ALC Marketing & Program Evaluation Survey Analysis

Elma Garcia, Hanna Grossman, Citlally Reynoso, Faith Tsang

3/8/2020

Abstract

With a decrease in enrollment in similar programs, the UCLA Extension American Language Center (UNEX-ALC) wants to identify ways to see where they stand. Additionally, UCLA Extension ALC wants to evaluate their program as a whole, which involves making surveys that provide useful and easily interpretable information for future decisions. To meet these two goals, a survey has been given to some students at UCLA Extension ALC that will provide more information for both of these aspects. The survey consists of seven open-ended questions and one ordinal question, as a previous project by STATS 141SL students determined that open-ended questions would be a helpful change that would provide more information as compared to the old survey, which consisted of many close-ended questions. The focus of the survey is to provide information regarding demographics, why students chose to attend this specific program, and what they plan to do after the program is completed. The program evaluation survey targets pre-university students, mid-career professionals, and certificate program students who have at least an intermediate level in English proficiency. We first started with data cleaning, then did an exploratory data analysis which involved looking at word frequencies and word clouds for each question. We used semantic networks and sentiment analysis to determine marketing techniques in boosting admissions and strategies to improve the quality of the program. We looked at which responses were often repeated to determine where modifications are needed. As a result, we concluded that the survey as a whole needs revision to include a multitude of different types of questions in order to increase survey response turnout, and that the program should prioritize its most popular attractions such as the prestigious reputation of the school. We also found that current students enjoy the environment and curriculum that UCLA Extension ALC offers. Challenges in our project included not having enough data and the data coming from multiple surveys. We recommend giving out an updated survey implementing our findings to gain further insight regarding marketing.

Clear statement of the problems to be solved

There are two main problems that we are looking to solve through this survey:

1. We want to use our findings to create the best possible survey for future years. We hope to create a survey for next year that garners the greatest response rates, while also providing the best quality of responses.

2. We will use our results to suggest marketing techniques that will allow UCLA Extension to improve the program and attract more students. Programs such as the UCLA Extension ALC that teach English have had lower enrollment due to a variety of factors in recent years, so UCLA hopes to understand what it is about the UCLA extension ALC that causes students to choose it as the place to study, as well as where students are coming from, their goals after studying at ALC, and whether students feel that the program is meeting their needs.

Variables of the study and how they were measured

We had eight columns of answers from the survey:

- Q1: How did you learn about the UCLA Extension American Language Center program?
- Q2: Why did you choose to study at the UCLA Extension American Language Center?
- Q3: What do you like best about your American Language Center program?*
- Q4: What would improve your experience at ALC?*
- Q5: What was your experience learning or studying foreign languages before you came to the American Language Center?
- Q6: Please describe your work experience before you came to the American Language Center.
- Q7: What are your most important goals after you complete your studies at the American Language Center?
- Q8: Would you recommend the ALC program to a friend? (numeric)**
- Q8_text: Would you recommend the ALC program to a friend? (text)**

*It is important to note that this data was collected from multiple surveys. This mainly affected questions 3 and 4. Various surveys asked these questions in different ways. For example, for question 3, some students were asked "What did you like best about your professor" or "What did you like best about this class" rather than the overarching question "What do you like best about your American Language Center Program." This therefore yielded slightly different responses depending on the version of the survey a given student received. All of these similar versions of questions 3 and 4 were merged together and analyzed as if they were from the same question.

**It is also important to note that although questions 1-7 were open ended, question 8 was ordinal, with options 1-5:
- 1. Definitely would not recommend
- 2. Probably would not recommend
- 3. Not sure
- 4. Probably recommend
- 5. Strongly recommend

Data set:

Our data set had 175 rows and 9 columns. However, about ninety rows of data only contained answers for Question 3, Question 4, and Question 8 as they came from a separate survey that did not ask the other questions, causing the other columns had NA values. Additionally, some of our observations had useless responses, some of which we identified by eye and deleted. Some of the responses from the survey that only had students answer Question 3, 4, and 8 had "NULL" as the responses, so we cleaned out these responses through adding words such as "null" and "nothing" to our stop words.

Exploratory data analysis

Data cleaning

First, we read in the data using the package readxl. We then removed unusable rows that had useless responses, which we identified by eye and deleted. We renamed the variables, which originally had the whole question written out, and labeled them from Q1 to Q8. We also had a variable we called Q8_text, which was text that corresponded to number levels that were the response for Q8. We also turned Q8 into a factor.

Next, we replaced all NA values with empty strings. We combined all question answers for each open-ended question into one string, resulting in eight strings total. If we had not replaced all NA values with empty strings, we could not concatenate strings with NAs without errors occuring. We then made sure to remove all special characters, and then turned each of our eight strings into corpus format. Once we did this, we removed punctuation, turned all letters to lowercase, removed numbers, and removed all basic stopwords in the English language. We created extra stop words after looking at the resulting word frequencies, and we established a set of stop words for each question. Extra stop words included words such as "null", "can", "also", and "just." Finally, we removed white spaces.

When doing data cleaning for sentiment analysis and semantic networks, we did not combine all question answers into one string, but instead kept each row separate to maintain the context for such analyses.

Word frequencies

We created word frequency bar plots and word clouds to get a better feel for which words occurred most frequently for each question. We first transformed each question's data in corpus form into a document term matrix. We then turned this data into a matrix and then took the row sums to get word frequencies. Finally, we created a data frame that contained the unique words in the left column and the frequencies of each word in the right column, sorted by descending frequency.

For each word frequency bar plot, we showed the top ten words that appeared most frequently for each question. In addition to bar plots, we produced word clouds to get an interesting visualization of how often words appeared. The word clouds can contain up to one hundred words.

By this exploratory analysis, we found that we needed context. After all, we need to know how words were connected to better understand what students are more commonly answering in response to each question. For example, from the word frequencies for Question 4, we can see that common words include such as "like", "good", "always", and "clearly," but we cannot understand how these words are connected even though they seem to have a positive connotation at a glance. Hence, we chose to use two types of statistical analyses that can be used for text mining: sentiment analysis and semantic networks. We chose to use sentiment analysis on Question 3 and Question 4, which had the highest word frequencies, in order to expand our understanding about the attitude students had toward UCLA Extension ALC. Semantic networks can allow us to better understand the context that words come in by showing words that have semantic connections.

Statistical analysis used to answer the research question

Sentiment analysis

Sentiment analysis measures the tendency of people's opinion and determines whether their attitude towards the product in question is positive or negative. There is sentiment detection where subjective expressions such as opinions, beliefs, and views are kept and objective information is rejected. After, the subjective expressions must be classified. It helps us optimize marketing strategies by understanding what the customers want to see in the product. Therefore, we are able to adjust any changes we see fit in order to have a more successful product.

Semantic networks

One type of analysis we used was semantic networks. Such networks resemble neural networks and consist of nodes and edges. Each node represents a word, and each edge has a weight that shows how often the two connected words appear together. Semantic similarity is related to spatial proximity in the semantic networks. Hence, unlike looking at word frequencies, which looks at each word separately and loses context completely, using semantic networks will enable us to understand the context that words appear in and better understand the responses from the survey.

Results and recommendations

Question 1: How did you learn about the UCLA Extension American Language Center program?

We recommend making this question multiple choice, as it had few answers repeated many times amongst students. We believe this will minimize response fatigue, thus increasing response rate on the survey as a whole. In addition, this data type will allow for the best analysis of this response.

We see below the exploratory data analysis word cloud and the neural networks plots for this question.
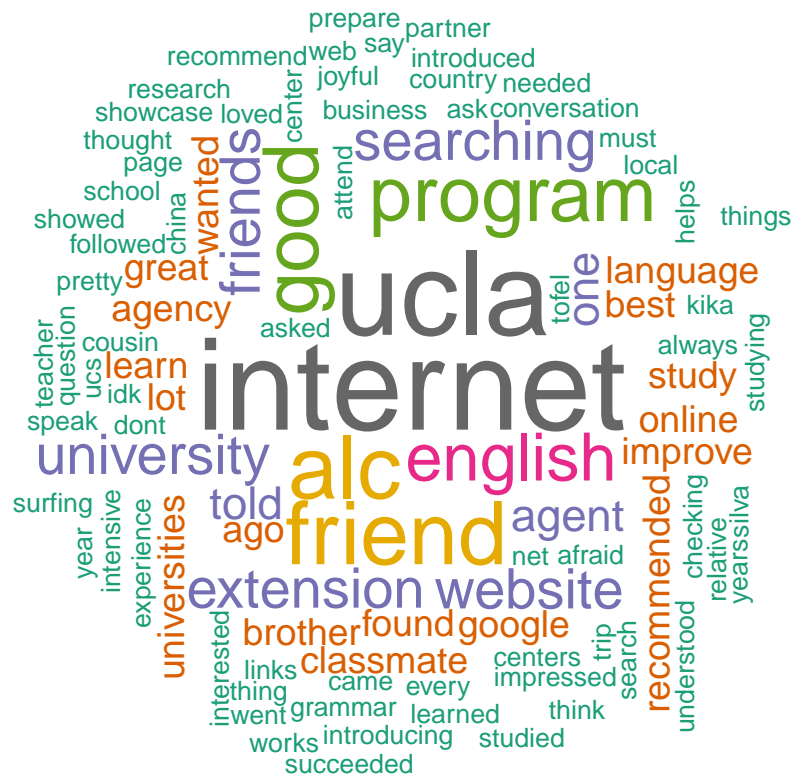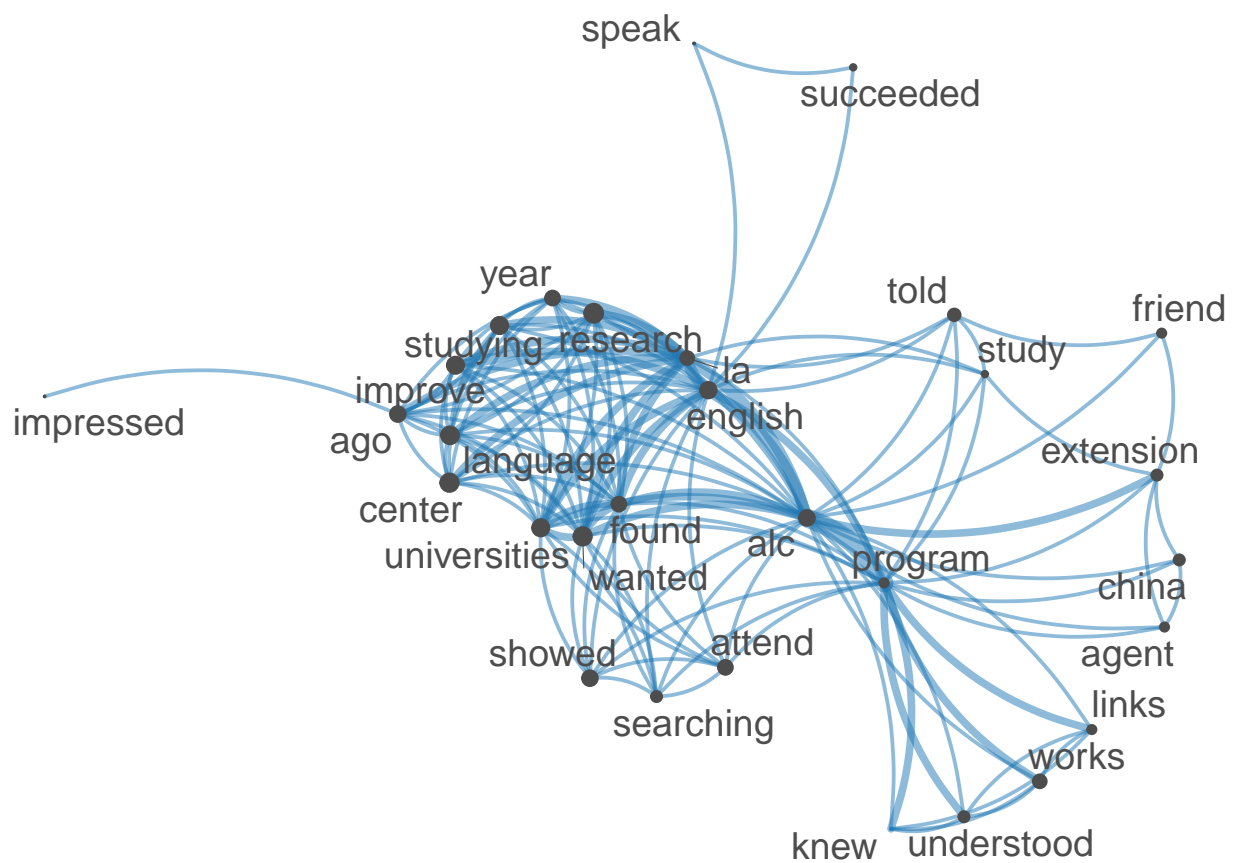

Figure 1: Word Cloud for Question 1

Figure 2: Semantic Network for Question 1

Through these visualizations, we were able to identify the top answer choices:

- Family member
- Friends
- Internet
- Agent
- Previously visited UCLA

We would also include an "other" option, and if a student chooses "other", they could then type in their response.

Question 2: Why did you choose to study at the UCLA Extension American Language Center?

We recommend either keeping this question open ended, or making it select all that apply. If the goal is to receive many varied responses, it is helpful to keep this question open ended. This may be beneficial to marketing, as it can yield a variety of reasons students choose this program, allowing for marketing material to target these reasons. However, if the goal instead is to see more specifically which reasons are important to students, and to analyze this quantitatively, select all that apply may be more useful.

We did see in our analysis that there were a select top answer choices, seen below in the exploratory data analysis word frequencies and the neural networks plots for this question.
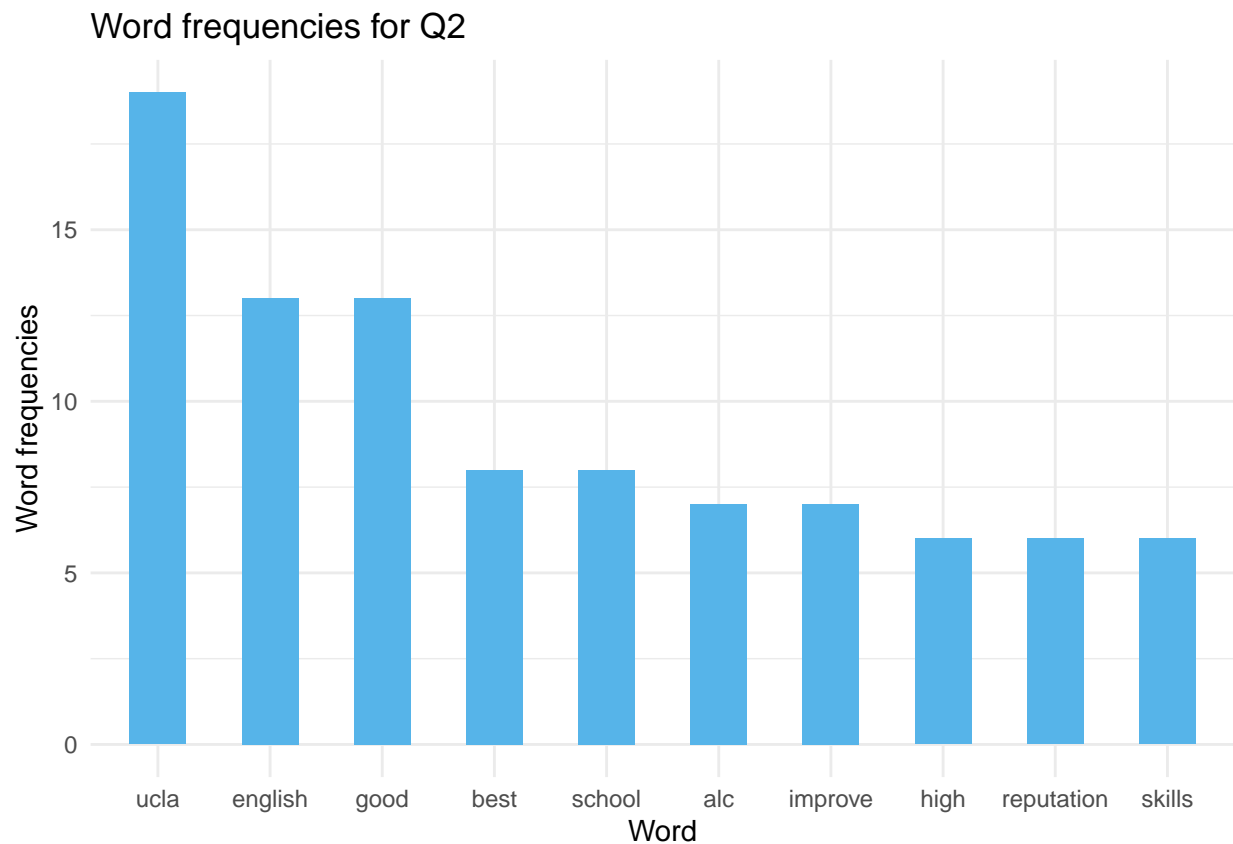

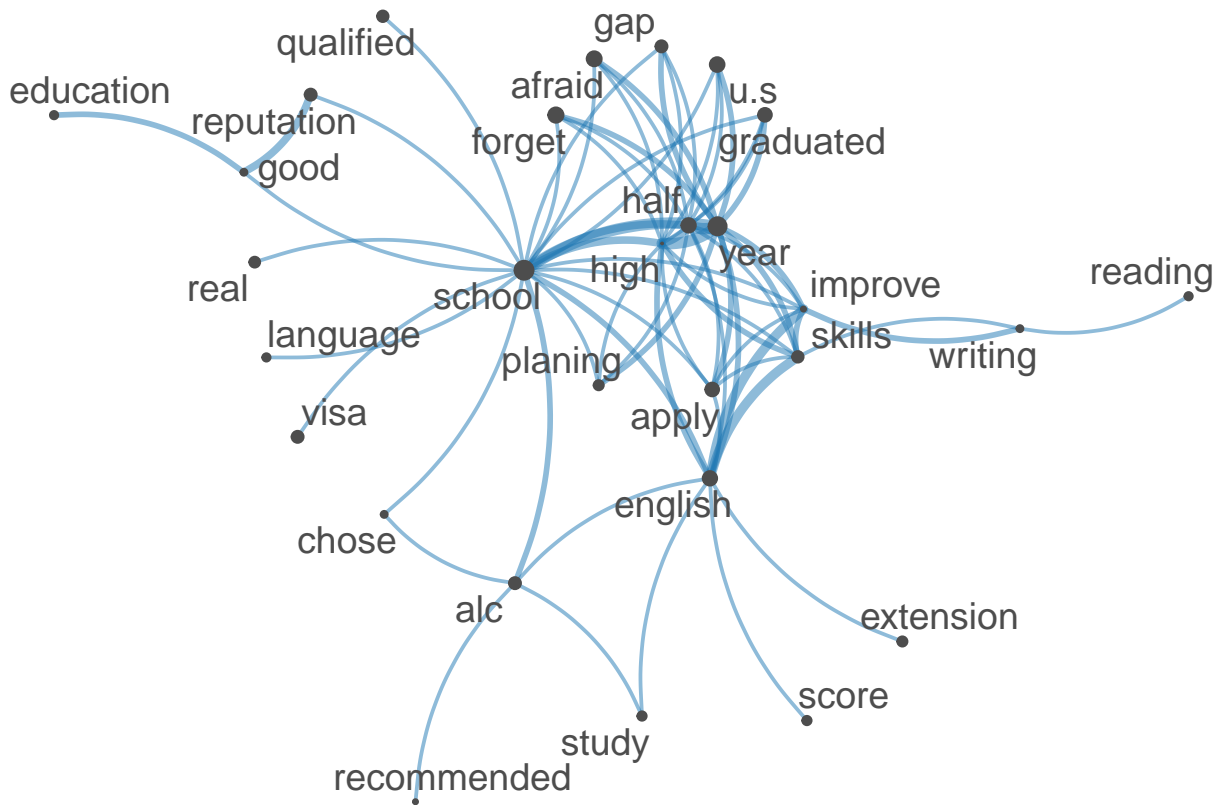Figure 3: Word Frequencies for Question 2

Figure 4: Semantic Network for Question 2

Through these visualizations, we see that the top answer choices are:

- To learn english (improve reading/writing skills)
- Resume/CV
- Desire to attend UCLA or another top university in the future
- Location (U.S. or Los Angeles)

Question 3: What do you like best about your American Language Center program?

We recommend keeping questions 3 open ended as it is now, as this question yielded a wide variety of responses, proving extremely helpful for marketing suggestions and improvements.

Word frequencies for Q3

Figure 5: Word Cloud with Positive Sentiment for Question 3

Question 3 asked students what they liked best about the ALC program so we performed a positive sentiment analysis for our marketing evaluation. The histogram above identifies the positive words used most frequently to describe the program.

Environment: The environment of the program and the way students feel when they are learning seems to play the biggest role when assessing the positive sentiments students associate with ALC. Students value teachers and the human qualities associated with them. Teachers who are nice, kind, and helpful nurture the kind of environment that students want to be a part of. This sort of environment is what garners the most positive reviews, and is most likely to lead to future recommendations.

Curriculum: The next factor which garners the most positive sentiments is the curriculum of the ALC program. Students appreciate the clarity of the information, what they learned, and the practice they received from the program. Students recognize and value the results that they see in their ability to communicate in English, and this translates into positive reviews towards the program.

Prestige: The last aspect of the program that received a lot of attention is the prestige that the ALC program has in connection to UCLA. Many students list this as a positive attribute because they know that listing this program in their CVs or resumes comes with a level of prestige that can positively impact their careers.

Question 4: What would improve your experience at ALC?

We recommend keeping questions 3 open ended as it is now, as this question yielded a wide variety of responses, proving extremely helpful for marketing suggestions and improvements.

Due to the nature of question 4, we thought it would be fit to do both a positive and a negative sentiment analysis.

The negative sentiment analysis tells us about the problems that students see with the already established qualities of the program. In other words, it tells us where students feel corrections should be made. The histogram and word cloud for the negative sentiment analysis shows that there are actually very little negative feelings expressed in response to this question. Students do not feel like there are currently any negative traits of the program that need to be improved, other than the fact that it is expensive.

The word cloud allows us to explore a larger amount of the negative words that came up in the survey, but most words appeared only once and seem to be more geared to understanding the personal experience of the individual student.

For marketing purposes, this means that the ALC program should focus its energy on improving the programs that are already established and that there are no concerns regarding correcting something about the program at this moment.
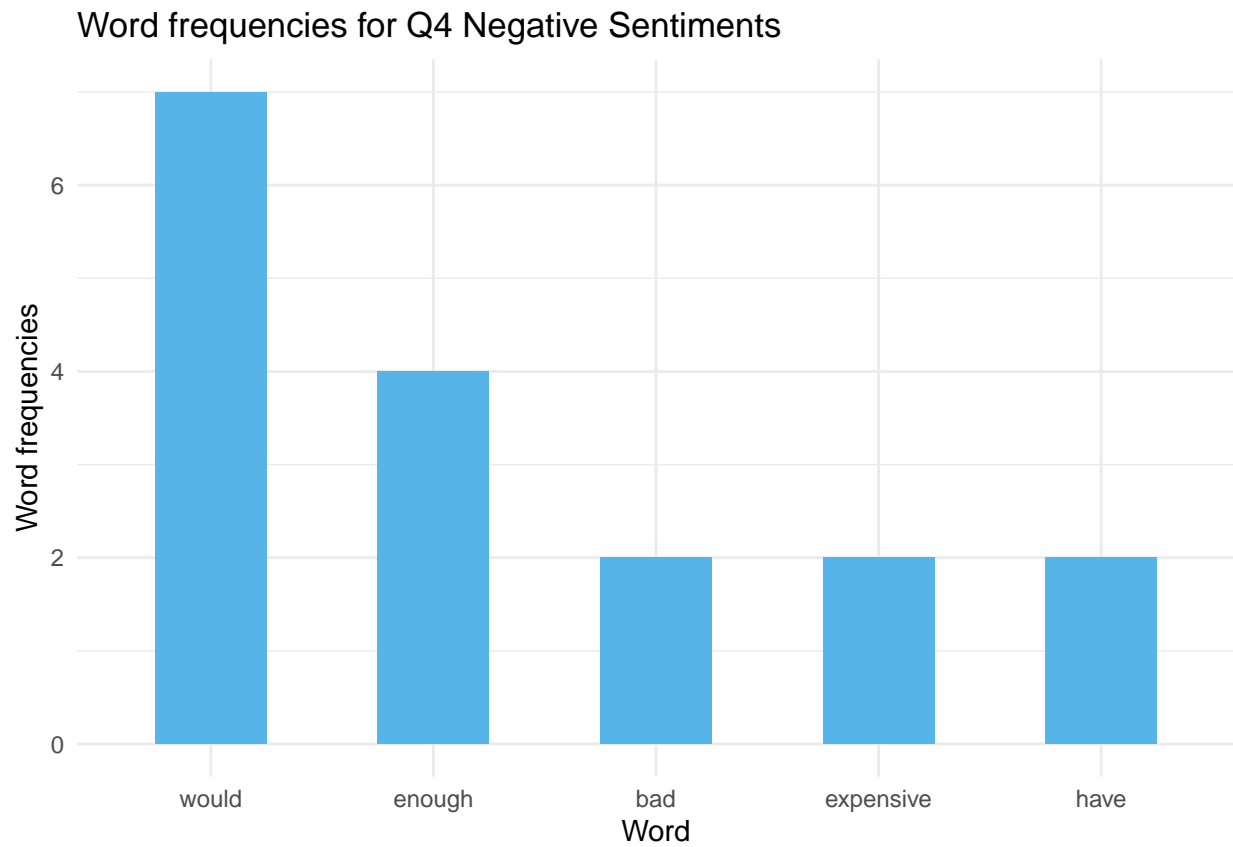
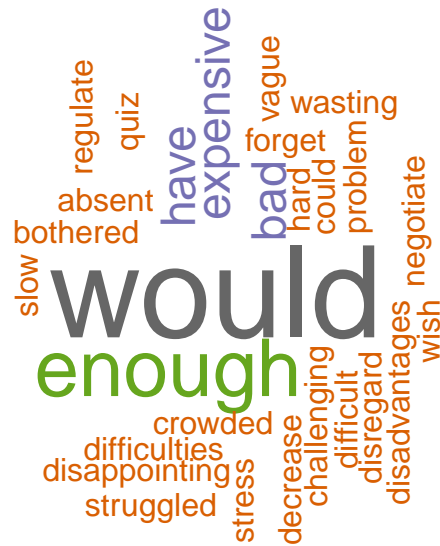Figure 6: Word Frequencies with Negative Sentiment for Question 4

Figure 7: Word Cloud with Negative Sentiment for Question 4

The positive sentiment analysis of this question tells us what positive changes or improvements students want to see incorporated into the program, and the students seem to have a general consensus.

If we take a look at the histogram of positive words used to answer this question, students are most interested in making ALC a more interactive program. They want to incorporate practice, volunteer opportunities, and more talking so as to make learning English a more immersive experience.

## Word frequencies for Q4 Positive Sentiments
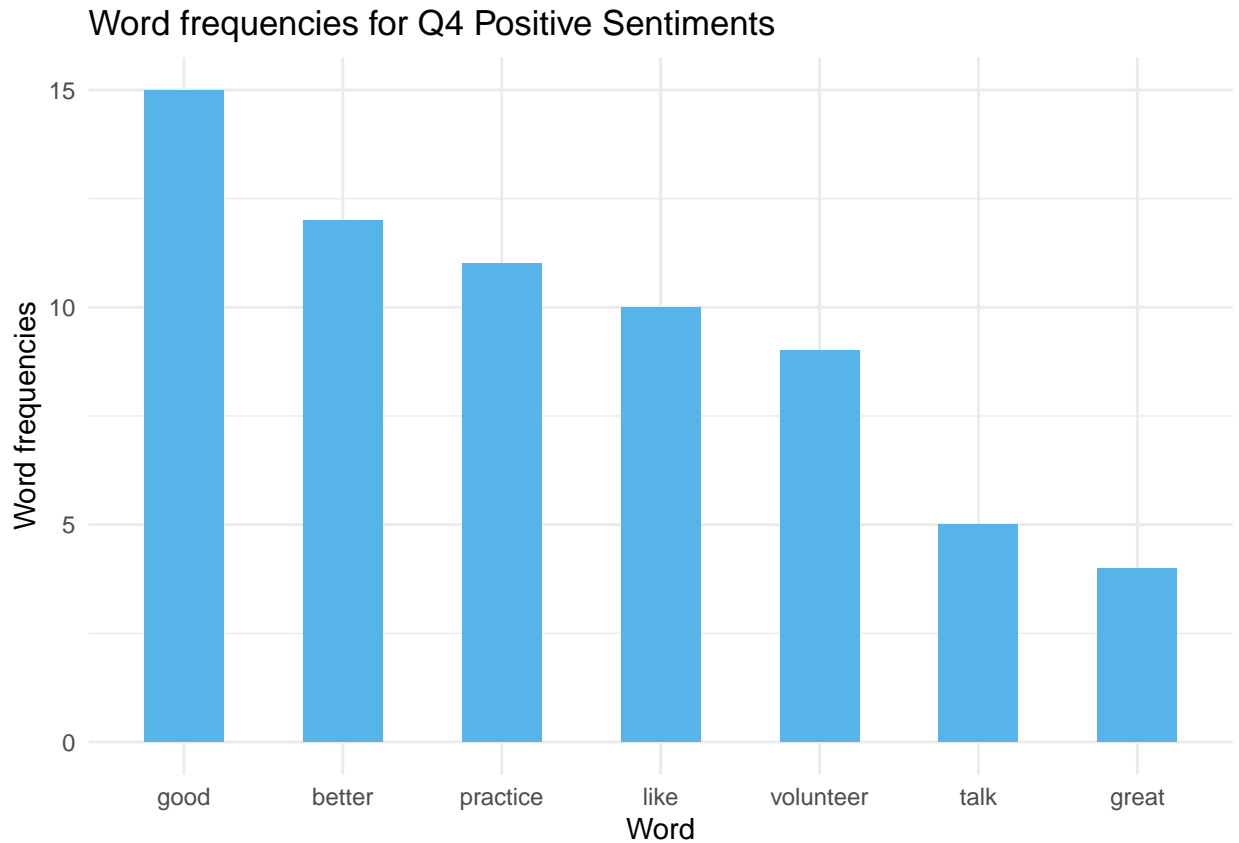
For more specific examples of the ways in which students wish to make ALC a more interactive program, we can take a look at the word cloud and the neural network connections that were formed between words.

From the word cloud we can gather some examples of activities that students think would make a positive addition to the program. Incorporating volunteering opportunities seems to be a very popular sentiment from the student as well as increasing the time used to practice verbal skills. If we take a closer look at the word cloud, it is also important to highlight that the words fun, humor, and comedy come up in the text analysis as separate ideas, but together make a powerful case for the incorporation of humor in the classroom.

Figure 8: Word Cloud with Positive Sentiment for Question 4

If we take a look at the neural network, this provides more of a map to the way in which words connect in the text. Something new that comes into the picture through the neural networks is social media. This was not apparent from the sentiment analysis because social media platforms do not have sentiment associations, but they are equally important to this question. We see that platforms such as Facebook, Instagram, and posting come up in close connection to the words website and links. It would be beneficial to incorporate links to platforms like facebook and instagram to the main ALC page because this will better communicate a sense of community to students. This idea is further supported by tying it back to the words gathered from the word cloud since social media can help convey the friendly nature of the campus which appears to be important to students.

The final pattern that we see reflected in both the negative sentiment analysis and the neural network, is a concern about the cost for the program. In the negative sentiment analysis we see that expensive is one of the most common words, in the positive analysis we see the word affordable, and in the neural network we see several links made to money. These words are concerning because they may differ students from enrolling in the ALC program, however if the website focuses on highlighting why the program is worth investing on, this might soothe the concerns of prospective students.
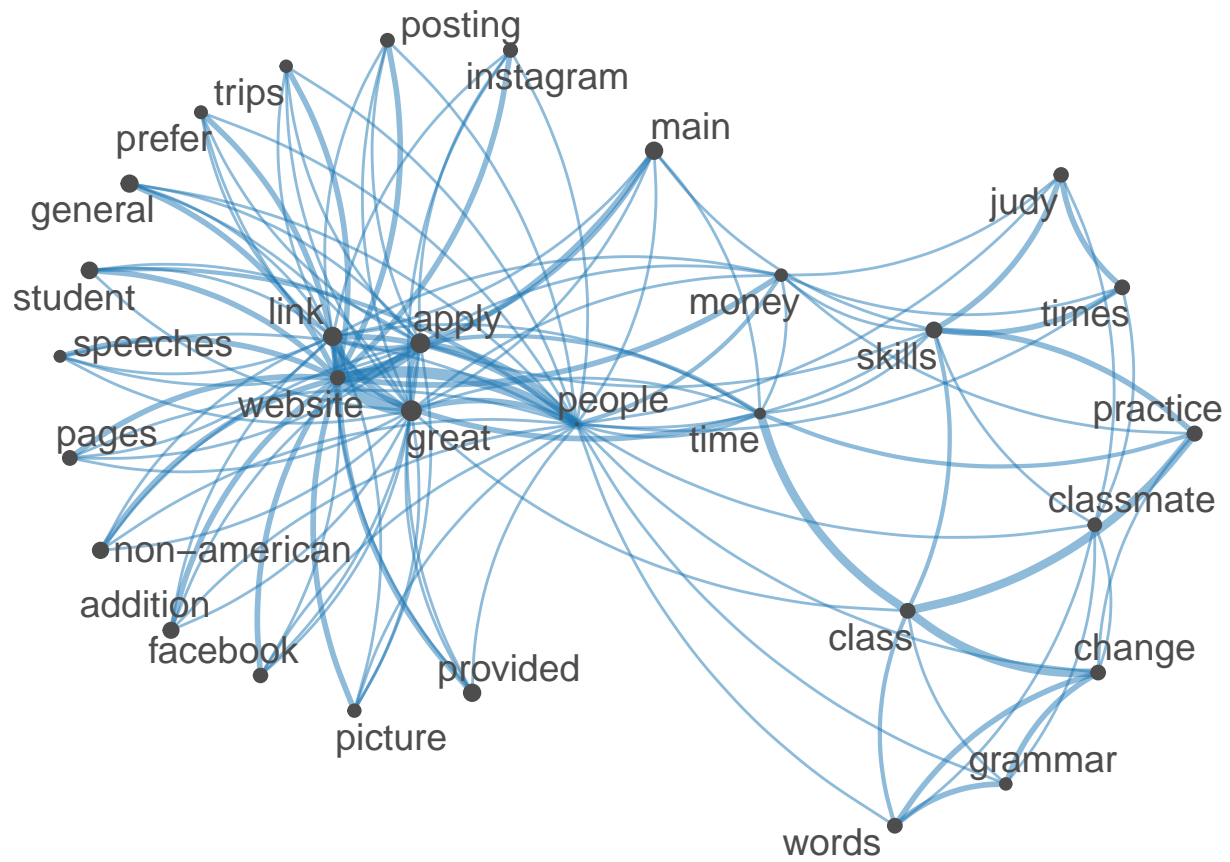
Question 5: What was your experience learning or studying foreign languages before you came to the American Language Center?

We recommend making this question select all that apply. We received many responses that did not answer the question, and therefore believe that students were confused about what the question was asking. By making this questions select all that apply, this question will become much more clear, allowing for a stronger analysis. We also recommend rewording this question to "What was your experience learning or studying english before you came to the American Language Center?" If further interested in students' exposure to other languages, we recommend making this a separate question.

We see the top responses below in the neural network plots for this question.
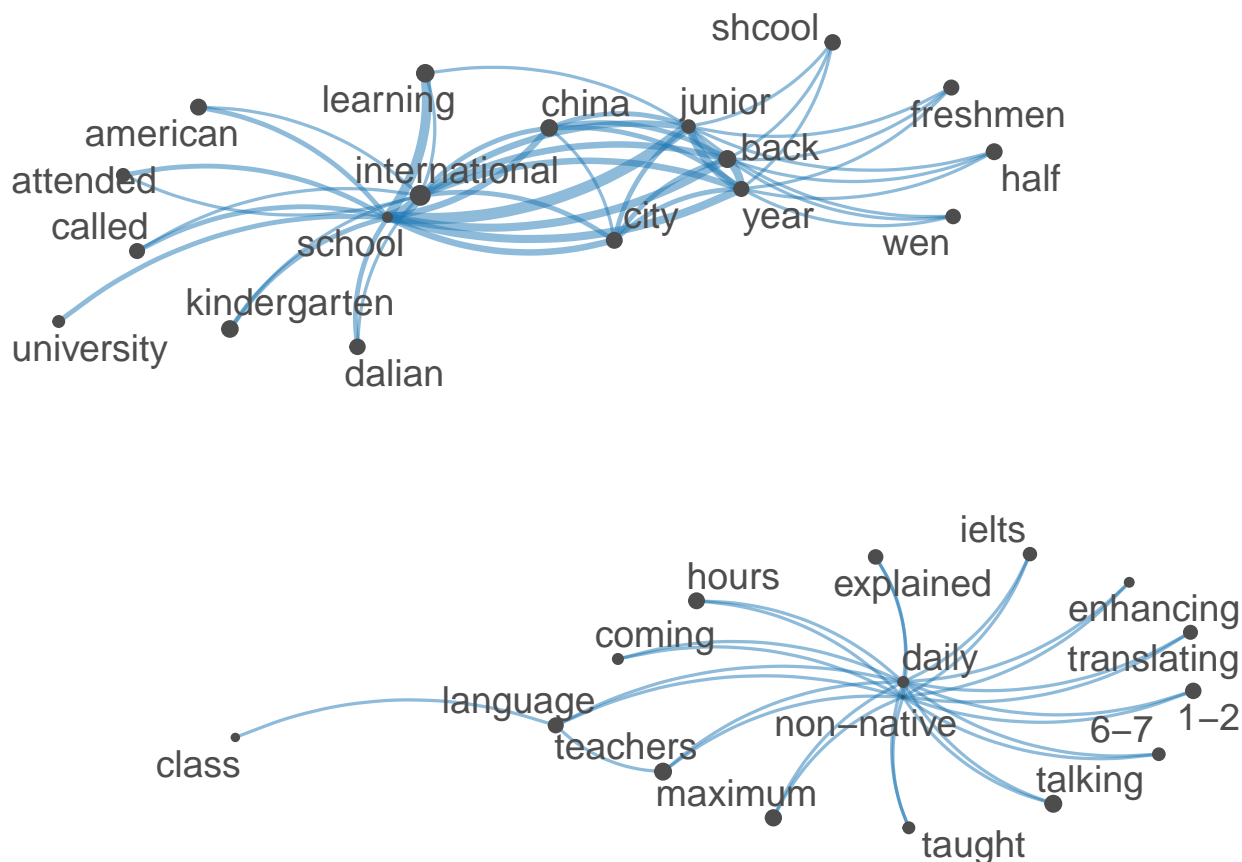

Figure 9: Semantic Network for Question 5

Therefore, we recommend making the answer choice options:

- I was exposed to english in school
- I was exposed to english in my work
- I was exposed to english through social media and/or television
- I have never been exposed to english before

We would also include an other option, and if a student chooses other, they could then type in their response.

Question 6: Please describe your work experience before you came to the American Language Center.

We recommend making this questions two parts: first asking if the student has worked before (yes/no). If yes, then what was their work experience? (open ended). We received many responses that did not answer the

question, showing that students may be confused by this question. We also received many responses saying the student had never worked before. We believe by making this two parts, it will clarify the question, and will therefore allow for a more successful analysis.
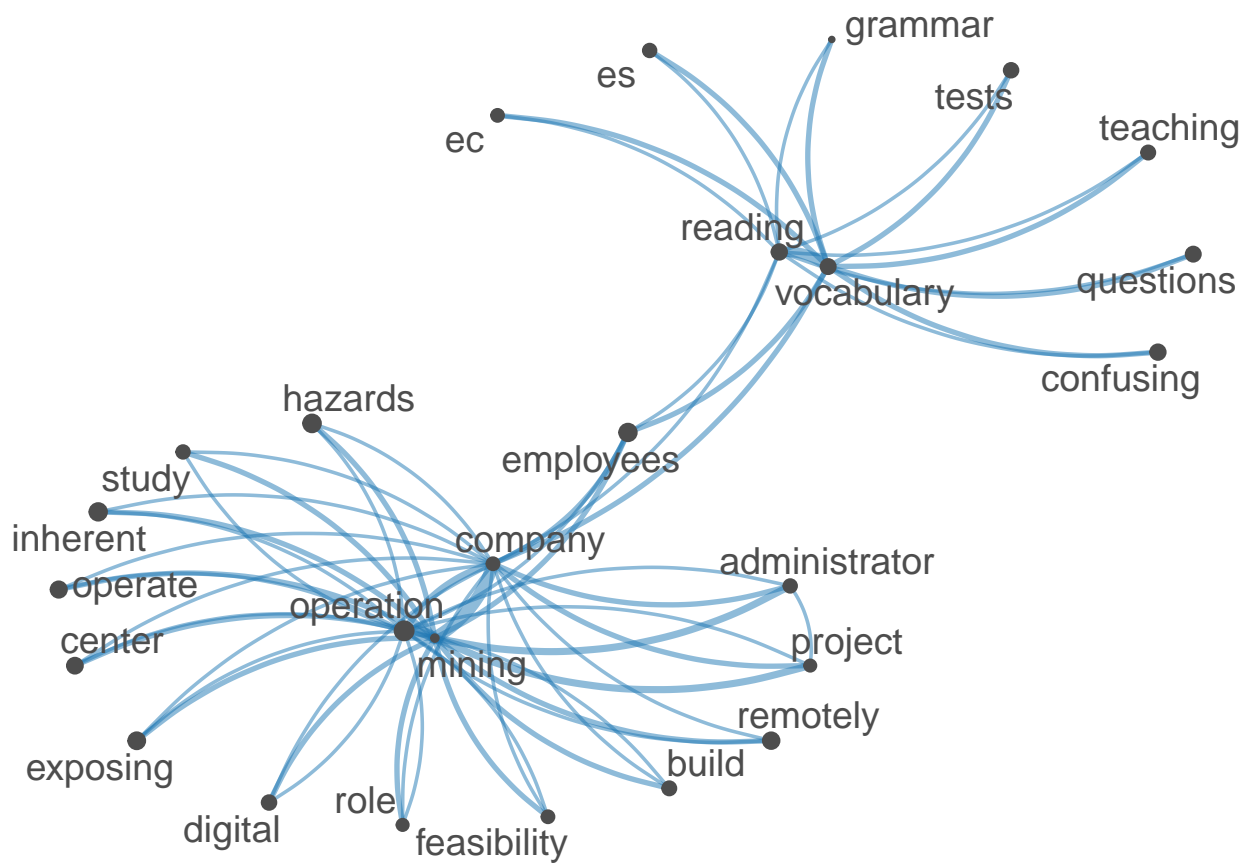
Figure 10: Word Cloud for Question 6

Figure 11: Semantic Network for Question 6

Question 7: What are your most important goals after you complete your studies at the American Language Center?

We recommend keeping question 7 open ended as it is now, as this questions yielded a wide variety of responses, many of which proved useful for the purpose of analysis.

We see the top responses below in the exploratory data analysis word frequencies and the neural networks plots for this question.

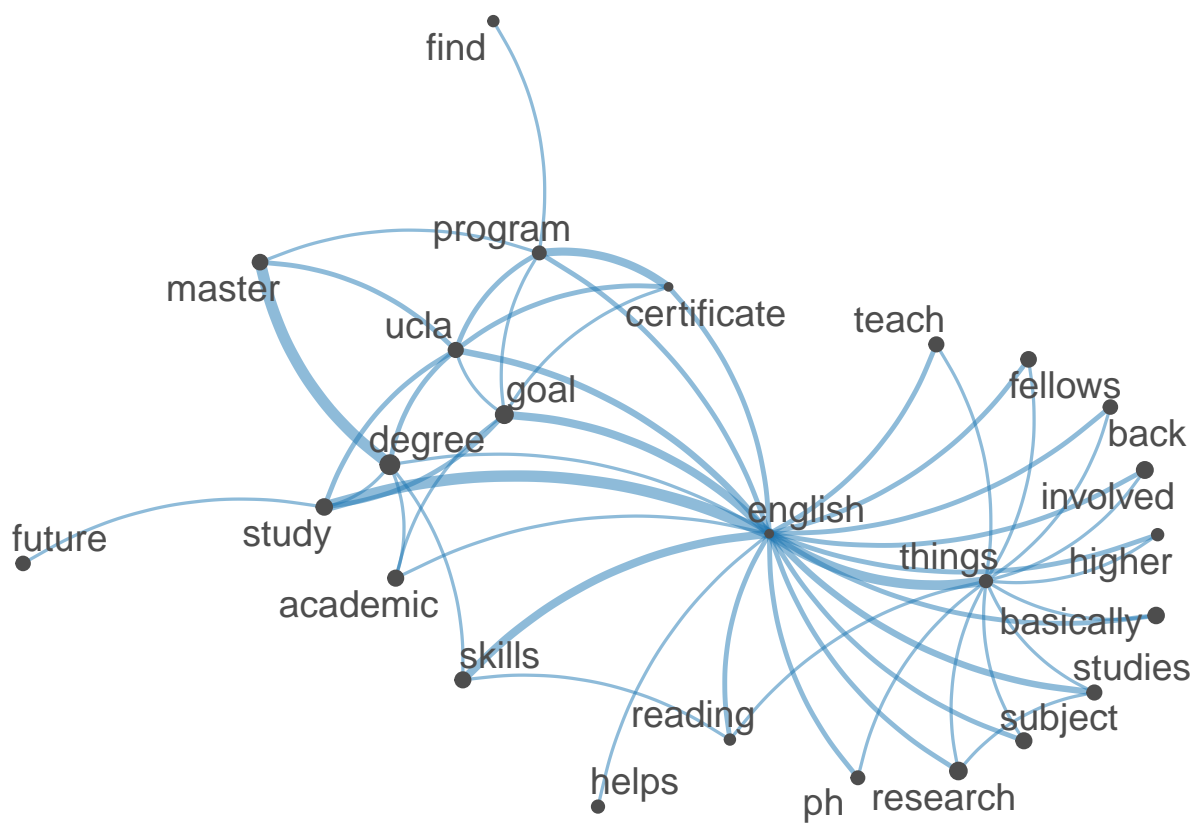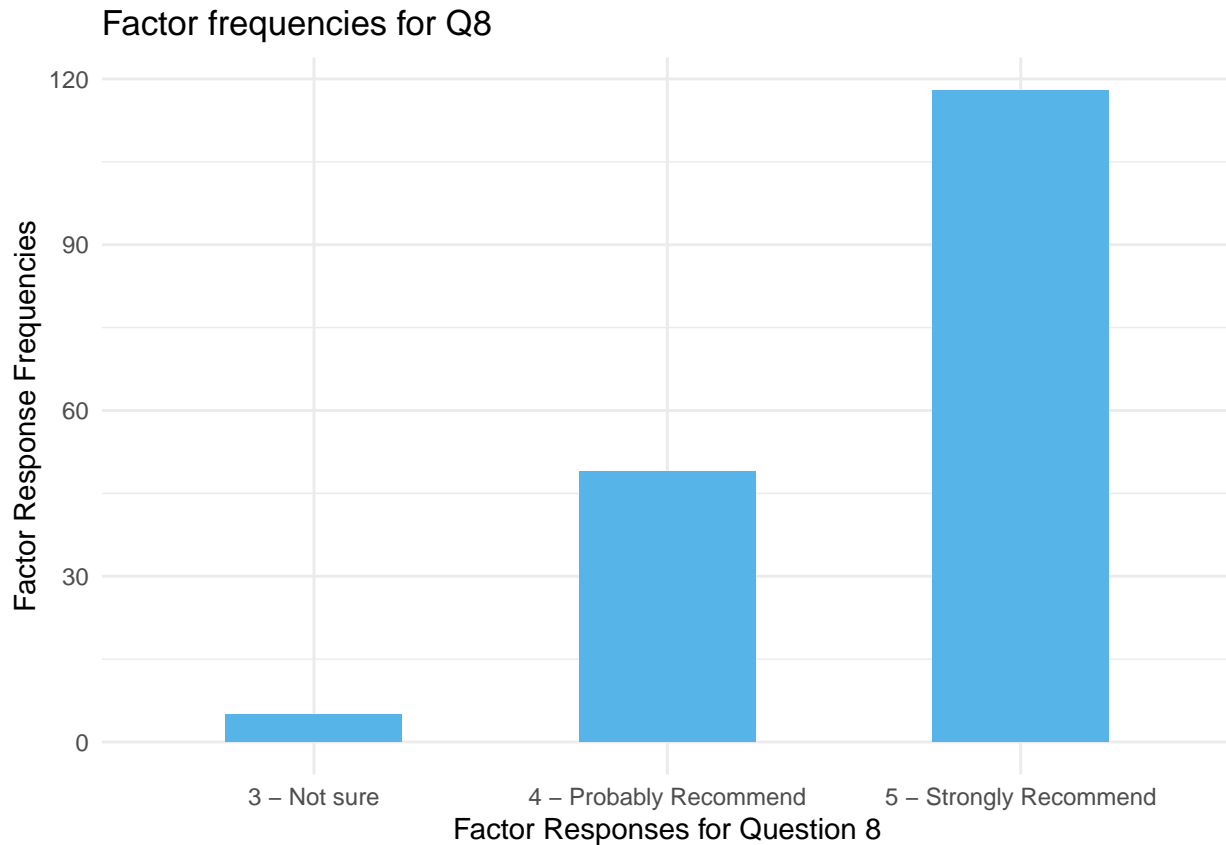
Figure 12: Word Cloud for Question 7

Figure 13: Semantic Network for Question 7

Here we saw a wide variety of responses including:

- Getting a bachelors degree
- Getting a masters degree
- Communicating with clients at work in English
- Gaining a certificate from UCLA Extension
- Continuing to improve english
- Making friends

Question 8: Would you recommend the ALC program to a friend?

We recommend keeping this question multiple choice, but changing the choices to "don't recommend", "not sure", and "recommend". With 3 options, rather than 5, we believe we would have a larger amount of students in each category, allowing us to analyze the differences between responses for each category more effectively.

Factor frequencies for Q8

Factor Response Frequencies

120

90

60

30

0

3 – Not sure 4 – Probably Recommend 5 – Strongly Recommend

Factor Responses for Question 8

Challenges of the study

It took some time to obtain the data during the beginning of the study, and we first had only about 10 complete rows of responses. We were able to manage an analysis with this available data, but it was difficult to provide actionable results due to the lack of responses. Dr. Thomas later provided more data, allowing for a richer analysis. However, we realized that the data came from a variety of surveys, each with slightly different questions. In addition, we wanted to analyze the data by category in question 8, but there were not enough observations in "not sure" to make a meaningful analysis. We found that the "probably recommend" and "strongly recommend" categories were too similar to find meaningful differences between the two categories.

Overall conclusions

The survey was able to provide a solid basis for how the program can continue to advance in the future. Through performing semantic networks and sentiment analysis, we were able to draw conclusions that could improve marketing strategies as well as be used to modify the survey for future years. We found that students enjoyed the environment teachers created for learning as well as the curriculum covered by courses at UCLA Extension ALC. Additionally, a big draw of UCLA Extension ALC is the prestige that comes with being able to put UCLA on CVs and resumes for future applications. We also found that students are interested in having volunteer opportunities, incorporation of humor in the classroom, as well as increasing talking opportunities. Furthermore, it would be beneficial to incorporate links to social media platforms like Facebook and Instagram to the main ALC page because this will better communicate a sense of community to students. As for survey changes, we recommend choosing a combination of question types to avoid response fatigue, as more responses will allow for an accurate analysis. In the upcoming years, this updated survey can be distributed to students at UCLA Extension ALC, allowing for further analysis to gain marketing insights.