

MA213 LAB Group 7

Section: C1

Abstract

This study analyzed the potential correlation between a student's language abilities and their choice of major. Increased language abilities can lead to more job opportunities and a larger variety of interests, so we looked to assess this relationship. For the study, we invited 207 Boston University students to fill out a survey regarding their major, language abilities, language learning environment, and reason for choosing their major. We found that the majority of students spoke two languages and this correlated with the large majority of students in the College of Arts and Sciences (CAS), which has a second language requirement. However, CAS is the largest college at Boston University, yielding the largest range of majors, so this correlation can not be proven solely on this fact. We were unable to find a distinctive correlation between language and choice of major in BU students.

Introduction

The diversity of students on a college campus allows for a large variety of languages and personal backgrounds represented. The questions of how majors are distributed among colleges and how language abilities could impact this are of interest to our group. Some specific topics of interest for us are the number of languages a student speaks, how fluent they are, what their major is, and why they chose their major, etc. Another subject of interest for us is the differentiation between stem and liberal arts majors in relation to language abilities. Looking at data from the pilot experiment though, we did not see any significant correlation with the small sample size. With the final study, we noticed trends between the number of languages spoken, the category of major/college, and a student's willingness to take additional language courses in relation to their learning environment.

Our survey questions can be broken into three groups: Demographics, quantitative results on the number of languages spoken, and qualitative results on the categorized major, learning environment, and willingness to take additional language courses. To see whether the amount of languages people speak influences one's decision to choose a major, we also need to see possible factors such as living background and personal will to learn additional languages. Table 1 presents a comprehensive overview of our survey questions, providing the key inquiries employed in our study.

Table 1: Survey Questions

Demographics	Quantitative	Qualitative
Year	Age	Rank your fluency in the language(s) other than your first language
Major	Number of languages spoken	(Only proficient)

<p>What is the most important reason for going to college for you?</p> <p>What language(s)? Do not include beginner languages.</p> <p>Why did you choose your major?</p>	<p>(fluently)</p>	<p>languages)*</p> <p>How would you best categorize your major?</p> <p>Which BU college are you currently a part of?</p> <p>Learning Environment</p> <p>Willingness to take additional language courses in college (not for HUB requirements)*</p>
--	-------------------	--

Noted: Questions marked with "*" are answered from scale 1 (very bad) to 5 (very good)

Table 2: Hypothesis

<p>Null Hypothesis: The number of languages a BU student speaks does not impact their decision of choosing a major.</p>
<p>Alternative Hypothesis: The number of languages a BU student speaks impacts their decision of choosing a major.</p>

Experimental Methods

We used qualitative and quantitative data in our survey and gave numerical scales and ranking for certain data sets. The population of interest for our study is the population of Boston University students.

Data collection

In pursuit of mitigating response bias and achieving a precise estimate, our data collection methodology entails the systematic utilization of a Google form distributed via QR codes. The survey dissemination targets students from different campus locations, including our friends and classmates, both in-person and online, encompassing a relatively random sampling approach.

Survey design

Our survey endeavors to receive responses from a diverse group of students affiliated with Boston University, with the objective of obtaining equitable and comprehensive data. To design the final survey, we administered a preliminary sample survey to 20 randomly selected students. This initial exercise revealed certain challenges that, if unaddressed, could potentially impact the precision of the final results.

Some issues we faced in collecting data for the pilot experiment included inconsistencies in responses and the dispersion of specific questions. Regarding the language people speak, although we received similar responses, variations in order and usage (Chinese/Mandarin) were present. To address this, we are considering implementing multiple-choice questions to enhance

the precision of the dataset. Additionally, when asking our respondents which BU college they are in, we faced challenges due to uncategorized responses. To solve this, we have simplified the response categories to only Stem/Humanity/Others, aiming to establish a clearer and more unified dataset for further analysis.

Our survey omitted potential extrinsic factors that might introduce ambiguity into the results. Regarding student's willingness to take additional language courses, we excluded participants compelled to undertake such courses due to hub requirements. This exclusion is particularly relevant for students within the College of Arts and Sciences (CAS), where a prerequisite of proficiency at the fourth-semester level or beyond in a language other than English is mandated.

In conclusion, the survey with less bias and confusion to our respondents can lead to a solid foundation for further analysis and interpretation.

Sample population and sample size

The study draws its sample from the entirety of Boston University's student body, encompassing respondents across diverse grade levels, including Freshman, Sophomore, Junior, Senior, and Graduate students. Alumni and graduate students were excluded from the study, resulting in a sample reflective of the undergraduate student demographic. The composition of the sample reveals a distribution across grade levels, with approximately 39.8% of respondents identified as Sophomore students, 27% as Freshmen, 18% as Juniors, and 13% as Seniors. The primary location for soliciting responses is at Marciano Dining Hall which contains more lower-class students residing on-campus and individuals sharing academic classes with our team. Consequently, the respondent pool is notably enhanced with Sophomores students, aligning with the dining hall's demographic composition and our team member's grades.

As mentioned before, the demographic that was interviewed was primarily sophomores, constituting the demographic makeup of our sample. The sample size, represented by the number of sophomores included, inherently carries certain characteristics that might not be reflective of the broader population. Moreover, the survey adds an additional layer of consideration. Conducting the survey in a busy environment, such as the dining hall, could have impacted the respondents. Many respondents were occupied with their time when asked to fill out the survey and might have caused them to be rushed and not complete the form with complete honesty. This introduces a source of bias that, while not altering the sample size itself, can affect the quality of the data collected.

Data visualization and descriptive analyses (details of the survey)

- 1) Relationship between Number of Languages Spoken and the respondent's categorized major

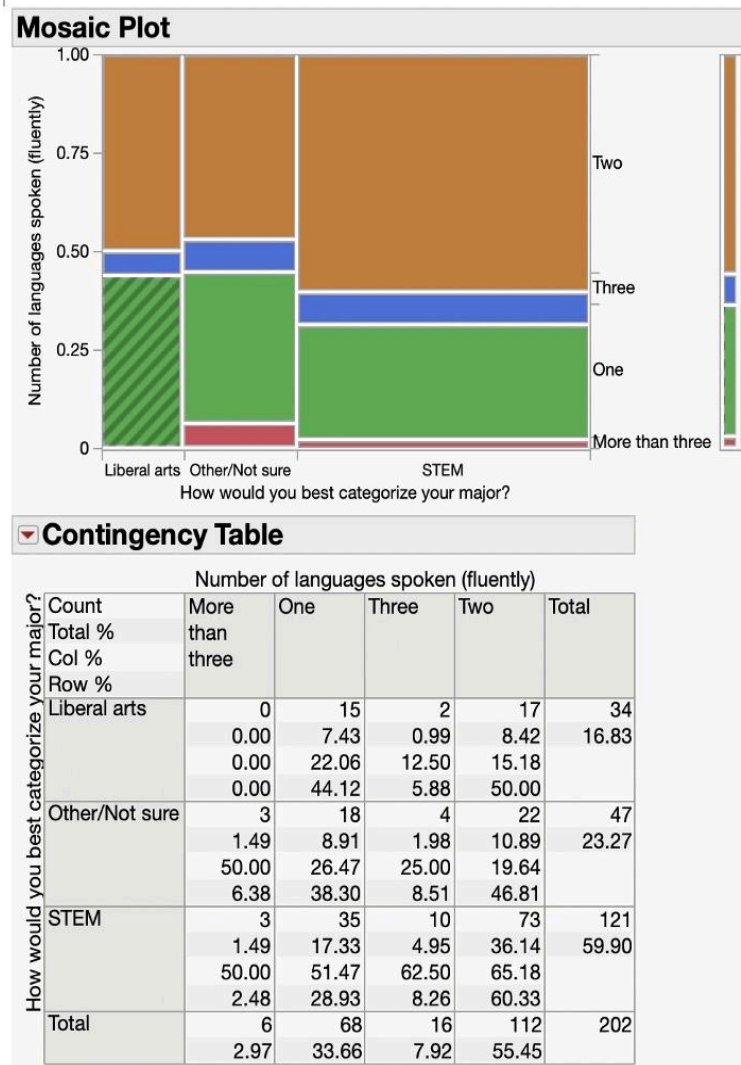


Figure 3: Distribution of relationship between number of languages

people speak and their categorized major

Within our dataset, 57.8% of students are pursuing STEM majors, 18.9% are enrolled in Liberal Arts programs, and 23.3% fall into the category of either "other" or are unsure of their major. Given the higher prevalence of STEM majors, it follows that the reported numbers of languages spoken may also exhibit a greater range than observed in the other two categories. Notably, the response indicating proficiency in two languages is consistently the most prevalent across various major categories. Across all major groups, individuals who speak two languages constitute the highest proportion, followed by those proficient in one language, three languages, and finally, individuals proficient in more than three languages. However, the specific proportions of these language proficiency categories vary within each major group. This observation suggests that the number of languages spoken may not significantly influence individuals' decision-making processes when selecting a college major.

Test statistic:

Fisher's Exact Test (two-sided)

$$Prob \leq P = 0.3791$$

Two-sided p-value=0.3791

In order to explore the null hypothesis that the number of languages spoken by a BU student bears no correlation (two variables are independent) with their choice of major, with the alternative hypothesis proposing either a positive or negative correlation (two variables are not independent) between the number of languages spoken and the selected major, our approach involved employing a statistical test to ascertain the p-value at the confidence level of 95%. The computed p-value is 0.3791, which is larger than $\alpha=0.05$. Consequently, we fail to reject our null hypothesis, indicating there is no correlation between these two variables, meaning the two variables are independent. This outcome aligns with our anticipated result, as there is no significant relationship between the number of languages spoken by students and their choice of majors.

2) Relationship between Number of Languages Spoken and Willingness to take an additional language course

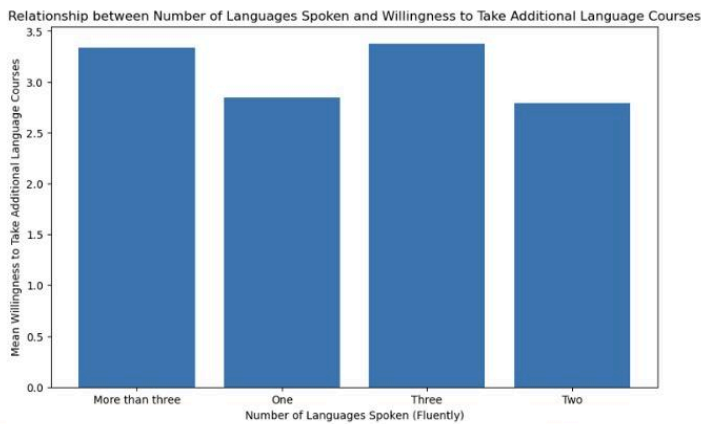


Figure 4: Relationship between Number of Languages Spoken and Willingness to take additional language courses

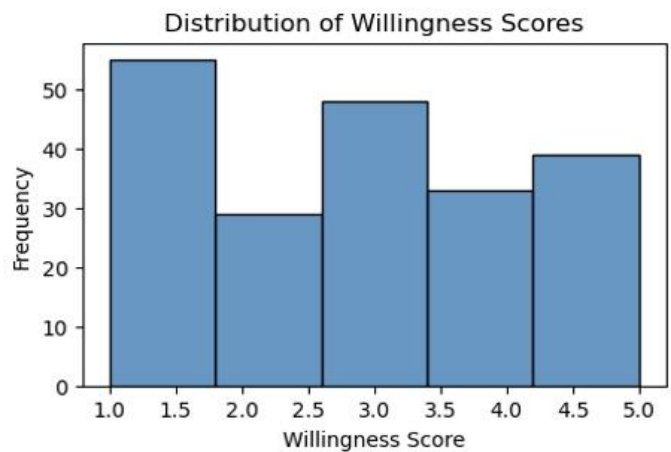


Figure 5: Distribution of Willingness Scores

There is not much difference between the number of languages spoken and the willingness for BU students to take another language. This could suggest that there is no correlation between the two. It also suggests that most people are equally likely to be willing to take another language regardless of the number of languages they previously know.

The willingness data was examined from 203 survey responses. The descriptive statistics show a mean willingness score of 2.86 with a standard deviation of 1.46. The scores range from 1 to 5, with quartile values of 1, 3, and 4. This shows a diverse distribution of responses.

An Ordinary Least Squares (OLS) regression was also performed to explore the relationship between the willingness data and the number of languages spoken. The overall model's R-squared value was 0.014. This indicates a limited ability to explain variability in willingness scores. The coefficients for different language proficiency categories, including one,

two, and three languages spoken, did not show significant associations with willingness scores ($p > 0.05$).

- Ordinary Least Squares (OLS) test was performed.
- R-squared value: 0.014
- coefficients for different language proficiency categories, including one, two, and three languages spoken, did not show significant associations with willingness scores ($p > 0.05$)

While the willingness data presents a diverse range of responses, further investigation is needed to address the issues identified in the t-test and explore potential factors influencing willingness scores.

3) Compare the data of the willingness of students to take additional language classes against their language learning background

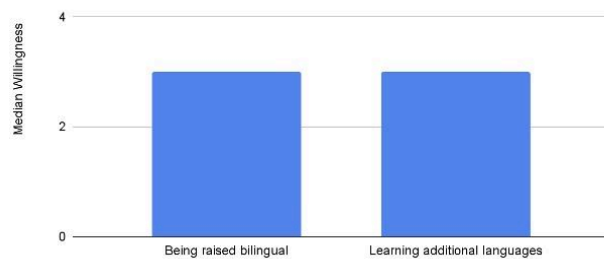


Figure 6: Relationship between the median willingness of taking additional language courses and students' learning background

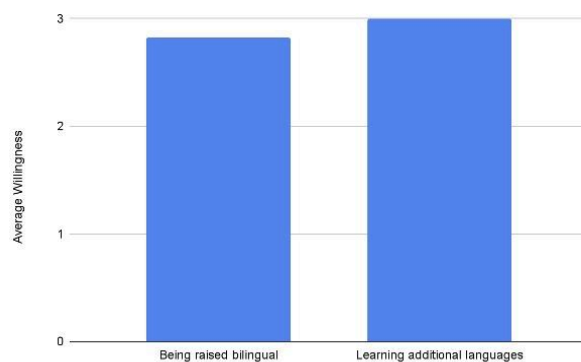


Figure 7: Relationship between the average willingness of taking additional language courses and students' learning background

These two graphs aggregate and compare the data of the willingness of students to take additional language classes against their language learning background. On average, students who were raised bilingual had a 2.8 (on a subjective scale of 1-5) willingness to take additional

language classes, while those who learned their second language outside the home had an average willingness of 3.

As a null hypothesis for these data sets we can say that $\mu(\text{raised bilingual}) = 2.5$ and $\mu(\text{learning additional}) = 2.5$

And as alternative hypothesis we have $\mu(\text{raised bilingual}) > 2.5$ and $\mu(\text{learning additional}) > 2.5$

For two one-tailed tests.

We want 95% confidence and so we will take .05 as alpha and thus our rejection region for both tests will be $z > 1.645$ as the sample is large

For the test statistics we have $z(\text{raised bilingual}) = (\bar{x} - 2.5) / (s/\sqrt{n}) = 0.3/.137 = 2.18$

And $z(\text{learning additional}) = (\bar{x} - 2.5) / (s/\sqrt{n}) = 0.5/.17 = 2.94$

For p values of $P(z > 2.18) = .02146$ and $P(z > 2.94) = 0.00164$. As both p values are less than 0.05 we reject the null in favor of the hypothesis, and say that the mean willingness of both populations is above the value of 2.5 on the subjective scale.

However, the median response for both populations of students was 3. Indicating that typically the entire student population has the same willingness to take additional language classes. We can say, however, that active study of additional languages outside the home predisposes students at least slightly towards being willing to take additional language classes at the university. To see whether this holds up to provisional scrutiny, we can compare the fluency in their second language against the willingness to take additional languages:

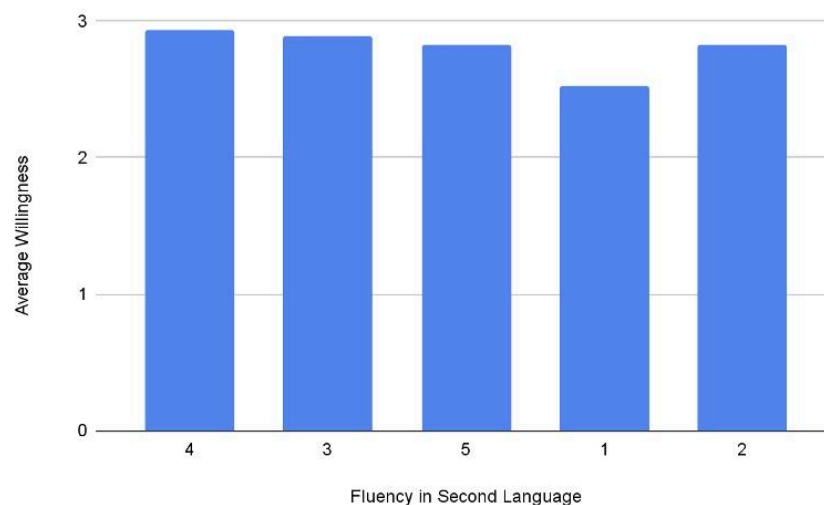


Figure 8: Relationship between the fluency in second language and student' average willingness of taking additional language courses

4) Relationship between the fluency in second language and students' college in BU

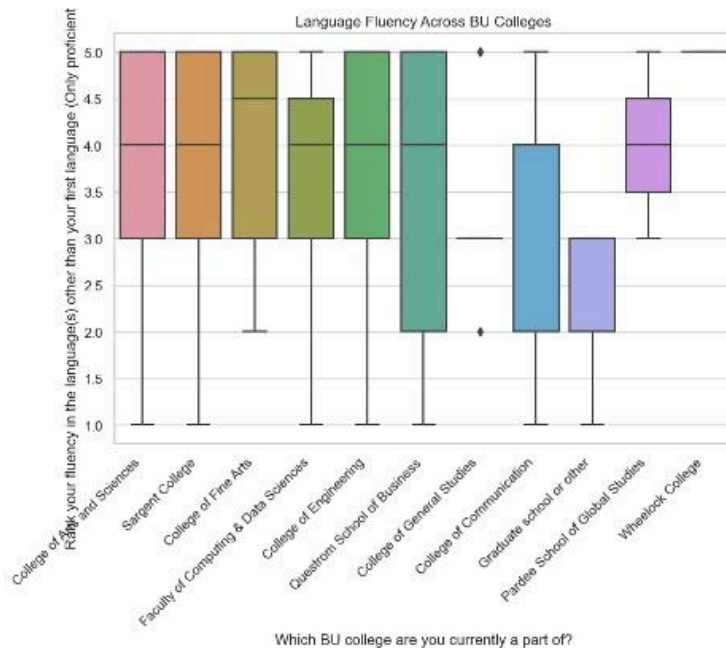


Figure 9: The relationship between the fluency in second language and students' college in BU

We can see that the average willingness to take additional language courses is lowest when fluency is 1, followed by 2 and 5, and highest at 3 and 4, indicating that past learning experiences may inform the willingness of students.

The box plot illustrates the distribution of language fluency levels across various BU colleges. The median fluency level varies across colleges, ranging from 3.23 in the College of Communication to 4.0 in the College of Fine Arts and the Pardee School of Global Studies. scores within each college.

The College of Arts and Sciences exhibits a median fluency of 4.0, with an IQR spanning from 3.0 to 5.0. Meanwhile, the College of Communication has a median fluency of 3.23, and its IQR extends from 2.0 to 4.0. The boxplot also highlights outliers in the data.

The data shows that there is not a strong relationship between the school a student at Boston University chooses to attend and their language fluency as most of the box plots are similar. This further supports the hypothesis that language fluency does not affect their major choice.

5) Normality distribution

In Figure 8, the data can be considered non-normally distributed as a substantial portion of data points deviate from the middle red line on the plot. Specifically, the quantitative variable measuring the willingness to acquire additional languages displays a pronounced left skew, notably diverging from the typical shape of a normal curve. The presence of this deviation suggests that the distribution of the data does not align closely with a normal distribution as the QQ plot indicates. In conclusion, we cannot say the data is normal. Instead, it may exhibit characteristics of another type of distribution or may be influenced by specific factors that result in non-normality.

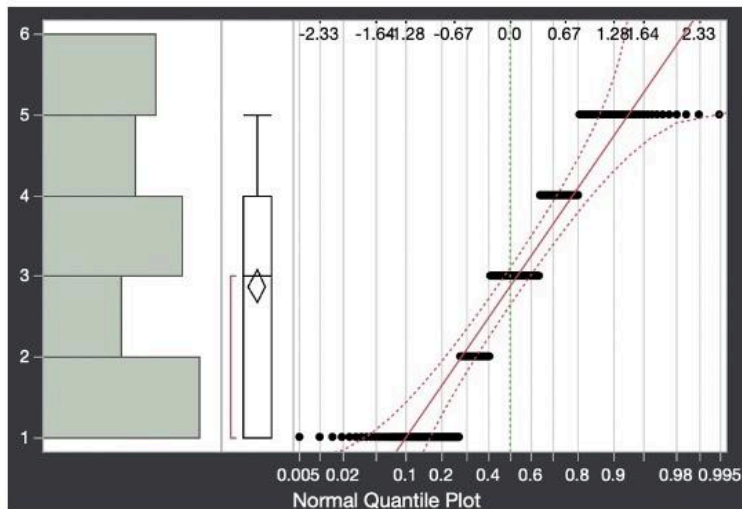


Figure 10: The normality distribution of the Willingness of taking additional courses

Discussion and Conclusion

Our study suggested that language proficiency doesn't significantly impact students' choice of major at Boston University. Willingness to take additional language courses showed no specific patterns based on language fluency. However, the variations in language proficiency observed across different colleges, suggested potential college-specific influences.

Hypothesis	Findings
<p><u>Null hypothesis:</u> The number of languages a BU student speaks does not impact their decision of choosing a major</p> <p><u>Alternative hypothesis:</u> The number of languages a BU student speaks impacts their decision of choosing a major.</p>	<p>$P > \alpha$</p> <p><u>Fail to reject the null hypothesis</u></p> <p>With a 95% certainty the mean willingness to take additional language classes of both students raised bilingually and those who learned additional languages in school are above 2.5 on a 1-5 subjective scale.</p>

Contributions

Stella Ye

Deliverable 1: majority writing and adding questions

Deliverable 2: brainstorming questions, survey creator, hypotheses

Deliverable 3: wrote parts of each section of analysis, adjusted survey

Deliverable 4: collected data points

Deliverable 5: abstract, introduction

Sohail Mohammed

Deliverable 1: organized the meeting and adding questions

Deliverable 2: Gave feedback and possible new questions to ask

Deliverable 4: Create graph on willingness to take another language and number of languages spoken, and analyzed it

Deliverable 5: preformed analysis and created a graph for the data

Adrianne Tsai

Deliverable 1: adding questions

Deliverable 2: brainstorming questions, survey creator, adding feedback

Deliverable 3: wrote parts of Experimental methods section and first graph analysis

Deliverable 4: added first two analysis, all Experimental method part

Deliverable 5: finish 1) , 2) and 6)

Egor Vazgryn

Deliverable 1: add possible additional questions

Deliverable 2: brainstorming questions, Analysis writeup

Deliverable 3: Added second graph and analysis

Deliverable 4: Added charts and analysis for willingness.

Deliverable 5: Hypothesis and test for mean willingness to take additional classes of both student populations

Appendix 1

Willingness to take additional language courses in college (not for HUB requirements)

Summary Statistics	
Mean	2.875
Std Dev	1.4663307
Std Err Mean	0.1036852
Upper 95% Mean	3.0794628
Lower 95% Mean	2.6705372
N	200
N Missing	4

Figure 1 Summary statistic of Willingness of taking additional languages

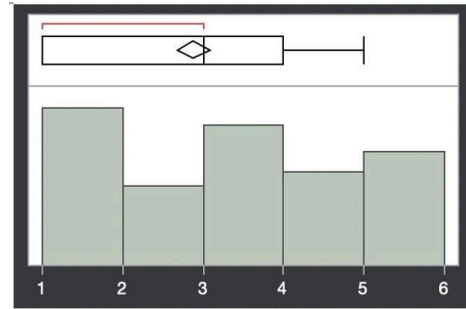


Figure 2: distrubution of the Willingness to take additional languages

This graph indicates there is no obvious preference for the sequence of responses, manifesting neither an ascending or descending trend across the 1-5 scale. Furthermore, it is evident that the response corresponding to a willingness score of 1 elicits the highest number of responses. Subsequently, willingness levels 3 and 5 also garner a noticeably higher number of responses compared to the remaining two levels of willingness. The overall distribution of the graph is skewed to the left, denoting that the median, denoted by a score of 3, is in closer proximity to the third quartile (Q3) at a score of 4 than it is to the first quartile (Q1) at a score of 1. Additionally, the median surpasses the mean, which stands at 2.875.

For a 95% Confidence Interval given the large sample (207 responses), we can say that with 95% confidence the true population average willingness to take additional language classes is given by

$$\bar{x} \pm z_{\alpha/2} (s/\sqrt{n}) = 2.87 \pm 1.96(1.466/14.45) = \text{between } 2.671 \text{ and } 3.068$$