + (0.0055 \* Exposed.To.English) + (0.0030 \* Self.Perceived.English)

#### Checking Assumptions of transformed subset model:

##   
## Shapiro-Wilk normality test  
##   
## data: residuals(trans.model)  
## W = 0.99463, p-value = 0.02189

A graph with numbers and lines

Description automatically generatedA graph of a graph

Description automatically generated

While the Shapiro-Wilk test still failed, the p-value has increased significantly, and both the historgram and QQ Residuals plot shows normality, so normality is assumed with the Box-Cox transformation.

A graph of a number of values

Description automatically generated with medium confidence

The model shows great linearity.

## Age Exposed.To.English Self.Perceived.English   
## 1.015889 1.132979 1.147822

According to the VIF test, our subset model has no to very slight multicollinearity, with values being just slightly above 1.

A graph of a number of values

Description automatically generated with medium confidence

The residuals vs. fitted plot shows homoscedasticity.

#### F-Test to compare two models:

## Analysis of Variance Table  
##   
## Model 1: trans.score ~ Age + Exposed.To.English + Self.Perceived.English  
## Model 2: trans.score ~ Age + Education + Exposed.To.English + Self.Perceived.English +   
## Age.of.Acquisition  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 646 33.388   
## 2 644 33.274 2 0.11414 1.1045 0.332

The F-statistic for the comparison was 1.1045, with a p-value of 0.332. Since the p-value was greater than our significance threshold of 0.05, we failed to reject the null hypothesis that the additional predictors do not significantly improve the model.

The smaller model is more appropriate for our purposes as the extra variables do not provide much effect on the outcome. The results indicate that the variables not included in our smaller model do not have a significant effect on the outcome.

The R2 and Adjusted R2 values are 0.1616 and 0.1577 respectively. The slight difference in those two values suggests that the variables included in our model are necessary and overfitting is not an issue. The values themselves are fairly low, which can mean thewre is another variable which was not included that may provide better predictability.

#### Leverage and outlier points:

A graph of a number of black dots

Description automatically generated with medium confidence

The residuals vs fitted plot shows some points which may be leverage points, so a refit of the model is done to remove said points.

Residuals vs Leverage Plot without outliers:

A graph of a number of individuals

Description automatically generated with medium confidence

##   
## Call:  
## lm(formula = trans.score ~ Age + Exposed.To.English + Self.Perceived.English,   
## data = best.eng)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.81786 -0.13711 -0.00195 0.13591 0.73435   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.2364355 0.0525702 99.608 < 2e-16 \*\*\*  
## Age 0.0099346 0.0016110 6.167 1.23e-09 \*\*\*  
## Exposed.To.English 0.0055241 0.0010150 5.442 7.48e-08 \*\*\*  
## Self.Perceived.English 0.0029642 0.0004982 5.950 4.41e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2273 on 646 degrees of freedom  
## Multiple R-squared: 0.1616, Adjusted R-squared: 0.1577   
## F-statistic: 41.51 on 3 and 646 DF, p-value: < 2.2e-16

##   
## Call:  
## lm(formula = LEXTALE.Test.Score ~ Age + Exposed.To.English +   
## Self.Perceived.English, data = new.data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -25.3161 -5.0053 -0.4223 4.4315 31.4900   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 48.68550 1.96076 24.830 < 2e-16 \*\*\*  
## Age 0.36509 0.06003 6.082 2.05e-09 \*\*\*  
## Exposed.To.English 0.16344 0.04111 3.976 7.83e-05 \*\*\*  
## Self.Perceived.English 0.11665 0.01874 6.225 8.75e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.127 on 633 degrees of freedom  
## Multiple R-squared: 0.1472, Adjusted R-squared: 0.1432   
## F-statistic: 36.43 on 3 and 633 DF, p-value: < 2.2e-16

#### Final Model Discussion:

The model has been refit by removing high leverage points and outliers, which has resulted in the finding that the model without the refitting is a better model. This is due to the R2 and Adjusted R2 values being higher, explaining more variance, as well as the F-statistic (41.51 vs 35.13) indicating a stronger predictor significance. Residual Standard Error being lower in the second model (.2273 vs .221) is not significant enough.

This final model is:

LEXTALE.Test.Score = 5.2393156 + 0.0099346 \* Age + 0.0055241 \* Exposed.To.English + 0.0031533 \* Self.Percieved.English

This model explains that the age of a participant, the amount of time they spend being exposed to English and their confidence level of language are the most important factors in language proficiency. None of the variables seem to be highly correlated.