

2016-01-01

# Analysis Of English Language Learner Student Reasoning About Measures Of Center And Variation: A Triangulation Of Quantitative And Qualitative Data

Lorena Galvan

University of Texas at El Paso, [lorenagalvan72@gmail.com](mailto:lorenagalvan72@gmail.com)

Follow this and additional works at: [https://digitalcommons.utep.edu/open\\_etd](https://digitalcommons.utep.edu/open_etd)

 Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), [English Language and Literature Commons](#), and the [Science and Mathematics Education Commons](#)

---

## Recommended Citation

Galvan, Lorena, "Analysis Of English Language Learner Student Reasoning About Measures Of Center And Variation: A Triangulation Of Quantitative And Qualitative Data" (2016). *Open Access Theses & Dissertations*. 846.  
[https://digitalcommons.utep.edu/open\\_etd/846](https://digitalcommons.utep.edu/open_etd/846)

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact [lweber@utep.edu](mailto:lweber@utep.edu).

ANALYSIS OF ENGLISH LANGUAGE LEARNER STUDENT REASONING ABOUT  
MEASURES OF CENTER AND VARIATION:  
A TRIANGULATION OF QUANTITATIVE AND QUALITATIVE DATA

LORENA GALVAN

Master's Program in Statistics

APPROVED:

---

Amy E. Wagler, Ph.D., Chair

---

Lawrence M. Lesser, Ph.D.

---

Maryse Jayasuriya, Ph.D.

---

Charles Ambler, Ph.D.

Dean of the Graduate School

Copyright ©

by

Lorena Galvan

2016

ANALYSIS OF ELL STUDENT REASONING ABOUT MEASURES OF  
CENTER AND VARIATION:  
A TRIANGULATION OF QUANTITATIVE AND QUALITATIVE DATA

by

LORENA GALVAN, BS

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Mathematical Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

May 2016

## **Acknowledgements**

I want to take this opportunity to express my gratitude and admiration to Dr. Amy Wagler from the Mathematical Science department at The University of Texas at El Paso. Dr. Wagler has been my mentor for the past year and working with her has been one of the best experiences I have had. I first met Dr. Amy Wagler in 2010 when I took her Introduction to Biostatistics course. Ever since that time I developed an interest in her teaching methods and her field of study. After the spring of 2014 I made an appointment to meet with her and she talked to me about some of the research topics she had been working on in the past. As a former ELL student myself, the topics that she presented to me were immediately of great interest for me. Working with Dr. Amy Wagler was the best choice since she is a great academic and personal inspiration.

I would also like to thank the other members of my committee: Dr. Lawrence Lesser of the Mathematical Science department whose previous research on educational statistics has been of great help for this project. Dr. Maryse Jayasuriya, Associate Dean of Liberal Arts and Associate Professor of English, for taking time out of her busy schedule to be part my committee. I was very fortunate to have great professors from diverse backgrounds giving me their valuable input and guidance throughout this project.

I would also like to thank other professors of the Mathematical science department with whom I took classes during my undergrad and grad student years: Dr. Ori Rosen, Dr. Joan Staniswalis, Dr. Hamide Dogan-Dunlap and Dr. Panagis Moschopoulos. I would also like to express my gratitude to Dr. Behzad Djafari-Rouhani, for whom I was a teaching assistant for a year.

To my family a big thank you for being there for me during this long journey and for never letting me down. I want to say thank you to my mom and dad for always supporting me during my long weeks of study and for being a great support in everything I have ever done in my life. Thank you to my brothers for being inspirations, especially my elder brother Christian for showing me the path of hard work.

Last but not least, from the bottom of my heart I would like to express my most sincere gratitude to my biggest pillars during this hard and long journey: my husband, Indika Mallawaarachchi who is a former student who graduated five years ago with his Master's in Statistics and who has taught me so much during these years of friendship and marriage, and to Dr. Joan Staniswalis, one of my biggest inspirations for doing what I do. Thank you so much for always having the correct advice and the kindest words even during the most difficult times of my journey. From the bottom of my heart to you all thank you!

## **Abstract**

The population of Spanish-speaking English Language Learner (ELL) students has been growing rapidly in the past 20 years. It has been estimated that by the year 2018, 1 out of 4 students in grades K-12 will be an ELL. There is a large body of research about assessment of Spanish-speaking ELL students in mathematics; however, there is little that focuses on statistics only. Statistics courses are frequently confused and/or classified under mathematics courses. However, students taking an introductory statistic course are required to know a new set of concepts, words and vocabulary that are not always used in mathematics courses. Statistics is not focused on numbers or formulas but rather a more complex set of words, such as mean, median, standard deviation, range, variance, among others.

In introductory statistics courses students are often asked to provide a written response and/ or choose from multiple choice answers, both of which require a deep understanding of statistical terminology. This study is focused on understanding the thought process students utilize to answer questions about measures of center and variation and analyzes the possible nature of difficulties English Language Learner students encounter when dealing with concepts about measures of center and variation that are commonly encountered in introductory level statistics courses. The approach of this study was to perform a triangulation of quantitative and qualitative data about the responses of Spanish-speaking ELL and those who are not ELLs (non-ELL) students about measures of center and variation. This is an appropriate focus since all students entering an introductory statistics course will encounter and utilize these new concepts and terminologies. A survey consisting of ten questions was constructed during the Fall 2011 and Spring 2012 semesters. The questions, about measures of center and variation, are presented at a variety of complexity levels and were classified as direct, context-heavy and graphical. The

study analyzes not only the final selection of answers provided by students, but also the explanation students provided on how and why they came up with their answers for each specific question.

The analysis indicated that ELL students have a harder time answering context-heavy questions correctly, that both populations of students seem to perform well when it comes to direct questions, and that graphical questions seem to benefit ELL students. These results were consistent with those findings from the qualitative part of the data. From the triangulation we can say that there is some evidence to say that Spanish-speaking ELL students have greater difficulty answering context-heavy questions correctly even if these involve terms that are usually known by them in a simpler context. We can also say that both ELL and non-ELL seems to perform better when questions are asked in a direct way and finally, there is some evidence to say that the usage of graphical representations may be beneficial to ELL students. There is also some evidence to say that scaffolding difficulty on the questions presented on an assessment may play an important role on students' performance.

## Table of Contents

Acknowledgements.....	iv
Abstract.....	vi
Table of Contents.....	viii
List of Tables.....	ix
List of Figures.....	x
Chapter 1: Introduction.....	1
Chapter 2: ELLs in the Mathematical Sciences.....	3
Chapter 3: Analysis.....	5
3.1 Student Population and setting.....	5
3.2 Methods.....	6
3.3 Analysis of Quantitative Data.....	9
3.4 Analysis of Qualitative Data.....	16
Chapter 4: Discussion.....	27
4.1 Qualitative Data Discussion.....	27
4.2 Individual Question Discussion.....	32
4.3 Conclusions.....	37
References.....	41
Appendix A - Flyer.....	44
Appendix B - Survey.....	45
Appendix C - Informed Consent Form.....	50
Appendix D1 – Transcripts from Interview I1.....	54
Appendix D2 – Transcripts from Interview I2.....	58
Appendix D3 – Transcripts from Interview I3.....	61
Appendix D4 – Transcripts from Interview I4.....	66
Appendix E – R-Code.....	70
Appendix F – Code of Qualitative Responses.....	73
Curriculum Vitae.....	76

## **List of Tables**

Table 3.1: Procedure and Time Frame for the Study.....	9
Table 3.2: Association between ELL status and Response Pattern.....	11
Table 3.3: Students' answers in percentages (%) by answer choice.....	12
Table 3.4: Percentage of Correct Responses between ELL and Non-ELL.....	16
Table 3.5: Summary of Individual Finding from Qualitative data Analysis .....	26
Table 4.1: Recommendations for Teaching.....	40

## List of Figures

Figure 3.1: Excerpt from Interview I1 .....	17
Figure 3.2: Excerpt from Interview I2 .....	18
Figure 3.3: Excerpt from Interview I3 .....	20
Figure 3.4: Excerpt from Interview I4 .....	21
Figure 3.5: Excerpt from Interview I3 .....	22
Figure 3.6: Excerpt from Interview I4 .....	23
Figure 3.7: Excerpt from Interview I1 .....	25
Figure 3.8: Excerpt from Interview I2 .....	25
Figure 4.1: Excerpt from Interview I4 .....	29
Figure 4.2: Excerpt from Interview I1 .....	29
Figure 4.3: Excerpt from Interview I1 .....	30
Figure 4.4: Excerpt from Interview I2 .....	30

## Chapter 1: Introduction

The Census Bureau estimates that approximately 5.5 million students in the United States are English language learners (ELLs) and 80% of these students speak Spanish (Cardenas-Hagan, 2010). This is of great interest especially in the United States Southwest border region and since Texas has the second largest ELL population after the state of California. There is a large body of research studies that have been conducted about assessment of ELLs in subjects such as reading, writing, and mathematics; however, little research has been conducted on ELLs in statistics with the exception of Lesser & Winsor 2009, Lesser, Wagler, Esquinca & Valenzuela 2013, Wagler, Lesser, González & Leal 2015, Lesser & Wagler 2016, and Lesser, Wagler & Salazar 2016. Due to the lexical ambiguity of many statistics terms, statistics is a challenging course not only for English language learners and especially for those whose first language is English (Lavy & Mashiach-Eizenberg, 2009; Kaplan, Fisher & Rogness, 2009; Lesser & Winsor, 2009; Kaplan, Fisher & Rogness, 2010; Kaplan, Rogness & Fisher, 2012). It is essential for statistics instructors to be able to communicate effectively with all students, ELL or non-ELL, and it should be acknowledged that ELL students must become adapted to new customs, culture, and a new language, while trying to learn a new material in a course.

Lesser and Winsor (2009) illustrated how some ELLs have difficulties understanding the language and concepts utilized in introductory statistics classes. Due to the No Child Left Behind Act of 2001, the inclusion of limited English proficient students and the provision of reasonable accommodations were mandated (Abedi, 2009; Menken, 2010). Therefore, there is a need for research about ELLs learning concepts frequently encountered in an introductory statistics course, such as measures of center and variation.

The intention of this study is to investigate how ELL status relates to student-generated responses about measures of center and variation commonly used in the statistics classroom. We not only want to investigate further whether there is an association between ELL status and how the student responds to the question, but the focus of this study is to identify the nature of confusion and difficulties ELL students may encounter in learning concepts on measures of center and variation. We want to investigate whether language plays a major role in students' assimilation of concepts commonly encountered in introductory statistics courses. We also want to make a side-by-side comparison via a triangulation of quantitative and qualitative data collected on student reasoning about measures of center and variation to analyze the performance of ELL vs Non-ELL students in the statistics classroom.

## Chapter 2: ELLs in the Mathematical Sciences

The population of ELLs in the United States has been growing rapidly in the past few years. According to Goldenberg (2008), the population of ELLs in K-12 public schools grew from 1 out of 20 in 1990 to 1 out of 9 only fifteen years later. With this fast growth, he argues that 1 out of 4 K-12 students in the United States will be an English Language learner by the year 2018. Even though not all of this population will attend college, there is still a growing population of ELLs in college. Just in Texas, which has the second highest population of ELLs (832,000 ELL students compared to California with 1.1 million ELL students), about 46% of Asian ELLs, 26% of Black ELLs and 15% of Hispanic ELLs enroll in a 4-year public college (Flores, Batalova, and Fix, 2012, p. 17).

Navigating academic language is an essential skill when taking college courses. If an ELL is more familiar with the everyday usage of English, it is very likely that he or she will struggle to understand the higher-level academic language used in college-level coursework. According to Cummins (1992), there are two proficiencies acquired when someone learns a new language: Basic Interpersonal Communicative Skills (BICS) and Cognitive Academic Language Proficiency (CALP). BICS are required for everyday communication such as reading, writing and listening, whereas CALP skills are necessary for an academic context. The later the person tries to learn a new language the harder it would be to acquire the necessary CALP skills to succeed in school since the academic registers are already built in their own language and being able to distinguish between BICS and CALP vocabulary becomes increasingly difficult. Lesser and Winsor (2009) state that “the challenge ELLs face is that the academic meaning of a term may be the same as the everyday meaning, different from everyday meaning, or not have an everyday counterpart at all” (p. 8). For the purpose of this research we would make the argument

for the need to assess ELL students in statistics focusing on the largest population of ELLs, native Spanish speakers.

Reading technical academic text is difficult for any student. According to August (2005), “Skilled readers can tolerate a small proportion of unknown words in a text without disruption of comprehension and can even infer the meanings of those words from sufficiently rich contexts” (p. 50). On the other hand, for ELLs the proportion of unknown words is longer and reading skills are likely less developed, thus making comprehension more difficult. In addition, context-heavy questions (questions that ask statistical concepts embedded in a data context) can complicate understanding of basic statistical concepts rather than aid them even though context is essential for developing statistical reasoning skills. If the reader is not skilled enough, a familiar concept may be misunderstood if it is presented in a different and/ or more complicated context.

Diane August also argues that ELLs are more likely to be diagnosed as learning disabled than non-ELLs. However, their poor performance might be due to their lack of English vocabulary rather than a learning disability. The poor performance may also be due to the lack of academic registers (deficiency in CALP), something that can be present not only in ELL but also in non-ELL students. In the context of statistics instruction, students may struggle more due to language rather than lack of understanding and be less inclined to study statistics further or be labelled as a low-achieving student. This study makes an effort to separate to what degree language proficiency affects how students respond to and understand a set of questions regarding measures of center and variation.

## **Chapter 3: Analysis**

### **3.1 Student Population and Setting**

The research participants were enrolled in a large research university or regional community college system in the Southwestern United States. In the region, approximately 80% of the student population at this university as well as the city population are Hispanic, and about 5% of the student population consists of Mexican nationals. This population is a good target for this type of research due to the large population of Spanish speakers. The majority of the students surveyed at this stage of the study were approximately between 17-22 years of age.

For the purpose of collecting the qualitative part of the data the targets were bilingual students in English and Spanish taking an introductory-level statistics course during the Fall 2015 semester. This time students were exclusively chosen from a large research university located in the Southwest region of the United States. Students from a local community college were not targets at this stage of the study, though they are reasonably very similar with regard to demographics to the students selected since the community college system is a feeder school for the research university. Two types of courses are common in undergraduate statistics. One type, a statistical literacy course, emphasizes understanding statistical results in research in the media and deemphasizes formulas and computation. In contrast, a statistical practitioner course trains students to use statistical analysis and produce statistical results. Students taking literacy and practitioner courses were contacted via email and once students showed interest in the survey, they were contacted by the investigator to take part in an interview.

### **3.2 Methods**

For the quantitative data collection, the survey was administered during the Fall 2011 and Spring 2012 semesters. There was an option to take the survey as a paper-based or an online version. Either version took no more than 10 minutes to administer. The survey was not mandatory, and students had the option to withdraw at any time. The survey was administered around 2/3 of the way into Fall 2011 semester and at the highway point of the Spring 2012 to ensure that the teacher had covered the topic references in the survey.

For the qualitative data collection, the first five questions of the survey were answered online by students and the remaining ten questions were answered during the interview that was videotaped and audiotaped. These interviews were videotaped in order to capture not only the verbal responses given by the students but their movements and gestures as well. The audiotape was utilized mainly as a backup in case the audio on the videotape was not clear enough. This data was collected during the months of November and December of the Fall 2015 semester, by this point students had covered most of the material needed for the questions. The five students interviewed were from 18 to 25 years old. Three out of five students identified themselves as ELL students and were classified as so based on a set of three questions as described below. Note that, for the purpose of the final analysis, one of the interview transcripts was removed from the analysis due to lack of explanation about the thought process behind the answers provided. When students was asked about the reason why she picked her answers she indicated that she was guessing or that she really did not know why she selected that choice. Students also indicated on several occasions that she had a lack of knowledge for statistical terms and that she was “bad” in statistics. The investigator encouraged the students to unpack about her reasoning behind her choices but student responses were the same. The student did not want to add any additional

comments and investigator did not want to cause a discomfort on the student asking for additional information. Of the four students interviewed, three indicated that their first language was Spanish. All of these students answered 8 or above to question 5 of the survey (Proficiency with English Language) and the time these students indicated was spent speaking Spanish during the month prior to the interview ranged from thirty to eighty percent of the time. One out of the 4 students interviewed indicated that his first language was English and his answer for question 5 of the survey was a 10, the maximum level in English proficiency and the percentage of Spanish-speaking time during the previous month was only 10%. Students 1 and 2 were students selected from a practitioner course and students 3 and 4 were selected from a literacy course. Students 1 and 2 were from the same course; however, they were not from the same class. They were taking the class from a different instructor on different days and times. Students 3 and 4, on the other hand, were selected from the same literacy class period.

Students were classified as ELL or non-ELL based on their answers to questions 3, 4 and 5. Questions 3 asked about the student's first spoken language; question 4 asked what percentage of time the student had spoken Spanish in the previous month; and question 5 (Interagency Language Roundtable (ILR) and American Council on the Teaching of Foreign Languages (ACTFL) scales respectively from 0-10). A principal components model was estimated using these three measures of language use. This resulted in a set of principal component scores that weighted each measure using the following formula:  $Score = 0.609 * First\ Language + 0.628 * \% \text{ time speaking Spanish} + 0.484 * ACTFL\ rating$ . The resulting scores were plotted and a division is noticeable around the scores less than or equal to approximately 30, which is also very near the median principal component score. This cut-off of 30 provided an operational definition for classifying the students as being ELL or non-ELL and is used in the

remainder of the analysis. Through this approach 128 students were classified as ELL and 219 were classified as non-ELL. Similarly, the three ELL students in the qualitative phase of the study who self-identified as ELLs were also classified as ELL using this method.

The ten questions that were given to the students during the interviews and quantitative survey were classified into 3 categories: 1) context heavy, 2) direct and 3) graphical. The classification of questions was straightforward. Questions were classified by two investigators that worked in the classification independently and then compared their results with 100% agreement. Context-heavy questions are those that may lead to multiple interpretations due to the context. Students may answer in one way or another depending on their interpretation of the context. Direct question are usually short questions that are easy to interpret and that do not lead to multiple interpretations: for example, questions that ask for definitions. Graphical questions may be context-heavy or direct, but they are characterized by the aid of a graphical representation to help in the interpretation of the questions. Below, Table 3.1 gives an overall timeline for the quantitative and qualitative stages of this study.

Table 3.1: Procedure and Time Frame for the Study

Procedure	Time Frame
Selection of Questions	Fall 2011
Quantitative Data Collection	Fall 2011-Spring 2012
Quantitative Data Analysis	September 2015
Qualitative Data Collection	November-December 2015
Classification of Questions	December 2015
Qualitative Data Analysis	January-March 2016
Triangulation of results	March-April 2016

### 3.3 Analysis of Quantitative Data.

The questions included in the scale were about measures of center and variation. Questions 6, 7 and 8 were about measures of center, questions 9, 11, 12, 13, and 15 were about measures of variation and questions number 10 included both. However, they differed with respect to how the questions were asked. Some questions were heavy in context and included a more complicated sentence structure, some questions were asked directly and the sentence structure was simple. Some of the context-heavy questions presented to students with a word problem where students had to be able to identify the concepts given within context. Other questions, classified as direct questions, asked straightforward definitions. For example, question 13 was classified as direct: this questions simply asked students to define average. Similarly, questions 14 and 15 asked students to define the words range and spread. Another classification was graphical questions (Appendix A). These were the questions that utilized some sort of

graphical representation to explain the questions. Examples of graphical questions include 10, 11 and 12 (Appendix A). Following the classification, it was of interest to investigate if there was an association between ELL status and the response pattern using a Pearson's Chi-squared test. The results of this analysis are listed on Table 3.2 where p-values are adjusted for multiplicity using a false discovery rate method (Benjamini and Hochberg, 1995).

We used alpha of 0.10 as our significance level. This level of significance is commonly used in behavioral studies where human responses are being measured. Questions 6, 7 and 8 were classified as context-heavy; we can see that question 6 showed a significant difference in the association between ELL status and response pattern. A p-value of 0.0170 and an adjusted p-value of 0.0830 were obtained. Question 8 also showed significance, the adjusted p-value of 0.0900 is fairly close to the adjusted p-value obtained from questions number 6. However, even when question 7 was classified under the same category, this question did not show significance. We can see in the table below that a p-value of 0.1800 and an adjusted p-value of 0.2500 were obtained from this question.

Questions 10, 11 and 12 were categorized as graphical questions. We obtained non-significant results from both of these questions p-values of 0.3500 and 0.1600 were obtained respectively. After the adjusted p-value was calculated the values became ever more non-significant. On the other hand, question 12, which was also categorized as graphical, showed a significant result. This means that for this questions ELL status seem to affect the way students responded to the questions. A p-value of 0.0280 was first obtained and after adjusting for multiplicity a borderline value of 0.090 was obtained.

Finally, questions 9, 13, 14 and 15 were categorized as direct questions. For this type of question, only question 9 showed significance a p-value of 0.015 and an adjusted p-value of

0.083 were obtained. This questions was classified as direct; however, it deferred from questions 13, 14 and 15 where students were asked for definitions. Questions 13 and 14 were highly non-significant; indicate that ELL status does not really affect students' responses for these questions specifically. From question 13 a p-value of 1.000 was obtained. Question 15 also presented a non-significant association between ELL status and response pattern with p-value and adjusted p-value of 0.104 and 0.520, respectively.

Table 3.2: Association between ELL status and Response Pattern

Question	Question Type	Terminology Used	P-value	P-value adjusted	Decision
6	Context	Average, mean	0.0170	0.0830	Significant
7	Context	Distribution, skewed, mean, median	0.1800	0.2500	Non-significant
8	Context	Median, distribution	0.0358	0.0900	Significant
9	Direct	Range	0.0150	0.0830	Significant
10	Graphical	Center, spread, distribution	0.3500	0.4400	Non-Significant
11	Graphical	Range	0.1600	0.2500	Non-significant
12	Graphical	Variability	0.0280	0.0900	Significant
13	Direct	Range	1.0000	1.0000	Non-significant
14	Direct	Average	0.5100	0.5700	Non-Significant
15	Direct	Spread	0.1040	0.5200	Non-Significant

After investigating association between ELL status and response pattern, a comparison of the correct response percentages between ELL and non-ELL groups using a 2-sample test for equality of proportions with continuity correction was conducted. The sample proportions for each response option are shown in Table 3.3 and the results comparing the proportion correct are shown in Table 3.4.

Table 3.3: Students' Answers in Percentages (%) by Answer Choice

Question	ELL								Non ELL							
	a	b	c	d	e	f	g	h	a	b	c	d	e	f	g	h
Q6	27.8	<b>30.9</b>	29.4	11.9					19.4	<b>47.9</b>	21.2	11.5				
Q7	28.6	<b>33.3</b>	38.1						20.1	<b>34.6</b>	45.3					
Q8	11.0	<b>24.4</b>	53.5	11.0					6.5	<b>35.9</b>	51.2	6.5				
Q9	11.2	7.2	12.0	<b>69.6</b>					8.5	3.8	6.1	<b>81.7</b>				
Q10	21.4	27.0	<b>22.2</b>	29.4					18.1	28.8	<b>27.4</b>	25.6				
Q11	18.5	<b>45.2</b>	18.5	17.7					18.2	<b>53.7</b>	15.4	12.6				
Q12	15.9	34.1	33.3	<b>16.7</b>					10.8	29.7	31.6	<b>27.8</b>				
Q13	4.0	<b>79.4</b>	4.0	10.3	1.6	0.8			6.1	<b>79.7</b>	5.7	7.5	0.5	0.5		
Q14	2.4	<b>58.4</b>	5.6	24.8	0.8	1.6	6.4		3.7	<b>62.6</b>	5.6	22.0	2.3	1.4	2.3	
Q15	51.2	6.4	<b>28.8</b>	4.0	0.8	1.6	2.4	4.8	39.7	4.2	<b>43.5</b>	1.9	1.9	1.4	2.8	4.7

\*Note: Boldfaced type signifies the correct answer.

We noticed that some of the questions regardless of being direct, context-heavy or graphical deal with the same or similar concepts. Below we can see an explanation of questions

that were grouped together according to the terminology they made reference to. For example, if several questions deal with the concept of range, those questions were grouped together for the purpose of discussion.

### **Terminology: Range**

Questions 9, 11, and 13 refer to the range. Question 9 gives a list of numbers and students are asked to find the range of the given list of numbers. Question 11 presents a histogram and asks students to agree or disagree with a given statement. Finally, question 13 asks for the definition of the term statistical term range. Question 9 was answered correctly by 69.6% of the ELL population and by 81.7% of the Non-ELL population. There seems to be no significant source of confusion for this question. There is no significant difference between the percentages of correctness for this answer as shown on the table. Question 11 was answered correctly by nearly 50% of both populations. 45.5% of ELL students and 53.7% of Non-ELL students answered this question correctly. Lastly, question 13 was answered correctly by a high percentage of students from both ELL and Non-ELL populations, around 79% of students respectively. Even though there is not an evident source of confusion for this question, around 10% of ELL and 7% of Non-ELL of students who answered this question incorrectly chose option D (the size of the data set).

### **Terminology: Mean**

Questions 6 and 14 talk about the mean or average. For question 6 there was a significant difference in correct response percentage favoring non-ELL students. Only 30.9% of ELL students compared to a 47.9% of Non-ELL students responded correctly to question 6. The incorrect answer that was chosen the most by both populations was option C. Option C states

that the most common number of children per household is 2.2. Both, ELL and Non-ELL students had a similar pattern response for question 14.

### **Terminology: Mean and Median**

Questions 7 and 8 refer to concepts of mean and median. The correct response percentage is considerably low for both questions. For question 7 most students from both populations chose option C. There seems to be confusion between the definitions of mean and median. About 38% of ELL and 45% Non-ELL opted for option C which is not enough information to tell which is which. Question 8 presents a very high percentage of students who chose option C. This question talks about giving extra points to students. Five more points were given to 10 students. Most students choose the option that states that the median of the new score distribution will be higher than the original score distribution.

### **Terminology: Mean, Median and IQR (Interquartile Range)**

Question 10 refers to measures for describing center and spread for a given distribution shown in a (histogram). The correct answer to this question is option B. Then median and IQR are the two measures that will most appropriately describe the center and spread of the given distribution. However, most students from the ELL population chose option D Mean and Standard deviation. Most students taking introductory level statistics courses are highly familiar with the terms mean and standard deviation. Among Non-ELL students, option B was the most popular.

### **Terminology: Variability**

Question 12 covers the concept of variability. On this question, students are asked to select the best measure to summarize the variability on a given data set. The correct answer for this question is option D since the interquartile range is the best measure to use to summarize the

variability on a data set because it's resistance to outliers. This question had the lowest percentage of correct response by both ELL and non-ELL students. Only 16.7 % of ELL and 27.8 % of non-ELL responded correctly to this question. Options B and C were the most popular wrong options among both populations.

### **Terminology: Spread**

Lastly, question 15 had 51.2% of ELL students choosing option A. Option A states that the correct definition of the term spread is to scatter, distribute, disperse, go apart or separate. On the other hand, non-ELL students had a 43.5% of correct response, but option A was the most popular answer after the correct answer which is option C. One reason for such confusion may be the association of the word spread to its everyday usage. If we think about the non-academic setting the word spread can have a different meaning and option A seems to fit some of those meanings.

In conclusion, we can say that there are some differences in the way ELL and non-ELL responded to some of the questions. We have questions where that difference is not statistically significant and some questions where the difference is statistically significant. However, it is complex to assess whether those differences are due to ELL status of the students or rather due to the difficulties in the assimilation of the statistical terms learned in an introductory level statistics class.

Table 3.4, compares the percentage of correct response between ELL and non-ELL. P-value was obtained using a proportion test. For the most part, in most of the questions there is not a significant difference; some questions have higher significance than others. Question 13 is of special interest since both the p-value and the adjusted p-value are practically one.

Table 3.4: Percentage of Correct Responses between ELL and Non-ELL

Question	Question Type	ELL	Non-ELL	P-value	P-value adjusted	Decision
6	Context	30.9	47.9	0.0031	0.0310	Difference
7	Context	33.3	34.6	0.9100	1.0000	No difference
8	Context	24.4	35.9	0.0360	0.0900	Difference
9	Direct	69.6	81.7	0.0150	0.0770	Difference
10	Graphical	22.2	27.4	0.3500	0.5000	No difference
11	Graphical	45.2	53.7	0.1600	0.2600	No difference
12	Graphical	16.7	27.8	0.0280	0.0900	Difference
13	Direct	79.4	79.7	1.0000	1.0000	No difference
14	Direct	58.4	62.6	0.5100	0.6400	No difference
15	Direct	28.8	43.5	0.0100	0.1040	No difference

### 3.4 Analysis of Qualitative Data.

#### Context-heavy questions

Questions 6, 7 and 8 were classified as context-heavy. Question 6 presented students with a word problem where students were given the sample mean in the context of the problem and based on that students had to pick the correct choice out of four possibilities. For this particular question, students 2 and 3 provided a fully correct answer. That means, they were able to correctly express their thought process behind their answer. See Figure 3.2 for an example. Student 2 indicated that his first language was English and that he spent only 10% of his non-academic time speaking Spanish outside of class. Student number 3 indicated that her first

language was Spanish; however, she indicated that she spent only 30% of her time speaking Spanish. She rated herself a 10 in question 5 of the survey (proficiency self-assessment) indicating that she had the highest proficiency level in the English Language. On the other hand, students 1 and 4 answered the question incorrectly. Student number 1 was unable to explain her thought process in a clear way. (See excerpt in Figure 3.1). Similarly, student number 4 had a clear context confusion on question 6. (See Appendix D4). We illustrate this confusion with statistical reasoning in context with Interviewee 1:

I1: Ok  
Ok?  
Ok, so which one do you think is the answer? A, B, C or D?  
I1: [after a pause] probably A.  
Probably A? So, you think that is a: half the household in the time have more than 2 children?  
Can you tell me a little bit about why you picked that answer?  
I1: mmm because if there is more than 2 then...I don't know that would compensate for the average and if the other half of the households have only one... I don't know it makes sense to me for the average without making any math. (Appendix D1-lines 5-13)

Figure 3.1: Excerpt from Interview I1

In contrast, Interviewee 2 did not experience confusion due to context:

I2: You want me to give you an answer what it means?

Yeah, I want you to like which one do you think is the answer for the question.

[Student takes a few more seconds to look at the paper].

I2: I would say B.

You would say B, so how did you come up with that answer? Or what makes you think that would be the answer for this question?

I2: because they divided the total amount of children by 50

Aha

I2: and then there are an average of 2.2 children in the town.

Ok, ok good. (Appendix D2-lines 3-12)

Figure 3.2: Excerpt from Interview I2

Question 7 talked about the concepts of mean and median as measures of center. Students were asked to determine which quantity was the mean and which quantity was the median for the top 1% of individual incomes. Contrary to question 6, this question was not only answered correctly or incorrectly but there were also responses not clearly correct or incorrect. Recall that the “between” classification was assigned when students were able to pick the correct choice but it was unclear how they got to that answer. During the interview processes, a great deal of confusion between the terms mean and median was evident. Examples of this can be seen on the Appendix (See Appendix D1-lines 21-23 D3-lines 31-33, 40, 43, 48-49, 55, D4-lines 21-23. There were also other issues related to question 7, for example, there was confusion of the usage of the term “measures” in statistics. This can be viewed in detail in Appendix D2-line 15.

Question 8 presented students with the concept of median. For the most part, students were able to understand the question and therefore, were able to correctly explain the thought

process behind their answer. The major problem that was detected here was the lack of academic terminology to express the students' ideas (i.e., Deficiencies in CALP). For the most part, students were able to explain the thought process behind their answer but the lack of proper academic register was clear. There was also a minor issue regarding confusion of academic terms. (Mean and median confusion) but this was not consistent with all the students. Examples of these are found in: Appendix D1-lines 29, 32, 38-39, D2-lines 29, 37, D3-lines 37 and D4-lines 28-30. But, for the most part, students provided clearer answers for question 8 than the other two questions that were classified as context-heavy.

### **Direct questions**

Questions 9, 13, 14 and 15 were classified as direct questions. Question 9 presented students with a short data set of ten numbers, then students were asked to calculate the range of the given data set. Students from a practitioner course responded correctly to this question. They explained that the question was straightforward since they have learned the definition of range in class. Students from this class seemed to have a clear register for the statistical term range. There was no statistical term confusion about the concept and there was no confusion between academic and everyday registers presented by the students in a literacy course. However, an interesting finding was that students from a more basic statistics level course were lacking the concept of range. Students 3 and 4 were not only unable to answer to the question but were not able to remember the statistical definition of range even after a several hints were given by the interviewer. (See Appendix D3, lines 38-49 and D4, lines 35-43). The following excerpts in Figures 3.3 and 3.4 illustrate when this happened.

Ok. Could you go ahead and read question 9?

[Silence, student reading the question]

I3: [nervous smile] I forgot about range.

[Pause]

So, is there a confusion there with the word range?

I3: yeah, I forgot how to calculate range.

Ok, so you don't remember the process how to calculate range of a data set?

I3: yeah.

So, which one would you think is the answer? Like, how do you... Do you remember if there is any... like, the process when they give you a data set? Like, numbers?

I3: Oh my God, I am trying to remember. Cause, I remember, mean, median, mode... range is the one I am trying to remember but I can't.

Figure 3.3: Excerpt from Interview I3

After students 3 and 4 were given the definition of range they had no trouble recognizing the correct answer, but this was a clear case of deficiency in their cognitive academic language proficiency. It can be argued that this is due to academic maturity rather than level of English proficiency. Question 13 asked students to define the term range. After the investigator explained this concept to students 3 and 4, they were able to identify the correct response.

Ok. Ok, next question 9.

[Silence, student reading the question.]

I4: I honestly don't know, because I don't know what range is.

Ok, so ah, you don't know... ah, you don't remember the definition of range?

I4: I don't.

Ok, so whenever you have a data set, for example like the data set that you have here. [Pointing at the paper] that they give you like a set of numbers, do you remember the range is the difference between which number and which number?

I4: The first and the last?

Figure 3.4: Excerpt from Interview I4

After students 3 and 4 were given the definition of range they had no trouble recognizing the correct answer, but this was a clear case of deficiency in their cognitive academic language proficiency. It can be argued that this is due to academic maturity rather than level of English proficiency. Question 13 asked students to define the term range. After the investigator explained this concept to students 3 and 4, they were able to identify the correct response.

All students answered these two questions correctly and had no issues explaining their choice. Students that originally responded correctly to question 9 were able to make a connection between question 9 and question 13. Students who originally were lacking knowledge of the term range were able to retain the information provided and were also able to make a connection between question 9 and question 13. Question 14 was also answered correctly by all students. When students were asked why they picked that answer, all indicated that the definition of average was mean as they have learned in class. Some students went further in their explanation

and said that the mean was the “average of the data”. Below we can see excerpts from interviews 3 and 4 to illustrate this connection.

Question 13.

[Student is reading the question]

I3: B?

So, do you see any relationship between this question and any other questions?

I3: ah, yes, you were asking about range in [flipped the pages] number 9.

Number 9, ok. So, your answer will be B?

I3: Yeah.

So the difference between the highest and the lowest number in the data set.

I3: yeah.

Ok, question 14.

[Student reading the question]

I3: B.

Ok, so B, average is defined as mean, so ok. What makes you think that’s the answer for this question?

I3: Just knowing the definition of mean, median and mode.

So, it’s a definition that you learned in class that you remembered

I3: yeah.

Ok. (Appendix D3, lines 97-114)

Figure 3.5: Excerpt from Interview I3

Ok, ok number 13.

I4: Oh, well. [Silence students reading question]

This one is B. The difference between the highest and lowest values.

Ok, so do you see any relationship between this question and any other questions?

I4: oh yeah. The number 9.

Number 9. Ok, number 14.

[Silence, students reading the question]

I4: Average is, I am guessing mean, B. because usually they use it like what is the average mean of something, something.

Ok, so you pick this answer because it's more related to...

I4: the class.

Is more related to the statistical meaning of the word.

I4: yes. (Appendix D4, lines 88-100)

Figure 3.6: Excerpt from Interview I4

Question 15 was also classified under the direct questions category. However, this question represented a challenge for most students. Three out of the four students interviewed were able to come up with the correct answer, but their explanation was weak and vague. Most students tended to go with choices that had more statistical terms. None of the students considered E Butter, jam dip as a possible answer. Most students were undecided between options A, B, C, D and J.

### **Graphical questions**

Questions 10, 11 and 12 were categorized as graphical questions. Question 10 was answered correctly by three out of the four students. The 3 out of the 4 students who answered this question correctly indicated that their first language was Spanish and the percentage of time spent speaking Spanish ranged from 30% to 80% of the time. The student who answered this question incorrectly indicated that his first language was English and he had spent only 10% of the time speaking Spanish during the previous month. (Question 4 on the survey). Question 10 asked students to find which two measures would be most appropriate to describe center and spread for the distribution given. A graphical representation of the distribution was provided. It can be argued that the lack of understanding of the statistical usage of the term “measures” may have led student number 2 to the wrong answer. Once again, this is an example of confusion between academic and everyday registers. For the most part, the other three students were able to explain the thought processes that led them to their final choice. A minor graphical confusion was expressed by one of the students, but after clarification the student was able to pick the correct answer. See Appendix D1 for more details.

Questions 11 and 12 presented no specific patterns. Question 11 was answered correctly by 2 students 1 and 4, incorrectly by students 3 and “in between” by student 2. Question 12 was answered incorrectly by student 1, correctly by students 2 and 4 and “in between” by student 3. Even students who responded correctly or “in between” to question 11 presented a few problems with their Cognitive Academic Proficiency (CALP); for the most part these students were able to understand the question but their explanations were lacking the appropriate academic terms (See Appendix D1 and D2). Student 1 answered question 12 incorrectly. Student 1 indicated that she was unfamiliar with some of the terminology used in the context of the question; she also indicated that she was trying to guess the answer based on the terms that were familiar and/ or

make sense to her (See Appendix D1 lines 85, 88). A portion of the interview transcript of an ELL (Interviewee 1) illustrates how language background does not tend to affect their understanding of graphics-based questions but lack of knowledge about the associated statistical vocabulary did affect the response (Appendix D1, lines 81-89).

Ok. Now, the last page. Could you read question number 12?

[Silence: student reading the question].

I1: I think Standard deviation, because it's based on all the information. So, letter B.  
letter B. So letter B says the standard deviation; because it's based on all the information on the data set. So, why do you think this is the answer?

I1: mmm I was going to say because the other ones did not really make sense to me?

Ok, so the other ones are not really familiar? They don't... ok...

I1: aha.

Figure 3.7: Excerpt from Interview I1

Now, a portion of interviewee 2's transcript (a Non-ELL student) illustrates how a student with a strong command of the statistical register responded to this graphics-based question (Appendix D2, lines 64-67). Table 3.5 below, shows a summary of the individual findings from the qualitative data analysis.

Ok, ok. On the next page, you have question 12 it also refers to the graph.

I2: [student looks at the paper] Ok.

[Silence: student reading the question].

I2: [student flipped paper twice] It probably be D, just because it's more resistant to outliers.

Figure 3.8: Excerpt from Interview I2

Table 3.5: Summary of Individual Finding from Qualitative data Analysis

Student # 1  Practitioner   ELL			Student # 2  Practitioner   Non-ELL				
Q6*	W/W	<ul style="list-style-type: none"> <li>Deficiency in Cognitive Academic Language Proficiency (CALP): student had trouble explaining thought process using the appropriate academic terminology</li> <li>Statistical concept confusion: student had a confusion between the terms mean and median</li> <li>Misconception of concept due to context. Student had difficulty recognizing terms she indicated she was familiar with due to the context they were presented in.</li> <li>Role of context: Student seemed to have difficulty with context-heavy questions.</li> </ul>	Q6*	C	<ul style="list-style-type: none"> <li>Academic and everyday register confusion: Student had difficulty disassociating the academic and everyday meanings of the word “measures”</li> <li>Deficiency in Cognitive Academic Proficiency (CALP): Student had some difficulty explaining his ideas using the correct academic terminology.</li> </ul>		
Q7*	W		Q7*	C			
Q8*	C		Q8*	C			
Q9 <sup>#</sup>	C		Q9 <sup>#</sup>	C			
Q10 <sup>&amp;</sup>	C		Q10 <sup>&amp;</sup>	W			
Q11 <sup>&amp;</sup>	C		Q11 <sup>&amp;</sup>	B			
Q12 <sup>&amp;</sup>	W		Q12 <sup>&amp;</sup>	C			
Q13 <sup>#</sup>	C		Q13 <sup>#</sup>	C			
Q14 <sup>#</sup>	C		Q14 <sup>#</sup>	C			
Q15 <sup>#</sup>	B		Q15 <sup>#</sup>	B			
Student # 3  Literacy   ELL			Students # 4   Literacy   Non-ELL				
Q6*	W/C		<ul style="list-style-type: none"> <li>Deficiency in CALP: Student had trouble explaining her thought process using the appropriate terminology. Student expressed she lacked the register for Range.</li> <li>Role of context: Student seemed to have trouble with context heavy questions.</li> <li>Confusion about “intact” statistical phases: Students had trouble making the connection between spelled out words and their Acronyms.</li> <li>Statistical concept confusion: There was a clear confusion between the terms mean and median.</li> <li>Transfer between academic registers: confusion between the mathematical meaning of center and the statistical meaning of center of a data set.</li> </ul>	Q6*		W	<ul style="list-style-type: none"> <li>Confusion between academic and everyday registers: difficulty disassociating the academic and everyday meaning of variability.</li> <li>Deficiency in CALP: Lack of registers for concepts such as: outliers and range.</li> <li>Role of context: Student showed some difficulty with context heavy questions.</li> <li>Multi-modal representations: Student indicated her preference for an auditory representation of the questions.</li> <li>Misconception of concept due to context</li> </ul>
Q7*	W			Q7*		B	
Q8*	B			Q8*		C	
Q9 <sup>#</sup>	W			Q9 <sup>#</sup>		W	
Q10 <sup>&amp;</sup>	C	Q10 <sup>&amp;</sup>		C			
Q11 <sup>&amp;</sup>	W	Q11 <sup>&amp;</sup>		C			
Q12 <sup>&amp;</sup>	B	Q12 <sup>&amp;</sup>		C			
Q13 <sup>#</sup>	C	Q13 <sup>#</sup>		C			
Q14 <sup>#</sup>	C	Q14 <sup>#</sup>		C			
Q15 <sup>#</sup>	W	Q15 <sup>#</sup>		B			

\*=context heavy, &=graphical, and #=direct question types; W=wrong, C=correct and B=’in between’

## **Chapter 4: Discussion**

### **4.1 Qualitative Data Discussion**

#### **Context-heavy questions**

When comparing ELL to non-ELL students the following was noticed: Non-ELL students performed better in the context-heavy category than ELL students. This result is in accordance to the quantitative results. Questions 6, 7 and 8 were classified as context-heavy. Especially in the case of question 6, ELL students had a greater difficulty answering this question correctly. Context-heavy questions were particularly challenging for ELL students because of their complicated structure. Question 6 was particularly challenging for ELL students because some of the terms utilized in the questions had more than one possible interpretation. For example, when students read the “most common number of children” they interpreted most common as the number that repeats the most. ELL students indicated that they were familiar with computing the sample mean of a data set; however, once this concept was taken out of the context of a formula, students had trouble recognizing it. This was a clear confusion between academic and everyday registers due to context. Some statistical terms created confusion arose on question 7. Some students were confused between the terms mean and median. For the most part students were able to recognize which terms were being addressed in this question and were able to pick the correct answer. However, it was clear that for ELL students this was a more challenging task. Both of the ELL students interviewed either read the questions twice or wanted to revisit these questions at the end of the interview.

Question 8 also showed some every day and academic register confusion. The word “adding” confused some of the students. For the most part students were able to pick the correct

answer after reading the question carefully. For the most part students were able to recognize that the term adding here does not relate to the term adding in mathematics which means to make it higher or increasing something. Some students corrected themselves on the spot after picking the option of the distribution being higher than.

The main two issues detected for this type of questions was a deficiency in CALP (Cognitive Academic Language Proficiency) and the constant confusion between academic and everyday registers. There were also some instances where students had confusion between statistical registers such as mean and median, but for the most part when this type of confusion arose students were able to correct themselves within a few seconds or after they revisited the question.

### **Graphical questions**

In the case of graphical questions ELL students performed slightly better than their Non-ELL peers. A very similar result was obtained with the reanalysis of the quantitative data. Recall that questions 10, 11, and 12 were classified as graphical questions. According to Table 3.3, there was no significant difference between ELL and Non-ELL percent correct for questions 10 and 11 and there was a significant difference between ELL and Non-ELL percent correct for question 12. This difference favored ELL students. For the category of direct questions, both populations' responses were very similar. All students responded correctly to questions 13 and 14 regardless of ELL status. These results backup the findings for the quantitative data; Tables 3.2 and 3.4 adjusted p-values for the chi-square test and the proportion test are very high, making the difference in responses highly non-significant.

Question 9 presented approximately the same pattern of responses for both ELL and Non-ELL; this finding also agrees with our finding for the quantitative part of the data. However,

there was evidence that indicates that students taking a practitioners' course performed better on these questions. Students taking a literacy course indicated that they were unfamiliar with the term "range" and perhaps this concept is not emphasized in the statistical literacy course.

Ok. Ok, next question 9.

[Silence, student reading the question.]

I4: I honestly don't know, because I don't know what range is.

Ok, so ah, you don't know... ah, you don't remember the definition of range?

I4: I don't.

Ok, so whenever you have a data set, for example like the data set that you have here. [Pointing at the paper] that they give you like a set of numbers, do you remember the range is the difference between which number and which number?

I4: The first and the last? (Appendix D4, lines 35-43)

Figure 4.1: Excerpt from Interview I4

Could you read question 9?

[student reading the question]. [Pointing at the paper].

I1: I think its D.

D, ok so 30? So, how did you come up with that answer?

I1: mmm it was asking me the range so I just ah, the difference between 62 and 32.

(Appendix D1, lines 42-46)

Figure 4.2: Excerpt from Interview I1

Non-ELL students had slightly better explanations for question 15, for the qualitative part of the data. According to our previous finding on the quantitative data, there is not a significant difference in the proportion correct for ELL and Non-ELL students on this specific question. The following are portions of a Non-ELL and ELL students with regard to their responses to questions 15. Both had the correct response; however, their explanation was different.

Ok, could you please read the last question 15?

[Silence: student readying the question].

I1: I think C.

Ok. Range or difference between two points as in point spread?

I1: aha

So, can you tell me a little bit about why did you pick that answer?

I1: mmm I picked it because... I actually I do have trouble with this meaning in my class

[student smiles] with this work, but I don't know it seem like... the other ones... I know it's not butter or spreadsheet [student points at the paper]... I don't know it seems like the closest that I can figure out. (Appendix D1, lines 98-107)

Figure 4.3: Excerpt from Interview I1

Ok. The last one number 15.

[Silence: student reading the question].

I2: spread would be C. Range is just like this difference of denotes this numbers. (Appendix D2, Lines 86-88)

Figure 4.4: Excerpt from Interview I2

## **Graphical questions**

Students were presented with three graphical questions: questions 10, 11 and 12. According to our findings using the qualitative part of the data (interviews), ELL students had more reasonable explanations than Non-ELL students for these questions. According to our results from the quantitative data, the difference in the proportion correct for question 10 is not significant. Question 11 had a very similar pattern response from both ELL and Non-ELL students. This backs up the findings from the quantitative data analysis where it was found that there was not a significant difference in the percentage of correct response between ELL and Non-ELL.

Finally, question 12 presented students with the concept of variability; for this specific question it was clear that Non-ELL students could better describe variability than their ELL peers in the qualitative study results. This is in accordance with our finding from the quantitative data where we can see that there is in fact a difference in the percentage of correct response between ELL and Non-ELL students

## **Direct questions**

Direct questions appeared to provide little difficulty to both populations of students. ELL and non-ELL students performed considerably well when asked questions that were very simple in context, not only those questions with a simpler context but also those questions whose multiple choice answers were short and direct as well. For the most part, ELL and non- ELL students were able to perform well on these questions; however, the presence of confusion between academic and every day registers was detected in both ELL and non-ELL students. ELL students as well as non-ELL students had an easier time understanding these types of questions, and therefore their explanation on how and why they chose their answer was clear; however, the

lack of academic terminology was evident. There was also a persistent confusion between the terms mean and median among ELLs. Most direct questions were easy for most students; however, question 15, which was classified as direct, represented a challenge for some students. The students tended to provide explanations that were “in between” or wrong for this question. Three out of the four students interviewed were able to come up with the correct answer, but their explanation was weak and vague. Some students picked the correct choice but they indicated that they were guessing or they were picking that option over the others because the others did not make statistical sense to them. More than one student was confused between several choices. Two of the students indicated that they had trouble understanding that concept in their classes. Over all, this question was confusing for all students. (See Appendix D1-lines 104-107, D2- line 88, D3-lines 117, 120-121, 123-124, 126 and D4- lines 102-103, 106, 108) and warrants further consideration as to the reason why.

#### **4.2 Individual Question Discussion**

In this section we will focus on the discussion of those questions that were flagged with a significant difference. See tables 3.2 and 3.3.

##### **Question 6 (Sample mean in the context)**

For these questions students were asked to select which of the statements was true given that the average number of children per household was 2.2. It is also given in the context of the questions that the total number of children in the town was divided by 50. It was specified in the question that 50 was the number of households in the town. If we look at the correct percentage for ELL students vs Non-ELL students we can see that 30.9 % of ELL responded correctly to this question versus 47.9 % of correct response from their Non-ELL peers. This is then, a significant difference in the correct response percentage indicating that ELL’s had a harder time

answering this question correctly. Interviews provided some valuable insight on why this may be an issue for ELL students. During the process of the interview ELL students seemed to have a harder time understanding sample mean within the context of the question. Some students indicated that they were familiar with the definition of mean (or average) but indicated that the question was very confusing. If we look at table 3.4 we can see that student 1 and student 3 who were classified as ELL answered this questions incorrectly. Students were given the opportunity to revisit any of the questions at any time and students 1 and 3 went back to question 6 at the end of the interview. Student 1 was unable to answer this question correctly after the second attempt. Student 3, on the other hand, responded correctly on the second try. Question 6 was particularly challenging for ELL students due to the context. ELL students had a harder time than Non-ELL students making the connection between the formulaic forms of the sample mean and the usage of this same concept within context. ELL students indicated that this question did not really make sense to them because they could not explain where 2.2 children came from. Students were trying to rationalize the fact that 2.2 was not a whole number. Both of the ELL students chose option A most of the households in the town have more than 2 children. The reason behind this answer according to ELL students interviewed was that 2.2 could not represent the number of children. ELL students used the term “compensate” in several occasions to explain that 2.2 was not a possible number. Students indicated that the idea behind this number was that if some houses had 1 child and other houses had 3 for example, then 2.2 would “compensate” for the lack of one child in one house and the excess of one child in the other house. Also, there is some evidence to think that the word average created a source of confusion: some students may have confused the everyday meaning of the word average with its statistical meaning which is defined as mean. The word average has a different meaning in everyday usage. For example, it can mean

common, not extraordinary, standard, and not outstanding. The association of the word average and its meaning outside of the statistical context was a probable source of confusion. This shows that ELL students have a harder time answering context-heavy questions correctly than their non-ELL peers.

**Question 8 (Finding Median after adding 5 point to the top scores)**

This question asked students what would happen to the median of a data set if 5 extra points were to be awarded to the 10 students with higher test scores. If we look at the quantitative analysis we can see that there is a significant difference in the responses between ELL and non-ELL. Only 24.4 % of ELL students responded correctly to this questions versus 35.9% of non-ELL. During the interviews, for the most part ELL students were able to respond correctly to this item. However, ELL students had a harder time explaining the reasoning behind their choice. There was an apparent lack of academic terminology in their explanation as well as confusion between academic and every day registers. For example, Student 3 was confused about whether the new score would be the same or higher. The fact of “adding” 5 points was confusing. There was a clear confusion between the statistical meaning of adding 5 points and the everyday register for the word “adding” which means to make it bigger or to make something go higher. Student 3 was able to pick the correct answer to this questions but the explanation for it remained ambiguous. On the other hand, non-ELL students were able to explain how the median is not affected by adding points to the last observations. Some deficiency in CALP was detected but for the most part students were able to express the correct idea using everyday language. The results from the quantitative analysis may explain why most ELL students chose the incorrect option (quantitative data) C higher than. There is a clear confusion between everyday registers and academic register and perhaps a deficiency in CALPS.

### **Question 9 (Finding the range of a list of numbers)**

Question 9 asked students to determine the range of a given data set. According to table 03 there is a significant difference in the responses of ELL versus Non-ELL. After performing the reanalysis of the quantitative data we can see that the correct response percentage of ELL is 69.6% versus 81.7% for non-ELL students. Question 9 was classified as a direct question. The concept of range is presented in a direct and clean way. During the process of the interviews an interesting finding was discovered. Students taking a practitioner's course had no trouble answering these questions correctly. Students taking this type of course were able to answer these questions regardless of their ELL status. Students 1 and 3 were able to explain the thought process behind their choice and had no trouble expressing their ideas using the correct academic terminology. On the other hand, students taking a literacy course were unfamiliar with the definition of "range;" therefore, they were unable to answer the questions. Both students in literacy courses were unable to answer this question correctly regardless of their ELL status. Students 3 and 4 indicated that they were unfamiliar with the concept and were unable to give a correct response even after a brief explanation of it. This shows how the lack of academic registers may lead to a poor performance during assessment. It may also lead us to think that academic maturity may be a confounding factor in measuring students' performance while answering assessment about measures of center and variation.

### **Question 12 (Variability of a data set)**

For this particular question, students were asked to identify the best measure to summarize the variability in a data set. From the quantitative data analysis we can see that the proportion correct for ELL's versus non-ELL is 16.7% versus 27.8 % respectively. Options B and C were the most popular incorrect choices for both ELL and non-ELL. Option B indicates

that the standard deviation is the best measure to summarize the variability on a data set because it contains all the information of the data set. Option c indicates that the standard deviation is the proper measure to use because it's the most commonly used to summarize variability. After analyzing the qualitative part of the data we gained some insight on why ELL students seemed to have a harder time answering this particular question correctly.

ELL students had a hard time understanding the meaning of variability. Both ELL students seemed to lack the proper academic register for the word "variability;" there was a clear confusion between academic and everyday registers as well. One of the ELL students indicated her confusion between the word "variability" and the word "variety". Given this confusion ELL students tend to pick the answer that contains familiar words or concepts commonly used in class. Both students 1 and 3 opted for options containing the term "standard deviation" in the answer. When they were asked about their choice they indicated that it was the term they were more familiar with and therefore, the one that made more statistical sense to them. Non-ELL students also showed some difficulty with this question, Student 2 indicated his confusion about the meaning of the word "measures" statistically speaking. There was a clear confusion between academic and everyday register of the word measures. For this particular question ELL students also had a hard time switching from spelled out words to their acronyms. ELL students seemed to have a clear concept of what IQR was but had a harder time understanding this concept when the spelt-out word is used. Perhaps, it may be a possibility that they did not choose correctly because they had a hard time making the connection between IQR and interquartile range.

### **4.3 Conclusions**

From the triangulation of qualitative and quantitative data, we can conclude that there is some evidence to say that ELL students have a harder time answering context-heavy questions correctly. It was found that ELL students have a harder time making the transition from formulaic to contextual settings for some of the concepts. In addition to this, there is some evidence to say that ELL students experience more difficulty making the connection between fully spelled out terms and their acronyms. There was a clear confusion between IQR and Interquartile Range. ELL students indicated that they knew what IQR stands for but were unsure when IQR was presented as Interquartile Range. The directly worded questions provided approximately the same level of difficulty for ELLs and non-ELLs. This result is consistent with previous research in mathematics education (Monarez, 2012).

#### **Limitations.**

The following describe the limitations associated with this study:

- We had a lower response rate than expected. Students were contacted via email. Some students did not respond after completing the first 5 questions on the survey. Therefore, we had fewer candidates to choose from for the final interviews.
- One interview had to be eliminated from the final qualitative analysis. Student number 4 was eliminated from the final analysis. The interview was rather short due to the fact that for the majority of the question this student did not provide an explanation for the choices made. Student indicated on several occasion that she was guessing or just picking the answers at random. When asked to expand on the topic the student responded that she did not know what else to say and did not really know much about the concepts. The investigator did not

want to elaborate more on this and respected the student's decision to not add more to the interview.

- Students were told at all times during the interview that they were allowed to ask questions and ask for full or partial translation of any of the questions. However, even when it was evident that students had a hard time understanding a concept, students did not ask to be provided with a full or partial translation.
- For the last three interviews there was a space limitation. The last three interviews were held during the last week of school before finals. Previous interviews were held during the month of November and the time and location was fixed and agreed on beforehand. For the last three interviews the investigator and students had to switch locations due to the fact that study rooms at the library were full. A study room at the library had been originally reserved for this purpose but due to the busy week of finals groups of 3 or more people were given priority. This may have caused some discomfort for the students.
- Order of the questions: Some students indicated that the very first question on the survey was confusing and hard to read. This may have caused a block for students to correctly answer the rest of the questions on the survey.

### **Recommendations for Teaching**

From the triangulation of quantitative and qualitative data there is some evidence of a significant difference (for some items) between the responses of ELL versus Non-ELL students on measures of center and variation. The qualitative part of the data provided valuable insight on some latent issues students from both populations may experience when dealing with concepts about measures of center and variations commonly encountered in an introductory-level statistics course. From the triangulation it can be argued that ELL students have a harder time correctly

answering context-heavy questions. It can also be said that ELL students have a harder time making the connection between acronyms and formulaic form and their spelled-out counterparts. There is as well some evidence to say that ELL students may benefit from the usage of graphics. It was also evident that both ELL and non-ELL students had a much better performance on direct questions. Simple structured questions had a higher probability of being responded correctly by both populations. Overall, the questions that showed a higher difference in the correct response percentage were those questions classified as context-heavy questions. Question 6 was of particular interest since both ELL students expressed their discomfort when they encountered this as the first question on the survey. This may lead us to think that the order in which questions are presented on a survey may have an influence on a student's performance. We may need to consider this while creating assessment not only for ELL but also for non-ELL students in the future. Table 4.1 provides teaching recommendations supported by the findings of this study.

Table 4.1 Recommendations for Teaching

Teaching Recommendation	Question	Evidence
Use vocabulary activities (Lesser & Winsor, 2009)	6,8	There is evidence of confusion between academic terms (mean and median) and also between the academic and everyday registers of the words.
Introduce new idea conceptually first so that ELLs do not focus on procedural knowledge. Context to formulaic.	6	There evidence of confusion of the term sample mean when this was taken out of context. ELL indicated that they know the formula for sample mean, had a hard time identifying the sample mean within a heavy context.
Make sure students acquire the appropriate knowledge of terms commonly used in an introductory statistics course.	9	Students in literacy courses presented the lack of this register. “Range” was an unknown term for students taking a literacy course. Deficiency in CALPS.
Emphasize the difference between every day and academic registers.	12	Students seemed to be confusing the everyday meaning of the words variability and measures with their academic meaning.
Make acronyms explicit IQR= Interquartile Range	12	ELL students seemed to have a hard time matching the spelled out words with their acronyms, specifically in the case of IQR

## References

- Abedi, J., & Lord, C. (2001). The Language Factor in Mathematics Tests. *Applied Measurement in Education, 14*(3), 219-234.
- Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P. & Witmer, J. (2010). Guidelines for Assessment and Instruction in Statistics Education (GAISE): College Report. USA: American Statistical Association.
- August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The Critical Role of Vocabulary Development for English Language Learners. *Learning Disabilities Research and Practice Learning Disabil Res Pract, 20*(1), 50-57.
- Brown, C. L. (2005). Equity of Literacy-Based Math Performance Assessments for English Language Learners. *Bilingual Research Journal, 29*(2), 337-363.
- Cummins, J. (1992). Language Proficiency, bilingualism, and academic achievement. In P. A. Richard Amato & M. A. Snow (eds.), *The multicultural Classroom: Readings for content area teachers* (pp. 16–26), Reading, MA: Addison Wesley.
- Cardenas-Hagan, E. (2010). Response to intervention: Implications for Spanish-speaking English language learners. *Perspective on Language and Literacy, 36*(2), 1-13.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford Press.
- Delmas, R., Garfield, J., Ooms, A. & Chance, B. (2006). Assessing Students' Conceptual Understanding After a First Course in Statistics. *Statistics Education Research Journal, 6*(2), 28-58.
- Flores, S. M., Batalova, J. & Fix, M. (2012). *The Educational Trajectories of English Lan*

*guage Learners in Texas*. Migration Policy Institute, Washington, DC.

Gal, I., & Garfield, J. B. (1997). *The assessment challenge in statistics education*. Amsterdam: IOS Press.

Garfield, J. B., & Gal, I. (1999). Assessment and Statistics Education: Current Challenges and Directions. *Int Statistical Rev International Statistical Review*, 67(1), 1-12.

Garrison, L. G. & Mora, J. K. (2005). Adapting Mathematics Instruction for English-Language Learners: The Language-Concept Connection. In L. Ortiz-Franco, N. G. Hernandez, & Y. De La Cruz (Eds.), *Changing the Faces of Mathematics: Perspectives on Latinos* (pp. 35–48). Reston, VA: National Council of Teachers of Mathematics.

Goldenberg, C. (2008). Teaching English Language Learners: What the research does– and does not –say. *American Educator*, 33(2), 8–44.

Kaplan, J. J., Rogness, N. T., & Fisher, D. G. (2011). Lexical ambiguity: Making a case against spread. *Teaching Statistics*, 34(2), 56-60.

Lesser, L. M. & Winsor, M. S. (2009). English language learners in introductory statistics: Lessons learned from an exploratory case study of two pre-service teachers. *Statistics Education Research Journal*, 8(2), 5-32.

Lesser, L. & Wagler, A. (April 2016). Tools for assessing readability of statistics teaching materials. *Journal of Computers in Mathematics and Science Teaching*, 35(2), 153-171

Lesser, L., Wagler, A., & Salazar, B. (in press; expected Nov. 2016). Flipping between languages? An exploratory analysis of the usage by Spanish-speaking English language learner tertiary students of a bilingual probability applet. *Statistics Education Research Journal*, 15(2),

- Lesser, L. M., Wagler, A., Esquinca, A. & Valenzuela, G. (2013) Survey of Native English Speakers and Spanish-Speaking English Language Learners in Tertiary Introductory Statistics.
- Onwuegbuzie, A. J., & Leech, N. L. (2003). Assessment in Statistics Courses: More than a tool for evaluation. *Assessment & Evaluation in Higher Education*, 28(2), 115-127.
- Thompson, D. R. & Rubenstein, R. N. (2000). Learning Mathematics Vocabulary: Potential Pitfalls and Instructional Strategies. *Mathematics Teacher*, 93(7), 568–574.
- Young, J. W., Holtzman, S., & Steinberg, J. (2011). Score Comparability For Language Minority Students On The Content Assessments Used By Two States. *ETS Research Report Series*, 2011(2), I-21.
- Monarrez, A. (2012). *Analysis of Differential Item Functioning on Selected Items Assessing Conceptual Knowledge of Descriptive Statistics for Spanish-speaking ELL and Non-ELL College Students*.
- Salazar, B. (2012). *Qualitative Analysis of the Usage by Spanish-speaking English Language Learners of Resources for Learning Probability*.
- Wagler, A., Lesser, L., González, A., & Leal, L. (2015). Assessing the lexico-grammatical characteristics of a corpus of college-level statistics textbooks: Implications for instruction and practice. *The Journal of Technical Writing and Communication*, 45(1), 31-56.

## Appendix A-Flyer

# **Interview Study In Statistics**

- If you are a Spanish-speaking student enrolled in a statistics course at UTEP for the Fall 2015 semester.
- You would like to participate in an interview study that could help us better understand how to teach measures of center and variation concepts in introductory statistics courses more effectively to students.
- Conducted by a graduate student researcher able to listen and speak in English and Spanish.
- In our analysis reporting, we will not use your name or any other information that could identify you.
- The interview will take only 10 to 15 minutes of your time and a possible very brief email or phone follow-up later to clarify anything that was not clear.
- In appreciation for your time, you will be rewarded with a \$10 Visa gift card.

If you are interested contact Lorena Galvan to schedule an interview:

[lgalvan3@miners.utep.edu](mailto:lgalvan3@miners.utep.edu) or at (915)-740-3063

## Appendix B- Survey

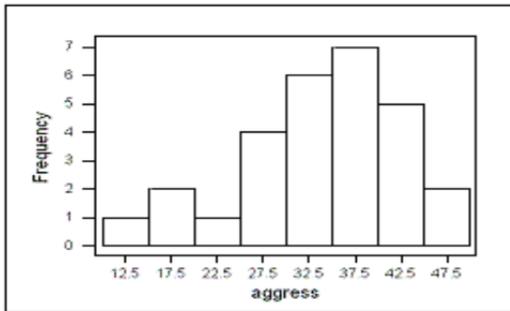
1. What year in school are you? a) freshman b) sophomore c) junior d) senior e) graduate student
2. What is your age? \_\_\_\_\_
3. What is your first language spoken? a) English b) Spanish c) other: \_\_\_\_\_
4. Estimate what percentage of time in the past month you spoke Spanish. Give a number between 0 and 100% \_\_\_\_\_
5. Using the (0-10) scale below, give the number that best describes your proficiency with the English language \_\_\_\_\_

Level	Description
10	Able to speak like an educated native speaker
9	Able to speak with a great deal of fluency, grammatical accuracy, precision of vocabulary and idiomaticity
8	Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal or informal conversations
7	Able to satisfy most work requirements and show some ability to communicate on concrete topics
6	Able to satisfy routine social demands and limited work requirements
5	Able to satisfy most survival needs and limited social demands
4	Able to satisfy most survival needs and some limited social demands
3	Able to satisfy most survival needs and minimum courtesy requirements
2	Able to satisfy immediate need with learned utterances
1	Able to operate in only a very limited capacity
0	Unable to function in English

6. The school committee of a small town wanted to determine the average number of children per household in their town. They divided the total number of children in the town by 50, the total number of households. Which of the following statements must be true if the average children per household is 2.2 children?

- a. Half the households in the town have more than 2 children.
  - b. There are a total of 110 children in the town.
  - c. The most common number of children in a household is 2.2.
  - d. None of the above.
7. The distribution of the top 1% of individual incomes in the US is strongly skewed to the right. In 1997, the two measures of center for the top 1% of individual incomes were \$330,000 and \$675,000. Which number represents the mean income of the top 1% and which number represents the median income of the top 1%? Choose the best answer.
- e. \$330,000 is the mean and \$675,000 is the median.
  - f. \$330,000 is the median and \$675,000 is the mean.
  - g. Not enough information to tell which is which.
8. You give a test to 100 students and determine the median score. After grading the test, you realize that the 10 students with the highest scores did exceptionally well. You decide to award these 10 students a bonus of 5 more points. The median of the new score distribution will be \_\_\_\_\_ that of the original score distribution.
- a. lower than
  - b. equal to
  - c. higher than
  - d. depending on skewness, higher or lower than
9. Find the range of the following list: 31, 36, 48, 50, 50, 53, 54, 56, 60, and 62.
- a. 62
  - b. 5
  - c. 10
  - d. 30
10. A test to measure aggressive tendencies was given to a group of teenage boys who were members of a street gang. The test is scored from 10 to 60, with a high score indicating

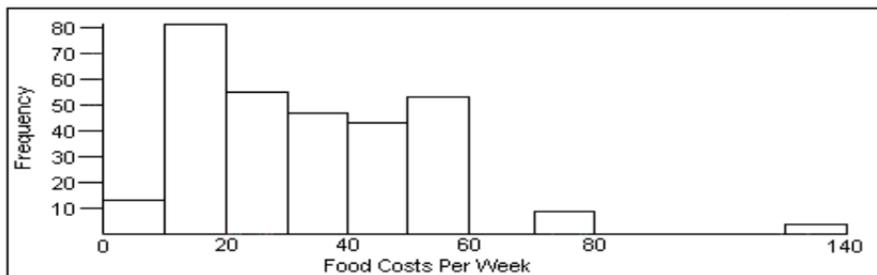
more aggression. The histogram represents the results for these 28 boys. Which two measures would be most appropriate to describe center and spread for this distribution?



- a. Range and mean
- b. Mean and median
- c. Median and IQR
- d. Mean and standard deviation

**Items 11 and 12 refer to the following situation:**

This is a distribution of how much money was spent per week for a random sample of college students.



11. The range for this distribution is \$132.50. Indicate your agreement or disagreement with the following statement: The range is not a useful summary of the variability of this data set.
- a. Agree, it is too vague.
  - b. Agree, it is too easily influenced by outliers.
  - c. Agree, it does not use information on the center of the data.

- d. Disagree, a range of \$132.50 is a good measure of variability because students are apt to spend any amount of money between \$0 and \$132.50.

12. What is the best measure to use to summarize the variability of this data set?

- a. Range, because it tells you the overall spread of the data.
- b. Standard deviation, because it is based on all the information in the data set.
- c. Standard deviation, because it is the most commonly used measure of variability.
- d. Interquartile range, because it is resistant to outliers.

13. Range is defined as:

- a. The set of all output values produced by a function.
- b. The difference between the highest value and the lowest value in a set.
- c. The numbered place in the list of ordered values.
- d. The size of a data set.
- e. The distance a charged particle travels before stopping.
- f. The set of notes a musical instrument can play, or that are used in a piece of music.

14. Average is defined as:

- a. Ordinary, Normal, typical, mediocre, not extraordinary, common, neither outstanding nor poor, standard
- b. Mean
- c. Median or in the middle
- d. Overall summary on something, general value that represents most of the data, overall outcome
- e. Mode, most common number
- f. Majority
- g. A value we can use to compare one person's performance to the group

15. Spread is defined as:

- a. To scatter distribute, disperse; to go apart or separate; to extend over a larger space
- b. To distribute in a thin layer, smear or cover evenly
- c. Range or difference between two numbers as in point spread
- d. Extend, open out as in spread his wings
- e. Butter, jam dip
- f. Spreadsheet
- g. A large group
- h. Graph

## Appendix C- Informed Consent Form

### University of Texas at El Paso (UTEP) Institutional Review Board Informed Consent Form for Research Involving Human Subjects

---

**Protocol Title:** Analysis of ELL student response about measures of center and variation for ELL and non-ELL students.

**Principal Investigator:** Lorena Galvan and Amy Wagler

**UTEP:** Mathematical Sciences Department

---

#### 1. Introduction

---

You are being asked to take part voluntarily in the research project described below. Before agreeing to take part in this research study, it is important that you read the consent form that describes the study. Please ask the study researcher or the study staff to explain any words or information that you do not clearly understand.

#### 2. Why is this study being done?

---

You have been asked to take part in a research study about how communication and language affect learning in statistics in an introductory statistics class. The main purpose of this study is to understand better what strengths and difficulties bilingual students have with language and learning measures of center and variation in an introductory statistics course. You are being asked to be in the study because you are currently registered in an introductory statistics class and you are bilingual in English and Spanish. If you decided to participate in this study, your involvement will last about approximately 10-15 minutes.

#### 3. What is involved in the study?

---

If you agree to take part in this study, you will be interviewed in an office on the UTEP campus by an interviewer who is bilingual. The interview will last about 10-15. Prior to the interview you will be handed a short survey containing 5 questions. During the interview process questions on measures of center and variation will be asked. The interview will be videotaped and audiotaped for analyzing purposes, but will be kept confidential (see #13).

#### 4. What are the risks and discomforts of the study?

---

There are no known risks associated with this research. However, there might be a minimal discomfort due to the recording of videotape and audiotape.

### **5. What will happen if I am injured in this study?**

---

The University of Texas at El Paso and its affiliates do not offer to pay for or cover the cost of medical treatment for research related illness or injury. No funds have been set aside to pay or reimburse you in the event of such injury or illness. You will not give up any of your legal rights by signing this consent form. You should report any such injury to Lorena Galvan at (915-740-3063) or [lgalvan3@miners.utep.edu](mailto:lgalvan3@miners.utep.edu) and to the UTEP Institutional Review Board (IRB) at (915-747-8841) or [irb.orsp@utep.edu](mailto:irb.orsp@utep.edu).

### **6. Are there benefits to taking part in this study?**

---

There will be a monetary compensation in the form of a \$10 Visa Gift Card for your participating on the study. Furthermore, you may gain a further understanding of measures of center and variation concepts. This research may assist us in gaining significant data as to how accommodations made for English language learners may aid their comprehension in an introductory statistics course, and this may be useful to students taught in the future.

### **7. What other options are there?**

---

You have the option not to take part in this study. There will be no penalties involved if you choose not to take part in this study.

### **8. Who is paying for this study?**

---

This study does not receive funding from any association.

### **9. What are my costs?**

---

There are no costs to you beyond the time you spend in the interview

### **11. What if I want to withdraw or am asked to withdraw from this study?**

---

Taking part in this study is voluntary. You have the right to choose not to take part in this study. If you do not take part in the study, there will be no penalty.

If you choose to take part, you have the right to stop at any time. However, we encourage you to talk to a member of the research group so that they know why you are leaving the study. If there are any new findings during the study that may affect whether you want to continue to take part, you will be told about them.

## **12. Who do I call if I have questions or problems?**

---

You may ask any questions you have now. If you have questions later, you may contact Dr. Wagler at (915-747-6847; [awagler2@utep.edu](mailto:awagler2@utep.edu)). If you have questions or concerns about your participation as a research subject, please contact the UTEP Institutional Review Board (IRB) at (915-747-8841) or [irb.orsp@utep.edu](mailto:irb.orsp@utep.edu).

## **13. What about confidentiality?**

---

1. Your part in this study is confidential. You will not be addressed by or asked your name at any time throughout the interview and, in any case, a pseudonym will be used for the interview transcripts. No footage would ever be shown or shared with anyone outside the research team unless the excerpt has your face blurred out or not visible. The raw videotaped interviews will be kept in a locked file cabinet in a locked office and seen only by the investigator and the research advisors. Interview transcripts will not use your name or any identifying information.

2. Every effort will be made to keep your information confidential. Your personal information may be disclosed if required by law. Organizations that may inspect and/or copy your research records for quality assurance and data analysis include, but are not necessarily limited to:

- The sponsor or an agent for the sponsor
- Department of Health and Human Services
- UTEP Institutional Review Board

The results of this research study may be presented at professional meetings or in publications, but only in a way that does not reveal your identity.

## **14. Authorization Statement**

---

I have read each page of this paper about the study (or it was read to me). I know that being in this study is voluntary and I choose to be in this study. I know I can stop being in this study without penalty. I will get a copy of this consent form now and can get information on results of the study later if I wish.

Participant Name: \_\_\_\_\_ Date: \_\_\_\_\_

Participant Signature: \_\_\_\_\_ Time: \_\_\_\_\_

Consent form explained/witnessed by: \_\_\_\_\_

Signature

Printed name: \_\_\_\_\_

Date: \_\_\_\_\_ Time: \_\_\_\_\_

## Appendix D1- Interview Transcript I1

- 1 Ok, we are going to go ahead and start the interview. If you please can read question 6.
- 2 I1: out load?
- 3 No, you can just read it to yourself.
- 4 [Silence: students taking time to read the question].
- 5 I1: Ok
- 6 Ok?
- 7 Ok, so which one do you think is the answer? A, B, C or D?
- 8 I1: [after a pause] probably A.
- 9 Probably A? So, you think that is a: half the household in the time have more than 2 children?
- 10 Can you tell me a little bit about why you picked that answer?
- 11 I1: mmm because if there is more than 2 then...I don't know that would compensate for the
- 12 average and if the other half of the households have only one... I don't know it makes sense to
- 13 me for the average without making any math.
- 14 Ok, ok now could you read question 7 now?
- 15 [Silence, students taking time to read the question. Students points at the paper while reading.
- 16 Interviewer taking notes].
- 17 I1: ok, I think the answer is letter A.
- 18 Ok, letter A, so the first one is the mean and 675,000 is the median.
- 19 I1: aha
- 20 Ok, so can you tell me a little bit about how did you came up with that answer?
- 21 I1: mm because it says that its skewed to the right, so then, I picture in the graph, because of the
- 22 values I would say it says 330,000 is lower than the other number.. I don't know... I think it is
- 23 what it is.
- 24 Ok, ok. Do you have any question regarding this question? Something that you don't understand
- 25 in the context of the question?
- 26 I1: mmm... [Pause] I don't know, I think I get the idea.
- 27 Me: ok, Ok. Could you please read now question 8?
- 28 [Student reading the question].
- 29 I1: ok, It could be higher, letter C.
- 30 Letter C? So, ah, so how did you like...how did you come up with the answer that it will be
- 31 higher or why do you think I will be higher?
- 32 I1: No, wait, I am changing it! Its equal to...[student laughs]
- 33 It's equal? So, your answer is?
- 34 I1: it is B.
- 35 Ok, B.
- 36 I1: yes, B.
- 37 Ok, so why did you change your answer?

38 I1: I don't know, I was thinking about mean, but then I was like ok median once I looked at it  
39 twice. Like, Ok I doesn't matter... all those changes.  
40 Oh ok, so at the beginning you were thinking about mean... oh ok, ok.  
41 I1: [saying yes with her head and smiling].  
42 Could you read question 9?  
43 [student reading the question]. [Pointing at the paper].  
44 I1: I think its D.  
45 D, ok so 30? So, how did you come up with that answer?  
46 I1: mmm it was asking me the range so I just ah, the difference between 62 and 32.  
47 Ok, there is one the back question 10.  
48 [Student flipping the paper to see the next answer. Student flipped the paper and reads the  
49 question].  
50 I1: mmm... I am not sure about this one.  
51 Ok. Is there anything that you have question about? Anything on the question that is not clear?  
52 [Student makes a pause and read the question again]. [Points at the paper].  
53 I1: like a... like in the graph the values on the x-axis... like I don't... [Confusion face]  
54 Ok, the values on the x-axis these ones right here? [Shows the paper to the student]  
55 I1: yes, aha.  
56 These ones stand for aggression.  
57 I1: oh that's why I was looking at the values between 10 and 60.  
58 I think this one is... Do you see the small point it goes from 12.5 to 47. 5 yeah.  
59 I1: ok... [Pause] to describe center is talking about the center of the graph. [Pause] so... I  
60 think... Is letter C.  
61 Letter C?  
62 I1: aha  
63 So, why did you pick letter C?  
64 I1: because... if I look at the center of the graph... I mean is not telling me the mean or the range  
65 but equal to more like the median, just because I mean is like the medium value and then IQR  
66 because I think of the values and that makes more sense to me.  
67 Ah, ok. Now, this graph right here will be for questions 11 and 12  
68 I1: ah ok.  
69 Could you please read question 11?  
70 [student takes time to read the question].  
71 I1: ah, I think B  
72 Oh ok, B?  
73 I1: aha, it is too easily influenced by outliers.  
74 Ok, ok. Something that you want to add on this question? So, how did you... your reasoning  
75 behind your answer?  
76 [student looks at the paper again].  
77 I1: because it seems... I don't know. Because it says more students they spend up to 80.

78 Ok.  
79 I1: like, adding on people that spend a lot more that may not be like the focus, what you are  
80 trying to figure out... I mean, what they graph is for I guess.  
81 Ok. Now, the last page. Could you read question 12?  
82 [Silence: student reading the question].  
83 I1: I think Standard deviation, because it's based on all the information. So, letter B.  
84 Letter B. So letter B says the standard deviation; because it's based on all the information on the  
85 data set. So, why do you think this is the answer?  
86 I1: mmm I was going to say because the other ones did not really make sense to me?  
87 Ok, so the other ones are not really familiar? They don't... ok...  
88 I1: aha.  
89 Could you please read number 13?  
90 [Silence: student reading the question].  
91 I1: I think its letter B. The difference between the highest and the lowest value.  
92 Ok, ok, could you please read number 14?  
93 [Silence: student reading the question].  
94 I1: oh, ah, I would say B. yeah.  
95 Letter B, ok. Why did you pick that answer that average is define as mean?  
96 I1: mmm because... I think that is the meaning of mean that the mean is the average of all data  
97 set.  
98 Ok, could you please read the last question 15?  
99 [Silence: student readying the question].  
100 I1: I think C.  
101 Ok. Range or difference between two points as in point spread?  
102 I1: aha  
103 So, can you tell me a little bit about why did you pick that answer?  
104 I1: mmm I picked it because... I actually I do have trouble with this meaning in my class  
105 [student smiles] with this work, but I don't know it seem like... the other ones... I know it's not  
106 butter or spreadsheet [student points at the paper]... I don't know it seems like the closest that I  
107 can figure out.  
108 Ok, ok. That is it. Thank you so much, than you so much for your time.  
109 [Shake hands with student interview stands up].  
110 Do you have any questions?  
111 I1: mmm no...  
112 Any question that you would like to read again.  
113 [Interviewer sits down again since student is flipping through the papers again].  
114 I1: mm maybe, just the first one.  
115 Me: Ok.  
116 I1: Just because I remember it was the first one and I was kind of... [Student laughs] nervous.  
117 Ok, you can go ahead and read it again.

118 [Silence: Student is reading the question].  
119 I1: Ok, ok I think my answer is still the same.  
120 Ok, so what was your answer for number 6? It was?  
121 I1: it was A  
122 A, so half the household in the town have more than 2 children. Ok, can you tell me again why  
123 that was your answer?  
124 I1: oh, why did I pick that one?  
125 Aha...  
126 I1: mmm, well, because... I don't know... I find it hard to believe others. The exact same  
127 number, like 110 children in town or because if they divided the number of children by 50  
128 houses like, that would say that there is specifically 2? Like, 2 children? Two childs maybe? I  
129 don't know... I still think is A.  
130 You still think is A?... ok.  
131 [Student was shaking her head saying yes].  
132 Ok, thank you so much for you time again and I am going to go ahead and turn off the camera.  
133 [Interviewer thank the students and they shake hands].  
134  
135 End...  
136

## Appendix D2- Interview Transcript I2

- 1 Ok, so we are going to go ahead and start with the questions. Can you read question 6?  
2 [Student reading the question. Paper on the table].  
3 I2: You want me to give you an answer what it means?  
4 Yeah, I want you to like which one do you think is the answer for the question.  
5 [Student takes a few more seconds to look at the paper].  
6 I2: I would say B.  
7 You would say B, so how did you come up with that answer? Or what makes you think that  
8 would be the answer for this question?  
9 I2: because they divided the total amount of children by 50  
10 Aha  
11 I2: and then there are an average of 2.2 children in the town.  
12 Ok, ok good.  
13 Can you go ahead and read question 7 for me?  
14 [Student reading the question].  
15 I2: What does it mean by the two measures? [Student pointing at the paper]  
16 The two measures of center... kind of like the two things that they give your for information you  
17 know to kind of have some information about the 1 percent. The top 1 percent.  
18 I2: Ok.  
19 [Student looks at the paper for a few more seconds and makes some hand gestures of  
20 calculation].  
21 I2: I probably say it was going to be B, just because it says it skewed right, so the median should  
22 be lower than the skew will be.  
23 Ok, so your answer will be B, what was it that you say because it's skewed to the...  
24 I2: because it says it is skewed to the right so that means that the answer should be, I mean the  
25 median should be lower than what the actual mean is.  
26 What the mean is... [shaking head in agreement]. Ok.  
27 Ok, so next one is number 8.  
28 [Silence: subject reading the question].  
29 I2: Wo, I think It would be B, because the median cannot really change just because the outliers  
30 they scores were getting higher.  
31 Oh, ok so, you think that it would be ah...  
32 I2: equal to...  
33 Equal to, so the mean will remain the same.  
34 I2: [student says yes shaking his head, but a few second later he corrects]  
35 I2: no, no the median  
36 I mean, the median, so you said that it was because?  
37 I2: just because the scores were higher does not mean that the total was affected I guess.  
38 Ah ok, ok. Can you go ahead and read question 9 please?

39 [Silence: student reading the question].  
40 I2: number 9 would be ah D.  
41 Number 9 will be D, ah, 30. So how did you come up with that answer?  
42 I2: subtract 32 from 62.  
43 Ok, so you subtract the...  
44 I2: The highest and lowest.  
45 The highest and lowest, ok.  
46 Ok, on the other side you have question 10.  
47 [Both interviewer and subject flipped pages]  
48 And that would be refereeing to the graph.  
49 I2: this one right here? (Student pointing at the paper)  
50 [Silence: student reading the question].  
51 I2: I think the answer for 10 will be B, mean and median.  
52 Mean and median. Ok, so, why do you think it would be mean and median?  
53 I2: WO, so the mean would show how much is the same if it high it will show that there is more  
54 aggression in the in the boys and median will be I guess depending on the data if its high then it  
55 will lean more towards them being more aggressive.  
56 Ok, now the next two questions are going to be refereeing to this graph right here [interviewer  
57 showing her paper to the student]  
58 Can you go ahead and read number 11?  
59 I2: ok.  
60 [Silence: student reading the question].  
61 I2: I put it on B since the outlier because since the outlier is a lot greater and it's very little.  
62 Ok, so your answer will be that you agree, because it's easily influenced by outliers?  
63 I2: yeah.  
64 Ok, ok. On the next page, you have question 12 it also refers to the graph.  
65 I2: [student looks at the paper] Ok.  
66 [Silence: student reading the question].  
67 I2: [student flipped paper twice] It probably be D, just because it's more resistant to outliers.  
68 Ok, ok good. Can you go ahead and read question 13?  
69 [Silence: students reading the question].  
70 I2: 13 would be B.  
71 So how did you, how did you come up with that answer? Do you see any relationship between  
72 this question and any other question that you have previously answered?  
73 I2: yeah, it will be ah, number 11 (pointing at the paper) because it's asking for the highest and  
74 lowest values.  
75 Ok, so your answer for 13 would be... B right?  
76 I2: aha  
77 The difference between the highest and the lowest values in a set, that is what defines the range  
78 right?

79 I2: aha.  
80 Ok. Can you go ahead and read question 14?  
81 [Silence: student reading the question].  
82 I2: average I mean 14 would be B, the mean.  
83 So B for mean. Ok. So, why do you think B is the answer for this question?  
84 I2: because, ah I think mean is the... You get all the data you add it up and you divide it by how  
85 many set there are. That is the same thing for the mean.  
86 Ok. The last one number 15.  
87 [Silence: student reading the question].  
88 I2: spread would be C. Range is just like this difference of denotes this numbers.  
89 Ok, ok. Do you have any additional comments on any of the questions? Any questions that you  
90 want to go back and read it again? Or...  
91 I2: No [confident laugh].  
92 Ok, ok. Thank you so much for your time and I appreciate it ok. [Shake hands with subject]  
93 I2: Thank you.  
94  
95 End...  
96

### Appendix D3- Interview Transcript I3

1 Ok, we are going to go ahead and start the interview. Could you please ah, ready question 6?  
2 I3: Just to myself right? I don't have to read it out loud or anything?  
3 Yeah, just to yourself.  
4 [Silence, student is reading the question]  
5 I3: mmm, C?  
6 C? Ah, so, C would be the most common number of children in a household is 2.2. So, why do  
7 you think that is the answer? Why did you pick that one instead of the other ones?  
8 I3: Because of the last statement where it says ah, if the average children per household are 2.2  
9 children, so that is the average so that is about how many children are in the household...2.2.  
10 Ok, ok  
11 I3: [Nervous smile]  
12 Is there anything in the question that is confusing?  
13 I3: Ah, where it says they divided the total number of children in the town by 50  
14 Oh ok.  
15 I3: I did not really get where they get that 50 from, that's what confuses me.  
16 Oh ok,  
17 I3: Oh, the total number of households is that 50? I really just don't know where they got that  
18 number from? 50? Is that the number of households?  
19 Ah, yeah.  
20 I3: Ah, Ok.  
21 Yeah, it's going to the total number  
22 I3: Ok.  
23 Ok, so your final answer for 6 will be... which one again?  
24 I3: C  
25 C? Ok.  
26 Could you go ahead and read question 7 please?  
27 [Silence, student reading the question]  
28 I3: I will say A.  
29 Ok. So, what makes you pick that answer over the other ones?  
30 I3: ah, because it says that the center of the 1% was 330,000 to 675 so, if you actually do it the  
31 center is the mean, so, it will be the 330 and then the median the 675.  
32 Ok. Question 8.  
33 [Student readying the question]  
34 I3: B.  
35 B? So, Do you think that the new score would... the new distribution would remain equal to?  
36 I3: yeah, because it is distributed to the same amount of students.  
37 Ok. Could you go ahead and read question 9?  
38 [Silence, student reading the question]

39 I3: [nervous smile] I forgot about range.  
40 [Pause]  
41 So, is there a confusing there with the word range?  
42 I3: yeah, I forgot how to calculate range.  
43 Ok, so you don't remember the process how to calculate range of a data set?  
44 I3: yeah.  
45 So, which one would you think is the answer? Like, how do you... Do you remember if there is  
46 any... like, the process when they give you a data set? Like, numbers?  
47 I3: Oh my God, I am trying to remember. Cause, I remember, mean, median, mode... range is  
48 the one I am trying to remember but I can't.  
49 [Students is looking at the paper again.]  
50 [Pause]  
51 Do you want me to clarify a little bit about the range?  
52 I3: [Nervous laugh] yeah!  
53 Ok, so basically the range when they give you a set of numbers is going to be the difference...  
54 I3: between each number?  
55 Is going to be the difference between the largest... [Pause to see if the student remembers]  
56 I3: Oh, [students still looks confuse]  
57 And the smallest observation right?  
58 I3: Oh, that's right. [Nervous laugh]  
59 So, based on that which one is going to be you're a...  
60 I3: Ah, D.  
61 D?  
62 I3: Aha.  
63 Ok. So, for this question you had... ah, is not really some concept that is very familiar so far.  
64 [Students says yes with her head]  
65 Ok, on the other side you have question 10 and is going to be related to the graph.  
66 [Student reading the question]  
67 I3: I would say D, not wait hold on! C.  
68 C?  
69 I3: yeah.  
70 Ok, so what makes you pick mean and IQR?  
71 I3: So, to find the center you would need to find the mean and then the spread there is an  
72 equation for IQR and then it helps you because you use the standard deviation and the mean for  
73 that.  
74 Ok, good.  
75 The next two questions, questions 11 and 12 are going to be related to this graph right here  
76 [pointing at the paper]  
77 [Student reading the next question]  
78 I3: I would say A.

79 A?  
80 [Students says yes with her head]  
81 Ok, so you agree with the statement and then you think that it's too vague right?  
82 I3: yeah, like you can't... you won't be able to find much with just the range, because like you  
83 said it is just the biggest number and the smallest number subtracted.  
84 Ok. Could you go ahead and read question 12?  
85 [Student reading the question]  
86 I3: the interquartile range is that to get the range of 4 part is that what it is? Or for D?  
87 What do you mean?  
88 I3: Is that where you get the range form that...  
89 Is the same as the IQR, yeah.  
90 [Student smiles in a content way]  
91 It is just the letter instead of spelling out the whole word.  
92 I3: ok, I go with D.  
93 Ok, you go with D. Can you tell me a little bit about why you picked D?  
94 I3: just because you get all the numbers to get the IQR so it will be better for you to understand  
95 where you get all those numbers.  
96 Question 13.  
97 [Student is reading the question]  
98 I3: B?  
99 So, do you see any relationship between this question and any other questions?  
100 I3: ah, yes, you were asking about range in [flipped the pages] number 9.  
101 Number 9, ok. So, your answer will be B?  
102 I3: Yeah.  
103 So the difference between the highest and the lowest number in the data set.  
104 I3: yeah.  
105 Ok, question 14.  
106 [Student reading the question]  
107 I3: B.  
108 Ok, so B, average is defined as mean, so ok. What makes you think that's the answer for this  
109 question?  
110 I3: Just knowing the definition of mean, median and mode.  
111 So, it's a definition that you learned in class that you remembered  
112 I3: yeah.  
113 Ok.  
114 Question 15, the last one.  
115 [Student reading the question]  
116 I3: I wanna say H is a graph. [Confused look]  
117 H? Spread is defined as a... as a graph? So, what makes you think that this is the answer for this  
118 question?

119 I3: just reading the other definition... I don't know, it doesn't click. I am trying to remember  
120 when we use the word spread in Statistics.  
121 Ok, so any of the other options seem like a reasonable answer? Other than a...  
122 I3: I would say C. I am between H and C.  
123 Ok, so you are between H and C. Ok, so if you..., which one do you think is.  
124 I3: or G, I don't know there's a lot of answers [nervous laugh] I think I am gonna go with G  
125 now. Spread of a large. Spread of a large group.  
126 Ok.  
127 I3: because it's picking from... yeah, I think G.  
128 Ok, so you are going to pick G. Ok.  
129 Ok, so for this question you had more than one option that made sense to the definition of spread.  
130 [Student shakes her head saying yes]  
131 Ok, but if you had to think specifically statistics or in the context of your class so that would be  
132 ah, your answer will be ah...  
133 I3: G.  
134 Ah, G. Ok, do you have any further questions? Or any question that you want to go back and  
135 read again? Or maybe...  
136 I3: I wanna go back to 6.  
137 You want to go back to 6? Ok.  
138 [Student reading question 6 again]  
139 I3: yeah, I am going to change it to B, not that I know what that number 50 stands for.  
140 Ok, so, for number 6 what was your previous answer?  
141 I3: It was C.  
142 Ok, it was C and then you are going to change it that to B right. Ok, so what make decided that  
143 you are going to change your answer for number 6?  
144 I3: just finally understanding just clicking in my head what that number 50 stands for.  
145 Ok.  
146 I3: Because I just didn't know where that was coming from.  
147 Ok, so letter B will be there is a total of 110 children in the town.  
148 I3: yeah, because they do it by 50 and that how... [Talking very low]  
149 Ok, any other question that you want to maybe read again?  
150 I3: yeah, I am just going to read 7 again now.  
151 [Student reading question 7 again]  
152 I3: I am going to change it from A to B instead.  
153 Ok, so your previous answer was A right? So now you are changing from A to B. then what  
154 made you decided to change from A to B?  
155 I3: just because I remember we went, we were going over the definition of mean and median ah,  
156 no I am sorry it was today. [Speaking low tone. ] No, I am going to stick to A.  
157 You are going to stick to A. Ok.  
158 I3: yeah, and then the other ones. That's it.

159 That's it. Ok. Do you have any overall questions? Or any comments?  
160 I3: no, just need to study for my statistics final more. [Content laugh]I realized that.  
161 Ok, thank you so much for your time. I am going to turn off the camera.  
162 I3: ok.  
163  
164  
165 End...  
166

## Appendix D4- Interview Transcript I4

1 You may start, go ahead with number 6.  
2 I4: Ok.  
3 [Silence: student reading the question]  
4 I4: I think it's A, because, they already gave us... [Pause] ok, hold on. [Pause] because the  
5 average is 2.2 so then would it be more than 2 in each house? I am guessing.  
6 So, which one do you think is the answer? Which one did you say is the answer?  
7 I4: A.  
8 A, half the household in the town have more than 2 children.  
9 So, why did you think that this is the most reasonable one out of the 4 options?  
10 I4: Just, it says that ah, mmm, the average children per household are 2.2 children so I am  
11 guessing there are more than 2 in each one.  
12 Ok, so that would be A. Your answer will be A for that one? Right?  
13 I4: Yes.  
14 Ok, so can you go ahead and read question 7.  
15 I4: Ok. [Student read the question out loud, Student indicated that it was easier for her to read the  
16 questions out loud]  
17 I4: I guess is B.  
18 You guessing is B. Ah...  
19 I4: Because, those incomes... I think like the overall... ah, I don't know.  
20 Ok, so for example A and B are very similar. What makes you pick B over A?  
21 I4: because I think that ah, the median is 33 ah, 33 and 30 thousand. I don't know why. I am just  
22 guessing that is the median.  
23 Ok.  
24 I4: but I don't know why.  
25 Ok, ah, can you read question 8? You can just read it to yourself and then tell me which one you  
26 think it's the answer.  
27 [Silence, student reading the question]  
28 I4: I guess its B I am guessing, because it does not say ah, like, he is going to award the 10  
29 students but it does not say where the 5 points are gone. Or are they gone to their grade?  
30 Because, there is only 10 so I am guessing it is not going to go that much high because not all of  
31 them got high scores out of a 100.  
32 Ok, so you mean he is just given those 5 points to only 10 students?  
33 I4: Yeah, so not everyone is getting it, so not all the scores go high.  
34 Ok, so you think it will remain equal to?  
35 I4: Yes.  
36 Ok. Ok, next question 9.  
37 [Silence, student reading the question].  
38 I4: I honestly don't know, because I don't know what range is.

39 Ok, so ah, you don't know... ah, you don't remember the definition of range?  
40 I4: I don't.  
41 Ok, so whenever you have a data set, for example like the data set that you have here. [Pointing  
42 at the paper] that they give you like a set of numbers, do you remember the range is the  
43 difference between which number and which number?  
44 I4: The first and the last?  
45 Yes, so that means it will be the difference between the largest and the smallest?  
46 I4: Yeah.  
47 Ok, so based on that which one would be your answer?  
48 I4: [silence, students looked at the question again] 30 D.  
49 Ok. So, this concept range is it something that do you remember seen it in class? Or is it  
50 something that you don't remember?  
51 I4: I don't remember. I remember seeing median, mean and all those but I don't remember range  
52 that why I couldn't recall it.  
53 Ok. Ok, on the next page there will be question 10 and question 10 it will refer to the graph.  
54 I4: Do you want me to read it out loud?  
55 You can just read it to yourself, but if you need to read it out loud you can read it out loud.  
56 [Silence, student reading the question]  
57 I4: I think I am going to go with C, because we did histograms and he [instructor] told us that I  
58 think IQR is the formula if I am, if I am remember it right.  
59 Ok.  
60 I4: And we used it for it.  
61 So, you used that formula for the IQR?  
62 Yeah. So I am guessing it will be C because it is a histogram and I recall.  
63 Ok, the next two questions questions 11 and 12 they will refer to this graph right here[showing  
64 paper to student]  
65 [Silence, student reading question]  
66 I4: I think I am going to go with... ok, I am confused between C and D, but I don't know if it is.  
67 Is there any concept of any word that you don't understand in the question? Anything that is not  
68 clear?  
69 I4: Maybe variability like the variety of it? Or what is the meaning of like?  
70 No, it's how much it varies what is the variation in the data.  
71 I4: Ok.  
72 I4: And then the answer what is outliers?  
73 Outliers mean when you have very extreme observations, when you have for example, when you  
74 have a data set where for example, students in the freshmen year and then most of the people are  
75 between a certain ages and there are some very extreme observations for example you have  
76 people form 18, 20, 19, 25 and then you have people in the 40's or 50's it will be an observation  
77 that is very far apart from the most of the data.

78 I4: Ok, so I will go with B, because the data 140 dollars is far from the data. I guess that makes  
79 more sense.  
80 Ok so B is you agree with the statement it's so easily influenced by outliers.  
81 I4: Yes.  
82 Ok, number 12.  
83 [Silence, students reading]  
84 I4: I think its D as well, because the resistance to outliers refers to that too though.  
85 Which one?  
86 I4: D  
87 D? Ok, the interquartile range because it is resistant to outliers?  
88 I4: Aha.  
89 Ok, ok number 13.  
90 I4: Oh, well. [Silence students reading question]  
91 This one is B. The difference between the highest and lowest values.  
92 Ok, so do you see any relationship between this question and any other questions?  
93 I4: oh yeah. The number 9.  
94 Number 9. Ok, number 14.  
95 [Silence, students reading the question]  
96 I4: Average is, I am guessing mean, B. because usually they use it like what is the average mean  
97 of something, something.  
98 Ok, so you pick this answer because it's more related to...  
99 I4: the class.  
100 Is more related to the statistical meaning of the word.  
101 I4: yes.  
102 Ok, the last one number 15.  
103 I4: I am going to go with C, because, I mean, all the other ones beside probably, probably like F  
104 or something are related to statistics.  
105 Ok, so C right? Spread is defined as range or difference two numbers as a point in spread. Ok,  
106 and you said that you picked C over the other ones because?  
107 I4: because it is more statistical or...statistics related  
108 Statistics related.  
109 I4: I don't think that they would like butter, Jam... I don't think.  
110 Ok, do you have any questions? Any questions that you want to go back and read again? Or any  
111 over all questions?  
112 I4: ah, no. I mean I learned from it because I did not know what range was or stuff like that or  
113 outliers.  
114 Ok. Ok.  
115 I4: Thank you.  
116 I4: is that all?

117 Yeah, if you don't want to go back to read any questions or if you don't have any questions any  
118 comments.  
119 I4: well, no just that.  
120 Then, thank you so much for your time I appreciate it.  
121  
122 End...  
123

## Appendix E- R code

```
setwd("C:\\Users\\lorena\\Documents\\THESIS MATERIALS")
dat=read.csv("survey_data2.csv",header=T,sep=",")
head(dat)
table(dat[,3])
library(xtable)
print(xtable((table(dat$Q1)/length(dat$Q1))*100), type="latex", file="output.tex")
print(xtable(table(dat$Q2)), type="latex", file="output.tex")
hist(dat$Q2)
table(dat$Q3)
table(dat$Q4)
#question 4 is the percentage of time spent speaking spanish the last month.
hist(dat$Q4,breaks=c(0,20,40,60,80,100),col=blues9)
table(dat$Q5)

cor.e=cor(dat[,c(3,4,5)],use="pairwise")
pc.ell=princomp(cor.e,321)
summary(pc.ell)

eigen(cor.e)
#principal components analysis. Combination of q3,q4,q5 to create a new criteria for ell or non-ell
pc.scores=.609*dat[,3]+.628*dat[,4]-.484*dat[,5]
head(pc.scores)
summary(pc.scores)
hist(pc.scores)
dat$scores=pc.scores
table(dat[,3][25<dat$scores&dat$scores<=30])

table(dat[,3],dat[,4])
nonELL1=ifelse(dat$Q3==1&dat$Q5==10,1,0)
nonELL2=ifelse(dat$Q5==10,1,0)
#this is the criteria we are going to use to categorize as ell or non ell
nonELL3=ifelse(dat$scores<=30,"non_ELL","ELL")
table(nonELL3)
nELL=128
non=219
prop.table(table(dat$Q6,nonELL3),margin=2)
```

```
res=prop.table(table(dat$Q6,nonELL3),margin=2)
write.table(res,file = "tempres.csv")
```

```
prop.table(table(dat$Q7,nonELL3),margin=2)
prop.table(table(dat$Q8,nonELL3),margin=2)
prop.table(table(dat$Q9,nonELL3),margin=2)
rprop.table(table(dat$Q10,nonELL3),margin=2)
prop.table(table(dat$Q11,nonELL3),margin=2)
rprop.table(table(dat$Q12,nonELL3),margin=2)
prop.table(table(dat$Q13,nonELL3),margin=2)
prop.table(table(dat$Q14,nonELL3),margin=2)
prop.table(table(dat$Q15,nonELL3),margin=2)
dat$Q6c=ifelse(dat$Q6==2,0,1)
dat$Q7c=ifelse(dat$Q7==2,0,1)
dat$Q8c=ifelse(dat$Q8==2,0,1)
dat$Q9c=ifelse(dat$Q9==4,0,1)
dat$Q10c=ifelse(dat$Q10==3,0,1)
dat$Q11c=ifelse(dat$Q11==2,0,1)
dat$Q12c=ifelse(dat$Q12==4,0,1)
dat$Q13c=ifelse(dat$Q13==2,0,1)
dat$Q14c=ifelse(dat$Q14==2,0,1)
dat$Q15c=ifelse(dat$Q15==3,0,1)
items=dat[,6:15]
head(items)
```

```
#is there an association between ell status and response pattern
```

```
chisq.test(table(nonELL3,dat$Q6))
chisq.test(table(nonELL3,dat$Q7))
chisq.test(table(nonELL3,dat$Q8))
chisq.test(table(nonELL3,dat$Q9))
chisq.test(table(nonELL3,dat$Q10))
chisq.test(table(nonELL3,dat$Q11))
chisq.test(table(nonELL3,dat$Q12))
chisq.test(table(nonELL3,dat$Q13))
chisq.test(table(nonELL3,dat$Q14))
chisq.test(table(nonELL3,dat$Q15))
#difference in proportion correct or not for items 6-15
prop.test(table(nonELL3,dat$Q6c))
prop.test(table(nonELL3,dat$Q7c))
prop.test(table(nonELL3,dat$Q8c))
```

```

prop.test(table(nonELL3,dat$Q9c))
prop.test(table(nonELL3,dat$Q10c))
prop.test(table(nonELL3,dat$Q11c))
prop.test(table(nonELL3,dat$Q12c))
prop.test(table(nonELL3,dat$Q13c))
prop.test(table(nonELL3,dat$Q14c))
write.table(prop.test(table(nonELL3,dat$Q15c)))
#run the p-values from the above analyses through p.adjust
p.vslues.chisq=c()
p.values.prop=c()
p.adjust(p.values.chisq,method="sidak")
p.adjust(p.values.prop,method="sidak")
write.table()
#adjusted p-values
p.values.chisq=c(0.01667,0.1771,0.03581,0.01535,0.348,0.1594,0.02763,1,0.5139,0.0522)
p.values.prop=c(0.003075,0.9079,0.03581,0.01535,0.348,0.1594,0.02763,1,0.5139,0.05222)
data.frame(pval.csq=p.values.chisq,pval.cs.adj=p.adjust(p.values.chisq,method="fdr"),
pval.prop=p.values.prop,pval.prop.adj=p.adjust(p.values.prop,method="fdr"))

```

## Appendix F-Code of Qualitative Responses

Theme	Student #/Page Number/Line in Transcript
Confusion between academic and everyday registers	Student # 1/Page Number:4/Line in Transcript: 104-107 Student # 2/Page Number:1/Line in Transcript: 15 Student # 3/Page Number:2/Line in Transcript: 37 Student # 3/Page Number:5/Line in Transcript: 120-121 Student # 4/Page Number:1/Line in Transcript: 19 Student # 4/Page Number:3/Line in Transcript: 66,68 Student # 4/Page Number:1/Line in Transcript: 4-5 Student # 1/Page Number:1/Line in Transcript: 10-11
Confusion about “intact” statistical phrases	Student # 2/Page Number:2/Line in Transcript: 53-55 Student # 3/Page Number:2/Line in Transcript: 87,89 Student # 4/Page Number:1/Line in Transcript: 4-5 Student # 4/Page Number:1/Line in Transcript: 21-22 Student # 4/Page Number:3/Line in Transcript: 71
Deficiencies in CALP (Cognitive Academic Language Proficiency)	Student # 1/Page Number:3/Line in Transcript: 83,86 Student # 1/Page Number:1/Line in Transcript: 21-23 Student # 1/Page Number:3/Line in Transcript: 79-80 Student # 1/Page Number:4/Line in Transcript: 104-107 Student # 2/Page Number:1/Line in Transcript: 15 Student # 2/Page Number:2/Line in Transcript: 29,37 Student # 2/Page Number:2/Line in Transcript: 61 Student # 2/Page Number:3/Line in Transcript: 88 Student # 3/Page Number:2/Line in Transcript: 37 Student # 3/Page Number:2/Line in Transcript: 40,43,48-49,55 Student # 3/Page Number:3/Line in Transcript: 83-85 Student # 3/Page Number:5/Line in Transcript: 117,120-121 Student # 3/ Page Number:5/ Line in Transcript:123-125,126 Student # 4/Page Number:1/Line in Transcript: 28-30 Student # 4/Page Number:2/Line in Transcript: 37,39,43,50-51 Student # 4/Page Number:3/Line in Transcript: 71 Student # 4/Page Number:1/Line in Transcript: 15-16 Student #4 /Page Number:3/Line in Transcript: 95-96 Student #4 /Page Number:4/Line in Transcript: 102-103,106,108
Role of Context	Student # 1/Page Number:1/Line in Transcript: 11-13 Student # 1/Page Number:5/Line in Transcript: 126-129 Student # 2/Page Number:2/Line in Transcript: 53-55 Student # 3/Page Number:1/Line in Transcript: 9-10,16,18-19

	<p>Student # 3/Page Number:5/Line in Transcript: 125-126</p> <p>Student # 4/Page Number:4/Line in Transcript: 102-103,106</p>
Graphics Confusion	<p>Student # 1/Page Number:1/Line in Transcript: 21-23</p> <p>Student # 1/Page Number:2/Line in Transcript: 53,57,59-60</p> <p>Student # 3/Page Number:5/Line in Transcript: 117</p>
Student Practices/Beliefs	<p>Student # 1/Page Number:1/Line in Transcript: 11-13</p> <p>Student # 1/Page Number:3/Line in Transcript: 86</p> <p>Student # 1/Page Number:5/Line in Transcript: 126-129</p> <p>Student # 3/Page Number:5/Line in Transcript: 120-121,123,125-126</p> <p>Student # 4/Page Number:2/Line in Transcript: 54</p> <p>Student # 4/Page Number:4/Line in Transcript: 102-103,106,108</p>
Statistical Concept Confusion	<p>Student # 1/Page Number:1/Line in Transcript: 21-23</p> <p>Student # 1/Page Number:2/Line in Transcript: 29,32,38-39</p> <p>Student # 1/Page Number:3/Line in Transcript: 83</p> <p>Student # 1/Page Number:4/Line in Transcript: 104-107</p> <p>Student # 2/Page Number:2/Line in Transcript: 53-55</p> <p>Student # 3/Page Number:2/Line in Transcript: 31-32,40,43,48-49,55</p> <p>Student #3 /Page Number:3/Line in Transcript: 72-74</p> <p>Student # 3/Page Number:5/Line in Transcript: 156-157</p> <p>Student # 4/Page Number:1/Line in Transcript: 21-23</p> <p>Student # 4/Page Number:2/Line in Transcript: 37,50-51</p>
Transfer between Academic Registers (stats vs other academic field)	<p>Student # 1/Page Number:2/Line in Transcript: 32,38-39</p> <p>Student # 1/Page Number:4/Line in Transcript: 104-107</p> <p>Student # 2/Page Number:1/Line in Transcript: 15</p> <p>Student # 3/Page Number:1/Line in Transcript: 9-10,14,16,18-19</p> <p>Student #3 /Page Number:2/Line in Transcript: 37</p> <p>Student #3 /Page Number:3/Line in Transcript: 72-74</p> <p>Student # 4/Page Number:1/Line in Transcript: 4-5,10-11</p>
Student Background	<p>Student # 3/Page Number:1/Line in Transcript: 3</p> <p>Student # 4/Page Number:1/Line in Transcript: 15-16</p> <p>Student # 1/Page Number:4/Line in Transcript: 104-107</p>
Multi-modal Representation	<p>Student # 4/Page Number:1/Line in Transcript: 15-16</p> <p>Student # 4/Page Number:2/Line in Transcript: 54</p>
Misinterpretation of concepts due to context	<p>Student # 1/Page Number:1/Line in Transcript: 11-13</p> <p>Student # 1/Page Number:5/Line in Transcript: 126-129</p> <p>Student # 2/Page Number:2/Line in Transcript: 53-55</p> <p>Student #3 /Page Number:1/Line in Transcript: 9-10,14,16,18-19</p> <p>Student # 3/Page Number:2/Line in Transcript: 31-32</p> <p>Student # 3/Page Number:3/Line in Transcript: 83-85</p> <p>Student #4 /Page Number:1/Line in Transcript: 4-5,10-11</p>

Timing or order of questions	Student # 1/Page Number:4/Line in Transcript: 114,116 Student #3 /Page Number:5/Line in Transcript: 137 Student #3 /Page Number:6/Line in Transcript: 151
------------------------------	---

## **Curriculum Vitae**

Lorena Galvan was born in El Paso, Texas. However, she spent the first 23 years of her life living in the neighbor city of Juarez, Mexico. Daughter of Victor Manuel Galvan-Zapata and Maria Rico, she graduated from Collegio de Bachilleres # 6, Cd. Juarez, Mexico in 2004. During the Fall of 2004 she was enrolled as a full-time student in an ELL course at El Paso Community College. One semester after that, she was enrolled as a full-time student taking freshman-level classes towards her degree in Mathematics. She graduated from El Paso Community College in 2007 with an Associates of Art and that same year enrolled at The University of Texas at El Paso to continue with her B.S in mathematics. Education was always a passion of hers, and during her undergraduate years Lorena worked as private tutor for different schools and private companies. She graduated with a B.S in Mathematics and a minor in Secondary Education in 2009. She entered the graduate program in Statistics in 2010. After completing her first year in the program, she took a break from her studies to dedicate time to her newborn daughter and husband and joined the tutoring center at El Paso Community College as a part-time academic Mathematics tutor and a bilingual GED instructor. Lorena worked at El Paso Community College from 2011 to 2014. In 2014 she reenrolled in the Statistics program at The University of Texas at El Paso and received her M.S in Statistics in May of 2016.

Contact Information: [Lorenagalvan72@gmail.com](mailto:Lorenagalvan72@gmail.com)

This thesis was typed by Lorena Galvan.