English Proficiency, Citizenship, and Their Impact on Household Income: A Comprehensive Analysis Within the States of New Jersey, New York, and Pennsylvania

Ria Aleksanyan and Steve Ta

2024-08-16

I. Introduction

As the United States becomes increasingly diverse, it is glaringly obvious that immigrants and non-native English-speaking communities may not get the same opportunities as native-born citizens. Some of the many challenges immigrants face when moving to America, include navigating language barriers and working around legal statuses. Our assumption is that American-born U.S. citizens with the highest levels of English proficiency earn higher incomes than other groups of people, since they have access to more well-paying job opportunities and less legal restrictions placed on their working statuses. If such is the case, America, albeit diverse, does not provide an equitable foundation for people of varying backgrounds, which would only further contribute to income inequality, discrimination, and lack of inclusivity within communities. By investigating the relationship between English proficiency and citizenship status and their combined impact on household income in the states of New York, New Jersey, and Pennsylvania, we can demonstrate the need to implement policies that would address income disparities and promote economic equity among diverse populations. With the help of such policies, we can work toward reducing poverty and improving the overall quality of life for immigrants and non-native English-speaking communities.

II. Ideal Experiment

Realizing the limitations both in practicality and ethics—such as some jobs having low-entry pays but growing higher after long-term employment or the near impossibility of obtaining a sample of individuals with identical characteristics—in an ideal world, this experiment would be divided into five phases:

Phase 1: Sample Collection

We would begin by collecting a random sample of equal proportions of individuals with the following citizenship statuses:

– U.S. Citizen (Born in the U.S.)

– U.S. Citizen (Born outside of the U.S.)

– U.S. Citizen by Naturalization

– Non-U.S. Citizen

These participants would be selected with **no English proficiency** and would have controlled characteristics, including **age, education, and work experience**.

Phase 2: English Language Proficiency Classes

Participants would then be divided into two equal parts within each citizenship group:

– Treatment Group: Receives intensive English courses focusing on all language skills.

– Control Group: Receives no additional language instruction.

Phase 3: Post-Intervention Assessment

After completing the intervention, both the treatment and control groups would undergo an **English proficiency test**. Participants would then be scored on a **1-4 scale** based on their test results:

– 1: Very Well

– 2: Well

– 3: Not Well

– 4: Not Very Well

Phase 4: Employment Placement

Next, participants would be placed in **equivalent entry-level jobs** across various sectors, specifically those that require English language skills.

Phase 5: Final Assessment

Finally, after **ten years** of employment, we would collect **household income data** from all participants to evaluate the long-term impact of English proficiency and citizenship status on their economic outcomes.

III. Data Overview

1. Data Inspection

During the course of this project, we accessed the *2019 ACS PUMS DATA DICTIONARY*, released to public access on October 15, 2020, by the U.S. Census Bureau. We used three variables from the dataset:

– ENG: Ability to speak English. Discrete variable:

```
- b .N/A (less than 5 years old/speaks only English)
- 1 .Very well
- 2 .Well
- 3 .Not well
- 4 .Not at all
```

– CIT: Citizenship status. Discrete variable.

```
- 1. Born in the U.S.
- 2. Born in Puerto Rico, Guam, the U.S. Virgin Islands, or the Northern Marianas
- 3. Born abroad of American parent(s)
- 4. U.S. citizen by naturalization
- 5. Not a citizen of the U.S.
```

– HINCP: Household Income (in U.S. Dollar). Continuous variable:

```
- bbbbbbb .N/A(GQ/vacant)
- 0 .No household income
- -59999 .Loss of -$59,999 or more
- -59998..-1 .Loss of $1 to -$59,998
- 1..9999999 .Total household income in dollars (Components are rounded)
```

2. Casual Graph

The relationship between household income (HINCP), citizenship status (CIT), and English proficiency (ENG) can be effectively visualized through a causal graph:

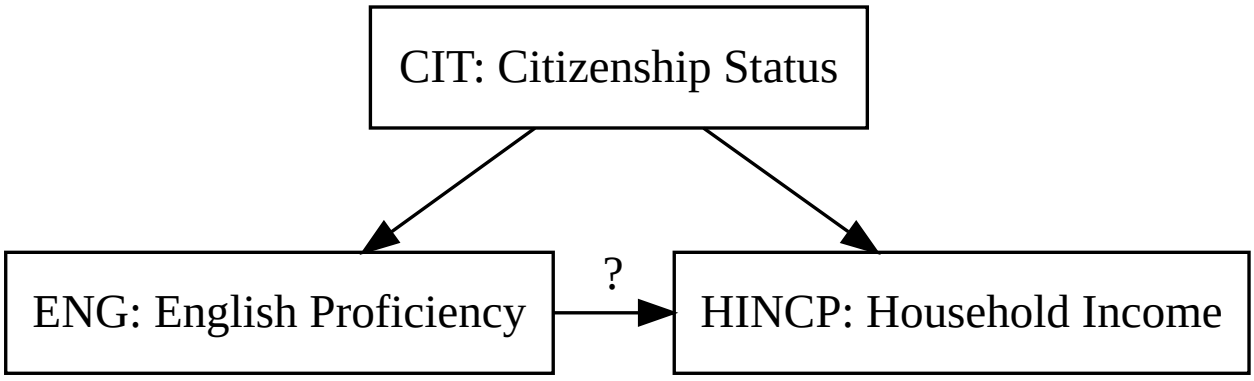

Figure 1: Casual Graph between HINCP, CIT, ENG

This graph represents the hypothesized causal pathways between the variables, guiding our analysis and interpretation of how English proficiency and citizenship status impact household income. In this context, citizenship status serves as a confounder, influencing both English proficiency and household income. Citizenship status can affect access to education and resources, which in turn influences English proficiency. Additionally, it can directly impact the types of jobs individuals can obtain, given that many workplaces enforce strict policies regarding the legal status of employees. This, in turn, affects their income and ultimately their household income.

We are particularly interested in examining whether English proficiency affects household income while accounting for the confounding effect of citizenship status.

3. Read the Data

For the purpose of this project, we made specific decisions to read in the dataset regarding the variables:

- ENG Variable: We only included individuals with a marked English proficiency level, meaning we excluded any N/A values.
- CIT Variable: We combined groups 2 and 3 due to the similarity in their household income statistics based on English proficiency. Additionally, individuals born abroad to American parents are granted U.S. citizenship by their parents, which provided further justification for grouping these two categories together. For more information, consult the U.S. State Department's guidelines on Acquisition of U.S. Citizenship for a Child Born Abroad.
- HINCP: We only included individuals with an income of 1 or higher. This decision was made to exclude cases with zero or negative income, which could skew the analysis. Such values might represent temporary financial setbacks, non-reporting, or other anomalies that don't reflect a stable household income

4. Data Overview

The table below shows numbers of individuals in each citizenship status and English proficiency that we used in this project:

```
##
##              Born in the US Born outside of the US (US Citizen)
##   Very Well            1715                                 311
##   Well                  194                                 161
```

```
##   Not Well               62                                    88
##   Not At All              6                                    31
##
##              Not a US Citizen US Citizenship by Naturalization
##   Very Well               703                                2098
##   Well                    437                                1140
##   Not Well                384                                 570
##   Not At All              144                                 114
```

IV. Graphs

1. Histogram Analysis

Using the HINCP variable with each facets of ENG variable, we computed the histogram of household income:

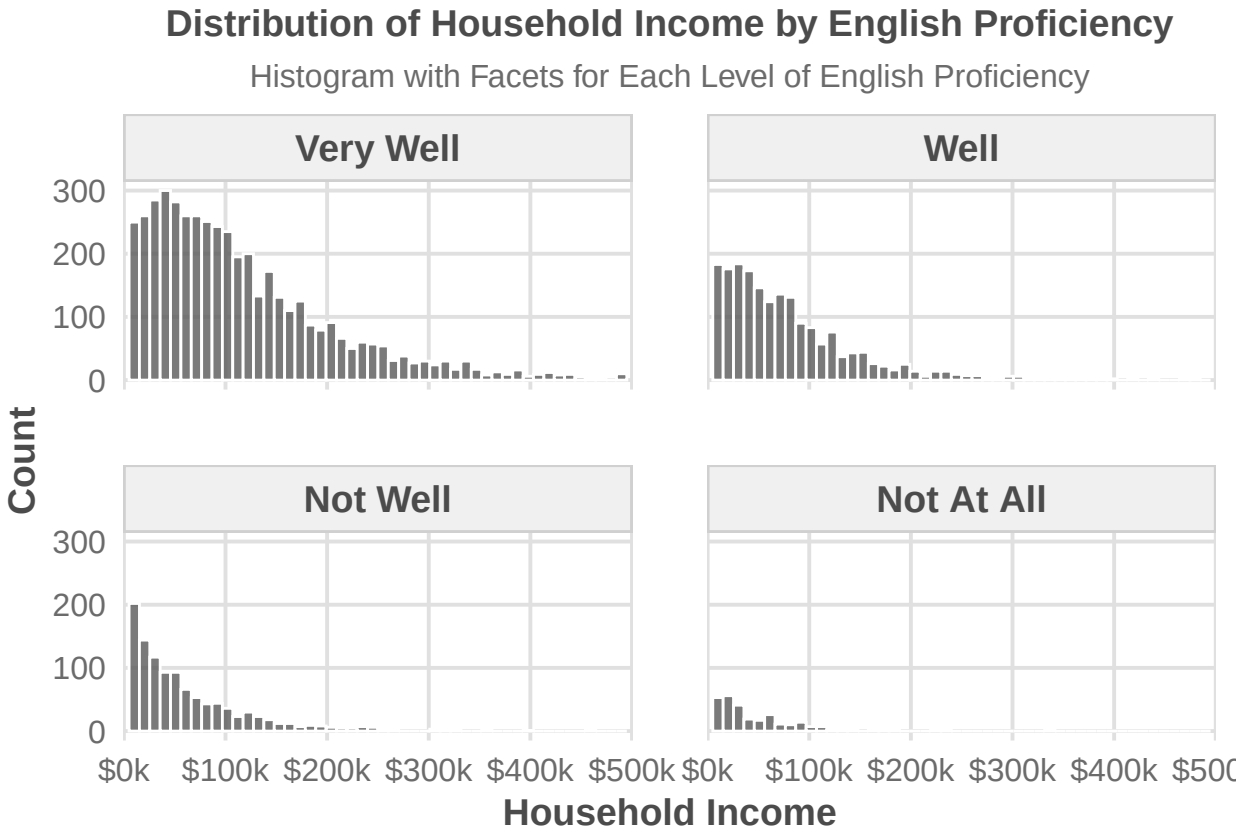

Figure 2: Histogram of Household Income by English Proficiency Levels

We also used the HINCP variable to compute the mean and median for each ENG levels:

```
## # A tibble: 4 x 3
##   ENG        Mean_Income Median
##   <fct>            <dbl>  <dbl>
## 1 Very Well      131248.  94000
## 2 Well            82372.  61360
## 3 Not Well        63362.  43075
## 4 Not At All      51363.  31200
```

The analysis of the histogram and summary statistics indicates a positive correlation between English proficiency and household income. Specifically, both the mean and median household income demonstrate an upward trend as English proficiency improves, as outlined in the summary table. The histogram further reveals an increase in income variability corresponding to higher levels of English proficiency, evidenced by the broader range of household income. This pattern suggests that individuals with higher English proficiency not only tend to earn more but also experience greater variability in their income distribution, likely reflecting broader access to diverse and higher-paying job opportunities.

2. Plot Analysis

Using the HINCP variable with each ENG levels for each facet of citizenship status, we computed the jitter plot of household income:

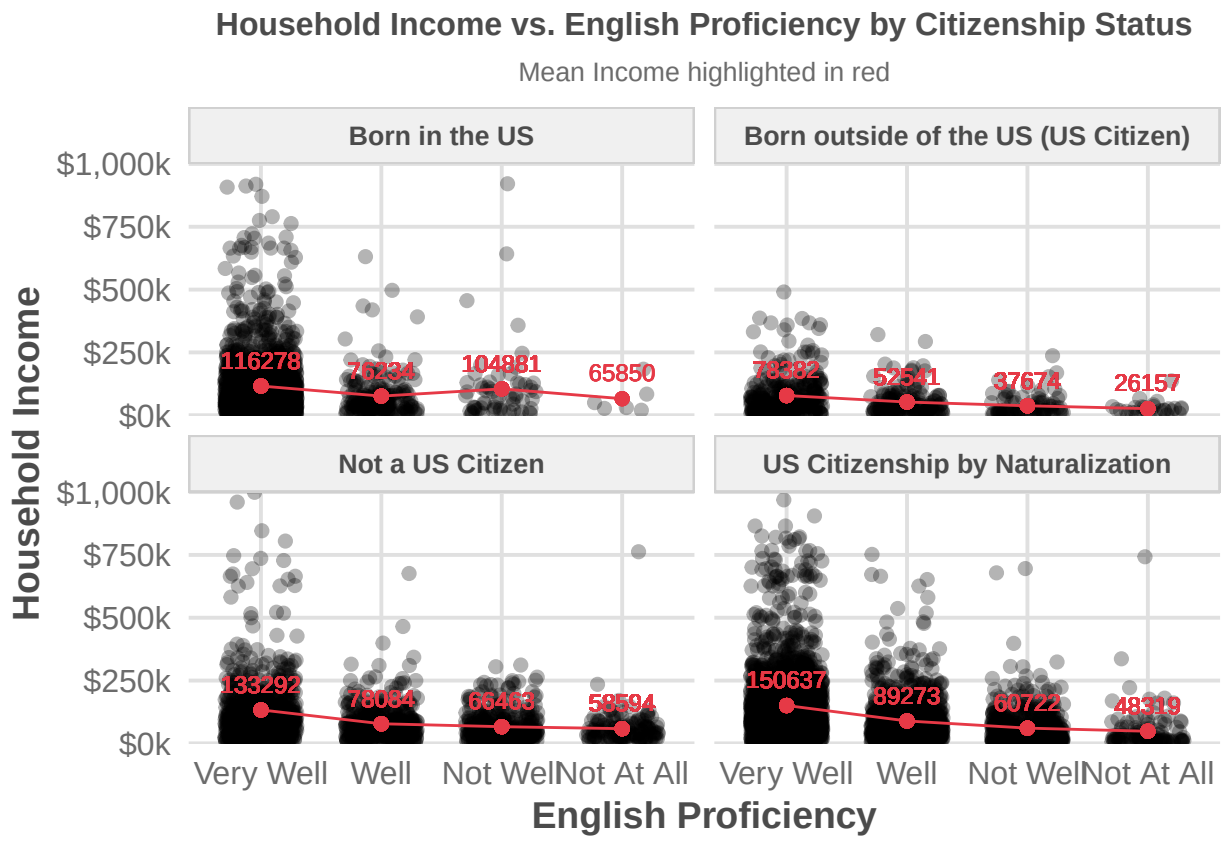

Figure 3: Jitter Plot of Household Income by each English Proficiency Levels with Each Facet Citizenship Status

We also used the HINCP variable to compute the mean and median for each ENG levels for each facet of citizenship status:

```
## # A tibble: 16 x 4
## # Groups:   citizenship_status [4]
##    citizenship_status                ENG        Mean_Income Median
##    <chr>                             <fct>            <dbl>  <dbl>
##  1 Born in the US                    Very Well      116278.  78000
##  2 Born in the US                    Well            76234.  48300
##  3 Born in the US                    Not Well       104881.  67960
```

```
##  4 Born in the US                        Not At All      65850   40350
##  5 Born outside of the US (US Citizen) Very Well       78382.  54500
##  6 Born outside of the US (US Citizen) Well            52541.  35000
##  7 Born outside of the US (US Citizen) Not Well        37674.  17500
##  8 Born outside of the US (US Citizen) Not At All      26157.  20000
##  9 Not a US Citizen                      Very Well     133292.  99100
## 10 Not a US Citizen                      Well           78084.  62000
## 11 Not a US Citizen                      Not Well       66463.  46950
## 12 Not a US Citizen                      Not At All     58594.  45100
## 13 US Citizenship by Naturalization    Very Well     150637. 110000
## 14 US Citizenship by Naturalization    Well           89273.  68400
## 15 US Citizenship by Naturalization    Not Well       60722.  40000
## 16 US Citizenship by Naturalization    Not At All     48319.  20540
```

The jitter plot, combined with summary statistics, suggests a strong association between English proficiency, citizenship status, and household income. The analysis reveals that higher English proficiency and U.S. citizenship are generally linked to higher household income. Each citizenship facet exhibits an upward trend in household income as English proficiency increases.

However, an anomaly is observed in the "Born in the U.S." facet, where the "Not Well" English proficiency group disrupts this trend due to inconsistencies in the mean and median incomes. This deviation is likely due to a small sample size and the presence of significant outliers, which skew the data. Despite this inconsistency, the overall trend remains valid.

The data also indicate that individuals who are U.S. citizens, either by birth or naturalization, typically earn more than those born outside the U.S. or non-citizens, particularly in the lower English proficiency categories ("Not Well" and "Not At All"). This can be attributed to the greater access to educational opportunities and resources available to U.S. citizens, which are often restricted or less accessible to non-citizens.

Interestingly, the household income for non-citizens and naturalized U.S. citizens who have English proficiency of "Very Well" or "Well" is unexpectedly higher than for those born in the U.S. This can be explained by the fact that many non-citizens and naturalized citizens often come to the U.S. with extensive educational and financial backgrounds, frequently arriving through visas[1] for advanced studies or high-paying careers. Additionally, immigration policies such as the EB-5 Immigrant Investor Program, which grants U.S. green card (in which they can later on petitioning for U.S Citizenship) in exchange for substantial investments, or the EB-1 and EB-2 visas for individuals with extraordinary abilities, may contribute to this trend. Consequently, the "Very Well" and "Well" English proficiency groups within these populations show particularly high household incomes.

In conclusion, the overall trend supports the notion that higher English proficiency and U.S. citizenship are associated with higher household income, despite certain deviations in specific subgroups.

[1]For reference, we suggest browsing the U.S. Citizenship and Immigration Services on mentioned visas' policies: https://www.uscis.gov/working-in-the-united-states/permanent-workers/employment-based-immigration-first-preference-eb-1

V. Regression Analysis

Using the HINCP as the dependent variable and CIT & ENG as independent variables, we produced a multivariable linear regression model represents the relationship between household income $\bar{Y}$ and independent variables related to citizenship status and English proficiency:

```
##
## Call:
## lm(formula = HINCP ~ citizenship_status + ENG, data = acs_data)
##
## Coefficients:
##                                          (Intercept)
##                                               119625
## citizenship_statusBorn outside of the US (US Citizen)
##                                               -27327
##                     citizenship_statusNot a US Citizen
##                                                15279
##    citizenship_statusUS Citizenship by Naturalization
##                                                25672
##                                              ENGWell
##                                               -53580
##                                           ENGNot Well
##                                               -72654
##                                         ENGNot At All
##                                               -82770
```

this model shows $\bar{Y} = a_0 + b_1\bar{X}_1 + b_2\bar{X}_2 + b_3\bar{X}_3 + b_4\bar{X}_4 + b_5\bar{X}_5 + b_6\bar{X}_6$ in which:

- a_0: x-intercept (Born in the U.S. and Have "Very Well" English Proficiency)
- b_1: Born outside of the US (U.S. Citizen)
- b_2: Not a U.S. Citizen
- b_3: U.S. Citizenship by Naturalization
- b_4: Having "Well" English Proficiency
- b_5: Having "Not Well" English Proficiency
- b_6: Having "Not At All" English Proficiency

All independent variables X_1 through X_6 are binary (0 or 1), indicating the presence or absence of the characteristic described by each coefficient.

The regression model suggests that individuals born in the United States with "Very Well" English proficiency tend to have higher household incomes, as indicated by the intercept a_0. The coefficients associated with "Not Well" and "Not At All" English proficiency (b_5 and b_6) demonstrate a strong negative effect on household income, implying that lower English proficiency significantly reduces expected earnings.

Although the coefficients for non-citizens and naturalized U.S. citizens (b_2 and b_3) indicate a positive impact on household income, it is important to consider the previously discussed skewness in the data regarding the "Very Well" and "Well" English proficiency levels among non-citizens. Despite this, the overall conclusion of the regression model remains valid: lower English proficiency is associated with lower household income, regardless of citizenship status.

VI. Confidence Interval

Given the interest in understanding how English proficiency impacts household income, particularly for individuals born in the U.S. and those who are not U.S. citizens, we employed a bootstrapping method. We generated 5000 bootstrap samples, with the sample size matching the number of individuals in each English proficiency category ("Very Well" and "Not At All"). This approach allowed us to estimate the 95% confidence interval for the mean household income within these groups.

Born in the U.S. :

```
##   English_Proficiency  Lower_CI  Upper_CI
## 1           Very Well 109837.80 122955.16
## 2          Not At All  55333.82  77561.82
```

Not a U.S. Citizen :

```
##   English_Proficiency  Lower_CI  Upper_CI
## 1           Very Well 123785.59 143168.73
## 2          Not At All  55876.32  61490.97
```

The 95% confidence intervals provide a range within which we are 95% confident that the true mean household income for each level of English proficiency, specifically for individuals born in the U.S. and non-U.S. citizens, lies. The non-overlapping nature of these intervals strongly suggests that English proficiency significantly impacts household income.

Although the confidence intervals for non-U.S. citizens exhibit both a higher upper bound and a lower lower bound, this variability can be attributed to the previously discussed skewness in the data, particularly regarding the "Very Well" and "Well" English proficiency levels among non-citizens. Notably, the confidence interval for individuals born in the U.S. with "Not at All" English proficiency shows higher bounds compared to non-citizens. This indicates that non-citizens with limited English proficiency tend to earn less than their U.S.-born counterparts, emphasizing the influence of both citizenship and language skills on income.

VII. Possible Policies

After getting a better understanding of the relationship between English proficiency, citizenship status, and household income, there are several policies we could propose to address income disparities and help promote an inclusive and equitable environment within American communities. For example, we could advocate for accessible, free English language courses in schools and workplaces for immigrants and non-native speakers to help improve their English proficiency, and thus increase their employability. Additionally, we could provide targeted job training programs, where immigrants and non-native speakers are provided with free workshops to help them develop specific job skills tailored to their local community's workplace demands. Finally, we could advocate for more anti-discrimination laws, which would prevent discrimination based on language proficiency and citizenship status during the hiring process, ensuring that everyone has equal access to job opportunities.

VIII. Conclusions & Limitations

Our research had many limitations, such as the lack of accounting for other confounding variables, which may have influenced household income. Examples of such factors include education level, work experience, sex, etc. Additionally, it is important to note that the data collected is based only on three states, so it may not accurately represent the entire immigrant and non-native English speaker population in America. Finally, the ACS data collected is from 2019, so it may not reflect recent changes in immigrant and labor policies.

That being said, our conclusion is that it appears as though English proficiency and citizenship status have a significant impact on household income, which specifically affects immigrant and non-native English-speaking communities in New York, New Jersey, and Pennsylvania. There is a strong positive correlation between English proficiency and household income, even after accounting for the confounding variable of one's citizenship status. The income disparities we observe through our graphs suggest that America does not provide equitable economic opportunities to everyone. There are language barriers and legal restrictions in place, which heavily contribute to income inequality and discrimination. To address these issues, it is essential to implement targeted policies that enhance English language education, provide legal support, and promote inclusive hiring practices. By doing so, we can create a more just and equitable society, where everyone, regardless of their background, has the opportunity to achieve financial stability and social mobility.

REFERENCES

[1]. U.S. Census Bureau. (2020). *American Community Survey (ACS) 2019 Public Use Microdata Sample (PUMS) Data Dictionary*. Retrieved from [https://www2.census.gov/programs-surveys/acs/tech_docs/pums/data_dict/PUMS_Data_Dictionary_2019.pdf]