

# Analysis of Sociocultural Adaptation

Leung, Yvonne  
Shih, Hsiu-Yu (Sherry)  
Northeastern University  
Apr 15, 2025



# Data Overview-Total 60 Entries

---

Variable Name	Description
participant.ID	A unique identifier assigned to each student
Native.language	The student's native language ('Chinese', 'Spanish', 'Vietnamese', 'Ukrainian', 'ran', 'Portuguese', 'oriya', 'Korean', 'Russian', 'Greek', 'French')
High.School.Language	The primary language of instruction at the student's high school('Chinese', 'English', 'Spanish', 'Vietnamese', 'Greek')
Time.in.the.US	How long the student has lived in the U.S. ('Less than one month', 'One to three months', 'Three months to one year', 'One to two years', 'More than two years')
X.Pre.1 to X.Pre.21	Self-rated adaptation items before the experience (Likert scale: 1 = Strongly Disagree to 7 = Strongly Agree)
X.Post.1 to X.Post.21	Same items as Pre-Survey, rated after the experience (Likert scale: 1 = Strongly Disagree to 7 = Strongly Agree)



# Data Cleaning and Preprocessing

---

- Native.language was recoded into two categories: "Chinese" vs. "Other" for Native Language. Factors were set with "Other" as the reference.
- High.School.Language was recoded into two categories: "English" vs. "Other". Factors were set with "Other" as the reference.
- Time.in.the.US was grouped into: "Less than 1 year", "1-2 years", and "More than 2 years". Factors were set with "Less than 1 year" as the reference.



# Missing Value Imputation

---

- Missing values were thoughtfully filled in using each student's own average—unless they skipped all pre-survey items, in which case we couldn't include them.



# Summative Score Creation and Reshaping

---

- We gave each student a "Pre-Sum" and a "Post-Sum" score by adding up how they rated their adaptation on a series of questions before and after a period of time.
- From this, we calculated a simple but powerful number: the "Change Score" =  $\text{Post-Sum} - \text{Pre-Sum}$ . This told us how much each student improved.
- We reshaped the data to allow comparison at both the individual and group levels across time. (long format)



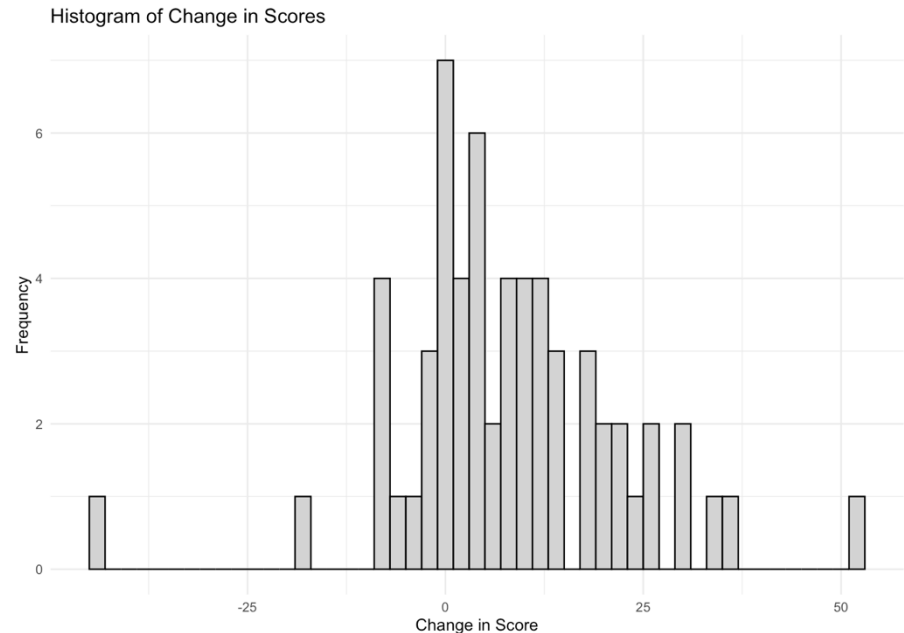
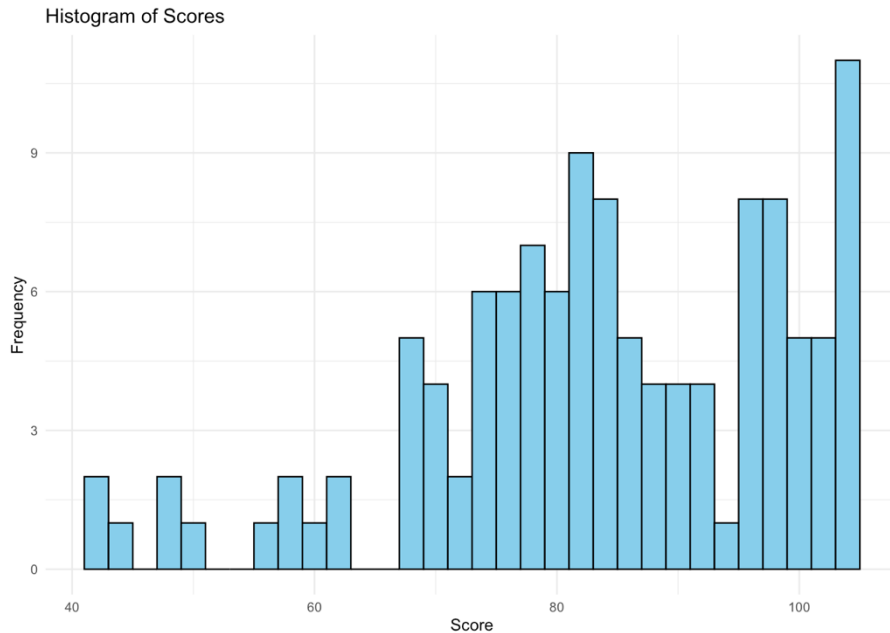
# Comparison between original dataset and cleaned dataset

---

Aspect	Original Dataset	Cleaned Dataset
Participants	60 entries	59 entries (1 removed with all missing Pre responses)
Variables	54 columns	54 columns
Key Variables	Pre & Post scores across 21 adaptation items	Same
Missing Data	Present in multiple Pre/Post items	Imputed using each participant's row-wise mean
Categorical Variables	Text-based (with inconsistent formatting)	Recoded (e.g., "chinese" → "Chinese", grouped into factors)
Derived Variables	Not included	Pre_Sum, Post_Sum, and change_score added
Data Structure	Wide format	Wide and reshaped into long format for time-based modeling
Skewness	Not assessed	Examined for raw, log, sqrt, and change score distributions

# Score Transformation and Skewness Comparison

- Before modeling, we examined the distribution of the scores:

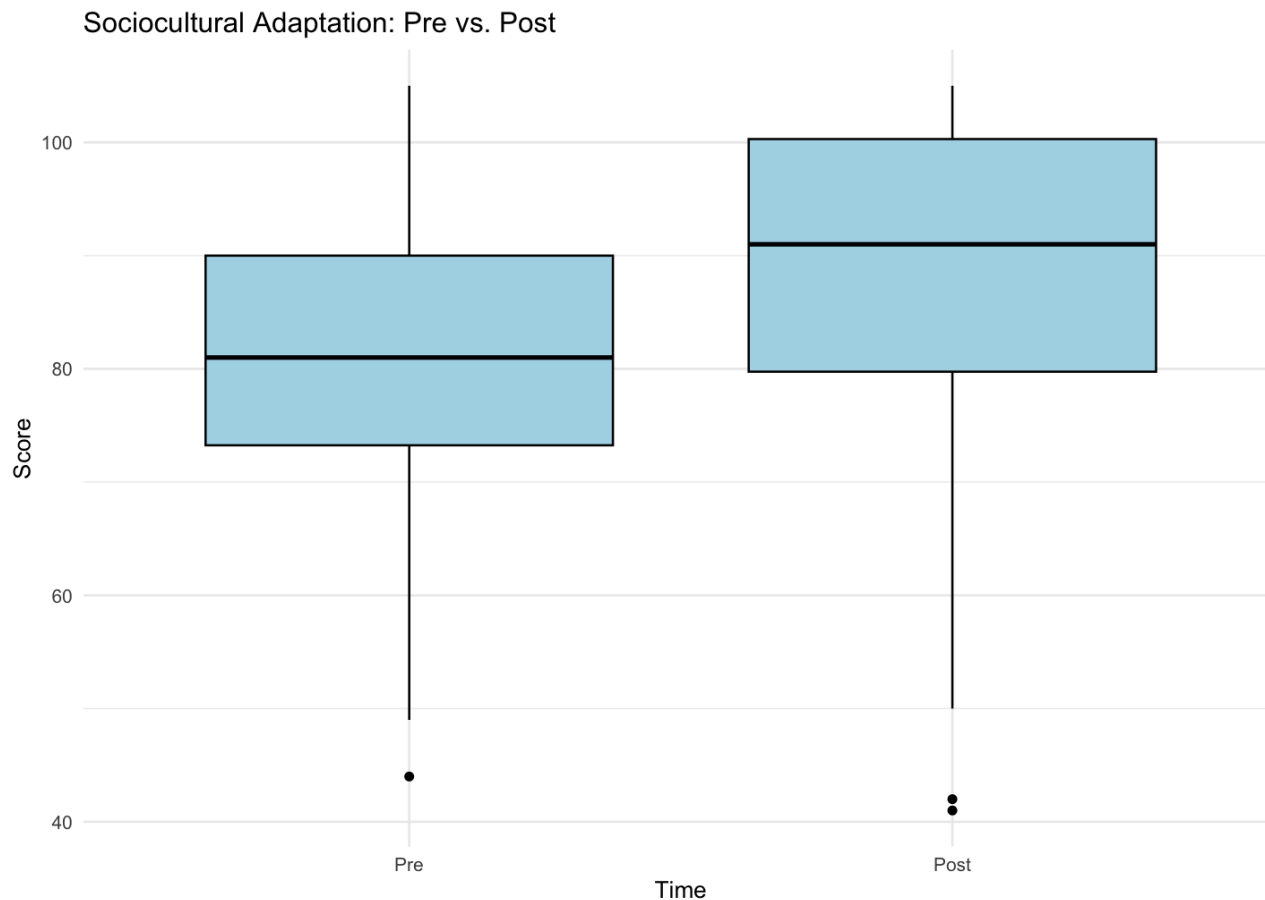


	Score	Change Score
Skewness	-0.688	-0.147
Interpretation	moderately left-skewed, meaning more students rated themselves higher with fewer low scores. not severely skewed and still acceptable for analysis	very close to symmetric

# Social Adaptation Score: Pre vs. Post

---

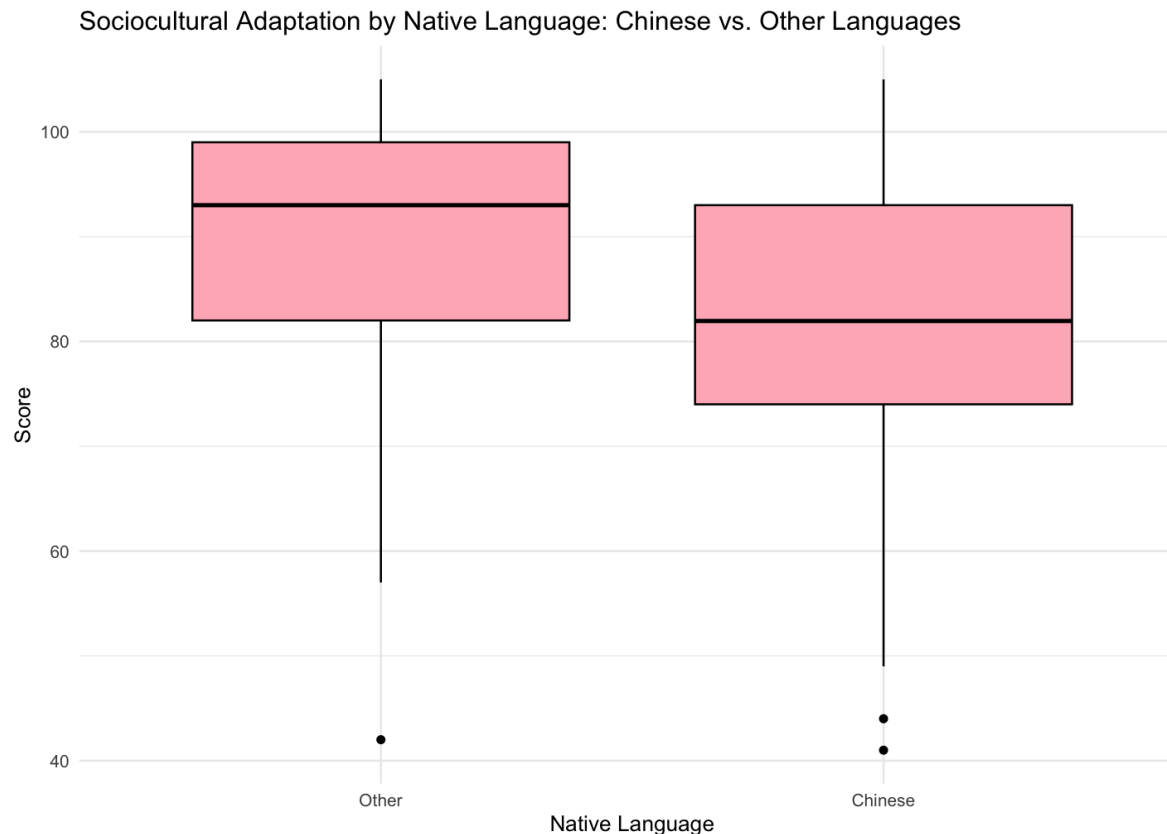
- Most students feel more socioculturally adapted over time.





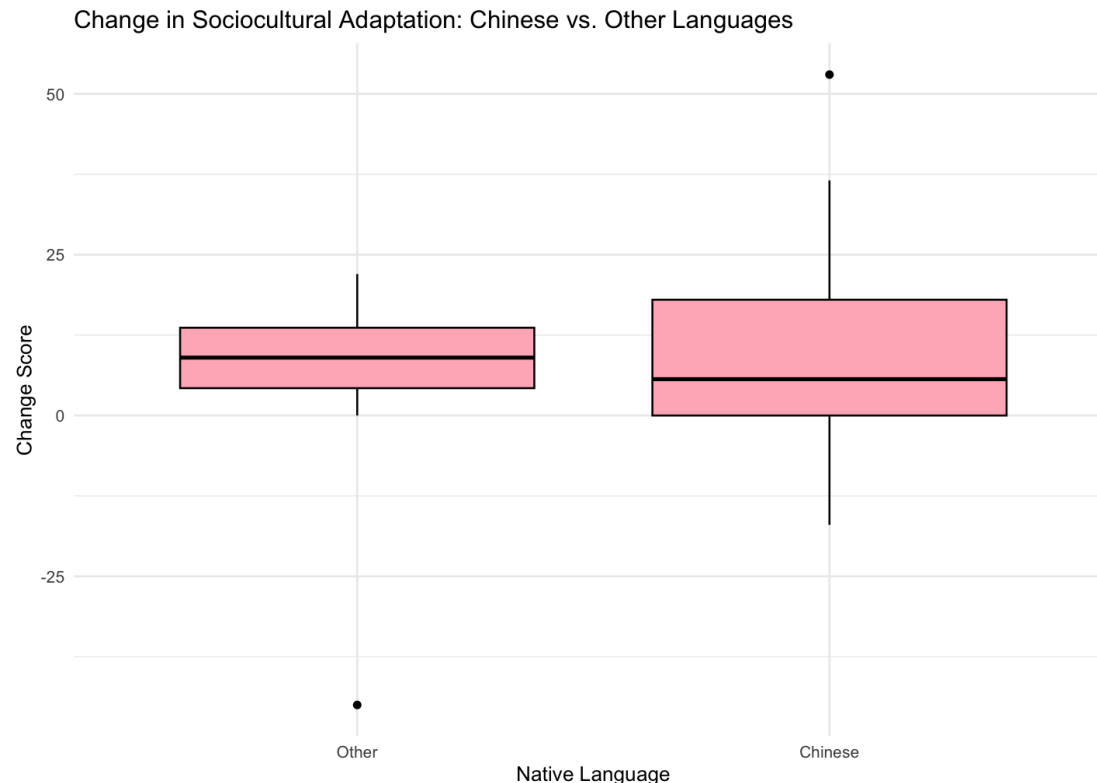
# Social Adaptation Score by Native Language: Chinese vs. Other Languages

- On average, students from other language backgrounds rated themselves more adapted than their Chinese-speaking peers.
- Chinese-speaking students may face more challenges in adapting socially or culturally.



# Social Adaptation Change by Native Language: Chinese vs. Other Languages

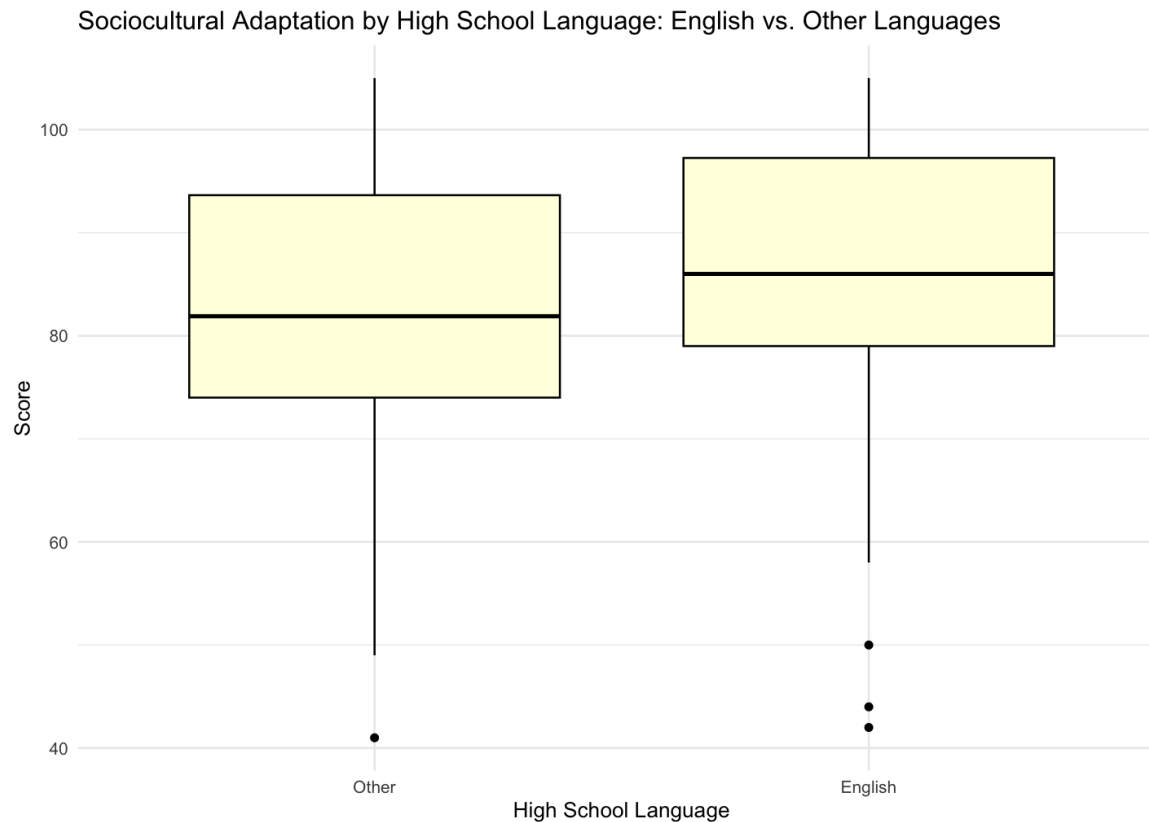
- Students from non-Chinese language backgrounds tend to show consistent, positive growth in adaptation. Their change scores cluster closely together, suggesting a relatively uniform experience of improvement.
- Students who speak Chinese as their native language demonstrate a much wider range of outcomes. While some report strong gains, others show little improvement—or even a decline—in adaptation scores over time.



# Social Adaptation Score by High School Language: English vs. Other Languages

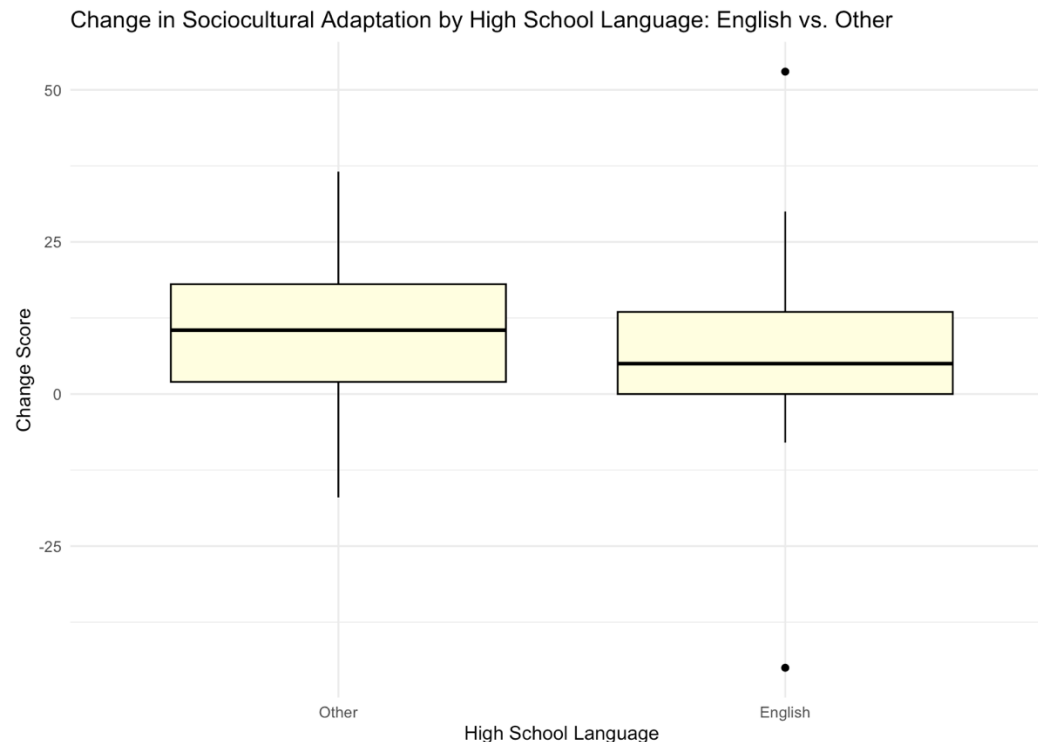
---

- The difference between students who went to high schools where English or not is mild. In short, this factor alone isn't driving big differences.



# Social Adaptation Change by High School Language: English vs. Other Languages

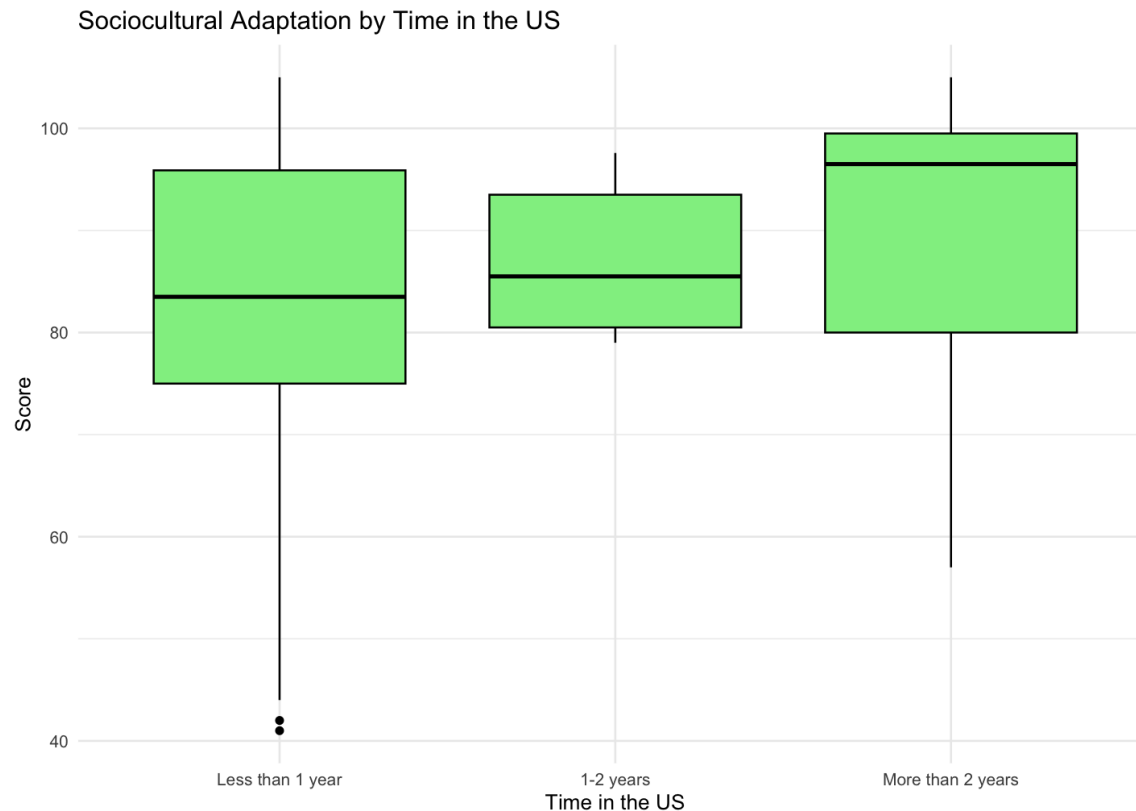
- The median change score is slightly higher for students who studied in non-English high schools. This suggests that these students, on average, reported more growth in sociocultural adaptation over time.
- However, the distribution of scores is relatively similar between the two groups. Both show a wide range of experiences, from students with strong gains to those with minimal or even negative changes.



# Social Adaptation Score by Time in the US

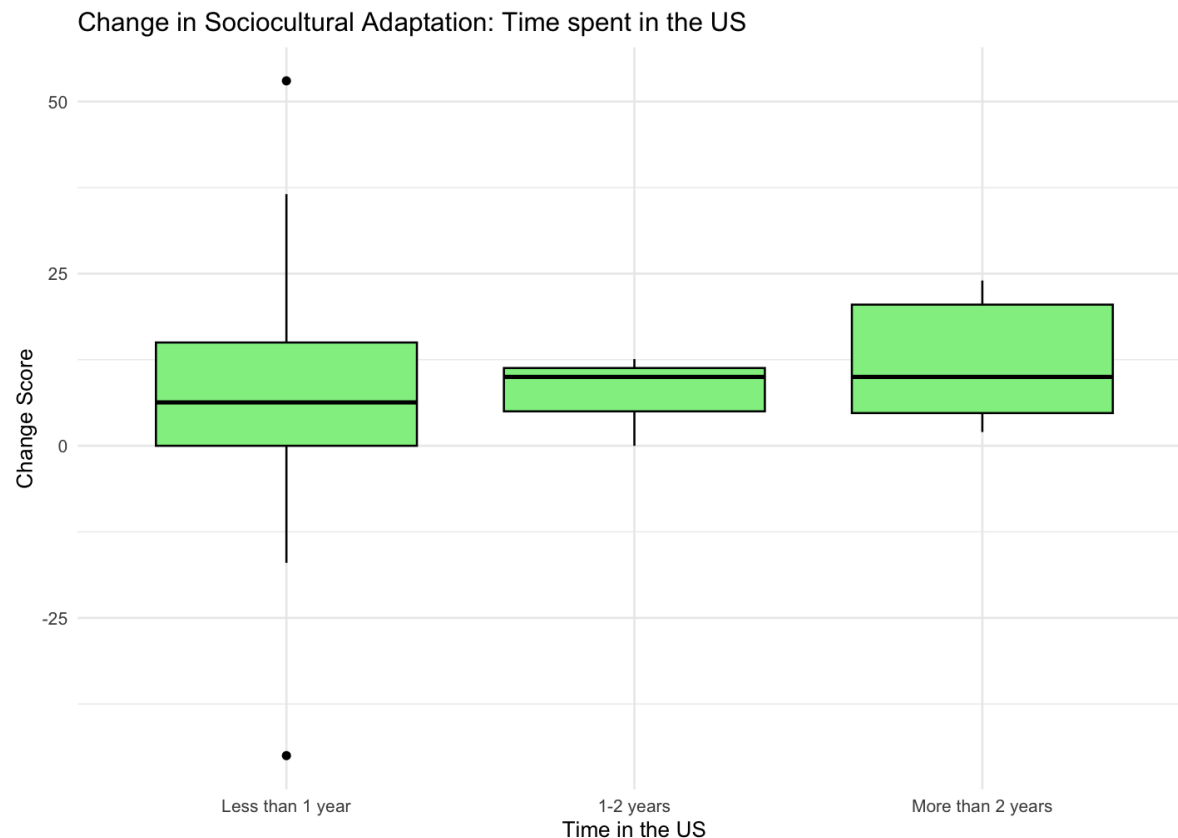
---

- The difference of social adaptation between the time in the US were not dramatic.
- Time in the US 'more than 2 years' group scoring a bit higher — the difference isn't dramatic.



# Social Adaptation Change by Time in the US

- Students who have been in the U.S. for more than 2 years show the highest median improvement in adaptation scores.
- Those in the 1–2 years category show a modest gain, with a relatively narrow range of improvement.
- Students with less than 1 year in the U.S. display the widest range of outcomes.



# Modeling Sociocultural Change

---

We ran five models to identify what predicted change in adaptation.

- 1) Model 1: Linear Regression on Change Score
- 2) Model 2: Linear Regression on Change Score with Interaction
- 3) Model 3: Linear Regression on Scores (long format)
- 4) Model 4: Linear Regression on Scores with Interaction (long format)
- 5) Mixed-Effects Model



# Model 1: Linear Regression on Change Score

---

What predicts someone improving more over time?

- Students who started off lower (lower Pre\_Sum) showed more growth.
- Language background or time in the U.S. didn't strongly explain who improved.
- $R^2$  was 0.138, meaning about 14% of variation in improvement is explained by these factors.





## Model 2: Linear Regression on Change Score with Interaction

---

Does Language Moderate Starting Score?

We added an interaction: does the effect of Pre\_Sum depend on whether you speak Chinese?

- The interaction wasn't significant, so Chinese vs. Other didn't change the relationship between starting point and improvement.
- Model fit stayed about the same.



# Model 3: Linear Regression on Scores (long format)

---

We then looked at all scores across time (Pre and Post together), asking: are scores higher after time passes?

- Time” variable was significant — students rated themselves higher after some time in the program.
- Chinese speakers still reported lower scores overall.
- $R^2$  was 0.102 — a bit less explanatory power than the earlier model but still meaningful.



# Model 4: Linear Regression on Scores with Interaction (long format)

---

Do Chinese Students Improve Differently?

We added an interaction between Time and Native Language. Asking: do Chinese speakers grow differently than others?

- The interaction wasn't significant, and the model fit didn't improve much.



# Mixed-Effects Model

---

We accounted for the fact that each participant gave two scores (Pre and Post).

We used a model that lets each person have their own baseline.

- “Time” was again significant — students improved in social adaption
- This model confirmed that the improvement wasn’t just due to chance or specific individuals. It was consistent across participants.



# Model Comparison

Model	Purpose	Significant Findings	Adjusted R <sup>2</sup>
<b>Model 1</b> Linear Regression on Change Score	Examine what predicts improvement over time	<b>Lower Pre_Sum → more improvement (p &lt; .01)</b>	0.138
<b>Model 2</b> Linear Regression + Interaction	Check if Pre_Sum effect differs by language	Interaction not significant	0.137
<b>Model 3</b> Linear Regression on Raw Scores	Check effect of time (Pre vs. Post) on scores	<b>Post scores significantly higher (p &lt; .01)</b>	0.102
<b>Model 4</b> Interaction Between Time & Language	Test if improvement over time differs by language	Interaction not significant	0.095
<b>Model 5</b> Mixed-Effects Model	Account for repeated measures within participants	Time significant (p < .001)	-



# Conclusion

---

This analysis demonstrates that students in our program experienced meaningful growth in sociocultural adaptation over time. Two key models offer complementary insights:

- Model 1 shows that students who began with lower adaptation scores experienced the greatest improvement, suggesting our support systems are especially impactful for those who need it most.
- Model 3 confirms that, on average, adaptation scores increased significantly from pre- to post-assessment, highlighting the overall effectiveness of the program.
- While language background and time in the U.S. had modest effects, they did not strongly alter the trajectory of adaptation. These findings emphasize the value of early support and indicate that continued programmatic efforts should focus on identifying and empowering students who start with lower confidence or adaptation readiness.



# Thank you!

---