

TOEFL Test Validity: Student Perceptions of Purpose and Preparation Techniques

J. H. Stringer

2018

#### Abstract:

As the number of international students wishing to attend Western English speaking universities has increased recently, the importance of accurately gauging their ability to be successful at these institutions has also increased. As a result high stakes exams such as Test of English as a Foreign Language by the Educational Testing Institute has become one of the major means of determining these students English skill level and thus giving insight into how successful they will be in an English speaking university setting. The People's Republic of China is one of the main sources of these international students and TOEFL is the main English language proficiency test chosen by these students.

Despite the growing significance of these Chinese students preparing for Western university, they are under studied in the education literature in general and specifically how they prepare for university entrance. This research is specifically concerned with what Chinese students believe about the TOEFL exam and how they prepare for it and what the consequences for their education are and what does this say about the validity of the TOEFL exam.

In order to exam these questions, a case study design was conducted with one group of Chinese students in an international program using an American curriculum. This research was a mixed methods approach combining quantitative methods, examining the relationship between students grade point average and TOEFL scores as well as collecting closed survey questions, and qualitative methods, open ended survey questions and semi structured interviews. The intention was to find out what the students thought about the TOEFL, what they did to prepare and does this vary by ability level.

The results of this survey found evidence to suggest high stakes testing has had some impact on some of these students' beliefs and how they approach university. A strong correlation was found between top ranked and low ranked students grades and TOEFL scores, but mid ranked students had a very limited correlation between TOEFL scores and grades. These students were found to be those who tended to practice testwise strategies which were of limited values

in increasing language proficiency and had more erroneous beliefs about educational expectations. This research suggests that further studies be conducted with this type of students to see how TOEFL and other high stakes language proficiency tests are impacting international education.

## Acknowledgment

I would first like to thank Carol Cruickshank and Stephanie Atkinson from the University of Sunderland for advising me throughout the thesis writing process as well as all the other faculty members at the University of Sunderland who have helped me for the last three years. I would also like to thank Leo McArdle and Shi Yuan for providing the management support that allowed this project to proceed. I would also like to thank Bai Xue and Liu Lixia for providing expertise on the Chinese language, culture and educational system and to Thomas Fittipaldi for helping with survey administration. I would also like to thank all the students and counselors who agreed to participate in the project making this possible. Finally I would like to thank my sister Dorothy for being my academic role model and always believing in me.

## Table of Contents

---

Abstract: 3

Acknowledgment: 5

Chapter 1: Introduction:

1.1: Rationale and Context: 9

1.2 Aims and Significance: 14

1.3 Research Questions: 17

1.4 Research Context: 19

Chapter 2: Literature Review

2.1 Purpose: 21

2.2 Washback Effects: 21

2.3 Student Attitudes and Beliefs about Standardized Tests: 24

2.4 High stakes Testing: 25

2.5 Test taking strategies: 27

2.6.1 TOEFL Test: Development 29

2.6.2 TOEFL Test: Validity Argument by ETS: 30

2.6.3 TOEFL Test: Validation Post Release: 33

2.6.4 TOEFL Test: Validation by Outside Research: 36

2.7 Conclusion: 37

Chapter 3: Methodology

3.1 Introduction: 39

3.2 Replication Studies and Context: 40

3.3 Case Study Methodology: 41

3.4 Quantitative and Qualitative Methods: 43

3.5 Setting and Participants: 44

3.6 Ethical Issues: 47

3.7 Piloting Research Instruments: 48

3.8 Focus Group: 49

3.9 Survey Construction: 52

3.10 Data Validity: 55

3.11 Semi-Structured interviews: 55

3.12 Data Analysis Method: 57

3.13 Conclusion: 58

#### Chapter 4: Data Analysis

4.1 Introduction: 60

4.2 GPA and TOEFL Comparison: 61

4.3 English Education: 67

4.4 Experiences with TOEFL Test: 72

4.5 Students Intentions and Concerns: 74

4.6 Student beliefs about TOEFL: 88

4.7 Student Study Techniques: 92

#### Chapter 5: Conclusion

5.1 Introduction: 102

5.2 Conclusions by Research Questions: 103

5.2.1 What do Chinese students believe the TOEFL iBT measures?: 103

5.2.2 To What Extent do Chinese Students think the TOEFL iBT prepares them for Western University?: 104

5.2.3 How do Students Vary by Ability Level as defined by TOEFL Scores?: 105

5.2.4 What do Chinese students think are the most effective ways to prepare for the TOEFL test?: 107

5.2.5 Do student's preparation techniques vary by ability level?: 108

5.3 Limitations of the Study: 108

5.4 Future Research: 109

Bibliography 111

Appendix A: Focus Group Questions: 119

Appendix B: Student Questionnaire: 122

Appendix C: Counselor Interview Questions: 126

Appendix D: Student Interview Questions: 127

Appendix E: Student Questionnaire Data: 129

Appendix F: Student TOEFL and APGPA Data: 141

Appendix G: Counselor Interview Transcripts: 144

Appendix H: Student Interview Transcripts: 153

## Chapter 1 Introduction

---

### 1.1: Rationale and Context

This study is concerned with the effect of high-stakes testing, specifically concentrating on the Test of English as a Foreign Language internet Based Test, here after referred to as TOEFL iBT. Specifically the study is looking at what effect the TOEFL iBT has on teaching practice and student beliefs in an international setting. There has been a dramatic increase in the number of international students, with students from the People's Republic of China making up a large and growing proportion, attending university in English speaking countries with the US being a particularly attractive choice (Nelson, 2013). Approximately 270,000 Chinese students attended university just in America; many more also went to Canada, Australia and the UK (Rahala 2015, p. 35).

University admissions have numerous requirements including high school transcripts; letter of recommendation and personal essays but an enormous amount of weight is given to the standardized tests, specifically the English as a Second Language tests both by the universities (Abel, 2002, Douglas, 1990) and by the student themselves (Ginther and Elder, December 2014). Chinese students in particular are acculturated in an environment where standardized test scores define success in life and many of those who apply to Western University carry that belief with them, despite the change in context (Wang, 2014). This study will attempt to shed light on what, if any affects, the importance of tests like the TOEFL have on teaching and the student's, whose future might depend on them, education. This study is looking into test validity, students belief's about education and the wash back effect that standardized tests have on teaching.

The TOEFL test has come under criticisms in the past with many arguing that the test had a serious validity problem. The specific issue at stake here is construct validity, that is, does the test measure what it purports to. Does the test measure something 'real' or does it measure your ability to take the test (Cohen, et. al. 2005, p. 110). Does the TOEFL give you an indication of the test takers actual English ability or does it give you an indication of how well they studied for the

test? This has been a response to past iteration's of the test. For example Menke (1998) found no correlation between student GPA and TOEFL scores and similarly Simmer (1998) summarizes numerous studies which failed to find any significant correlation between the test and the success at university. These criticisms lead the Canadian Psychological Association to issue a statement urging universities to reconsider using the TOEFL or to put less weight on the exam. The result was the development of the new TOEFL iBT created to ensure that the test would conceptualize English proficiency "more broadly as the ability to use complex knowledge and processes to achieve particular goals rather than narrowly as knowledge of linguistic forms or a skill" (Chapelle, C. et. al. 3008, p. 2). That is, the test needs to test the ability to use language even though that is more difficult to test.

When looking into the validity of the TOEFL iBT, there is a robust collection of material published by Educational Testing Services (ETS), the company which is responsible for producing the test. It was very specifically developed to respond to the criticism leveled against its previous iteration and ETS goes into great depth showing how the test was developed in order to deal with these specific concerns. Taylor and Angelis (2008) go through an iteration by iteration examination of the TOEFL since its inception in the 1960s showing how they responded to criticism each time. They specifically bring up the development of communicative competence through the 1980s and 1990s as the impetus for the development of the new test. Communicative competence being understanding of language proficiency as the interaction of knowledge context and other factor's to show whether you can communicate or not. More detail will be provided in the literature review, but the company shows how universities should use the New TOEFL test to make up for a perceived lack of understanding of the test by admissions officers and administrators (Chapelle, et. al., 2008). They showed the process by which they created the framework for the test with reference to the goals of the test and how they would be achieved (Jamieson, J., et. al. 2008). They detail the process by which they prototyped the assessment tasks with construct validity and communicative theory firmly in mind and how that leads them to further develop the test (Enrightm, et. al., 2008, Huff, 2008). Pearlman (2008) then

shows how the test was finalized with specific reference to the goals of the test in response to the criticisms of the test and Wang (2008) did a substantial statistical analysis of the work done to date to see if their underlying assumptions sound and if the goals of the project had been met. Assessing validity is an ongoing process so additional validation studies have been done by ETS looking into reliability across tests (Chen, June 2015) as well as comparing the TOEFL to other standardized English as a Second Language tests (Ginther and Elder, December 2014). Most importantly for the purposes of this study Malone and Montee (December 2014) looked into stakeholder's beliefs about the use of the test. Stakeholder's in this case being University administrators, instructors and students. This current study is largely inspired by this work as it lead to the consideration of students actively studying for the exam rather than those who have already completed the admissions process and were in university. That is, what do students who are actively preparing for or have very recently taken the test think as opposed to students who have already started university?

Outside of work produced by ETS one can find a variety of studies supporting the validity of the TOEFL iBT. It should be encouraging that independent sources find evidence to support the validity of the test. More detail will be provided in the literature review but Deluca, et. al. (2012) examined its construct validity to determine if additional factors could have an influence on the exam and left with an overall positive impression of its validity. Xin (2008) study showed that the exam assessed skills in a more naturalistic fashion than previous iterations. Gebril and Plakans (2014) analyzed one of the major changes to the examination, the use of tasks that required the integration of multiple skills, and found they were well constructed though there were some reliability issues in their scoring. Similarly, Knoch and Sitajalabhorn (March 2013) were positive about the integrated tasks with their main criticism being that there should be more of them. Slomp (February 2012) shows that the integrated tasks use the skills necessary for a university composition class.

So if there is a wealth of evidence as to the validity of the exam, why is this study necessary? Looking to Golden's 2011 article in Bloomberg Business Review provides a hint. In it

Golden is at a college fair for perspective students and he describes an encounter with a Chinese student who was planning to go to university in America and had been accepted to a number of well-respected institutions. This student gave his TOEFL and SAT scores as 2040 out of 2400 and 90 out of 120 on his TOEFL. Golden asked the student “Do you feel prepared to go to a university in America?” to which the student needed to ask a translator what Golden had said. If you ask someone who has worked with international student’s, especially from China, if this ring’s true or not, you are very likely to get a yes. ETS’ own data shows that university instructor’s don’t believe that the exam accurately measures the ability of student’s to perform in English (Malone and Montee, December 2014) while at the same believing that the tasks in the test accurately reflect the kind of (Enright, et. al., 2008) a statement supported by studies done by Cumming et. al. (2004, 2005). So how can a test accurately reflect the kinds of skills necessary for academic success in an English speaking academic setting while also not reflecting the student’s ability to succeed in an English speaking academic setting?

One possible answer to this gap in understanding is the existence and practice of standardized ESL test training centers. New Oriental Education is one of the largest of these centers and it is a massive presence all across cities as well as the airwaves and the internet in the People’s Republic of China. The company offers language training centers as well as online education and has a presence in primary and secondary schools. In one quarter of 2016 New Oriental earned more than 534.1 million US dollars (Marketline, 2016). This is just one of many companies offering these kinds of services so it is impossible to estimate the true scope of the industry itself. Test preparation services are not unusual in and of themselves and familiarizing oneself with an exam format and concentrating on the most likely content for a test is common sense.

However, according to Golden (2011) at New Oriental and similar companies the main focus of the training is on identifying testwise strategies, that is strategies for improving your score by taking advantage of test construction rather than becoming more familiar with the content. That is, by tackling the construct irrelevant aspects of the test. For example in order to pass the writing

portion of a test you might look at past examples of writing prompts and memorize canned answers rather than responding to the prompt with a genuine thought at that time. This sounds like it would be more difficult than just studying for the test but the techniques of these kinds of schools give their students the tools to take advantage of the test construction. For example the students might memorize an essay with the prompt 'who is someone you admire?' The essay could discuss Steve Jobs and his admirable characteristics. With a few modifications this canned essay could be used for any number of prompts. What role does technology have in the modern world? Describe someone who has had a major impact on the world. How have things changed since your parents' youth? Similarly, through statistical analysis you can predict the kinds of questions you are going to be asked on a speaking test. You then make a few modifications based on the questions you get but essentially answer with the memorized part. For example you might memorize a speech about your favorite place, a park in this case. If you get a question asking you to describe your favorite restaurant all you have to do is say your favorite restaurant is next to a park and then complete your memorized speech.

## 1.2 Aims and Significance

New Oriental Claims to be able to dramatically increase TOEFL and SAT scores with these testwise strategies. While academic research is currently scarce looking at Golden (2011), and their popularity (Marketline, 2016) suggests this could be true and suggests the need for further research. More research is needed in order to discern the specific impact these kinds of strategies and that will be one of the primary focuses of this study. ETS has released a study detailing study strategies and their effect on the TOEFL and SAT, specifically looking at students at New Oriental Education, but they specifically exclude the testwise strategies from their analysis (Liu, O. December 2014). This study will address this absence of knowledge for two specific reasons. The first is the issue of whether the significance of these high stakes test impacts teaching practice and whether student's attitudes towards education are affected by them potentially making life in an English speaking academic context more difficult.

Washback is commonly described as the effects of testing on teaching practice wither positive or negative (Barnes, November 2017). A positive effect could be that it encourages the teacher to fully cover all aspects of a curriculum, a negative effect could be an excess of focus on the specific material on the test at the expense other aspects of the subject as a whole. This study is primarily interested in the negative washback that might be associated with the TOEFL iBT, but will highlight any positives to be found. Highstakes testing has come under a great deal of criticism for this exact reason with teachers in America (Popham, November 2001) and the UK (Patton, July 2008) decrying the need to 'teach to the test'. In the US there has been a gradual increase in the significance of high stakes testing with the goal of increasing the efficiency, reliability, objectivity and accountability of educators. Many have argued that this demand for a reliable and objective assessment has led to a loss of a holistic method of assessing the whole student and narrowing the scope of teaching to just the material found on the test (Wall, D. 2000). By concentrating on the analyzable data from a standardized test you lose the individuality of the student and lose a wealth of educational possibilities of the classroom. There are numerous studies showing the negative washback effects on teaching practice which will be detailed in the literature review.

In the international education system the focus of the problem is less about subjects and more concentrated on the ESL exams like TOEFL. East Asian countries, China in particular have had an exam driven culture for most of its educational history and the holistic nature of the Western educational system is not well understood by stakeholders including students, parents and instructors (Rauhaui, April 2015). As a result, ESL teaching practice is often focused on improving scores over actual ability to communicate. Huang and Brown (2009) describe how test taking dominates teaching practice in all subjects including English in the People's Republic of China while Barnes (2016) shows an how the TOEFL test dominates English classes in Vietnam. Washback clearly has a major effect on teaching practice in international settings and this study will examine the extent to which it influences practice.

Students themselves and how they think of themselves as learners is also impacted by high

stakes testing. The importance that is put on to test results has the effect of shaping students self-perception and how they approach education. To make the most common sense claim first, when students are presented with a high stakes test they will prioritize that test and study for it in ways that they perceive to be most likely to lead to higher scores (Zhan and Wan, 2014). While this might seem like an obvious point, it opens up a window into understanding why students would choose to go to a training center like New Oriental rather than concentrating on improving their English skills overall. ETS makes the argument that the best way to study for the TOEFL test is through the improvement and use of English skills (Liu, December 2014) but as we have seen from Golden (2011) that is not the strategy Chinese students' use when preparing for the exam. Beyond that, high stakes testing affects how students see themselves. The number that they get influences whether they see themselves as a 'smart person' or as a 'good student' (Reay, and Wiliam, 1999). And finally that number affects how they will perceive what education is (Xie, May 2015). The risk here is that the students will go from this context where examination scores are the only thing that is perceived to be important and take it to a new context English speaking university context where it is wholly inappropriate. Students are preparing for an assessment that is unlike anything that they will be assessed with in the University context. They may never take another test like the TOEFL or SAT again so we can benefit from a greater understanding of the extent to which they are being influenced by these exams. Exploring these issues leads to the following research questions.

### 1.3 Research Questions

#### 1) What do Chinese High School students believe the TOEFL iBT measures?

This question goes to the issue of student's conception of the tests purpose. The study will examine whether or not student's beliefs about the test impact their sense of self, how they prepare for the test and what does this suggest about their attitudes towards education as a

whole. The underlying question is whether their beliefs about the test affect their success in the future.

- 2) To what extent do Chinese students think the TOEFL iBT prepares them for the Western universities?

Considering how much importance is put onto this high stakes test does this affect their perception of what education in a Western university context is like. If they think that the way that they prepare for an exam like the TOEFL is similar to how they should prepare for university courses they may encounter difficulties in the future. Is the negative washback from the test going to impact their education in the future? On the other hand the students might realize that the university admissions process and the standardized tests necessary for it are a very specific context and not reflective of what university will actually be like. In this case the washback effect might be minor or non-existent.

- 3) How do students vary by ability level?

Perhaps students who are of lower achievement are attracted to training schools out of a desire for a quick fix to their lower achievement. High achieving students could be taking advantage of additional educational services as a reflection of their greater motivation.

- 4) What do Chinese students think are the most effective ways to prepare for the TOEFL exam?

What are the most popular study techniques? Are testwise strategies the most popular or is that an assumption. To what extent are students using different kinds of study techniques for knowledge acquisition versus preparing for the test.

5) Do student's preparation techniques vary by ability level?

How are students preparing for the test and what is the reasoning behind that choice. Are they taking the advice of training schools like New Oriental or are they taking the methods that ETS says are the most effective for preparing for the exam.

#### 1.4 Research Context

In order to answer these questions a mixed methods approach will be taken involving both quantitative and qualitative data taken from multiple sources. This study will take place in a High School in a major city in the People's Republic of China. The targeted groups of students, 90 in total, are in their last or next to last year of study before university ranging in age from 16 to 18. The students study an American Advance placement curriculum as part of an international center in a public high school with the goal of attending university in the US, UK or other Western country. The students study for the TOEFL and SAT in their own time with most students attending additional courses after school and on weekends. The curriculum includes AP, pre-AP and non AP courses in maths, natural sciences, social sciences, humanities and academic English. This is a for profit education enterprise with high tuitions paid by the parents suggesting they are relatively wealthy. They intend to apply to Universities in the West and they are culturally and linguistically similar all being citizens of the People's Republic of China.

In order to answer these research questions through the use of three data collection tools looking at both qualitative and quantitative data. This is a strategy of triangulation to improve the validity of the study (Cohen, et. al., 2000 p. 113) as well as an attempt to access different levels of questioning to more deeply analyze the situation (Flick, 2009, p.445). Students will be randomly selected and asked to attend a focus group to discuss the research questions. From this data a questionnaire will be developed to get a larger survey of the students. Finally students will randomly be selected a requested to participate in an interview to get to a deeper understanding of the issue at hand.

In order to establish the need for this research it is necessary to examine the literature in question, specifically concerning the issues of standardized ESL tests, washback effects, high stakes tests and students beliefs, test taking techniques and how the TOEFL itself specifically works. Looking at what knowledge is available will show the significant gaps in our understanding and how this study can begin to fill those in.

### 2.1 Purpose

This literature review's purpose is to give a good summation on the current state of understanding of the arguments surrounding high stakes standardized testing in general, but specifically focusing on the TOEFL iBT's place in this argument. Through this exploration of the literature, some of the gaps in research will be explored showing the benefit of this study. The first step will be to look at the criticisms of high stakes testing both from the perspective of how it affects teaching practice and how it affects students and their sense of self. The effect of the local education culture and attitudes towards testing also need to be taken into account. The literature produced by ETS, both during the development of the test and after it was implemented, will be extensively examined in order to look into how they argue for the tests validity while also looking at validation studies done by independent researchers. Finally, the studies on ESL testing validation and practice with students who resemble the subjects of this study will be examined to show the necessity of additional research.

### 2.2 Washback Effects

The study of the effects of standardized testing on teaching, also called washback, goes hand in hand with the gradual increase in the importance of those tests. That is as these tests became 'high stakes' they had a greater and greater increase on teaching by design. Messick (May 1996) describes the concept of washback as the effect a test has on teaching practice either positive or negative. That is, would you have done this if the test did not exist? Positive washback could be an increase in focus on the part of instructors and negative could be a loss in the authenticity of the teaching. That is, teaching to the test as opposed to teaching the subject (Ibid., p. 4-5). Huddleston and Rockwell (2015) provide a good survey of the development of standardized tests and what their specific goals were. That is to make teaching more efficient, objective and

accountable. The work shows how standardized tests immediately impacted teaching practice, for example recitation of reading passages was once the most common form of reading teaching practice, but it was harder to test and seen as less objective leading to the dominance of silent reading (Ibid., p. 41). They detail throughout the work how a drive for objectivity gradually increased the importance of these tests leading to the practice of teaching to the test. For example, letters of recommendation were once seen as the most important tool for advancement at university or in the workplace, but the drive towards efficiency and objectivity since the beginning of the twentieth century drove the push towards standardized testing, a fact which was noted and criticized by some at the time (Ibid., p. 43). The criticisms of this drive are detailed in the study and include the idea that they isolate skills outside of the context of use, don't take into account student's individuality and don't reflect real life. Thus by adjusting teaching practice to the test they are limiting the student's education (Ibid, p.46). ETS provides a history of the development of the TOEFL from its inception in the 1960s describing its changes and development of the exam in response to these kinds of criticisms, for example developing integrated writing and reading tasks that might more accurately reflect real life situations (Taylor, and Angelis, 2008).

Many studies (Messick, May 1996; Wall, 2000; Kunnan, and Wu, 2010; Lee, and Coniam 2013; Pan, 2016) have been done detailing specific examples of wash back in teaching practice. Wall (2000) examines a variety of issues that can be generally described as wash back. Will teachers look at the tests "idiosyncrasies" in order to improve students' performance, i.e. commonly asked questions or repeated examples, rather than teach an overall understanding of the material? Teachers may lose the freedom to teach their classes in a way they feel is most effective for their classroom context and overvaluing the skills which are easiest to test rather than the most important ones (Ibid, p.500). Wall argues that through careful test design with the input of stakeholder's, namely instructors, with regular evaluation and innovation of the design can lead to positive washback, that is a positive impact on teaching practice rather than a

negative one (p. 507-8). Kunnan and Wu (2010) did an examination of the development of high stakes testing in Taiwan and the washback effects resulting. They found that due to a philosophy of taking stakeholders, teachers and students, experiences into account and including training in teaching methodology which would allow for the test to be used as a tool rather than a threat lead to positive washback in the classroom and positive feedback from instructors and students. Lee and Coniam (2013) found similar results in a case study of the introduction of an ESL teacher qualification exam in Hong Kong. By providing training and a rationale for decisions the test could be seen as a positive and lead to improvements in practice. On the other hand, Pan (2016) looks into negative washback in Taiwan. With the high stakes nature of testing in Taiwan the pressure to improve scores leads students to take any advantage they can resorting to testwise strategies to improve scores. It was found, in particular with lower achievers, that they were less interested in improving English proficiency the way commonly prescribed. That is they were not motivated to learn communicative English but rather wanted to stick to test oriented instruction (Ibid, p. 177-178). As a result, teaching practice bent to the desires of the students and the benefits higher results for teachers leading to teaching practice focusing on test preparation consisting of rote memorization and repeatedly doing practice tests (p. 180). That is, the kinds of techniques that are practiced by training centers like New Oriental. One of the major foci of this research project will be on student's expectations of teaching as a result of attending training schools like New Oriental so we need to establish the potential negative effects of washback.

### 2.3 Student Attitudes and Beliefs about Standardized Tests

The impact of high stakes testing on students, their attitudes, beliefs and behaviors is as great if not more so than teachers. As Lee and Coniam (2013) show, standardized tests affect students access to university, graduate school, professions promotions and more. The stakes are indeed high. Reay and Wiliam (1999) shows how high stakes testing literally comes to define the students in their minds. As his study shows the students came to think that the test showed, not their academic achievement level or the areas they needed to focus on in order to improve, but

rather it showed them that they were a kind of a person. Are you a good student or a bad student? Are you going to be a success or a failure (Ibid, p.352)? In a study of Chinese ESL students attending a university study abroad program in New Zealand, Li (2017) found that students who had rewarding experiences in an English speaking environment were more successful overall and were able to maintain a positive language learning identity. Through the collection of diaries and individual interviews, Li found that the students who had positive experiences described themselves in positive and proactive terms while those who did not lost motivation and described themselves as failures (Ibid, p. 120). This goes to show that positive experience and negative experiences shape students sense of self-worth and affect their motivation. This research is project is looking at how students beliefs are shaped by the TOEFL iBT, thus it is necessary to show the previous research that has been done on this subject.

#### 2.4 High stakes Testing

Student motivation is of primary importance and student motivation is affect significantly by high stakes testing. As Oroujilou and Vahedi (2011) show a positive attitude towards language learning is extremely important for becoming successful at language learning. Xu and Gao (2014) did a statistical analysis of learner motivation amongst ESL students attending Chinese universities. They found a correlation between student's who were primarily motivated by immediate achievement, high test scores, were those who were less likely to actively participate in class and communicate in English while those who had long term goals like developing fluency had greater self-confidence and more perceived progress (Ibid, p. 109-110). Jin (2011) in an examination of the College English Test, the university exam at public universities shows the importance of the perception of fairness from the test and what affect that had on student motivation.

Cheng and Deluca (2011) did a study of ESL test takers experiences, collecting qualitative data concerning the test takers reflections on the experience. They found that the students who questioned the fairness of the exam were most likely to be motivated by short term results and

to use testwise strategies to study for the exams (Ibid, 117). This shows how high stakes testing can influence student's beliefs and behaviors and shows that you need to take stakeholders beliefs into account when determining the validity of a particular assessment. If you do not than the construct irrelevancies in the exam can invalidate it. Another analysis of motivation of Chinese test takers by Hong and Peng (2008) found similar results. Students who valued the test and though it was fair were more motivated to study though they did not find a direct impact on test results. He and Shi (2008) examined an ESL test used by Canadian universities and found that students did not feel it could fairly test their English abilities so resorted to testwise strategies. For example, many of the writing prompts related to topics like Canadian politics which the students could not reasonably be expected to form an opinion about (Ibid, p. 140). Also because they had been accepted into university by passing standardized tests with techniques like memorizing writing passages they were unable to adjust to a course where they were expected to form opinions and follow standard writing techniques (Ibid, p. 137). This suggests that the effects of high stakes testing on student's behavior can have a negative impact on future performance. Once you learn to be successful through testwise strategies it will be more difficult to adjust to a new system. Wang (2004) relates one revealing anecdote in a case study of High Stakes test taking in China. He describes a group of students all taking an English writing test and having most of the class return an almost identical essay (Ibid, p. 4). While the case was declared to be a national scandal, it was apparently quite common. All he had to do was guess what the question on the exam was going to be and prepare his students for it. If the test is constructed in way where you can guess the questions it is certainly going to affect student preparation. The TOEFL iBT is the subject of this research project and it fits the definition of a high stakes test. As such the research on high stakes testing is relevant to this project.

## 2.5 Test taking strategies

Looking at different kinds of test taking strategies there are those which improve general knowledge and those that are specifically related to the construction of the test. Huang (2016)

describes these in detail and shows the importance for communication of which strategy students choose to take. By statistically analyzing data collected from a number of Taiwanese ESL students looking for associations in test preparation methods and behaviors he found a notable association in the data. Communicative strategies didn't positively associate with test results and testing strategies didn't positively associate with communication ability. Fox and Cheng's (2015) study of student experiences found similar results concerning test preparation. The pressure to get good scores motivated the students to primarily study the test rather than study for the test. Practice tests and training center courses were the most common form of test preparation and noted as a source of anxiety by most students (Ibid, p. 74). It is also important to examine the literature on the cultural, educational context in which the takes place, specifically on how East Asian educational culture, with the People's Republic of China as an example, being exam driven affect language learning. Cheng and Curtis (2010) also detail how exam driven educational culture has continued throughout the modern era of Chinese history leading up to today. As Huang and Brown (2009) describe, the largest number of ESL student in North America as from the PRC. Chinese culture emphasizes education and hard work are emphasized, but so is the passive acceptance of knowledge and an intensely exam driven culture. The exam is the only thing of importance for determining future success so all teaching practice and student behavior is directed to studying for it (Ibid, 645-646). When these students head to university in America or other western countries they find many aspects of learning difficult to adjust to including participating in class discussion, group work and lecture organization. This suggests that students from the PRC are especially likely to practice the kinds of testwise strategies which lead to the results they are expecting but harm the students' abilities to be successful in University itself. Zhan and Wan (2014) studied motivation in Chinese students in the Chinese university system through the use of a diary system. The results showed that students were primarily motivated by test results and practiced preparation strategies which they perceived to maximize their success even though they could articulate the idea that it did not actually improve their real ability to speak. One student even explicitly described how the tests decreased his motivation to actually

learn English (Ibid, p. 830). Another study of students in a similar context found that communicative language proficiency actually could harm the students' scores leading to stick to the most basic of material (Xie, 2015). The students would lose points for making mistakes while not gain anything for using more complex vocabulary, grammar etc. (Ibid, p. 30). This shows how test construction is both a product of culture and reinforces culture at the same time. In response to the demands of high stakes testing for Chinese students who are planning to attend domestic and international schools have grown in number. Cheng and Curtis (2010) detail how the testing schools have been able to improve scores through the use of testwise preparation strategies. As more students from China take the TOEFL and attend these kinds of training centers more and more questions about the validity of the exam are sure to arise. As student test taking strategies are the subject of this research project, it is clear that the literature on the subject is relevant to the project.

#### 2.6.1 TOEFL Test: Development

In order to understand the TOEFL's place in all this one needs to examine the process by which it was created. According to Chapelle et. al. (2008) the new TOEFL iBT was to be created from a language proficiency standpoint, that is, the goal was not merely to test language knowledge, but rather to be able to determine if could use language in a context which would be encountered in the real world outside of the test (Ibid, p. 2). In order to validate the exam an argument based approach would be taken in which a chain of reasoning is created. You start with the grounds, for example, if a student repeats the same phrases over and over again in a writing assignment they are probably unable to use the breadth of language for the task, with this being referred to as the warrant. The warrant needs backing, in this case teachers training and experience and a variety of studies on academic writing and ESL students. They then try to anticipate any rebuttals; maybe the writing assignment required highly technical vocabulary leading to the claim. That claim could be, this student needs a remedial writing class or even that

they are not capable of attending university (Ibid, p. 6-7). The research for this type of argumentative reasoning comes from Mislevy et. al. (2003). For validation and to ensure reliability the test required dual ground for validation, that is the grounds for the test task and student performance need the same warrants and rebuttals and must be able to with stand multiple kinds of inferences yielding the same results (Chapelle, et. al, 2008). So if a student has a performance like the example above it must follow that the student will receive 'X' score on the task, that same 'X' score on similar exams and it must be true that the student will have trouble in real life contexts. If these criteria are met than it is acceptable to use test scores in admissions or other important situations. The backing for the myriad of warrants used by ETS can be found in Chapelle, et. al. (2008).

#### 2.6.2 TOEFL Test: Validity Argument by ETS

The TOEFL validity argument can be found and summarized in Breining, et. al. (2011). They describe 6 propositions and provide evidence for each of them. The first is that the content of the test is related to and is exemplary of the activities as well as the written and spoken material that will be encountered at university. Their evidence comes from the examination of surveys of the use of English at universities. Secondly, the test and scores provide a real insight into English language ability with evidence being provided by pilot studies of tasks and the test as a whole. The test is supposed to show a student's ability to use academic language, the evidence for which is found through the discourse analysis of student written and oral responses. Additionally studies of student response to reading comprehension looked to whether students were using test wise strategies and found that they were relying on knowledge and understanding as figured by the test designers (Cohen and Utall, 2006). Success on the test is related to other gauges of English ability the evidence being students own perception, where students find themselves placed in college ESL classes and success at activities designed to mimic academic skills. Finally ETS argues that the test has positive effects on teaching and student behavior, also called positive wash back. These arguments will be further illustrated through the

examination of ETS' published material from both before and after the development of the test as presented below.

In Jamieson, et. al. (2008) a framework for creating the TOEFL iBT using a model of communicative competence is laid out. The goal being to create a test where the numbers resulting from the test could show the degree to which the student was genuinely competent in producing the English language. The argument for the validity of the work is based on Bachman and Palmer (1996) and their argument for the communicative competence. The new tests tasks would include new integrated tasks, i.e. writing something based of a lecture the students listened to, which could be argued to be more representative of real world and academic English use leading to greater construct validity. They would be testing real ability more closely than discrete tasks as you are more likely to be using multiple language skills in an academic setting than you would on a simple test (Guthrie and Mosenthal, 1987, Hale, et. al. 1996, Hansen, 1994). Important assumptions underlying warrants include the idea that language skills used in academic settings can be simulated and that the abilities and process underlying language construction can vary across contexts and explain a test score. Additionally they propose that these scores can be reasonably be interpreted by those who use them, admissions officers for example, to make important decisions (Ibid, p 59-60).

The process of developing evidence backing these warrants came first in looking at individual task development by Enright, M. (2008) followed by the development of each skill section (Enright, et al., 2008) and finally through the development of the prototype test (Huff, et al., 2008). The tasks were first developed and tested through expert evaluation, pilot testing and rubric development (Ibid, p. 142). There was a strong correlation found between the four language skills, reading, writing, speaking and listening. If all four language skills correlate with each other that suggests language competence (Messick, 1993). When combined with current language acquisition theory, (Bachman, and Palmer, 1989) and comparison to similar results on other types of language, as argues by Miller and Linn (2000) test it is suggested that there is validity to the exam. The outcome of the prototype test was also positive with task performance

being ranked positively in comparison to similar exams (Ibid, p. 222).

In final analysis ETS did a field test related to a number of issues with the test, several of which relate to issues of validity as identified in this review (Wang, L. et. al 2008). Specifically the study asks the explanation questions of whether or not the test is consistent with its theory of language proficiency, does performance vary by level of English proficiency. It also looks at extrapolation questions looking at how students rate themselves and how it will affect university placement. Additionally it looks at utilization questions, that is the score interpretable by admissions officers and will the test have positive wash back effects (Ibid, p. 261). Participants were recruited from more than 31 countries and were given either examples of the previous versions of the test or the new versions of the test. The participants were voluntary and self-selected sample, but they were approximately representative of the proportion of TOEFL test takers from their country. Age and education level were not provided, but educational aspiration was (Ibid, p. 270). No explanation for why this information was omitted is presented in the study. A background survey was created in order to correlate information in performance to the background of the student. In order to determine whether there was a relationship between test results and years of English study. After using a number of statistical analytic tools a statistically significant association was found between years of English study after 5 years of study though they not that the variance in test scores can only be partially explained by this measure. They suggest other factors such as motivation and study technique as being influential as well (Ibid, p. 293). For outcomes relevant to extrapolation two factors were analyzed. The first was by looking at placement in ESL programs. If a student was placed in an ESL class it was presumed that they needed additional ESL training which should correlate with their TOEFL score. Secondly they looked at self-reported English ability. They found a strong relationship between academic placement and TOEFL score and a moderate one with self-reported ability (Ibid, p. 297). According to Wang (April 2018) it is impossible to explore the utilization outcome until after the test has become operational but ETS will provide studies to link scores from the new test to previous versions and to fund studies in the future to see the impact that the test has on teaching

practice and student behavior. This test went live in 2009 and the studies done by ETS in its wake will be detailed in the next section. Chapelle (2008) summarizes the TOEFL argument for validity by showing how the evidence presented throughout ETS published materials provides support for the inferences that the Test is representative of an academic environment and its scores can be used to make inferences about a student's ability to be successful in said environment.

### 2.6.3 TOEFL Test: Validation Post Release

After the introduction of the TOEFL iBT ETS continued to research on test validity publishing numerous research reports since. In an effort to extend the validity argument Ling, et. al. (2014) did a study to determine whether or not the improvement in TOEFL scores found amongst a group of students actually reflected an improvement in English language ability. The study attempts to show how TOEFL scores are representative of improved proficiency at a series of English language learning programs in the United States. The study also looked into whether score improvements varied by background including national origin, specifically because they point to the fact that there is little research conducted on TOEFL, or other high-stake ESL standardized tests, and English language acquisition in Asia despite the large number of test takers (Ibid, p. 3-4). This fact will be further elucidated in the final section of this literature review. One school was a high school based in China taking a standard public school mandated curriculum as well as intensive TOEFL courses while the other was a US based intensive English course for international students. Students took a TOEFL iBT test at the beginning of the year and one at the end while also filling out a survey about the experience. Ling points out that while the test was not given in its normal context that is with no bearing on high-stakes decision making, it was seen as a proxy for ability improvement. The main take away is that even without taking intensive TOEFL courses you can improve your TOEFL score and thus TOEFL scores reflect improvements in English. However the study shows that there are no significant differences between the results from the English immersion program versus the intensive TOEFL program (p. 14). This ignores the assumption that TOEFL scores are a reflection of improvements in ESL ability

when it could be in one context but not in another.

Cohen and Upton (April 2006) specifically look at the effect of test preparation strategies on TOEFL scores. The belief being that if it turned out that test wise strategies were useful in significantly improving TOEFL scores than the tests validity would be called into question and if the reverse happened than the validity argument would be strengthened. In order to accomplish this goal test takers were recruited and they would then self-report what strategies they were using as they were taking the test based off of research showing that students would generally provide useful information into their cognitive process while they are working (Ibid, p.22-23). The argument being that because the students were not in a high-stakes situation while being familiar with the situation, and had been adequately prepped about the kind of data they were to provide the researchers they could honestly answer about their cognitive process. It could be argued that because it was not a high-stakes situation it would not be reflective of how students would think in this situation. After careful statistical analysis of the data collected the researches came to the following conclusions. The students used the same skills you would use in an academic language context, suggesting the test does measure what it purports to measure, but at the same time the students approached the test as a test taking task, looking at how to maximize scores rather than understand the text (Ibid, p. 118). In other words test wise strategies were a significant focus of the test takers, even in a situation that was not high-stakes. This suggests that test takers attitudes must be looked into further. ETS' work in this area will be examined next.

Stickler and Attali (January 2010) examined test taker attitudes towards the exam with the intention of filling in gaps in understanding of the exam as little work has been done on the current iteration of the test. The goal of the investigation was to discover if attitudes to the test had any relationship with the results of the exam and to see if there was any difference in this across cultural and national boundaries. Students from four countries, including China, were given questionnaires to determine what their attitudes towards the test were. The study found a slightly positive correlation between attitude and results and the mixed findings as to whether

there was variation by nationality suggesting the need for more study (Ibid, p. 15). Malone and Montee (December 2014) performed a similar study expanding stakeholders out to administrators and ESL teachers. They examined university students from South Korea, Germany and various Middle Eastern countries. They started with a focus group and developed a questionnaire from which they randomly set to ESL students from these countries and teachers of ESL at university. This was followed up with an exercise where students were given a TOEFL exam and were interviewed about their experience immediately afterwards. The results found that there was both dissatisfaction with the test and a belief that the test measured language ability. Students and instructors thought that the integrated writing tasks reflected real English proficiency providing strong evidence for validity. Finally the study showed that instructors and students found that tasks on the test were indicative of English proficiency while not thinking the test reflected projected performance. They suggest that more research is needed to understand this discrepancy.

#### 2.6.4 TOEFL Test: Validation by Outside Research

Besides ETS, a number of other researchers have looked at whether the TOEFL test has validity or not. These studies have had mostly positive results for the test. Deluca et. al (2013) did a study of a number of test takers to see if it had construct validity. They recruited people who were studying language acquisition and learning and had them register for the test. The participants thought that the test was successful for the most part in establishing construct validity supporting the validity of the test (Ibid, p. 673). They acknowledge that the test takers were not representative of test takers and the context was significantly different but claim it still added to the validity argument since they were bringing a kind of expertise to the exam. Similarly Gebril and Plakans (2014), writing instructors, found the integrated tasks to be good examples of academic English work. They surveyed university writing instructors, showing them examples of the TOEFL integrated tasks and asked them whether they were representative of academic English. The conclusion was that they were. Similarly Slomp (2012) through surveying the

literature on writing assessment, found that the integrated writing tasks followed sound test design. Knoch and Sitajalbhorn (2013) looked at the underlying assumptions related to integrated writing tasks, as found in the TOEFL, and take an argumentative based approach to suggest their validity. Xin (2008) looked at the framework for the new TOEFL iBT and provided an argument from the literature as to why the integrated tasks were superior to previous versions concentrating on the communicative understanding of grammar and vocabulary. He argues that as the tasks are superior to previous versions the test has greater validity. Riazi (2016) found similar results. Kim (2017) takes a more negative view of the test. Korean student's perception of the test was measured by looking at English language message boards finding that students were primarily looking into test wise strategies to beat the test. Barnes (2016) has a similarly mixed view of the TOEFL being one of the few studies on TOEFL and wash back. Through the study of a number of English classes in a variety of settings in Vietnam the researchers found that test preparation material was the dominant kind of text book and teaching practice was geared towards success on exams. This research is potentially calling into question the validity of the TOEFL iBT and as such a vigorous examination of the work done by ETS and outsiders is necessary to move forward.

## 2.7 Conclusion

As has been established through this review wash back is a well-studied phenomenon having both positive and negative effects. It can affect teaching practice as well as student behavior. Test validity is a serious concern and care needs to be taken to establish construct validity of a test and to try to prevent the dominance of test wise strategies. The TOEFL iBT was developed with this in mind and was rigorously validated in its construction and follow up studies from ETS and others have further strengthened the validity argument with some exceptions.

There still exists a major gap in understanding in that the TOEFL appears to have a strong case for validity, but the perceptions of many at universities do not believe it is a good predictor of actual ability. What's more much more is needed to look at what effect TOEFL has on wash

back, both from the perspective of teaching practice and student perception. That is what this study will be examining, looking at ESL students experience with the TOEFL test and their perception. Through focus groups, questionnaires and interviews these questions may be elucidated further.

### 3.1 Introduction

At its heart this research is about the how the TOEFL iBT affects the education and understandings in the specific context of high school students in the People's Republic of China who wish to study abroad in English language universities, and the impact the stakes of the exam have on students beliefs and actions and how that might impact teaching practice. This section will explain in detail the research methodology and design starting with the theoretical grounding of the work. Extensive examination of Educational Testing Services validation method as well as additional validation studies were examined in the literature review section but there will be additional discussion about the logic behind validation studies and the necessity for continued investigation into validity will be explored. The mixed methods approach will be justified in specific relation to the research questions of this project but also in terms of how they are related to the methodologies of previous studies. The research questions this study hopes to address are

- 1) What do Chinese High School Students believe the TOEFL measures?
- 2) To what extent do Chinese students think the TOEFL iBT prepares them for the Western universities?
- 3) How do students vary by ability level as defined by TOEFL score?
- 4) What do Chinese students think are the most effective ways to prepare for the TOEFL test?
- 5) Do student's preparation techniques vary by ability level?

Depending on the answers to these questions, the construct validity of the TOEFL iBT could be called into question. The research instruments used in this study are designed to answer these questions in order to question the extent of the validity of TOEFL iBT in the studies context. The

purpose of this work is to exam one case, one group of students and to look at the social context in which they form beliefs about study habits and standardized tests and take specific actions which can impact their education.

### 3.2 Replication Studies and Context

This study was initially conceived of as a replication study of the work ETS has done in order to validate the TOEFL iBT and encourage its use as one of the major high stakes exams used in university admissions around the world. The importance of replication studies cannot be overstated, especially considering the replication crisis in Psychology that has been in the news recently (Vashith, December 2017). Brandt (2014) argue for the importance of replication studies to ensure the validity of social research but argue that there isn't a clear enough definition of what constitutes replication and more specifically close replication. Any theoretical system needs to be tested and retested in order to ensure the validity of the argument (Cohen, et. al., 2005). The validity of a test is no different, if the test lacks construct validity that may be observable by repeating the same test in different contexts (Bayazidi and Saeb, Feb 2017; Nurcan and Mustafa, 2018). As it was explained in the literature review, most work on TOEFL validation, and ESL testing in general has been done on students already in university. In particular, Malone and Montee (2014) was the inspiration for this work and was done exclusively with international students who were already in college and from countries other than the People's Republic of China. This will not be a true replication study as the identical instruments will not be used, but this research has the same starting point, a focus group discussing high stakes testing leading to a survey, and will develop the instruments for the research project in a similar fashion though it will go to further extent doing follow up interviews to further validate the survey results. More detail about the logic of these decisions will follow in the discussion of research instruments.

### 3.3 Case Study Methodology

This work is best described as a case study, that is a method of examining a specific set of

circumstances in a rounded and deeper way (Breslen and Buchanan, 2008, p. 3). Case Studies are less characterized by a specific way of collecting data and more by what is being studied, that is a discrete group of subjects or phenomenon (Cohen, et. al. 2005, p. 181). In the words of one its main proponents (Yin, 2009, p. 4) case studies “allow investigators to retain the holistic and meaningful characteristics of real life event.” This is a particularly useful means of inquiry because as has been stated this is a group of research subjects who are highly underrepresented in the literature, thus a case study can help to illuminate some of the issues surrounding them as well as provide room for further research. All in all case study method is useful as the goal of this research is to gain a deep, well rounded understanding of a specific group of united by shared characteristics which will have some generalizability (Crewell, 2002.). The purpose of this case study is to ‘enlighten’ as Yin (2009) describes, that is to shed light on a situation with multiple causes and effects and a complex social context. Additionally, this is a case study which shares the characteristics of revelatory and representative case studies (IBib., p. 49). It is designed to describe an underrepresented group who a representative of similar cases allowing for some generalizability between this case and others that share the same setting.

This study is concerned with cause and effect, how do students beliefs impact their studies and how do their studies impact their TOEFL results. One could do an experimental design, using the scientific method to isolate specific variables during research, as some have argued it is the best way to establish cause and effect (Mertens, 2010; Maxwell, 2004). However a case study is more appropriate in this instance as it is looking at a number of variables and to see how the whole context of the education situation is developing. Further research may be possible in the future based off of this research. For example, if it is found that this group of students is likely to use test wise strategies, experimental design could be done on larger groups of similar students to further establish cause and effect. In this research setting, most of the variables are outside of the researcher’s control. This project is not an intervention but an analysis of the context.

Another potential method for case studies, which ultimately was not chosen, is doing participant observation. This involves observing subjects within a social context, gaining

understandings from the participants and pulling inferences from the observed subjects (Cohen, 2005). While this study does share the spirit of ethnography as Delmont (2002, p. 7) describes as “values the views, perspectives, opinions, prejudices and beliefs of the informants...and takes them seriously.” This method was not chosen for two main reasons. The first is if the research questions are specific enough, even those about somewhat ephemeral ‘beliefs’ that participant observation would not be required for, and secondly the time involved would have been more onerous on the students and researcher resulting in fewer participants and possibly less generalizability (Cohen, 2005, p. 157).

In the end a mixed methods, quantitative and qualitative approach was taken. While not the most common method for case studies, quantitative data can greatly illuminate a complex situation (Yin, 2009, p. 19). The methodology used follows.

### 3.4 Quantitative and Qualitative Methods

Quantitative and qualitative research methods have strengths and weaknesses. Quantitative research has the virtue of having objective numbers that can be compared, analyzed and examined for change over time. Associations between variables can be found and analyzed. Larger sample sizes can be examined and standardization is possible. At the same time, the greater breadth of data can lack depth and lead to superficial understandings. Complex problems and phenomena can be difficult to examine with strictly quantitative data (Mujis, 2004). For example, you can do a study about the educational achievement of male versus female students and find an association of some kind but a strictly quantitative analysis might not be able to examine the deeper question of why this association exists. On the other hand, qualitative research can provide much needed depth to the study of a research question, but the results are necessarily from a smaller sample size and interpretation can vary from researcher to researcher (de Marrais, 2004).

This study is looking into both student’s beliefs and their actions. As a result neither a strictly quantitative or qualitative approach is appropriate. The numbers students get on TOEFL

tests will be compared to choices in study habits, a quantitative method, but this research is trying to examine why these students made those choice by interviewing them about their beliefs, a qualitative method. Student's TOEFL results will be compared to their specific study habits quantitatively, but this is not the limit of the scope of this research. This research is also concerned with the 'why' of student beliefs and actions around the TOEFL test. As a result a mixed method's approach, combining quantitative and qualitative analysis, is most appropriate. As Mertens (2010) explains, if you are examining a complex social context and a complex question about social behaviors than a mixed method approach will give you a deeper understanding of the problem. Cohen (2005) describes the benefits of methodological triangulation, using multiple instruments on the same research subject, is best of a when a "holistic" view is being sought (Ibid., p. 115). That is, an approach where the student as a whole is the subject of the study, not just one particular action or result.

### 3.5: Setting and Participants

The setting of this research is a High School in a large city in the People's Republic of China. Specific details can be found in the introduction of this piece. This study is in part designed to access a group of students who have not been studied extensively, but are growing in number and importance for Western educational institutions as discussed in the literature review.

The targets of the study for the research are the Senior Two and Senior Three students, the equivalent of 11<sup>th</sup> and 12<sup>th</sup> grades in the United States, in the American international program at the school. Senior One students were not included in the study; Chinese high school consists of three years, because they have not begun to intensively prepare for the standardized tests needed for entrance into Western universities. As such they might be able to contribute to questions of impressions, but lack the hands on experience with TOEFL that this research is concerned with.

Total enrollment for senior 2 was 46 students and enrollment for senior 3 was 51. As many students as possible would be included in order to increase the generalizability of the study

(Mertens, 2010). That is the more students included in the study the more you could say that this was a population typical of schools like this and typical of the students who are attempting attending these programs with the hope of attending western universities. While not all students would be part of the qualitative research, all would be potential targets for said research. That is, students would be selected randomly with the intention of getting students to participate who are typically of their classmates (Ibid., p. 318).

The students in the target group are demographically similar. All students are citizens of the People's Republic of China. The difference in the number of male and female students is there, 55 female students to 42 male students, but is not substantial enough to be relevant to the project. Considering that gender is not a focus of this study versus their shared identity as perspective international students in China the differences in beliefs and achievement between male and female students will not be examined further, though there is potential for further research in this area. Additionally the students are of similar economic status as all are able to afford the steep tuition of the international program as well as international tuition at Western universities. This allows for greater generalizability to other similar programs in the People's Republic of China (Cohen, 2005).

The main way the target students will be divided for analysis purposes is by looking at high medium and low achievers as defined by results on the TOEFL and their Advance Placement courses grade point average. When sampling for the qualitative research instruments, care was taken to have representatives from each of the different achievement levels. Students who have different levels of achievement may have different beliefs about education and act differently in any educational setting (Gaudy and Spielberg, 1971; Shell, et. al., September 1995; Chesters and Daley, March 2017) as well as specifically in the Chinese school context (Wang et. al, April 2018). As such the student's names have been divided by their TOEFL scores into three roughly equal sized groups. TOEFL score is being used as a measure of academic achievement in this case as student behavior and how that affects the TOEFL score one of the focus of this test. Students will also be compared using their AP GPA in order to see what affect their actions and effects might

have on their school work, but they will be divided along their TOEFL score because that is the main focus of the work.

| Senior 2 TOEFL Scores |                    |
|-----------------------|--------------------|
| Scores                | Number of Students |
| High (91+)            | 11                 |
| Medium (76-90)        | 15                 |
| Low (41-74)           | 12                 |
| Has Not Taken         | 3                  |

Table 3.1 Senior 2 TOEFL Score

| Senior 3 TOEFL Scores |                    |
|-----------------------|--------------------|
| Scores                | Number of Students |
| High (101+)           | 17                 |
| Medium (85-100)       | 15                 |
| Low (35-84)           | 14                 |
| Has Not Taken         | 3                  |

Table 3.2 Senior 3 TOEFL Score

### 3.6 Ethical Issues

Overall this project is not concerned with particularly sensitive or personal subjects and all students participating will be over the age of 16. However there are still a few concerns to be addressed. The first potential issue is with students being perceived to be singled out due to academic performance. Students are keenly aware of who is a “good” student and there is risks

of making students feel stigmatized if they perceive they are being ranked (Black and Wiliam, 1998; Black et. al., June 2006). As such great care was taken in emphasizing the randomness of the selection for focus groups and interviews so the stratification of selection is not apparent.

Additionally students will might be reluctant or feel uncomfortable self-reporting grades (Ellis, et. al. 2017). As such, care has been taken to be able to gather accurate quantitative data, linked to students grades, but not have the students actually feel the need to put it down on a piece of paper that the teacher is going to se. Further details will be provided in the section on research instruments. Finally, the students might be on some level cognizant of their non-Chinese teacher's opinions about preparation techniques and high stakes standardized testing. In order to minimize this problem, it was strongly emphasized in all documents and discussions of the project that the student's statements would in no way affect their educational outcomes or their relationship with any staff member at the school. They were informed that their opinions were highly valued and would help to contribute to a greater understanding between teachers and students as well as improved teaching practice at the school. Any specific ethical issues related to each research instrument used will be detailed in the following sections. The principles of the British Educational Research Associations (2011) were rigorously adhered to at all stages of research.

### 3.7 Piloting Research Instruments

Piloting research instruments, focus group, survey, interview questions is a useful way to increase their validity, reliability and workability (Cohen 2005, p. 260). Additional care needs to be taken with non-native English speakers as there can be additional problems of understandability of the instruments (Gibbons, 2006; Mertens, 2010, p. 178). As such all research implements were looked over by two experienced, bilingual Chinese ESL teachers. In all cases the teachers had experience with the students who would be the target of the study. Several questions were reworded on the suggestions of these teachers in order to be more clearly understandable by the students. Additionally, all research questions were approved by Education

instructors at the University of Sunderland to ensure the understandability and usefulness of all research instruments. Changes were made to some questions as a result of these discussions in order to improve the understandability and clarity of the results.

### 3.8 Focus Group

“Focus groups involve organising and conducting group discussions in order to gain knowledge of the attitudes, beliefs, practices and values of participants on a specific topic” (Ahmed et. al. 2017, p. 1). The research of Malone and Montee (December 2014) that forms the inspiration for this research begins its investigation into test stakeholder beliefs with a focus group. Their justification was to allow for the open exchange of ideas between study participants leading to new avenues of thought the researchers had not considered. Focus groups can lead to new ideas which may not necessarily be found through more structured processes and can be a good jumping off point for further investigation (Mackey and Gass, 2005). Bloor et. al. (2001) describes focus groups as ideal for preliminary study of an issue for the development of more specific and focused research implements. This is the intention of this research to both examine ETS’ research methods as well as investigate these questions in their own right. To quote Malone and Montee, (December 2014, p. 6) “The purpose of the focus group was to gather information about participants’ attitudes towards the TOEFL iBT and identify topics to explore in subsequent phases of the study, specifically to develop questions for a large scale stakeholder’s survey.” The purpose of this researcher’s focus group is the same.

It was decided by the researcher that the focus group should have approximately six to eight members as that is considered to be an ideal number of participants to have a variety of opinions while not becoming too unwieldy or allowing some members to limit their participation (Bloor, M. 2001). Participation by students of high medium and low achievement was also considered to be very important. It was decided to only include the Senior 3 students at this point because the school had not been officially informed of the senior two TOEFL grades for

transcript purposes. The only way to divide Senior 2 by TOEFL score would have been to ask them to self-report which, as discussed previously, the researcher was trying to avoid. It was also believed by the researcher that the Senior 3s, having almost completed the university application process at this point would have the most perspective on the process and thus provide the best initial point for the development of further research instruments.

Students names were printed out and divided into their respective category and drawn from three names from each were drawn from a hat. It was announced to the students by English class that they could possibly expect an email message asking them to participate in a research project. The purpose of the research and its voluntary nature were emphasized to the students. Randomly chosen students, two from each achievement level were contacted through school email about participation in the focus group. If enough students did not respond within one day of being contacted further names were drawn until the sample size was sufficient. There was a fear of bias as higher achieving students might be more enthusiastic about participating in the project (Bloor, M. 2001) which the researchers hoped to minimize by recruiting from all three achievement levels. It proved much easier to recruit for the top tier achievement level as illustrated by table 3.3 though representatives from each level were found.

| Senior 3 Students Contact and Acceptance Rate |           |          |
|-----------------------------------------------|-----------|----------|
| Student TOEFL Achievement                     | Contacted | Accepted |
| High (101+)                                   | 4         | 2        |
| Medium (85-100)                               | 6         | 3        |
| Low (35-84)                                   | 8         | 1        |

Table 3.3 Student Contact and Acceptance Rate

The focus group took place over the course of one school period when the students were available and extensive notes were taken by the researcher. The questions can be found in Appendix A and the results were summarized by the researcher. A number of interesting results were found. The students came from a variety of academic interests including sciences, business and humanities decreasing the chance of bias if only one student interest was represented. The focus group questions were developed around three subjects of inquiry. The first is the Malone and Montee (2014) survey, the second were the researcher's beliefs about student's beliefs and practices and the third were questions designed to elicit unexpected outcomes. That is, specific information about student's beliefs about the test, Western universities, and preparation techniques could be found.

As the focus group lead to the creation of the surveys and interviews questions, the results of the focus group should be discussed here in the methodology section. Student's difficulties with the language were explored finding that academic writing and speaking were the main concerns of the students related to the test, but listening was considered a problem with general English use. The students expressed a general feeling of anxiety and unpreparedness for academics at Western university. The students had an overall positive feeling about the test even though they also said that they didn't think it was particularly helpful for developing language skills. Additionally they said they didn't really understand how they improved on parts of the test and their feeling when they left the test frequently did not match the score they got on the exam. The students said they thought that it was useful for simulating what they believe university education will be like, but also emphasized the significance of test-wise strategies and flaws in the test.

Outside classes specifically to study for TOEFL, with New Oriental being specifically named, were part of the student's experience as well as finding questions and advice about how to improve scores online were also mentioned. There was a sense of uneasiness about this topic and the students were quick to emphasize that they didn't think these strategies were helpful and thought some of the strategies they heard of students doing to prepare were unethical. It is

possible the students thought that the researcher would judge them for engaging in these kinds of behaviors. These were noted and would be emphasized on the anonymous survey where students might feel freer to express their opinions.

Major points of discussion from the focus group were noted by the researcher and combined with salient questions from Malone and Montee (December 2014) were connected to the project research questions. From there survey questions were developed as detailed below.

### 3.9 Survey Construction

It is useful to establish the purpose of a survey in order to keep it focused on your specific research questions (Mertens, 2010). As such the specific purpose behind the development of this research instruments was to quantify the extent to which certain beliefs related to the TOEFL are shared by the target population. From that, any association between the student's beliefs and their results can be identified. Taking the sample size into considerations, a semi structured survey was developed. It was believed that the sample size was sufficient to be able to derive meaningful quantitative data while open ended questions would allow unexpected patterns to emerge (Cohen, 2005). A wide variety of survey techniques were used in order to glean a wide variety data in order to more completely explore the research questions. This project is related to beliefs, a very broad question, and as such the more angles approached the more depth to the data could be discovered (Mertens, 2010).

For example, specific study preparation techniques had been identified in the focus group as well as from Malone and Montee (December 2014) but students might have other techniques which had not been previously identified. As such students were asked if they had done specific research techniques, but were also asked if they had any additional techniques. Thus common patterns could be quantified from both perspectives allowing data to come from targeted questions of the researcher, but also giving the students a chance to express their own subjective understanding of the question (Irwin and Stafford, 2016). The researchers understanding of the

subjective meaning of individual student's responses would be further validated through stage three of the process through interviews with students as detailed in section 3.7.

Another used in the survey was to use rank ordering. This is useful for examining beliefs because it gives the added dimension of the intensity that belief (Cohen, 2005). The students would be asked to rank their concerns about studying at university allowing us to see if there are any patterns in how strongly the students feel it. All the students may be concerned about writing papers and understanding accents, but this will allow us to see the intensity of that fear. Similarly, Likert scales were used to identify the intensity of belief (Cohen, 2005). A neutral option was not given so as to force students to take a stand on the issue (Neuman, 2013). The questions asked using the Likert scale specifically went to the student's beliefs about TOEFL and their English language ability and also shows the intensity of that belief. Specific demographic information would also be asked with the intention of finding any unexpected associations. For example, students were asked how many times they had taken the TOEFL test with the understanding that there might be an interesting association between number of times taken and specific study habits. See Appendix B for a copy of the survey.

One question that was not asked was what the student's grades and TOEFL results were. As previously discussed there was a reluctance to ask students to self-report results both for ethical considerations but there was also a desire to avoid self-reporting bias which could potentially invalidate the perceived associations (Mertens, 2010, p. 204). In order to accommodate this issue the students were assigned anonymous numbers by a colleague who would not be involved in the analysis of the data. Each number was linked to the official grades and TOEFL results. The survey was then distributed by the colleague who had anonymized the students. Each survey had a number linked to a student allowing the student's answers on the survey to be linked to their TOEFL results and school GPA.

The survey was distributed to all senior 2 and 3 students in attendance at the time. Students were given informed consent forms and given ample opportunity to choose not to participate in the survey. In the end, a total of  $n=90$  students participated including 41 senior 2

and 49 senior 3. The surveys and consent forms were collected and the data was quantified into excel spreadsheets. Open ended questions were analyzed for commonalities and when found categorized together. For example, under the question about the most important ways to study TOEFL some student responded 'TOEFL book', others said 'Study TOEFL book.' Others responses were less obvious in both how to categorize, this being one of the main reasons for the development of the final phase of research, the semi-structured interviews. The data was preliminarily categorized into Excel spreadsheets with changes made based off of the results of the interviews with students and college counselors.

### 3.10 Data Validity

Triangulation is a part of this project in order to enhance the validity of the research by confirming through multiple methods (Cohen, 2005, 253) as well as to increase the depth of information being explored (Mertens, 2010, p. 256). It is also worthwhile to get the input of multiple stakeholders who have relevant information on a subject (Neumans, 2014, p. 166 ) To this end two sets of interview questions were compiled, one targeting students and one targeting the college counselors who assist the students with matters including preparation for college applications including preparation for the TOEFL exam and other standardized tests.

### 3.11Semi-Structured Interviews

Senior 2 and 3 students would be asked to do interviews in order to add confirmation to the survey they had already completed. See Appendix C and D for all the questions asked. These interviews would be useful for two main reasons. Specifically, students would be asked to clarify common responses to open ended questions and students would be asked to elaborate upon any unusual responses found from the preliminary perusal of the data. This would allow greater certainty in the statistical analysis of the data. That is, if we were looking at test-wise strategies and students mentioned 'vocabularies' as an important study tool, greater certainty as to whether it would be appropriate to qualify the response as test wise or not.

Subjects were recruited in a similar manner as to the focus group. The school email system was used to recruit subjects from senior 2 and 3. Student's names were divided into groups of high, medium and low achievers in the same way as previously described and students were recruited from each group. Student response was much lower this time around, possibly because it was later in the school year as timing can have an effect on response rates (Graham, 2015). In the end only one student from each ability level volunteered for the interviews.

The interviews were done at the student's convenience. They were recorded and transcribed and copious notes were taken by the researcher. Through discussion of the questions asked greater confidence in the correct categorization of student survey responses were found and a few additional points which showed the need for further research as well as showing some of the limitations of this research were discovered. Further details can be found in the results section of this work.

The schools college counselors were asked to participate in the project primarily because of their specific expertise while not feeling in anyway beholden to the researcher. During the focus group and interviews with students some hesitance to discuss beliefs and strategies was detected. It is possible that the students could perceive that their teacher might perceive some of these actions or beliefs as being unethical and as such be vague or not answer as fully as they might. As such the college counselors were in a great position to assist with the research based on their unique relationship with the students.

As experts on this case, the counselors could provide the depth of knowledge necessary to interpret the student's response and provide further depth on this case study (Yins, 2009). Some background is necessary in order to explain the role college counselors play in these kinds of international school settings. There will usually be teachers from Western countries who take care of subject classes and Chinese college counselors who provide advice on preparing for standardized tests, TOEFL and SAT, and on college applications. Additionally, many students also attend outside counseling services and outside standardized test preparation centers. As such, the college counselors provided valuable insight for three reasons. The first is because they are

also second language English speakers having learned English in similar context as the students allowing them to provide valuable insight into how students translate specific beliefs and behaviors. The second is that their relationship with the students does not involve grading or disciplining students leading to more open communication. They also have knowledge of educational vocabulary and methods and so can be asked directly about concepts like 'test wise' strategies with limited explanation. Finally, the counselors have unique insight into the student's preparation for applications including what they do to prepare for TOEFL and why they do it. They are more likely to know what their outside counselors and outside classes are like because it is a regular focus of discussion for them.

### 3.12 Data Analysis Method

With data collection complete and the survey data organized based off of the student and counselor interviews data analysis can begin. This data is good for quantitative analysis because we are looking at numerical associations which we can create descriptive statistics from (Mertens, 2010, p. 405). A preliminary examination of the data was done looking for any associations that were suspected to exist, students who were low achievers might practice more test wise strategies, associations found by Malone and Montee (December 2014), students would feel the questions were fair, and looking for any unexpected patterns that would emerge from the data. So this is research where there are preexisting hypothesis are used in order to answer the research questions, but analyzing the data itself can lead to unexpected answers emerging (Mujis, 2004, p. 54).

As such, each question was divided up into which research question it could help answer and statistical analysis using Microsoft Excel was conducted looking at any associations and organize the data into graphic form. In the next section, each research question will be examined on a question by question basis looking at the patterns that have emerged from the data, seeing evidence for initial hypothesis as well as seeing whether Malone and Montee's (December 2014) are replicated.

### 3.13 Conclusion

As a case study (Yins, 2009) this research was done in order to elucidate this particular group of students understanding about the TOEFL to see what they do but also gain insight into why they think that. Data was collected at three stages in with the specific goal of answering research questions related to student beliefs about TOEFL iBT preparation. Initial ideas were developed from a focus group leading to the development of a survey distributed to all available senior 2 and 3 students. The answers to these questions would be used to examine the research questions after their construct validity was confirmed through interviews with college counselors and students. Combinations of quantitative and qualitative methods were used to enhance the validity of the data through triangulation (Cohen, 2005). Through the statistical analysis of the data at hand whatever patterns of association found could be examined on a question by question basis. The patterns found through said analysis will be detailed in the next section of the research.

## Chapter 4

### Data Analysis

---

#### 4.1 Introduction

The previous chapter concerned methodology, now the data resulting from this will be analyzed. This data analysis will consist of three main parts. First, the association between student's grade point average and TOEFL results, second, an analysis of the descriptive statistics provided by the student survey, and thirdly, a look at how the qualitative information provided by the student and counselors provides evidence for or calls into questions results from the quantitative analysis of the survey. The analysis of the association between grade point average and TOEFL score between different groups of students proved to be one of the most significant findings of the project and as such it shaped the analysis of all further data.

As previously discussed, the students who were the subject of this study, while demographically very similar, have very different levels of academic achievement and English language ability. Top achievers are basically fluent and the lowest achievers find basic comprehension difficult. This differentiation between students had a major impact on all aspects of the research and as such all data will be presented with this difference in mind throughout the chapter. Additionally all data will be presented as divided between the two year groups as it doesn't make sense to directly compare TOEFL scores of students who had an additional year of study to prepare.

#### 4.2 GPA and TOEFL Comparison

The first data point will be a comparison of the association between GPA and TOEFL for Senior 1 and 2 students. It makes intuitive sense that the stronger a student is in school, the better they should do on a test like TOEFL. In fact this is an argument made in favor of the validity of high stakes testing in general (Wragg, 2010, Darling-Hamond, 2014) and by ETS for the

TOEFL specifically (Chapelle, 2008; Breining, 2011; Ling, 2014). It is unlikely to be a perfect correlation, after all, a student might excel in maths and science but do poorly in English, but it is generally reasonable to assume that there should be some relationship between the two. As such Pearson's correlation coefficient between GPA and TOEFL score was drawn in Excel for all Senior 3 (Fig. 4.1) and all Senior 2 (Fig. 4.2). From this we can see the strength of an association, with more than .8 being very strong, .8-.5 being strong, .5 to .3 being moderate, .3 to .1 being modest and less than .1 as weak (Muijs, 2004, p. 145).

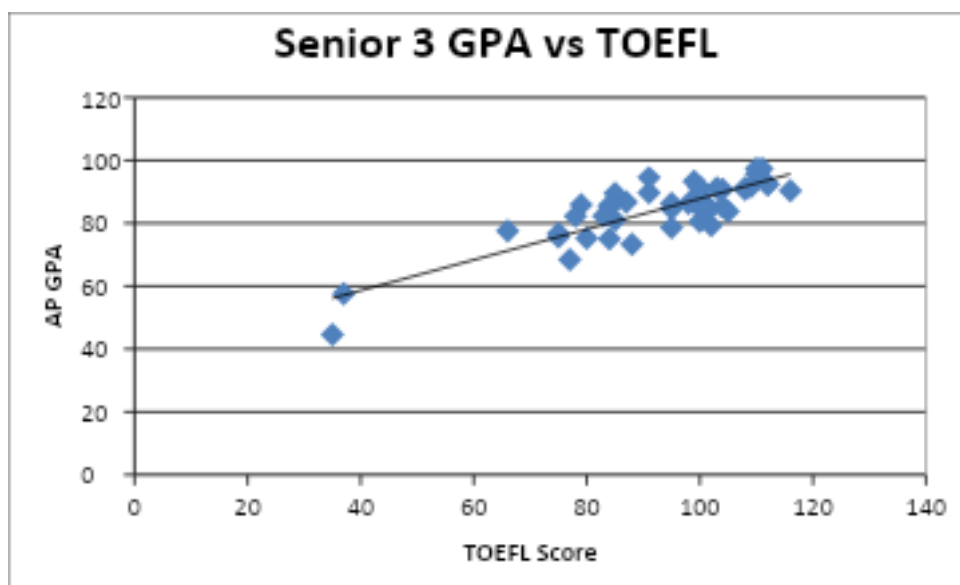


Figure 4.1 Association of .8468

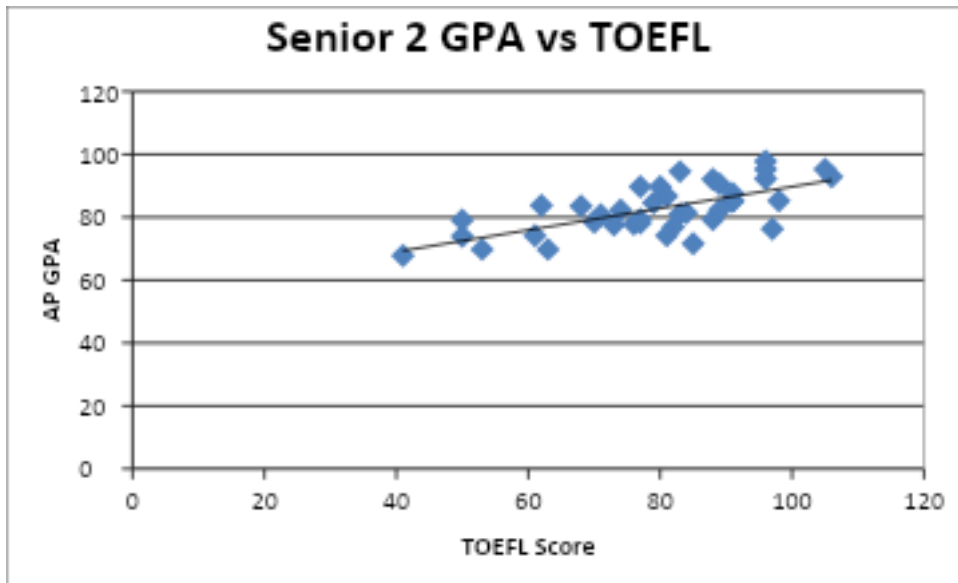


Figure 4.2 Association of . 6778

As expected there is a positive relationship between class results and TOEFL results. For senior 3 the results are very strong and the results for Senior 2 are moderate. Looking at these three data points gives evidence that in these two groups of students TOEFL reflects their actual academic ability and thus has construct validity. However, when the data gets parsed further new information comes to light. As previously described, the students were then divided by TOEFL score into top, middle and low achievers and the same statistic was calculated.

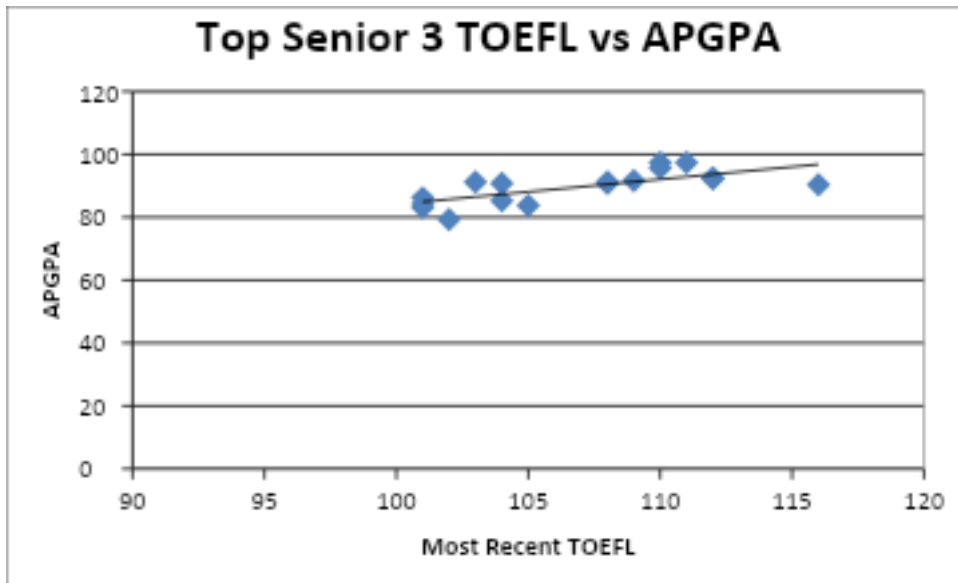


Figure 4.3 Association of 0.7117

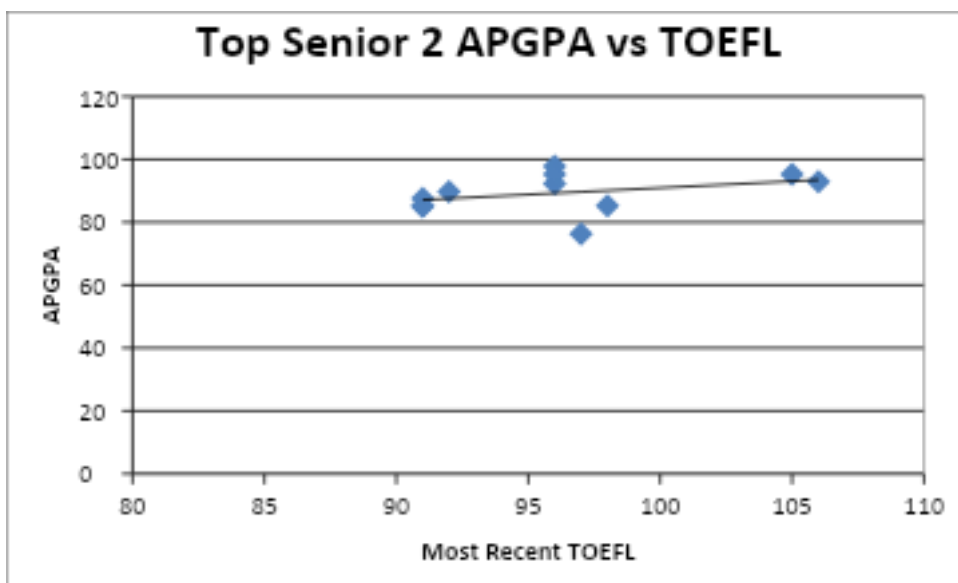


Figure 4.4 Association of 0.3570

For the top students in both classes there is still a positive association between GPA and TOEFL results ranging from strong for Senior 3 to moderate for Senior 2. The Senior 3 students may have had more time to study for TOEFL specifically explaining the stronger correlation or the Senior 2 students could be more maths and science focused.

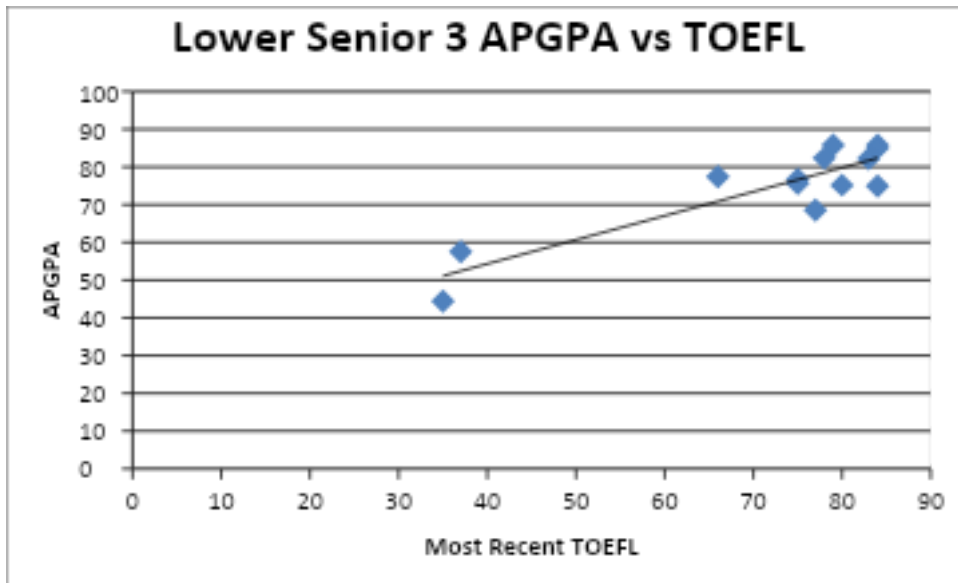


Figure 4.5: Association of 0.8912

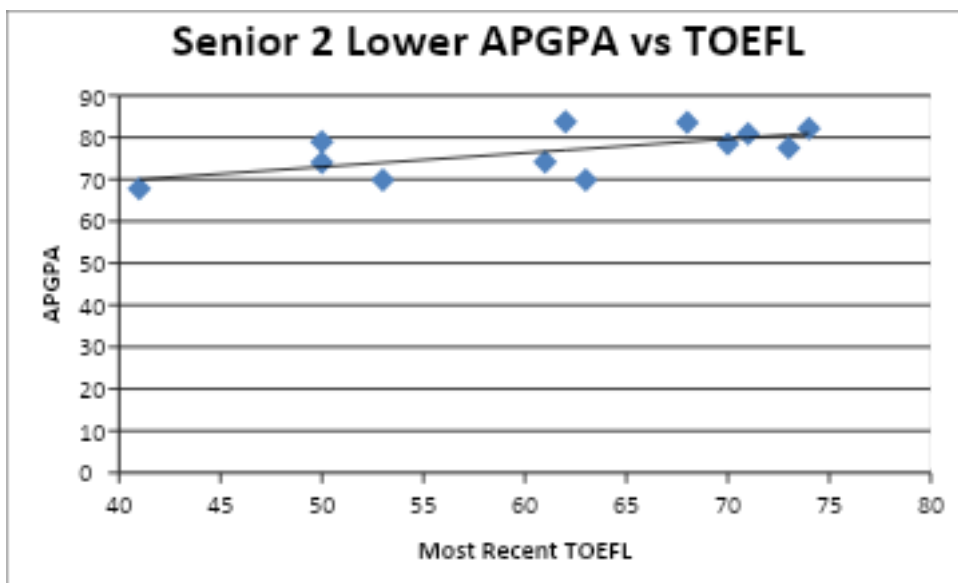


Figure 4.6 Association of 0.6388

Similarly, with the lower achieving students of both classes there is a very strong association for Senior 3 and a strong correlation for senior 2. This makes sense as one would expect students achieving poor results in school would do poorly on standardized tests. So far this is good evidence for the construct validity of the TOEFL test. Based on everything observed so far there is no reason to think there would be a significantly different result for middle

achieving students. Those who are moderate achievers in school should also be moderate achievers on the TOEFL test. However when the same statistic is applied to the middle achievers something surprising is discovered. As discussed in the literature review, the TOEFL validity hinges on it being an accurate measure of the student's ability to use English (Chapelle, et. al., 2008). As such there should be at least some correlation between students at all ability levels.

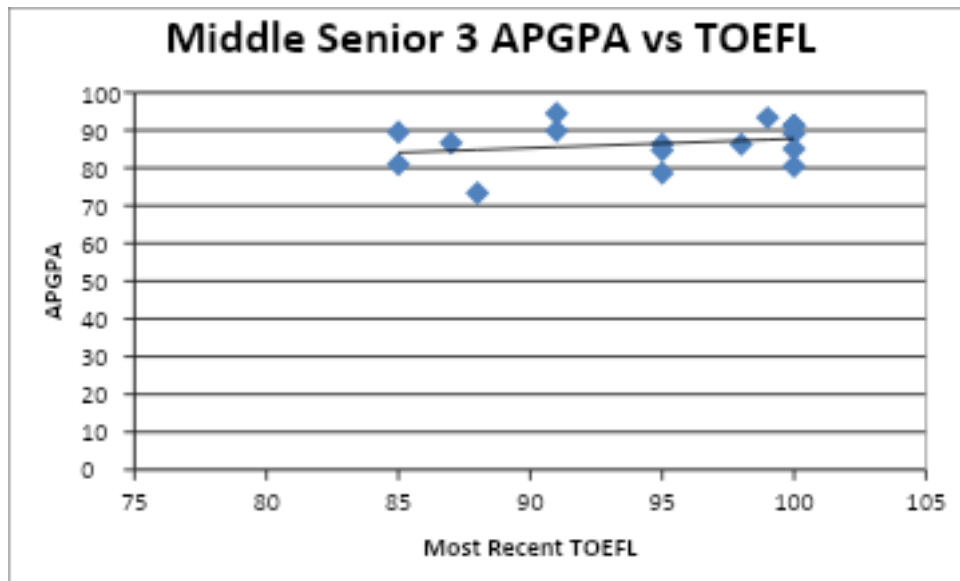


Figure 4.7: Association of 0.2443

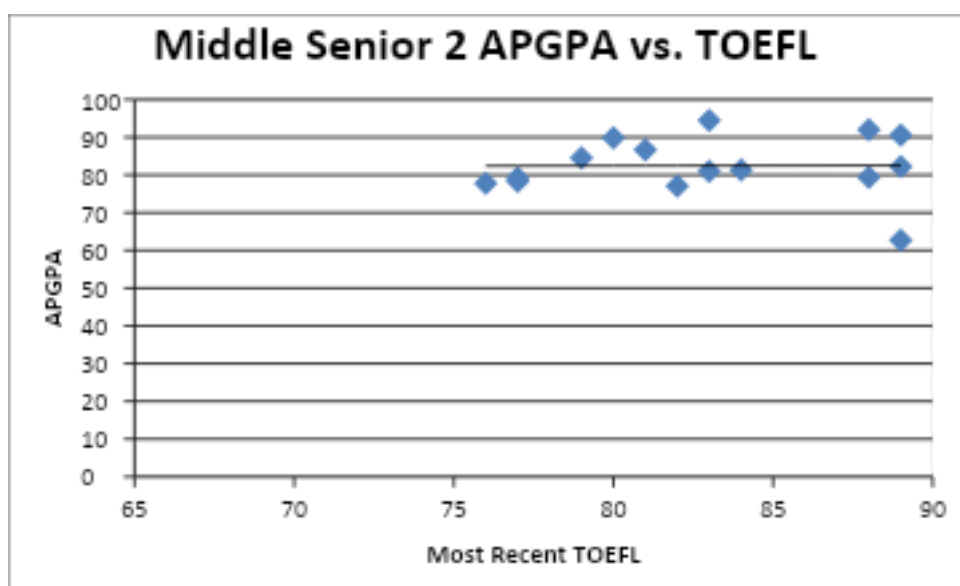


Figure 4.8 Association of 0.0054

The association drops all the way down to modest for Senior 3 and almost completely disappears for Senior 2. This suggests that there could be major differences in these different groups of students. As Such all data has been divided up by these categories and by further analyzing these descriptive statistics it may be possible to discover what those differences are. What's more, depending on what those differences are it may call into question the construct validity of the test. If this is a trend which can be extrapolated from this particular group of students than it could give pause to those who use the exam in admissions and other high stakes situations. Malone and Montee (2014) in their work with school admissions officers found an assumption on the administrator's part that TOEFL was a reflection of actual student ability. If that is not true for all students then that could affect the admissions process.

During student interviews, all three students when asked whether they thought the test reflected their English ability all three said no. It was a contentious issue during the focus group. The sentiment expressed was that it did not require improving actual English ability. As one student said, "I can ace the test if I learn the proper way" (Appendix H, p. 166). The distinction expressed by the students between studying for TOEFL and studying for regular classes was the difference between test wise strategies and improving English ability. "The test skills, your English ability is one aspect but your test skill is another aspect you need to know what is on the test." Of course this could also call into question the validity of the grades at this particular school. Further studies could be done at educationally and demographically similar educational centers or the TOEFL grades could be compared to standardized AP exam results. When comparing their AP classes to classes to study for TOEFL, another student said "(AP classes) we read the textbook, the teacher help you to understand, to inspire you, make you think, help you know, the critical thinking, for TOEFL it's just practice over and over and over again" (Appendix H, p. 157). That is one anecdote but it was a common sentiment in interviews and focus group.

An additional reason to separate the students into to high medium and low achievers comes from the counselor interviews. All four counselors confirmed that there would be major

differences between how different achieving students behaved and what they believed. Study habits were identified by all counselors, specifically general study skills versus test wise strategies, and beliefs about the value of study habits. This is corroborated by Jackson and Evans (2017) who say that students of different ability level have different beliefs and habits. In order to see if this is borne out by this research sample the survey data will be analyzed.

#### 4.3 English Education

The first part of the survey to be analyzed relates to the students English education and TOEFL. As such, the number of years studied what kinds of studies were engaged in and how this relates to TOEFL will be examined. As stated, data will be divided between different ability levels. As discussed in the literature reviews TOEFL training schools popular amongst Chinese students (Cheng and Curtis 2010; Zhan and Wan 2014;; Xie, 2015) and have significant washback effects.

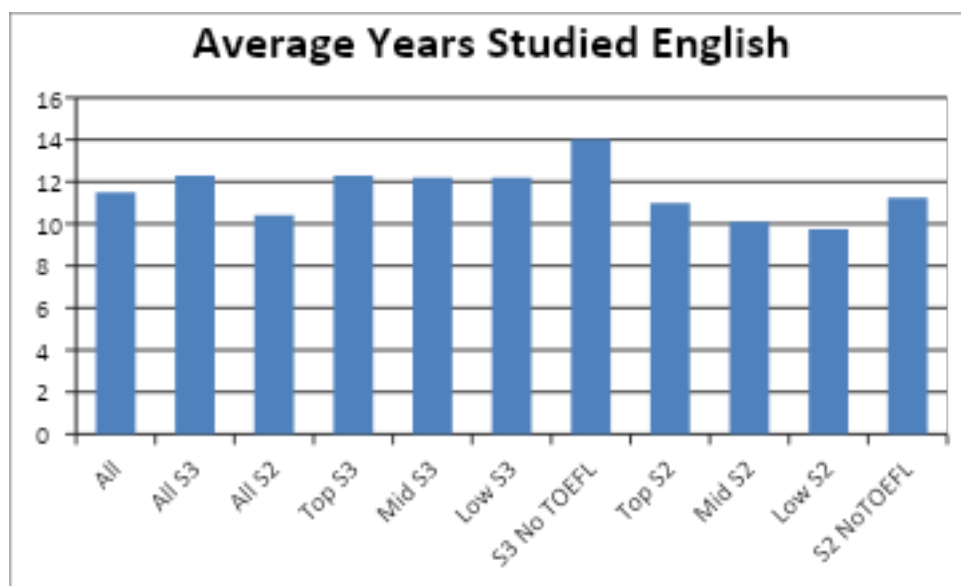


Figure 4.9 Average Years Studied English

English is not a new subject to any of these groups of students. Senior 3 three of course have had an additional year of study over senior 2 but no students are new to English. There is essentially no difference in years studied amongst achievement level among Senior 3 and only a

difference of one year between Top Senior 2 and Lower Senior 2. As stated before, these students are demographically similar and as such a difference in length of study is probably not a significant variable and so whatever differences resulting between these groups, achievement, attitudes and actions, are probably the result of the kind of education versus the length of education. All students interviewed said they had studied English since kindergarten so essentially their whole education.

It is worth noting that this is a similar result to Malone and Montee (2014, p. 16) who were asking University students. They asked on their survey “How long have you studied English?” with the categories, 1 to 6 as individual categories then ‘more than six’ as a category. The vast majority of students asked had studied for more than six years in their study suggesting they would have been better suited using more realistic categories. As Pardis and Jia (2017) show, length of education should be positively correlated with English outcomes, but as these students have similar length of study, other factors need to be considered.

During interviews students were asked how their education had changed over this length of time. All three gave a very similar answer saying they went from essentially learning vocabulary as young children, concentrating on grammar in primary school and switching to academic English and standardized test prep in high school. When asked if they thought there’s was a typical education they said not for Chinese students in general but probably for their fellow international students.

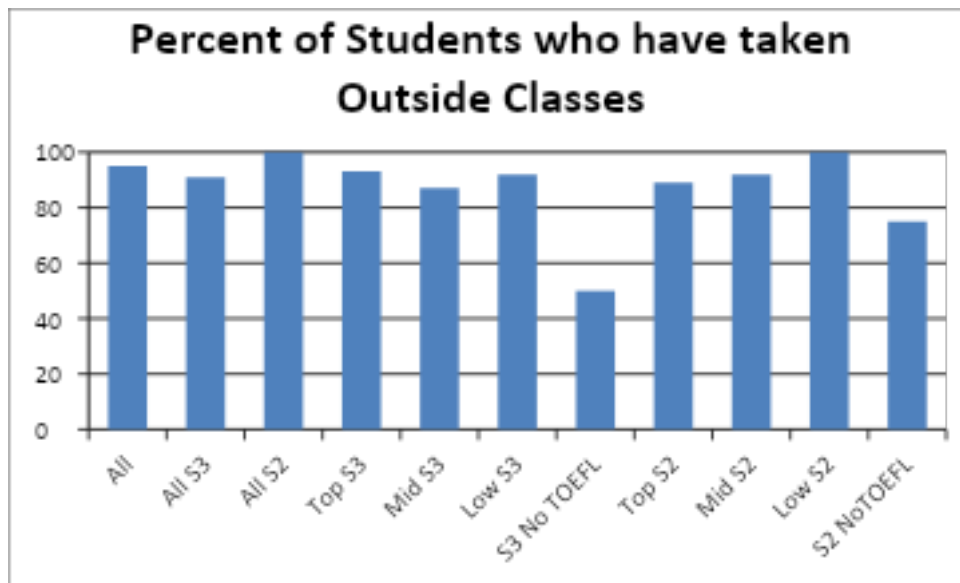


Figure 4.10 Percent of Students who have taken outside classes

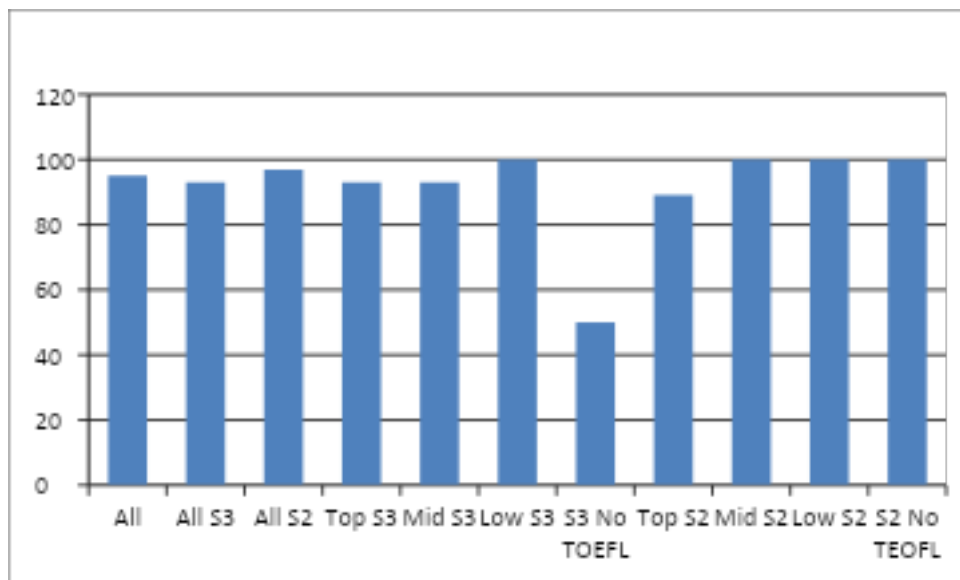


Figure 4.11 Percent Taken TOEFL Classes

In addition to length of time studying English, what kinds of English classes taken is be of interest. New Oriental and other training schools had been brought up earlier (Golden, 2011; Rauhala, 2015; Pan, 2016). If this group of students is typical of Chinese international students then they would be likely to be attending outside training schools. As figure 4.10 and 4.11 they are typical students. Virtually all students were taking outside classes. In interviews students were asked about outside classes and those given at New Oriental were reported to be popular

and New Oriental was mentioned by name several times. During counselor interviews it was even more explicit, with the role these classes play in student's education being strongly emphasized by all. When asked about who has the most influence on students beliefs about the test, one counselor said "their outside teachers, like the New Oriental, they have the authority because the students, I think once they pay for something that's valuable" (Appendix G, p. 152).

When asked about the difference between regular high school classes and those at training centers specific reference was made by students to an emphasis on test preparation versus learning an academic subject and critical thinking. Additionally, all four counselors emphasized the popularity of the training centers because of student's belief that they would raise their test scores. Considering ETS emphasis on TOEFL as a test of communicative competence, that is being able to measure students ability to use English outside of a test setting, (Taylor and Angelis, 2008) the emphasis on training centers teaching test wise strategies has a number of possible implications.

The first is that they believe that the testing centers are effective in raising their scores. That is a possibility beyond the scope of this project, but considering the popularity of the training centers that seems unlikely. The second possibility is that they are able to raise the scores of all students. However if this were the case then you would expect there to be either no association with any group of students or a strong association with all students. That is, either the test should have no relationship to grades of any kind or all students' benefit approximately equally inflating the TOEFL grades of all students proportionally. So why do top ranked and low ranked students have an association with their grades while middle range students do not?

One possibility can come from the counselor's interviews. When asked about whether they saw any difference in the study habits of different ability levels of students major differences were commented up by all counselors. Top students were referred to as active learners who were cognizant of different learning strategies and could think about their education while lower level learners were still working towards building basic study skills and had limited language ability. Middle range students were noted to be able to follow instruction but had limited ability to think

critically. As a result, these middle achieving students responded to extrinsic motivation that is strategies which had an immediate quantifiable results. As one of the counselors said “(Middle Achievers) they don’t have the active process of learning... the difference between top and low is the process, the top achievers take care initiative on their own learning, so when they read and listen they are active, their brain is active so they catch the words they don’t know and they learn, the medium ones they probably just go through, even they learn probably won’t notice they don’t learn, they probably won’t notice that” (Appendix G, p. 149). This is a view also supported by the literature, that there are significant differences in study habits between learners of different levels of achievement (Nonis and Hudson 2010; Lei 2015) as well as in the ability to think about study (Moore 2000; Terwel 2006; Hatcher 2006). In the next section whether there is positive reinforcement in terms of improvement in TOEFL score will be examined.

#### 4.4 Experiences with TOEFL Test

Repeatedly taking the TOEFL test is a popular strategy amongst this group of students. The students started taking the test in year 2 for the most part, meaning they were taking it two to three times a year on average. There were students who took the test up to 8 times over the course of two years. Additionally, taking practice tests in paper version, online or as a phone application was cited as one of the most popular preparation techniques by both counselors and students during interviews. TPO, TOEFL Practice Online was repeatedly brought up during multiple parts of each interview.

The achievement breakdown of the data is also worth noting. The highest average number of tests was from the middle range senior threes, who would have the most opportunity to take the test are, followed closely by the lower range Senior 3. This could suggest that these students are specifically concentrating on improving scores on test rather than generally improving their English ability. Repeated test taking is a strategy which is commonly engaged in with high stakes standardized test taking (Seddegh, R. et. al. 2016) and strictly repeating the test

has shown itself to be an effective way of increasing scores (Arendasy and Sommer 2017 ) and gives more opportunities to practice test wise strategies leading to higher scores (Heng-Tsung 2016).

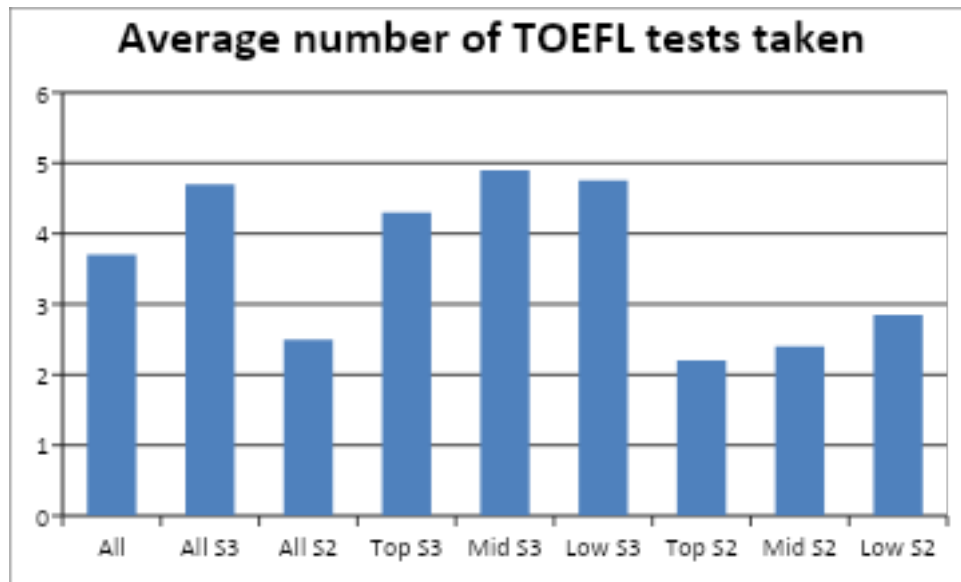


Figure 4.12 Average number of TOEFL tests taken

The question then is, do students see improvement in their scores suggesting that they could feel that they were being extrinsically rewarded for repeatedly taking TOEFL tests. From figure 4.13 you can see that an overwhelming number of students improved a lot or a little. Almost half of students reported their scores improving a lot. Middle Senior 2 would only have taken the test 2 times but almost half still reported their scores improved a lot. Middle Senior 3 also reported strong improvements which could suggest why they took the test so many times. This does not say why scores improved, but it does show that students could feel encouraged to repeatedly take TOEFL tests as they see continuous improvement. It is difficult to judge how much a TOEFL score should increase between tests or how many tests a student should take before the construct validity comes into question. ETS says that the test should reflect the ability to use language communicatively and be representative of how the student would perform in an academic setting (Chapelle, 2008, p. 347). This suggests further research needs to be done to see

what effect taking test eight times has on the students ability to use academic English and communicate in English.

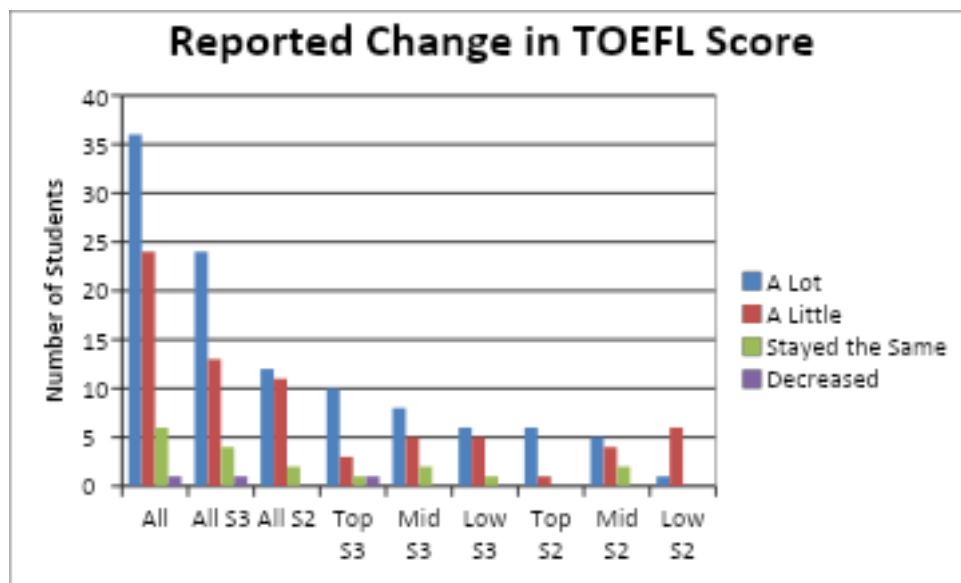


Figure 4.13 Reported Change in TOEFL score

In order to understand students test taking behavior it is necessary to look at why students are taking the test and what they are concerned about. With this information one can better understand why they take the actions they do and understand their motivations.

#### 4.5 Students Intentions and Concerns

When parsing the data from each student, their motivations are also important. The TOEFL is specifically designed to help universities discern whether students have the ability to succeed at an English speaking university (Chapelle, et. al. 2008) so looking at why students want to attend these universities and what their concerns about this experience are can shed some light into their actions.

Students were asked in the survey what their intended major was to be. This was asked to see in there was any differences in beliefs and actions between students who had different intended majors. It seems reasonable that there might be significant differences between students who are studying for math versus students who are studying for English. As such

students were asked their intended majors and they were grouped into four categories. All science, math and computer science majors were grouped together, as were business and economics and all humanities, arts and social sciences. The final category was students who were undecided. These were then divided up into students who were high, middle and low achievers in TOEFL for both Senior 1 and 2.

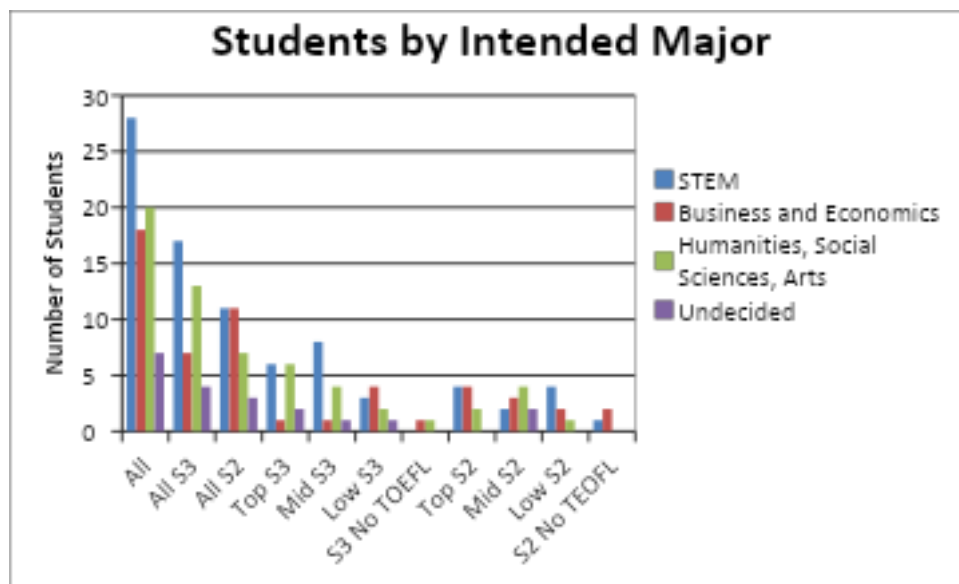


Figure 4.14 Students by intended major

The first take away is the prevalence of Science, Technology, Engineering and Maths (STEM) students. It seems reasonable that students who are especially interested in these subjects might be less likely to concentrate on an English test. Middle Senior 3 are particularly heavy on STEM subjects but Middle Senior 2 are actually heavier on humanities and social sciences and this is the group with the almost nonexistent association between TOEFL and APGPA. Top Senior 3 is almost equally divided between STEM and Humanities and that has one of the strongest associations between TOEFL and APGPA.

A similar question should be asked about why students are planning on studying in Western universities. Students were asked an open ended question about their motivations. All responses were grouped together into what appeared to be similar responses, for example, "Best

universities in the world” would be grouped with “good education method” under the umbrella term of quality of education. The responses lent themselves to this kind of sorting and there were no cases where it was unclear how a response should be sorted. According to Creswell (2002) this is an effective way of subjecting these kinds of open ended questions to statistical analysis. By organizing the data in this way it will allow us to see which groups of students have which motivations and what effect that might have had on their beliefs and actions in other respects.

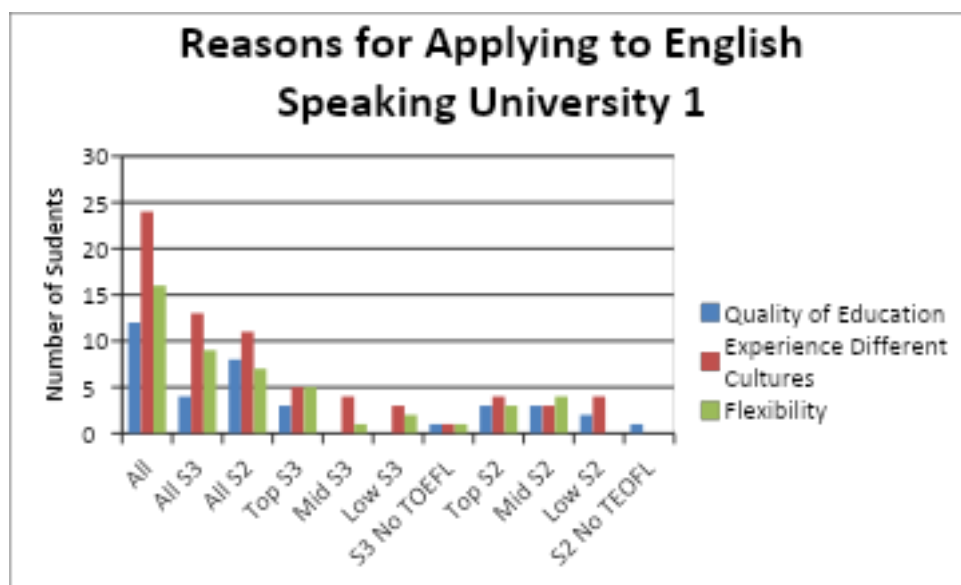


Figure 4.15 Reasons for applying to English speaking university part 1



Figure 4.16 Reasons for applying to English speaking university part 2

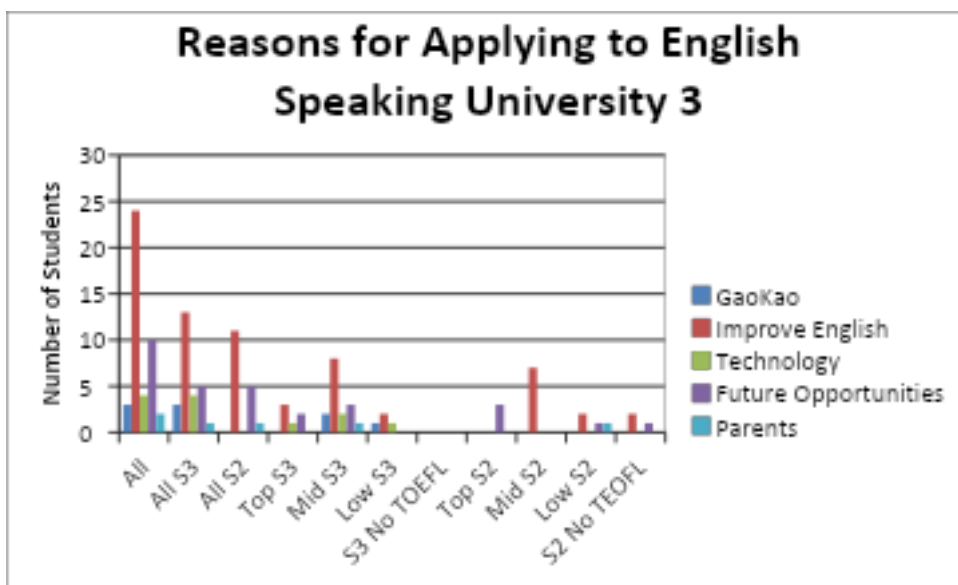


Figure 4.17 Reasons for applying to English speaking university part 3

It is worth considering the possibility of bias from an open ended question like this (Mertens, 2010, p. 189). There is a good chance that a student might answer the question in what they perceive to be the “right” way or in the way that makes them seem the best. Anonymity only goes so far as people are often unwilling to consciously answer a question in a way that might be true, but doesn’t paint them in what they perceive as a good light (Hanley,

2017). For example, considering the cost of attending college in Western universities (Cebolla-Boado and Hu, 2018), the intense competition for education in China (Du, 2013), the emphasis Chinese parents put on their children's education (Rajaram, 2013) and the prestige on Western diplomas when finding employment in China it seems questionable that a plurality of students really are most interested in studying at Western universities to experience different culture. This would be most useful as a tool for the further refinement of a survey. Another survey could be developed which would ask students to rate the significance of the top choices from this question.

One result worth noting is the number of unintelligible responses. These were responses which were not easily interpretable due to confusing English. These were overwhelmingly students who were of the lowest ability level which makes intuitive sense. This suggests that those who are of the lowest ability level according to TOEFL actually do have low ability (Chapelle, 2008).

In addition to looking at why students want to study at university it could also be useful to look at what these students are concerned about. Additionally it can be productive to see if there are any differences in concerns based on ability level according to TOEFL score.

The students were asked to rate their level of concern with a variety of potential problems they might encounter on a one to six scale with one being most worried and six being least worried. Three measurements were taken away from this question. The first is by averaging the results on the scale divided by different ability levels. So the lower the number the more worried the group of students is. This is similar to a method done by Malone and Montee (2014) where they averaged likert scale questions to show the extent of agreement with a question about TEOFL. The second is that the number of each choice will be examined to see the level of variation of choice. Is everyone worried about the same thing or is it more evenly distributed. It needs to be noted that a number of students filled out this part of the survey incorrectly, most commonly failing to rank the selections rather answering the question as a measure of degree. So many students showed they were very worried about specific multiple choices. These might

impact the results of the two measurements but it should inflate results rather than biasing them. In order to see if this is the case a third measurement will be an analysis of who made mistakes and how many.

| Ranked Level of Concern |    |                                 |                             |                                                       |                               |                         |                       |
|-------------------------|----|---------------------------------|-----------------------------|-------------------------------------------------------|-------------------------------|-------------------------|-----------------------|
| Students                | N= | Q1<br>Understanding<br>Lectures | Q2<br>Unfamiliar<br>Accents | Q3<br>Speaking<br>to<br>teachers<br>and<br>classmates | Q4<br>Amount<br>of<br>Reading | Q5<br>Writing<br>Papers | Q6<br>Taking<br>Tests |
| All                     | 75 | 3.8                             | 3.5                         | 3.7                                                   | 3.2                           | 3.1                     | 2.9                   |
| All S3                  | 44 | 3.9                             | 3.5                         | 3.3                                                   | 3.3                           | 3                       | 3                     |
| All S2                  | 31 | 3.7                             | 3.6                         | 4.4                                                   | 3.1                           | 3.2                     | 2.7                   |
| Top S3                  | 15 | 4.5                             | 3                           | 2.9                                                   | 3.6                           | 3.2                     | 4.3                   |
| Mid S3                  | 15 | 3.7                             | 3.9                         | 3.4                                                   | 3.2                           | 2.7                     | 2.6                   |
| Low S3                  | 12 | 3.4                             | 3.4                         | 3.9                                                   | 3.1                           | 3.1                     | 2                     |
| S3 No<br>TOEFL          | 2  | 4                               | 5                           | 1.5                                                   | 3                             | 3.5                     | 4                     |
| Top S2                  | 9  | 4                               | 4.1                         | 5.2                                                   | 2.5                           | 2.9                     | 3.4                   |
| Mid S2                  | 12 | 4.3                             | 3.6                         | 4.3                                                   | 3.3                           | 2.8                     | 2.3                   |
| Low S2                  | 6  | 3                               | 3.2                         | 2.4                                                   | 3                             | 3.5                     | 2.6                   |
| S2 No<br>TEOFL          | 4  | 2                               | 2.5                         | 5.25                                                  | 3.75                          | 4.25                    | 2.5                   |

Table 4.1 Ranked Level of Concern

Based on the averages observed, understanding lectures, Q1, is of little concern to the top Senior 3s, top and mid Senior 2. This could show the students are confident in their listening ability. Low Senior 2s are much lower than their class average and Low Senior 3s are somewhat lower than their class averages. This could suggest that these students are having a lot of trouble in class now and on TOEFL listening. Unfamiliar accents, Q2, are of little concern overall. Question 3, Speaking with teachers and classmates is all over the map. Senior 2s seem to have no issue of any kind with this while it is of greater concern to Senior 3 and particularly top Senior 3. The final three questions seem to be of greatest concern with amongst all students. This suggests that the students are think the biggest issue at university are reading, writing papers and taking tests. Considering how confident these students have shown themselves to be about tests, this could suggest that they are over emphasizing the significance of testing at university. The fact that the facts that they are worried about reading and writing papers, despite their standardized tests, could suggest that they don't feel the tests have prepared them or just that they are not prepared generally.

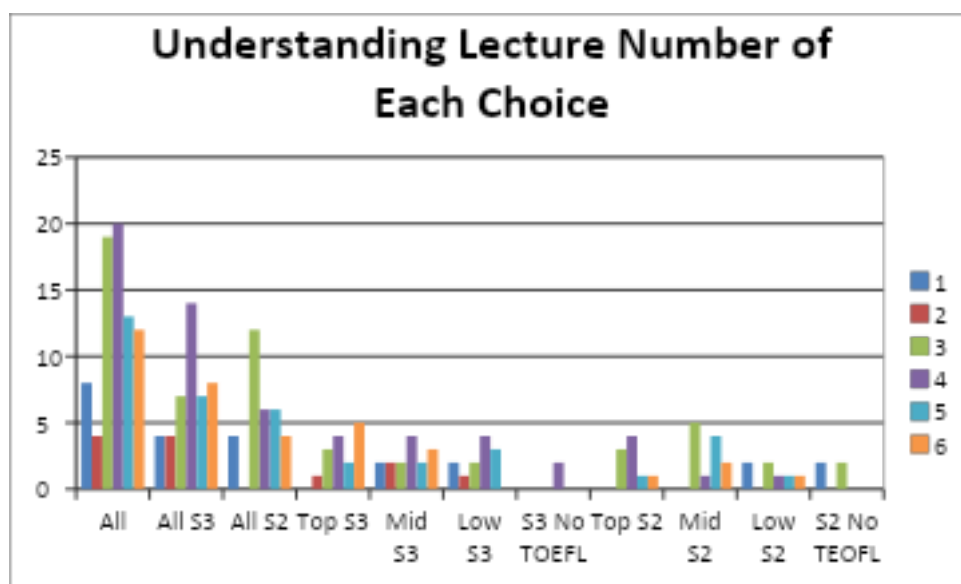


Figure 4.18 Understanding lecture number

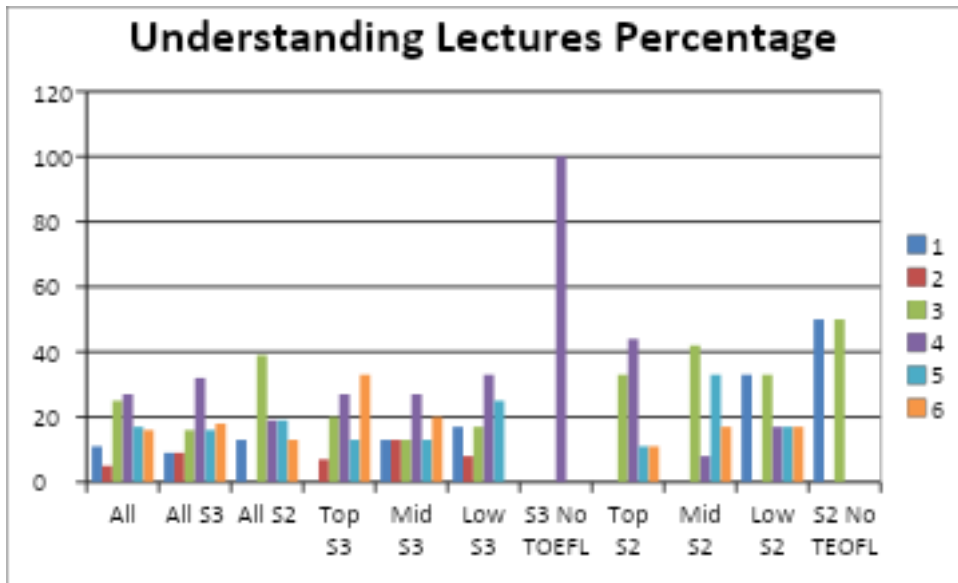


Figure 4.19 Understanding lecture Percent

These are how students ranked and what percent of students ranked understanding lectures as a concern.

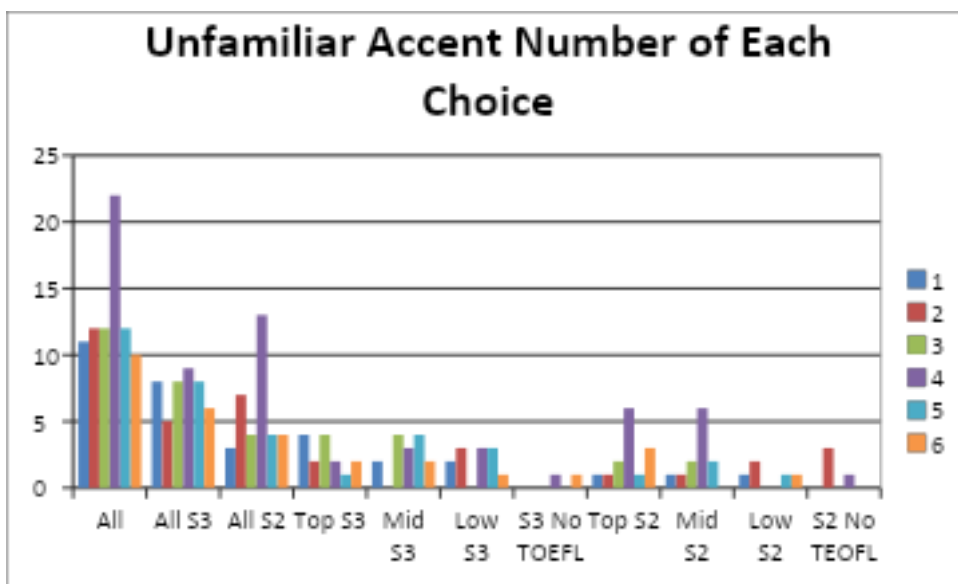


Figure 4.20 Unfamiliar Accent number

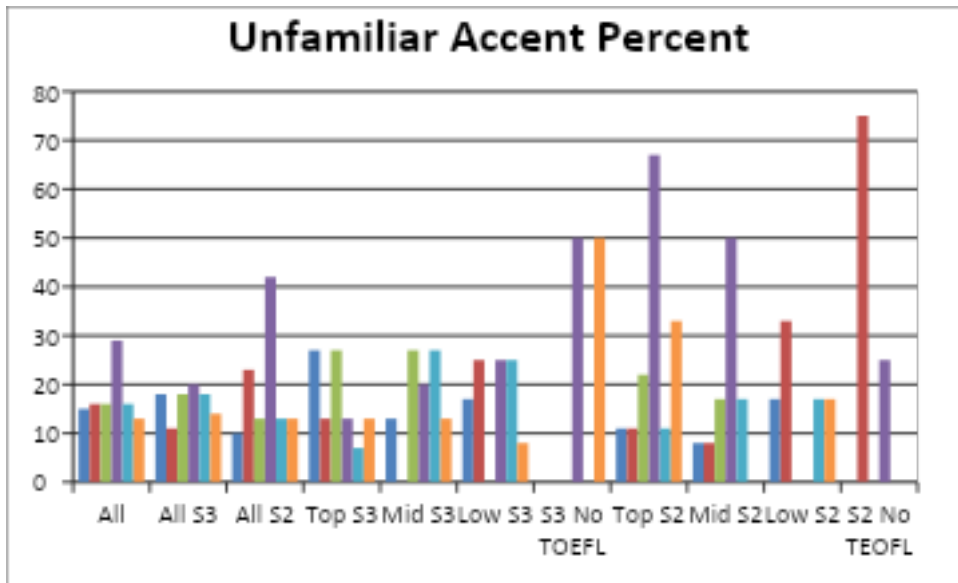


Figure 4.21 Unfamiliar accent percent

These are how student ranked and what percent of students ranked unfamiliar accents as a problem.

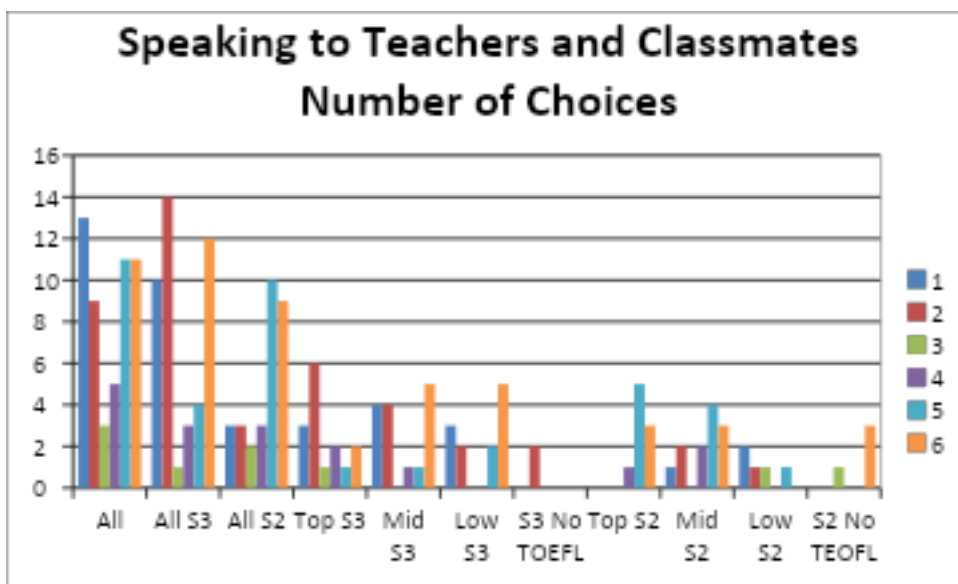


Figure 4.22 Speaking to teachers and classmates number

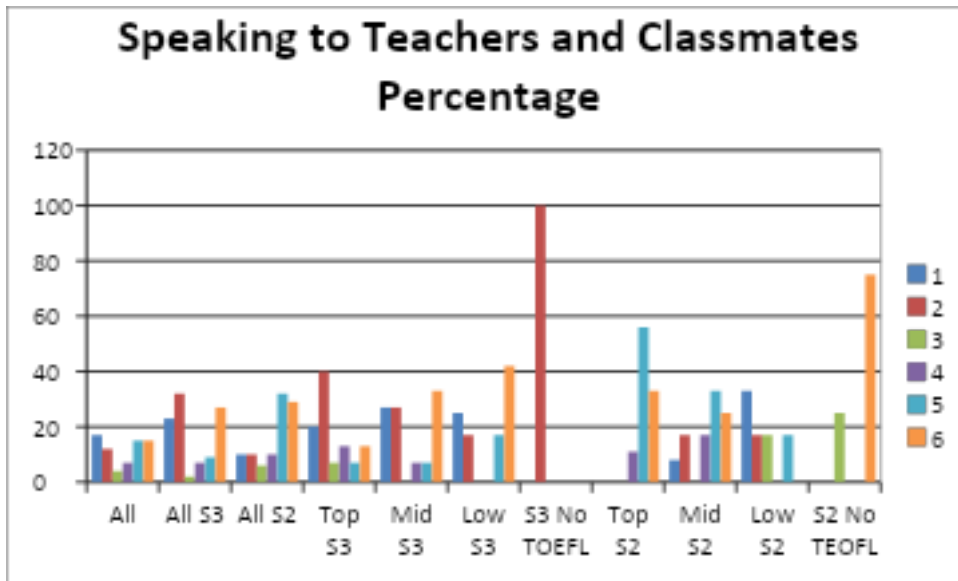


Figure 4.23 Speaking to teachers and classmates percent

These are how students ranked and what percent where concerned with speaking with teachers and classmates.

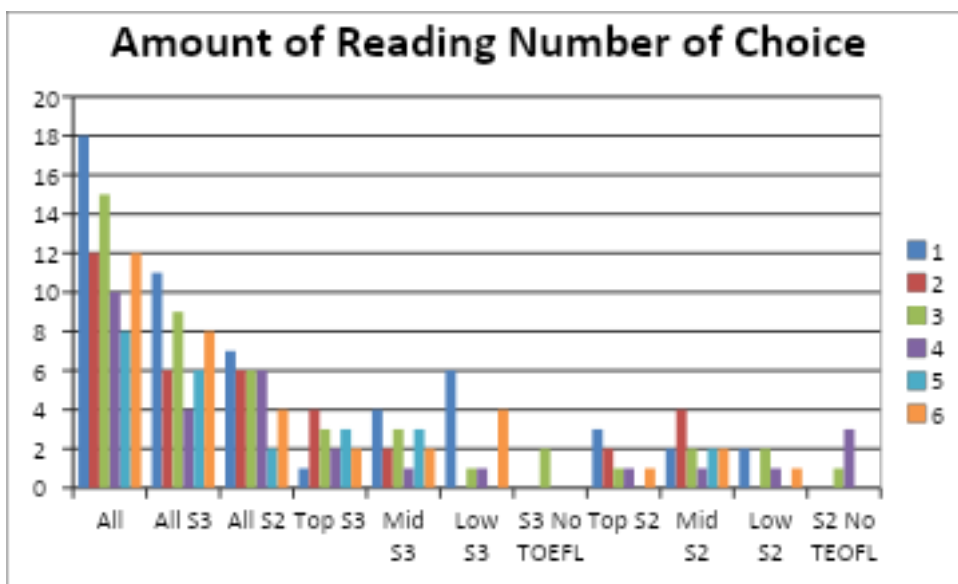


Figure 4.24 Amount of reading number

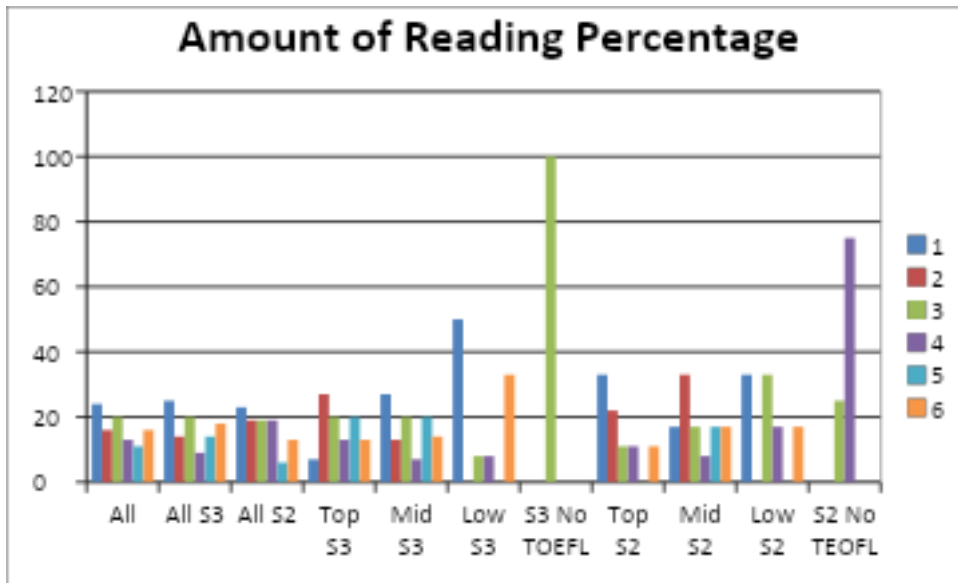


Figure 4.25 Amount of reading percent

How and what percent of students ranked the amount of reading as a concern.

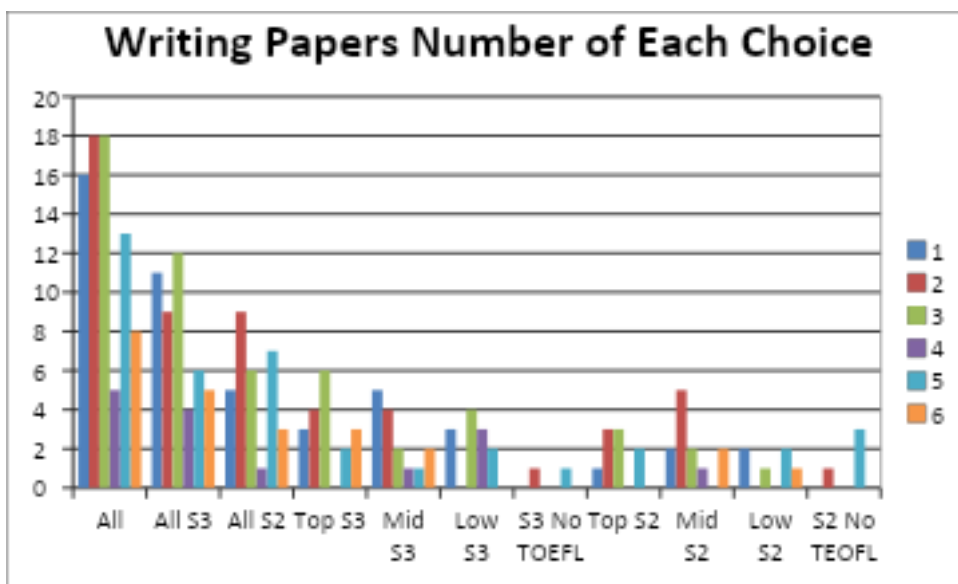


Figure 4.26 Writing papers number

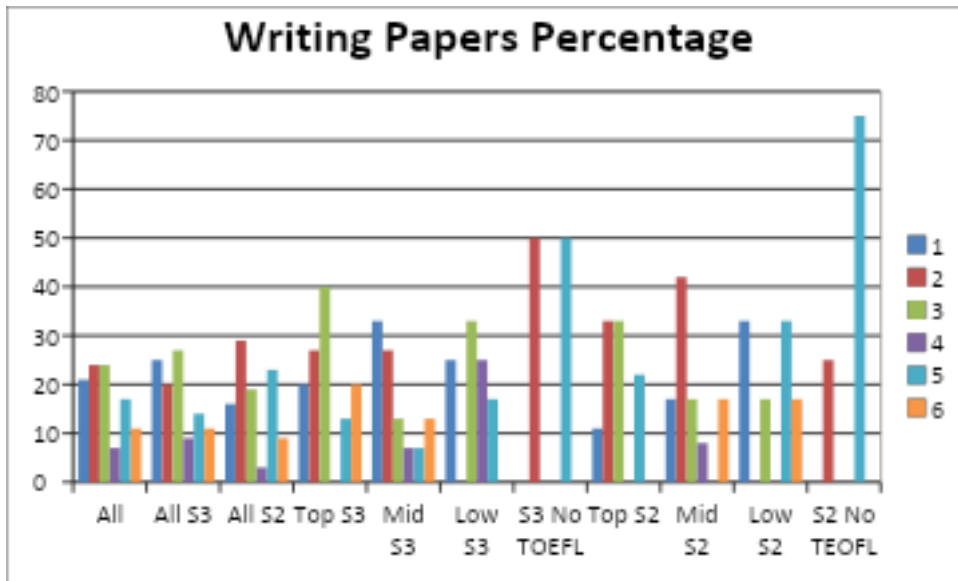


Figure 4.27 Writing papers percent

Writing papers as a total number of students and what percent of students who were concerned.

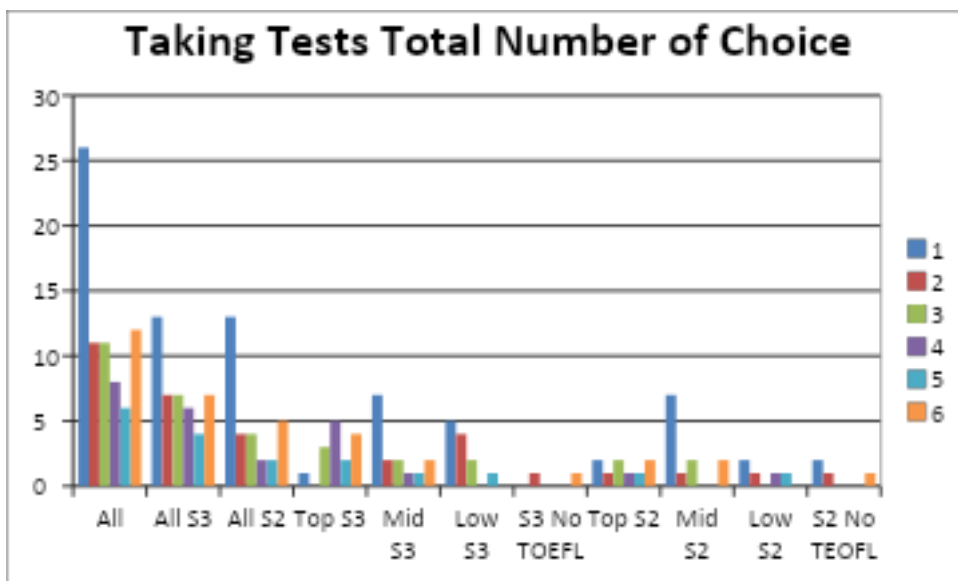


Figure 4.28 Taking test number

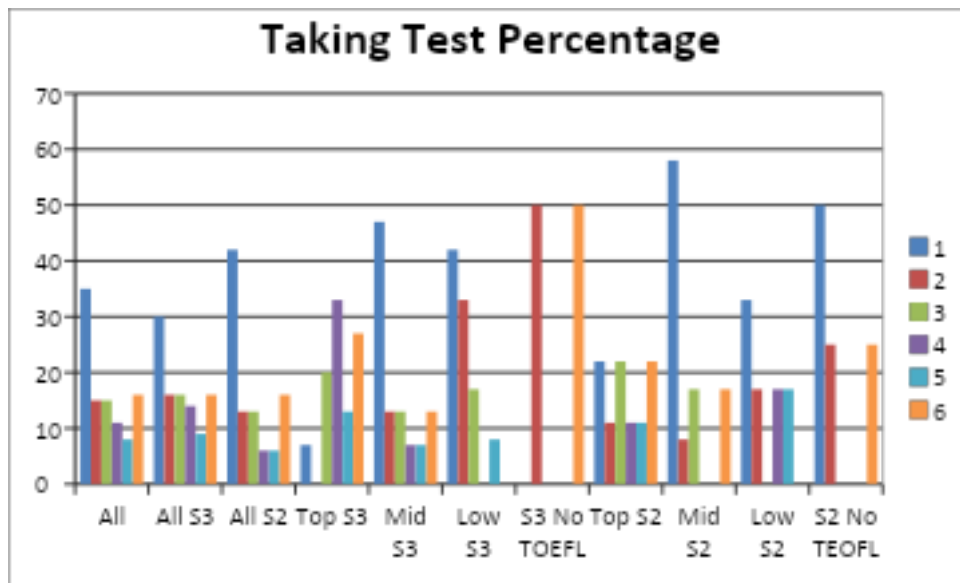


Figure 4.29 Taking test percent

Number and percent of students concerned with test taking.

The results of the ranking survey question have been illustrated by Figures 4.18 to 4.29. Each question has been illustrated according to the different groups of students both as a total number of each choice as well as a percentage of each group of students. So we can see not only how many Top Senior 3 ranked something but also which percentage of Top Senior 3 made that choice.

If both the average and the number and percentage of ranked are similar that could give further evidence of its importance and if they are different this could lead to different interpretations. Comparing these results to the averages finds a similar result for question 1; students are not particularly concerned about understanding lectures. So students could think that TOEFL listening has prepared them sufficiently or that their experience with lectures in their high school has prepared them for university. On the other hand the students could think that lectures are not important and they can get by using other study methods. Unfamiliar accents are fairly evenly distributed showing no particular pattern amongst groups of students. Speaking to teachers and students shows either great concern or little concern from all students so students are either very worried or not at all worried. Results are similar to the average for this question with Senior 3 much more worried than Senior 2. Similar results as averages are found for reading

and writing suggesting that the student don't feel confident about either skill. Test taking was ranked highest among all students. Interestingly, taking tests was ranked very high, the number one priority, for both Mid Senior 2 and 3. This could reinforce the idea that these students are particularly test focused and could be more inclined towards test wise strategies.

These results raise further questions about what the students are thinking about their education in general and the TOEFL in particular. In the next section a series of questions about the test were asked on a likert scale to elicit student opinion.

#### 4.6 Student beliefs about TOEFL

Similar to the averaging method in the ranking survey question, the Likert scale questions have also been averaged. Strongly agree was numbered as 1 and strongly disagree was rated as 4. Thus after a question has been averaged, the lower the result the greater the level of agreement has been found.

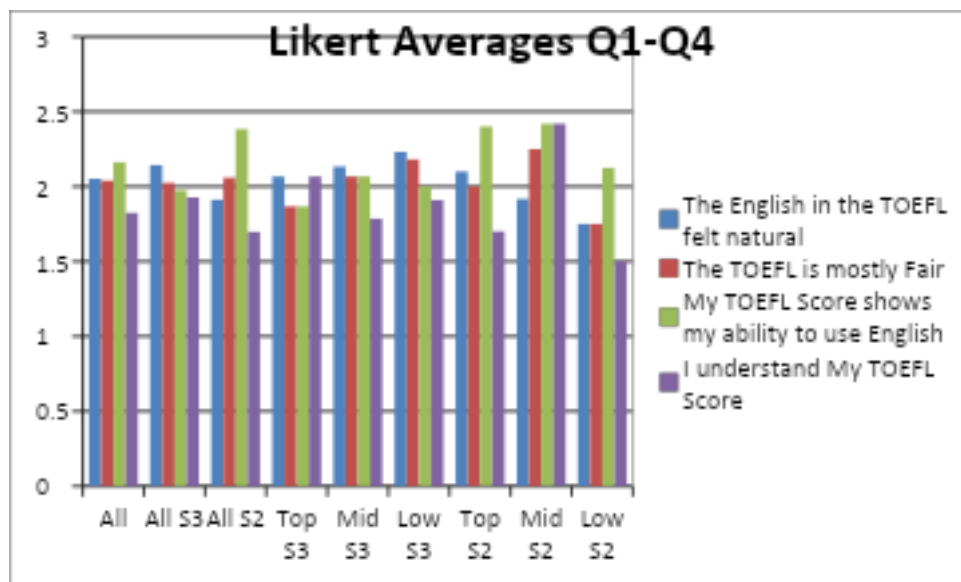


Figure 4.32 Likert averages Q1-Q4

Understanding TOEFL score was universal agreed upon. During interviews it was commonly commented upon that ETS provides very detailed breakdowns of student's results.

This gives students very specific guidance on what they need to be studying in order to improve their score, though not necessarily how to improve their English ability. The 'Did the English in TOEFL feel natural' question was directly adapted from Malone and Montee (2014) and the results were similarly agreed upon. The students of all levels consider the TOEFL fair, this is an interesting result as from the interviews as all students brought up the fact that the use of test wise strategies can lead to significantly higher scores. Combine this with the results of 'The TOEFL showed my ability to use English' and a complex picture develops. Students can think that the test is fair in one sense, perhaps that it is a reliable measure for universities while not actual being a valid measure of their English ability. So the test could be reliable while not being valid. This result runs contrary to ETS' validity argument (Jamieson et. al 2008) as well as the studies arguing about the importance of students beliefs about tests (Oroujilou and Vahedi, 2011; Jin, 2011; Xu and Gao, 2014). If students don't think the test reflects their abilities than the validity of the test could be called into question. This runs contrary to Cheng and de Luca (2011) who found that the students who questioned the fairness of the test were more likely to use test wise strategies, but their data were from open ended interviews which corroborates the open ended interview data form this project

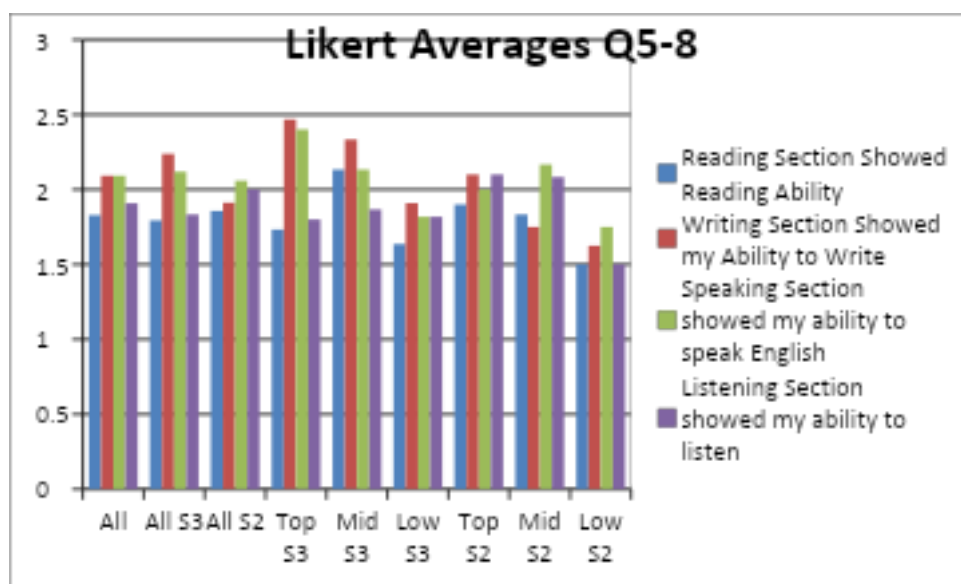


Figure 4.33 Likert averages Q5-8

The survey breaks this down further to see if the students have specific problems with

the tests English. The average for all students was around 2, meaning general agreement. Writing and speaking have the lowest approval. The speaking test has been commented up by ETS, Ling (2014) as there were concerns about the unnaturalness of speaking into a computer which runs contrary to communicative English (Chapelle, 2008). This was also commented upon by all students during interviews.

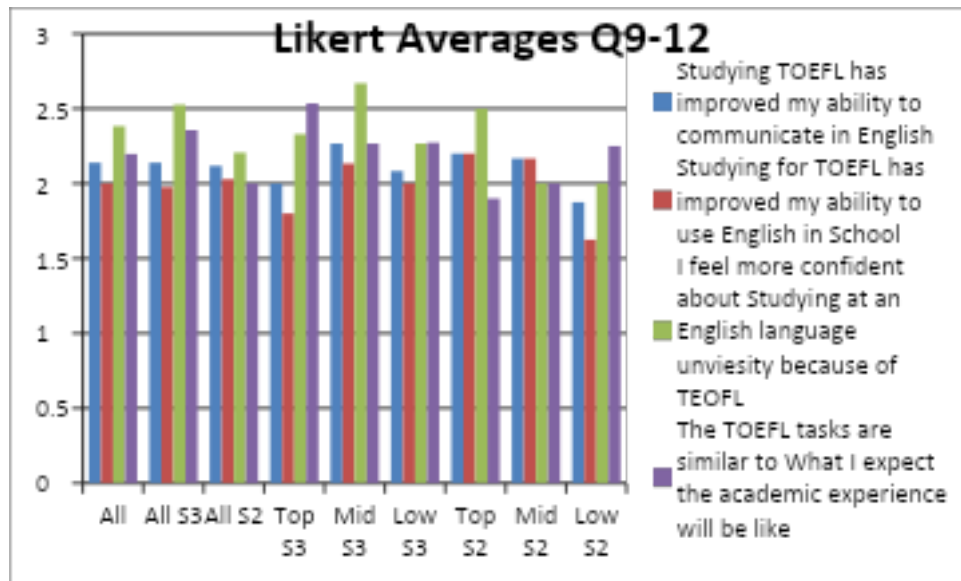


Figure 4.34 Likert averages Q9-12

While still close to the agreement, it is worth pointing out that Mid Senior 2 and 3 both are less likely to say that their ability to communicate in English has not improved as a result of studying for TOEFL. This corroborates the idea that these are students who are using test wise strategies as they won't necessarily gain communicative English as a result of their study. This is a point against TOEFL's argument for validity as they have consistently argued from the point of view of communicative English (Jamieson, 2008; Taylor and Angelis, 2008). There is general agreement that the TOEFL has improved academic English however the strongest disagreements found were about whether the student are more confident about studying at an English language university. If the students thought that their TOEFL score was an honest assessment of their English ability and their studies have improved that ability why wouldn't it make you more

confident about studying at university? This leads to the final two sections of the survey which are specifically looking at what students are doing to study for the exam which should give specific data about how beliefs and attitudes may have been converted into actions.

#### 4.7 Student Study Techniques

This part of the survey was divided into two parts. First are types of preparation techniques which students could check. As they could check as many as they had practiced, most surveys had many techniques checked. As such the data for this will be presented as a percent of each group presented. This will not show how significant these choices are, just how many had done them. As such students were also asked to provide a list of the student's perceived most important preparation techniques. The interviews with counselors and students were useful for grouping open choices and determining which choices were test wise strategies and which were designed to improve general English skills.

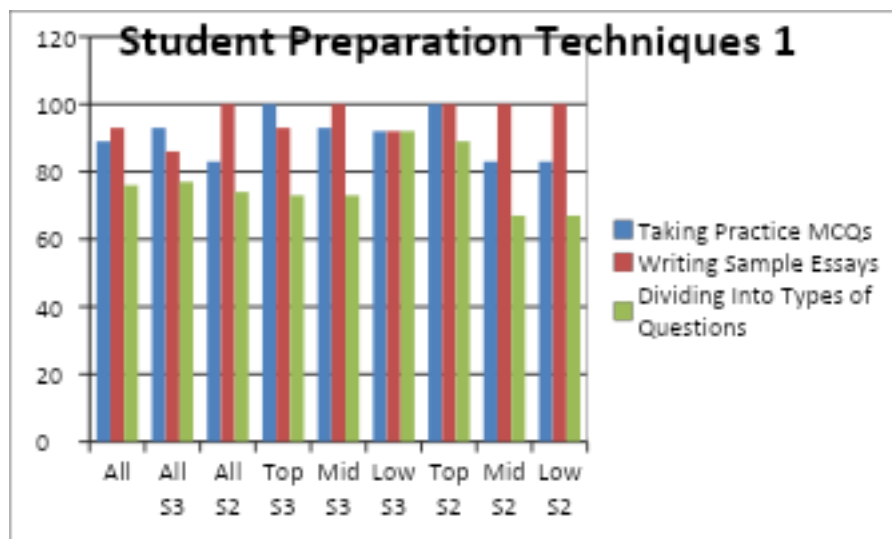


Figure 4.35 Student preparation techniques 1

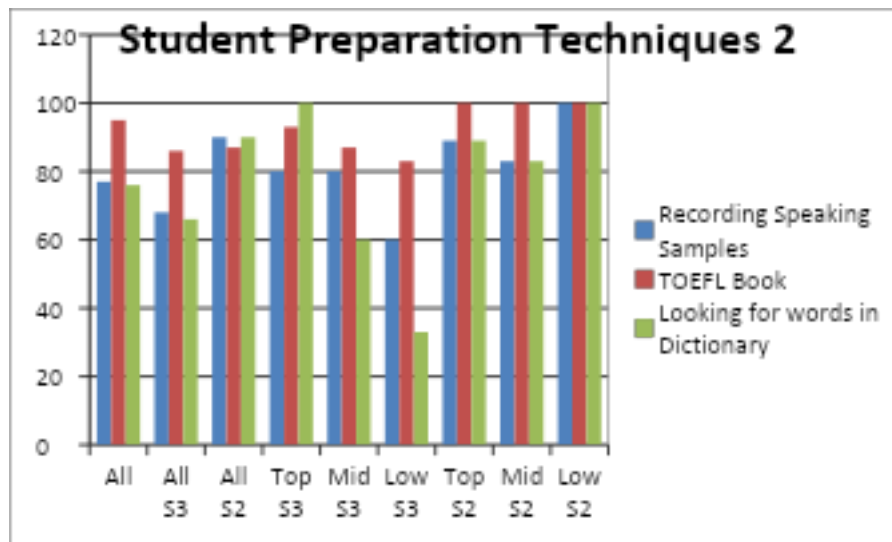


Figure 4.36 Student preparation techniques 2

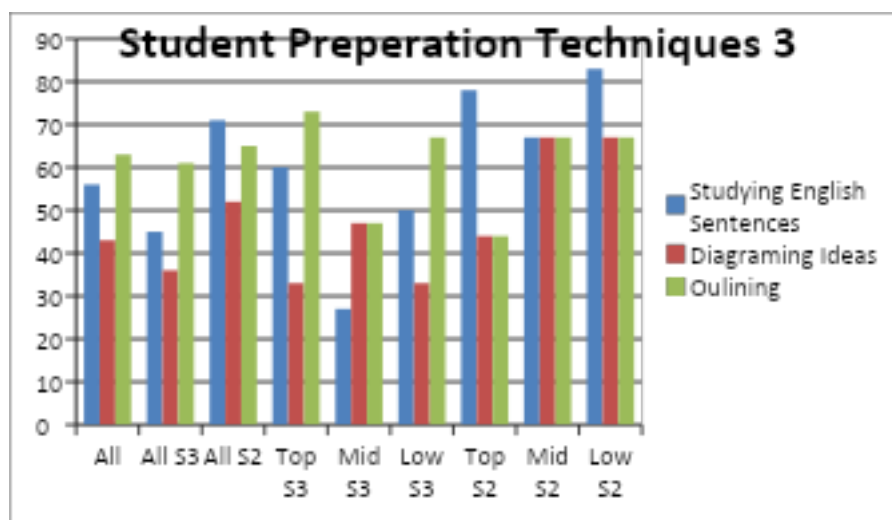


Figure 4.37 Student preparation techniques 3

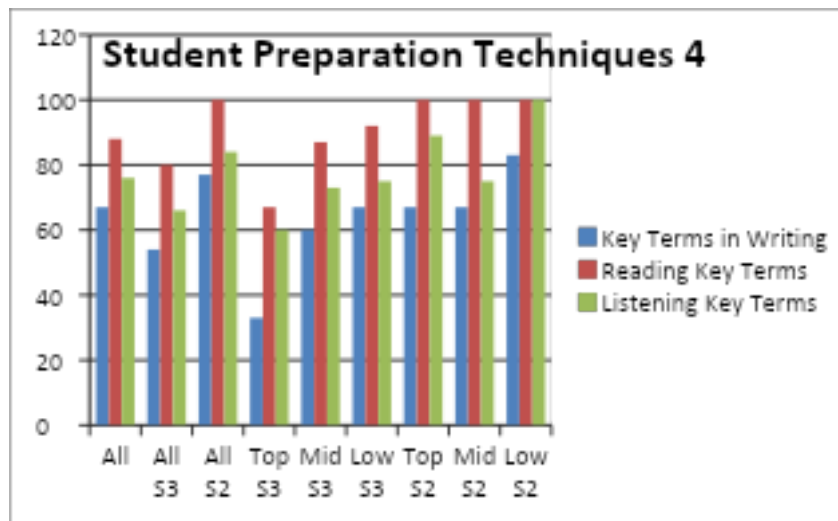


Figure 4.38 Student preparation techniques 4

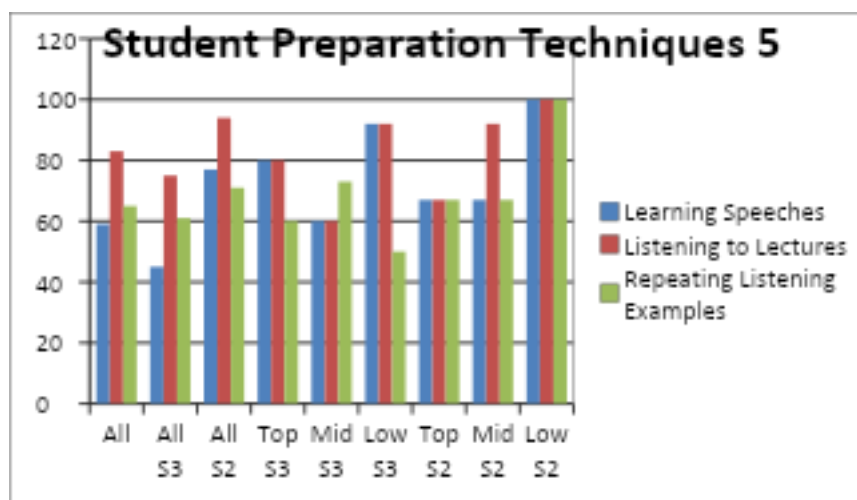


Figure 4.39 Student preparation techniques 5

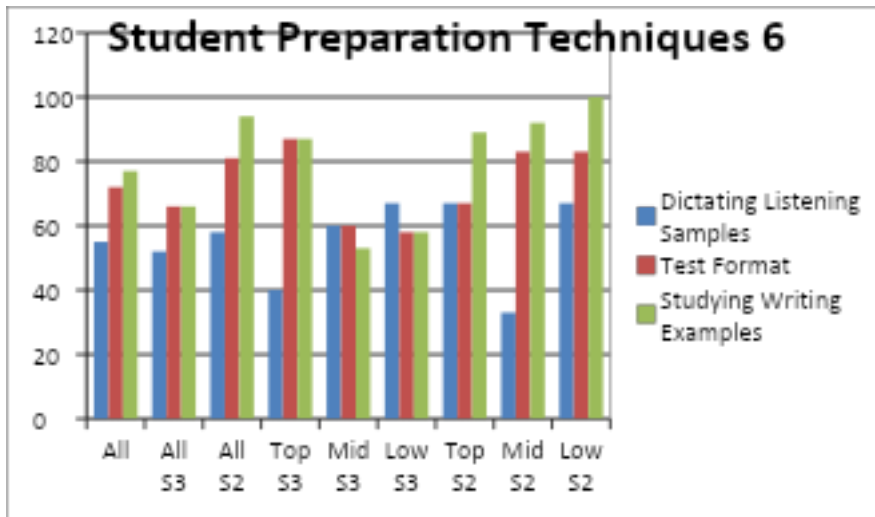


Figure 4.40 Student preparation techniques 6

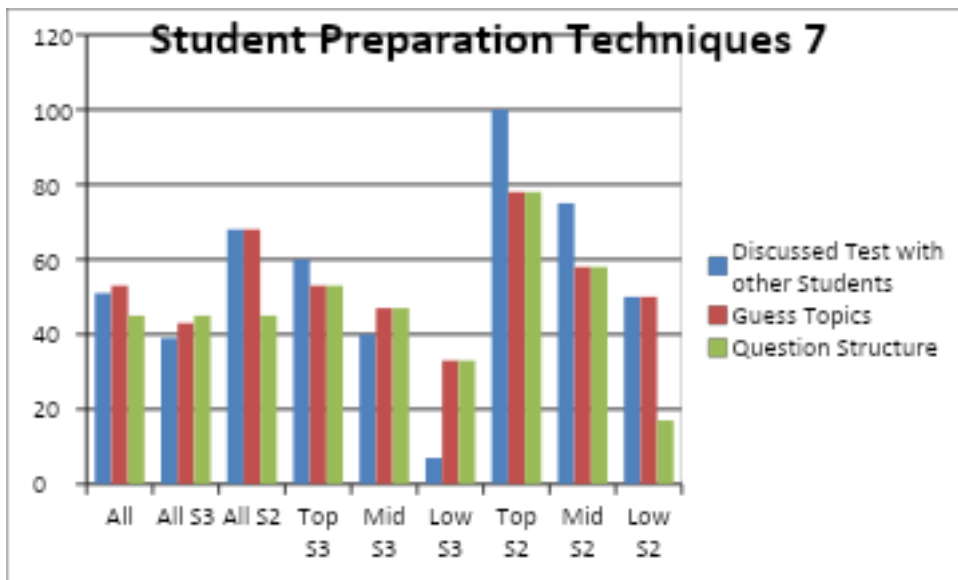


Figure 4.41 Student preparation techniques 7

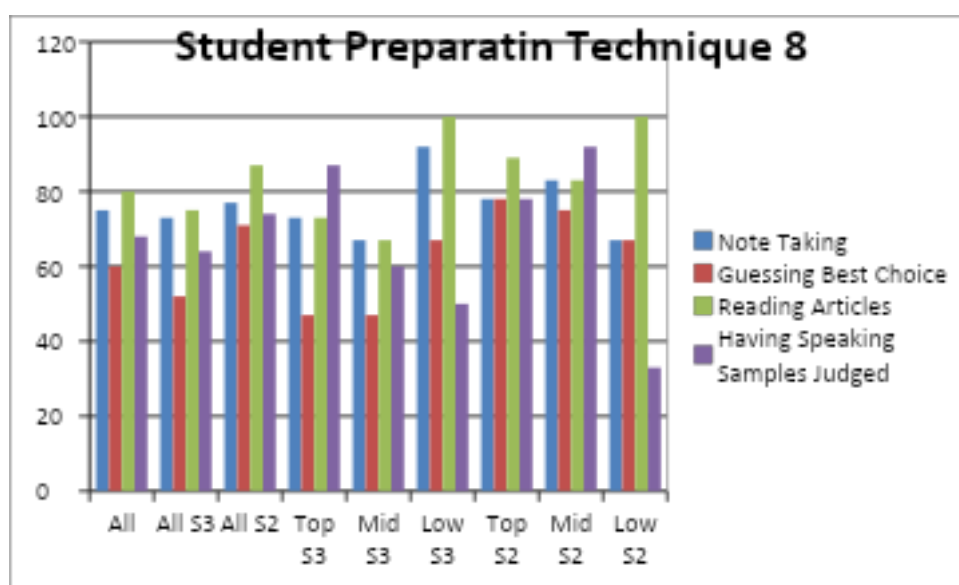


Figure 4.42 Student preparation techniques 8

Overall it is difficult to see much of a pattern strictly looking at this data. As Pan (2016) showed, students can use a variety of techniques for test preparation both test wise and to improve English skills. During the student focus group several students expressed the opinion that some of these techniques were unethical, suggesting students might underreport these kinds of behaviors. When asked about test wise strategies, both students and counselors said they were extremely common but this isn't necessarily seen in the data. For example students and counselors cited guessing techniques as being one of the main topics taught at training schools which almost all of the students attended. Hong and Peng (2008) and He and Shi (2008) commented on the prevalence of these techniques but also that the students knew they weren't considered to be the right way to study for the test. Considering that the survey was administered by the student's English teachers that might have impacted their answers.

From the results we can see a variety of techniques as being popular. Some are clearly test wise, learning speeches at nearly 60% of all students. Some are clearly techniques aimed at learning English, reading articles at 80%. On the other hand some could be either test wise or learning based depending on the student. Doing multiple choice questions can be a valid technique for improving English, but if it is done in a rote way merely learning how to answer

TOEFL questions than it would be a test wise strategy. The TPO, TOEFL Practice Online was very strongly emphasized in all interviews as being a favorite study technique and the way it is used by students, repeatedly doing the same questions over and over again, suggest it should be categorized as a test wise strategy. Similarly using a TOEFL study guide isn't inherently test wise if the practices are designed to encourage language acquisition, but based on interviews the primary use of the books, memorizing lists of vocabulary with their Chinese translations, suggests it is an extremely test wise strategy. The next step then is to look at the strength of beliefs about the efficacy of test preparation techniques.

Students were asked to name the three most important study techniques for doing well on TOEFL, either one mentioned in the survey previously or one of their own choices. The results were divided into 28 categories, though most only had a few responses. All results can be found at Appendix E. The most popular, anything with more than five responses, results is examined here in detail. The data has been illustrated by showing total number of selection for each choice and total number per student.

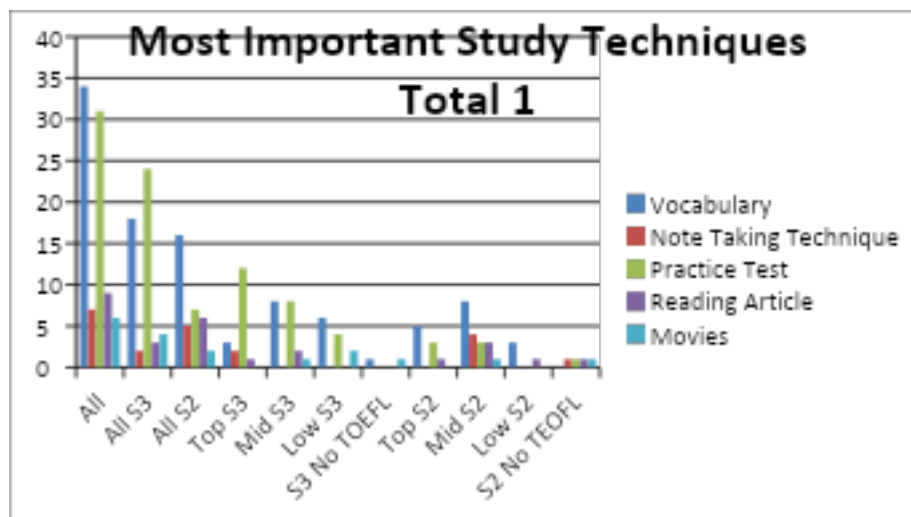


Figure 4.43 Most important study techniques total 1

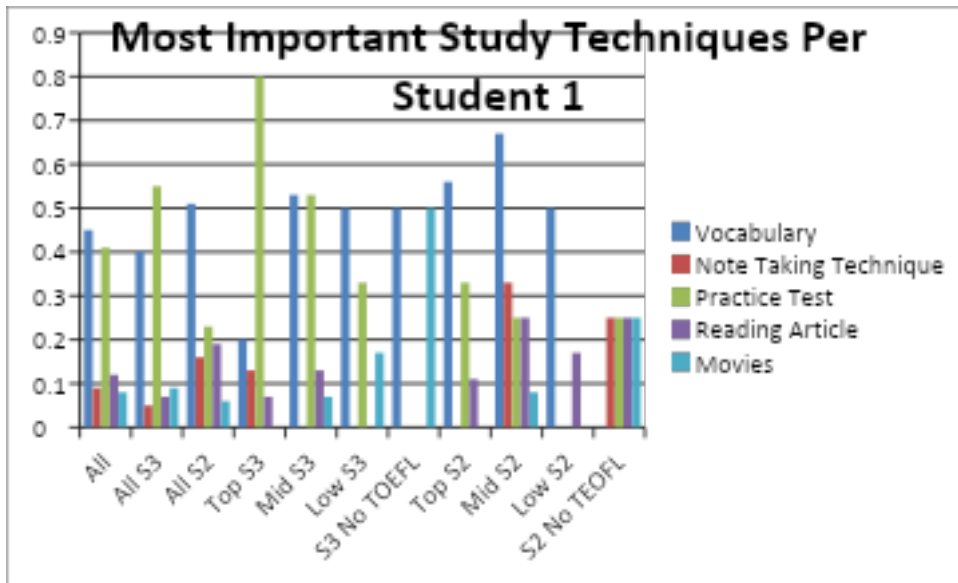


Figure 4.44 Most important study techniques per student 1

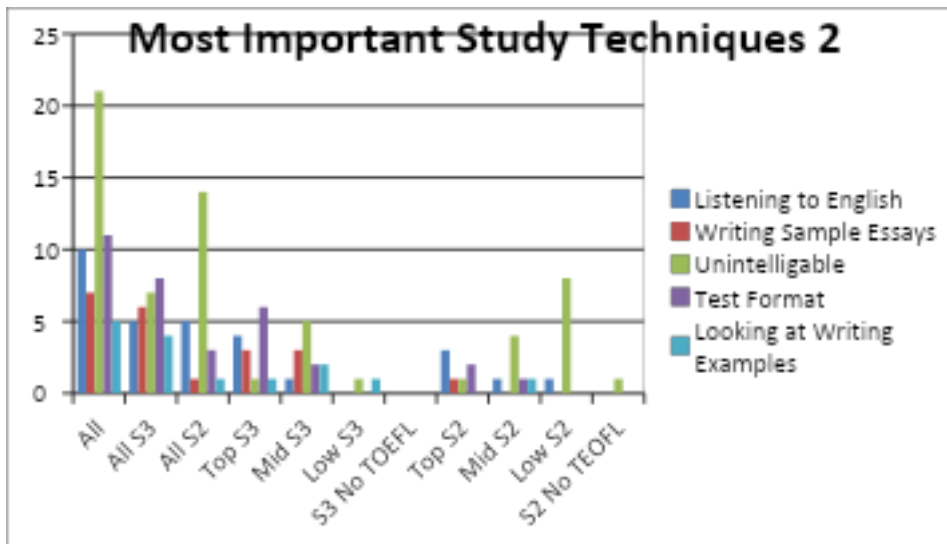


Figure 4.45 Most important study techniques total 2

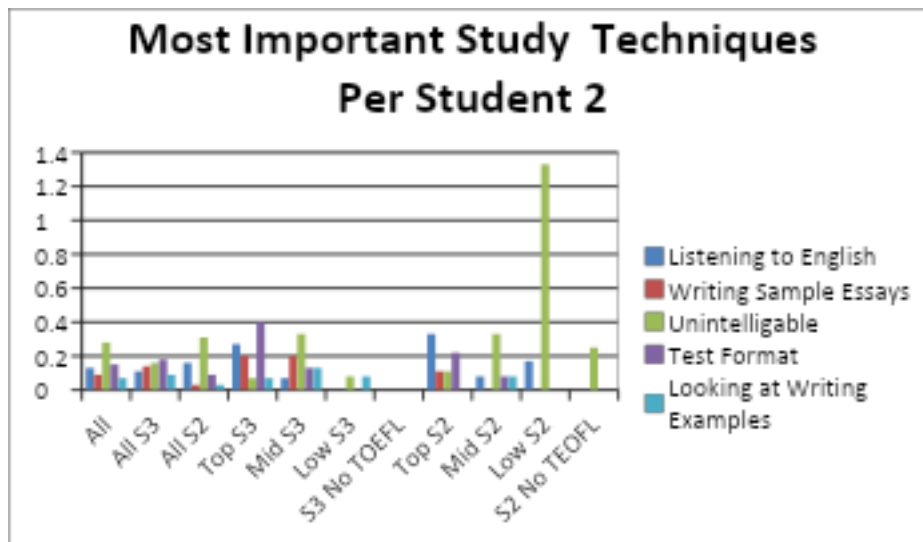


Figure 4.46 Most important study techniques per student 2

The biggest take away from looking at student selected most important study technique is the popularity of studying 'Vocabulary.' During counselor and student interviews this was specifically asked about and the unanimous consensus that students who wrote down Vocabulary or TOEFL book were specifically referring to memorizing the translation of vocabulary words. This was an unsurprising result for anyone with experience with Chinese high school students as the books in question are omnipresent and the students study technique is also very obvious. Looking at the data in more detail it seems particularly significant that Middle Senior 2 and 3 are both more reliant than others on 'Vocabulary'. Taking practice tests are also very popular and as discussed earlier is very likely to be used in a test wise way. It is possible that the top students are more likely use practice tests in a more educational sound way and so it is used more test wise by weaker students. Lei (2015) and Metzger, et. al (2018) top students are better able to use more sophisticated study strategies and thus use practice tests in an educational sound way.

These results will be classified into test wise, learning based or dependent on student. Vocabulary, Test Strategies and Test Format are test wise, Practice Tests and looking at writing samples are dependent on students and Listening to English, Writing Sample Essays, Reading Articles and Movies are considered learning based. From this data it is clear that test wise strategies are popular all around and considered very important by the students. It is also clear

that as the student's ability level declines the test wise strategies are more important. It is also possible that the student dependent results are more likely to be test wise depending on the student's ability level.

It is worth noting that most of these techniques relate to the reading and listening sections. During interviews when the writing and speaking section came up several study techniques came up, most of which were clearly test wise strategies. For example, students mentioned memorizing passages and writing generic essays that could be applied to any situation. As it stands, more information is needed about this as these techniques were too ill defined on the survey and were not prioritized by students enough to analyze their prevalence. This could be because they are not common or it could be that the students do not consider the writing and speaking sections as significant. Additionally, of all the techniques discussed so far these are the ones which are most likely to be perceived as plagiarism and thus considered more unethical. Copying essays could be seen by the students as particularly unethical and thus less likely to be admitted to.

From here a number of conclusions will be examined based on this project's research question. The limitations of the study will be considered and some avenues for future research will be explored.

## 5.1 Introduction

This case study has been conducted in a way to get quantitative data that is then interpreted with the lens of qualitative data with the goal being to use this case as an example that may be generalizable to other cases in further research (Campbell, 2015). This research has taken multiple instruments to try to get to the heart of a series of research questions concerning high stakes testing in general and the TOEFL in particular. The results have been presented in a way in which to clearly illustrate the questions which the project concerned itself with and allow others to draw additional conclusions as a quality case study analysis should (Yin, 2009). As Atmowardoyo (2018) shows, an example can provide a clear insight into one particular situation which may relate to other situations and encourage learning and further research.

The intention was by gaining insight into these students' beliefs and habits the need for more information could be shown. It was not believed that the TOEFL would be invalidated or a need to make major changes in how English as a second language is taught would be found. However, this research came from a place of genuine concern about how standardized tests are used and their role in international education. This research some of the assumptions many teachers who have experience with international education have about the system and about how it affects students beliefs, actions and preparedness for English university (Malone and Montee 2014; Li 2017; He and Shi 2008) have shown themselves to have some truth to them. The limitations of this will be clearly illustrated as well as a number of avenues of research that could provide an even clearer picture.

## 5.2 Conclusions by Research Questions

### 5.2.1 What do Chinese students believe the TOEFL iBT measures?

This group of students has a great deal of experience studying English, 12 and 10 years for Senior 3 and 2 respectively. In that time they have experienced both the Chinese, exam driven education system ( Cheng and Curtis, 2010) and have attended international school with high stakes tests, the US Advanced Placement subject exams as well as attending training centers, 80 to 90%, focusing on preparing for TOEFL and SAT. They also have a lot of experience with the test having taken it an average of four times.

The TOEFL was designed with academic English and communicative English in mind (Chapelle, 2008) and the students overall seem to have a positive outlook on the test, rating is highly for both academic and communicative English. They said that the test felt natural, the same as Malone and Montee (2014) suggesting they think it should measure their ability to communicate. During focus groups and interviews the students emphasized that the test was primarily academic and some parts, speaking in particular did not seem to relate to the ability to communicate in English. The students in general seemed to realize that the test with its integrated tasks and choice of topics and was intended to mimic real situations in university study (Enright, et. al 2008).

When looking at the what are you most concerned about studying at an English language university questions , understanding lectures was not seen as a major concern for students suggesting that the students think they have been well prepared for listening by the TOEFL. On the other hand, the student's writing papers and reading are of major concerns to almost all students suggesting they do not feel prepared. If that is the case it seems more likely that the students don't think lectures are important than they think they are ready for them. Considering that the number one choice for all students was taking tests, this could suggest that they think the TOEFL is a good model for what they will experience at university. While the specific tasks in the test are modeled after university lectures and readings the students are unlikely to experience high stakes standardized tests unless they plan to apply for post graduate education programs, i.e. law or medical school. So students are likely to think that the TOEFL is reflective of university experience, but do they think that it prepares them for university?

### 5.2.2 To What Extent do Chinese Students think the TOEFL iBT prepares them for Western University?

During the focus group and interviews students expressed numerous concerns about studying at an English speaking university both academically and socially. Overall they indicated they thought they would be successful at an English speaking university. Looking at the students reasons for applying to Western university, one of the most common choices, 25 students, was to improve English. This could suggest that the students feel prepared but also feel the need to continue their education, that is, the TOEFL is necessary but not sufficient.

To return to the Likert scale rankings of student's beliefs, the students say that the test reflects their ability to use English, so it would stand to reason that if you are successful at the test then you must be able to be successful at an English speaking university. However, the students rated it higher for academic English than any of the communicative category questions. This could suggest that they think it is reflective of future academic English study or study generally at university, except that the most poorly rated category for students was about their level of confidence about studying at Western universities because of studying for TOEFL. This could suggest that on an intellectual level they understand that the TOEFL is representative of tasks you might find at university, but they also suspect that the actual university experience is not like taking a TOEFL test.

Looking at the concerns about studying at English speaking university, the students may have an inaccurate view of what they will encounter at university. The overwhelming concern about taking test suggests that the student's erroneously believe they are prepared for English speaking university. Tests certainly are a part of the university experience, but it is unclear how the TOEFL could prepare students for research writing. The students do have a clear concern about the quantity of reading; however that could be a concern of any student preparing for university, not necessarily related to TOEFL reading.

### 5.2.3 How do Students Vary by Ability Level as defined by TOEFL Scores?

The most notable finding of this work is how clearly different kinds of students respond to standardized testing. As discussed earlier you would assume that generally good students should also do well on the TOEFL, medium students should do medium and poor students should do poorly. However, as we have seen in this context, good students and poor students make sense with their TOEFL and GPAs correlating, but those middle range students are all over the map. If the test scores and the grades do match than either the test or the grades are lacking in construct validity (Howard, et. al., 2016). While it is possible the grades at the site have a validity problem, the school has a student centered holistic learning policy which should provide a fuller picture of the students actual abilities (Bradford, 2016). Standardized tests on the other hand are coming under more and more scrutiny for their negative effects on students learning and the educational system as a whole (Nichols, 2012; Huddleston, 2015; Lewis and Hardy, 2015). As Dixon (2018, p. 46) says, "Rote memorization nearly always trumps critical thinking on standardized tests."

The argument is that these middle ranked students are distinguished from top students by their study attitude and from bottom students by their ability which leads to this divergence between APGPA and TOEFL results. As one of the counselors said in reference to these students, "I think they can understand and they can use English in the most of their times when they write the research paper they will organize their own ideas they still feel a little afraid of organizing their own language" (Appendix G, p. 146). During the focus group one student mentioned that middle students might have made a choice, concentrating on TOEFL over APGPA. That is they can use the language, but they don't feel confident to express themselves, they are looking for the "right" answer not trying to communicate information.

The best evidence for the influence of study attitude from the quantitative data comes from the students concerns about their future studies. The fact that the these students are so overwhelmingly test focused suggests that they are focusing their attitudes towards improving

test scores and think that this will be their best strategies going forward. While all students took the TOEFL multiple times and most students improved their scores as a result, this is particularly true of middle Senior 2 and 3. This could potentially reinforce these students' attitudes towards test taking acting as positive reinforcement for test taking strategies. This leads the next question about what these students do to prepare for the test.

#### 5.2.4 What do Chinese students think are the most effective ways to prepare for the TOEFL test?

This group of students does a wide variety of preparation techniques to study for the TOEFL exam. This makes sense as students often use a wide variety of techniques to prepare for exams, especially for as broad a skill as English. Any time a student reads a newspaper article in English they could arguably be preparing for the TOEFL exam. So some interpretation of student's responses is required. This is done partially through the triangulation of students surveys with counselor and student interviews but more work needs to be done to further refine our understandings.

The more important result from this research comes from the student's prioritization of study techniques. From this we can see that that test wise strategies are done by almost all students and they are the primary study technique for many. Specifically memorizing vocabulary translations and continuously doing practice tests. This lines up with the other literature (Golden, 2011; Zhan and Wan, 2014) suggesting students do a variety of techniques including both test wise strategies and English learning.

The other key piece of evidence comes from the student's discussion of their education history. Almost all students attended outside training schools, with most also attending one specifically to improve TOEFL. New Oriental came up in interviews and focus groups repeatedly with students and counselors describing its great popularity. Additionally, the counselors were specifically asked who they thought the students were most likely to take their education beliefs from and they all said that their parents were their primary influence and that their parents were primarily influenced by training school like New Oriental. During the discussion about different

kinds of students the counselors said that the top tier students were more independent minded, leading to our next question of whether the student's preparation techniques vary by ability level.

#### 5.2.5 Do student's preparation techniques vary by ability level?

As stated, the variety of techniques used by students make it difficult to parse out differences between groups of students. From the most important techniques section it is apparent that middle range students are more likely to do 'Vocabulary', the most common test wise strategy. Middle Senior 3 took more TOEFL test on average than anyone else and considering Senior 3s have had the most opportunities; this could suggest that mid ranked students would try to take tests repeatedly as a strategy to improve scores. Overall the instruments used in this research and the sample size have not allowed this question to be sufficiently answered. This will be discussed more in the next section on the limitations of the study.

#### 5.3 Limitations of the Study

As a case study, the first question is whether this case is representative of something larger. That is, by learning about this case have we learned anything that can be generalized? From the researchers experience this is a very typical case, in order to determine if this is in fact the truth more similar sites need to be examined to see if similar results can be found. One unanticipated problem was discovered over the course of the research. It turned out that the two year groups had significantly different educational cultures making it more difficult to compare them. This problem could be addressed by including students from multiple centers.

Another issue is whether this case's sample size was sufficient. While the overall population was sufficient in size for a case study, this was not necessarily the case once the sample was divided into multiple groups (Yin, 2009).

Writing the survey questions was a challenge as there was difficulty in writing choices in such a way as to elicit whether an action was test wise or not. For example, many student's

choose the answer 'Studied Writing Examples' but we don't know if they meant they memorized samples to use later or if they were learning about good writing technique. At the same time if the choice was written as 'Memorized Sample Passages' then students might be less likely truthfully answer. In general, there needs to be more focus on how students are studying for speaking and writing or on how students prioritize their study. This brings us to recommendations for future research.

#### 5.4 Future Research

This research issue requires further quantitative and qualitative research. This is a large enough issue that more breadth and more depth are required to fully examine the issue.

From the quantitative perspective, more studies are needed to be done across a wider variety of schools. Concurrent studies at multiple centers using the same research instruments could lead to greater sample sizes that can be examined more precisely. It would also help the generalizability argument further as the more places where these results are found; the more likely they are to be found elsewhere in China and variations between groups of students will be minimized. Additionally, based off of the results of this research the survey could be further refined to get more specifically differentiate between test wise and learning strategies as well as put more emphasis on writing and speaking preparation techniques.

From a qualitative perspective, two main things are needed, more individual detail and more research techniques to get to that detail. The first is that more detail is needed about exactly how students are using preparation techniques. Determining how students use practice tests can tell us more precisely whether a student is using it as a test wise strategy or not. By getting a better understanding of the student's motivations we can have a better understanding of how to classify them. In order to get this detail more techniques could be used such as ethnographic techniques or dairying to get a clearer description of these students and put them into a more developed context.

## Bibliography

---

- Avenia-Tapper, B. and Llosa, L. (2015) 'Construct relevant or irrelevant? The role of linguistic complexity in the assessment of English language learners' science knowledge', *Educational Assessment*, 20, p. 95-111.
- Ahmed, A. et. al. (2017) 'Optimising rigor in focus group analysis: Using content/ thematic and form/strtuctural approaches to understand British Somali's experiences of policing in London', *International Social Science Review*, 93(2), pp. 1-31.
- Arendasy, Ma nd Sommer, M. (2017) 'Reducing the effect size of the retest effect: Examining different approaches', *Intelligence*, 62, pp. 89-98.
- Atmowardoyo, H. (2018) 'Research methods in TEFL studies: Descriptive research, case study, error analysis and R&D', *Journal of Language Teaching and Research*, 9(1) pp. 197-204.
- Barnes, M. (2016) 'The washback of the TOEFL iBT in Vietnam', *Australian Journal of Teacher Education*, 41(7), p. 157-174.
- Bayazidi, A. and Saeb, F. (February 2017) 'Assessing reliability of two versions of vocabulary tests in Iranian context', *Advances in Language and Literary Studies*, 8(1), pp. 30-43.
- Bloor, M. et. al. (2001) *Focus groups in social research*, London: Sage Publications.
- Bradford, J., et. al. (2016) 'You can lead students to water, but you can't make them think: An assessment of student engagement and learning through student-centered teaching', *Journal of the Scholarship of Teaching and Learning*, 16(4), pp. 22-43.
- Brisay, M. (Winter 1994) 'Problems developing an alternative to the TOEFL', *TESL Canada Journal*, 12(1), p. 47-57.
- Brandt, M. et al. (2014) 'The replication recipe: What makes for a convincing replication?', *The Journal of Experimental Social Psychology*, 50(1), pp. 217-224.
- Breslin, M. and Buchanan, R. (2008) *On the case study method of research and teaching in Design*, MIT Press.
- British Educational Research Association (2011) *Ethical guidelines for educational research*, Available at: <https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf?noredirect=1> (Accessed August 18, 208).
- Bruce, E. and Hamp-Lyons, L. (2015) 'Opposing tensions of local and international standards for EAP writing programmes: who are we assessing fo?', 18, p. 64-77.
- Cansiz, N. and Cansiz, M. (2018) 'The validity and reliability study ir Turkish version of the sentiments, attitudesand concerns about inclusive education scale', *Kastamonu Education Journal*, 26(2), pp. 271-280.
- Cebolla-Boado, H. and Hu, Y. (2018) 'Why study abroad" sorting Chinese students across British universities', 39(3), pp. 365-380.
- Chen, J. and Sheehan, K. (June 2015) 'Analyzing and comparing reading stimulus materials across

the TOEFL iBT family of assessments', TOEFL iBT research report series and ETS research report No. 26 RR 15-08. *Educational Testing Service*: Princeton NJ.

Cheng, L. and Curtis, A. (2010) 'The realities of English language assessment and the Chinese learner in China and beyond.' in Cheng, L. and Curtis A. (ed.) *English language assessment and the Chinese learner*. New York and London: Routledge

Cheng, L. and DeLuca, C. (2011) 'Voices from test-takers: further evidence for language assessment validation and use', *Educational Assessment*, 16(104), p. 104-122.

Chesters, J and Daly, A. (March 2017) 'Do peer effects mediate the association between family socio-economic status and educational achievement', *Australian Journal of Social Issues*, 52(1), pp. 63-77.

Coffey, A. (1999) *The Ethnographic Self*, London: SAGE Publications.

Cohen, L. Manion, L. and Morrison K. (2000) *Research methods in education 5<sup>th</sup> edition*. London and New York: Routledge

Cohen, C., Manion, L. and Morrison, K. (2005) *Research methods in education*, London and New York: Routledge.

Cresswell, J. (2002) *Research design: Qualitative, quantitative, and mixed method approaches*. 2<sup>nd</sup> edn. Portland: Sage.

Crusan, D. et.al. (2016) 'Writing assessment literacy: surveying second language teachers' knowledge, beliefs, and practices', *Assessing Writing*, 28, p. 43-56.

Darling-Hammond, L., et. al. (2014) *Next generation assessment: Moving beyond the bubble test to support 21<sup>st</sup> century learning*, San Francisco: Jossey Bass.

Delamont, S. (2002) *Fieldwork in educational settings: Methods pitfalls and perspectives*. 2nd edn. New York: Routledge.

DeLuca, C. et.al. (2013) 'Putting testing researchers to the test: an exploratory study on the TOEFL iBT', *System*, 41, p. 663-676.

De Marrais, K. (2004) 'Qualitative interview studies: Learning through experience', in de Marais, K. and Lapan, S. (ed.) *Foundations for research: Methods of inquiry in education and the social sciences*, London: Lawrence Erlbaum Associates.

Dixon, B. 'The testing Emperor finally has no clothes', *Education Digest*, 83(9) pp. 33-43.

Doe, C. (2015) 'One teacher's take on using a 'test' for diagnostic purposes in the classroom', *Journal of English for Academic Purposes*, 18, p. 40-50.

Du, Ruiqing (2013) 'Gaokao in Chinese higher education to go or not to go', *Acta Universitatis Danubius: Communicatio*, 7(2), pp. 13-15.

Edele, A. et.al. (2015) 'Why bother with testing? The validity of immigrants' self-assessed language proficiency', *Social Science Research*, 52, p. 99-123.

Ellis, R., et. al. (2017) 'Improving learning analytics-combining observational and self-report data

- on student learning', *Educational Technology and Society*, 20(3) pp. 158-169.
- Flick, U. (2009) *An Introduction to Qualitative Research 5<sup>th</sup> Edition*. London: Sage Publications Limited.
- Fox, J and Cheng, L. (2015) 'Walk a mile in my shoes: stakeholder accounts of testing experience with a computer-administered test', *TESL Canada Journal*, 32(9), p. 65-86.
- Fulcher, G. (2000) 'The 'communicative' legacy in language testing', *System*, 28, p. 483-497.
- Gaudy, E. and Sydney, E. (1971) *Anxiety and educational achievement*, Chichester: Wiley.
- Gebril, A. and Plakans, L. (2014) 'Assembling validity evidence for assessing academic writing: rater reactions to integrated tasks', *Assessing Writing*, 21, p. 56-73.
- Gibbons, P. (2006) *Bridging discourses in the ESL classroom: Students, teachers and researchers*, London: Continuum.
- Ginter, A. and Elder, C. (December 2014) *A comparative investigation into understandings and uses of the TOEFL iBT test (sic), the International English Language Testing service (academic) test, and the Pearson Test of English for Graduate Admissions in the United States and Australia: a case study of two university contexts*. TOEFL iBT research report series and ETS research report No. 24 RR 14-42. Educational Testing Service: Princeton NJ
- Golden, D. (2011) 'U.S. college test prep is: a) turning Chinese kids into the latest threat to American competitiveness B) a hugely profitable legitimate business C) a crash course in cheating D) a source of revenue for American universities E) all of the above', *Bloomberg Business Review*, May 9-May 15, p.58-63.
- Graham, W. (2015) 'An empirical examination of the relationship between nonresponse rate and nonresponse bias', *Statistical Journal of the IAOS*, 31(2).
- Gu, L. and So, Y. (2015) 'Voices from stakeholders: what makes an academic English test 'international'', *Journal of English for Academic Purposes*, 18, p. 9-24.
- Gu, M. (2010) 'Identities constructed in difference: English language learners in China', *Journal of Pragmatics*, 42, p. 139-152.
- Hamp-Lyon, L. (2000) 'Social, professional and individual responsibility in language testing' *System*, 28, p. 579-591.
- Hanley, J. (2017) 'Correction of selection bias in survey data: Is the statistical cure worse than the bias?', *American Journal of Public Health*, 107(4), pp. 503-514.
- Hatcher, R. (2006) 'Class differentiation in education: Rational choices?', *British Journal of Sociology Educatn*, 19(1), 5-24.
- He, L and Shi, L. (2008) "ESL students' perceptions and experiences of standardized English writing tests." *Assessing Writing*, 13, p. 130-149
- Heng-Tsung, D. (2016) 'Test-taking strategies in L2 assessment: The test of English for international communication speaking test', *Perceptual & Motor Skills*, 123(1), pp. 64-93.
- Hong, E. and Peng, Y. (2008) 'Do Chinese students' perceptions of test value affect test

performance? Mediating role of motivational and metacognitive regulation in test preparation', 18, p. 499-512.

Howard, S., et. al. (2017) 'What are standardized literacy and numeracy tests testing? Evidence of the domain-general contributions to students' standardized educational tests', *British Journal of Educational Psychology*, 87(1), pp. 108-122.

Huang, H. (2016) 'Exploring strategy use in L2 speaking assessment', *System*, 63, p. 13-27.

Huddleston, A. and Rockwell, E. (2015) 'Assessment for the masses: a historical critique of high-stakes testing in reading', *Texas journal of literacy education*, 3(1), p. 38-49.

Irwin, C. and Stafford, E. (2016) *survey methods for educators: Collaborative survey*, Available at: <http://ies.ed.gov/ncee/edlabs> (Accessed 5 June 2018).

Jackson, N. and Evans, L. (2017) 'Self-reflection on differentiation: understanding how we taught in higher education', *Networks*, 19(1).

Jin, Y. (2011) 'Fundamental concerns in high-stakes language testing: the case of the college English test', *Journal of Pan-Pacific Association of Applied Linguistics*, 15(2), 71-83.

Johnson, R. and Riazi, A. (2017) 'Validation of a locally created and rated writing test used for placement in a higher education EFL program', *Assessing Writing*, 32, p. 85-104.

Kim, E. (2017) "The TOEFL iBT writing: Korean student's perceptions of the TOEFL iBT writing test." *Assessing writing*, 33, p. 1-11.

Knoch, U. and Sitajalbhorn, W. (2013) "A closer look at integrated writing tasks: towards a more focused definition for assessment purpose." *Assessing writing*, 18, p. 300-3008

Kunnan, A. and Wu, J. (2010) 'The language training and testing center, Taiwan', in Cheng, L. and Curtis A. (ed.) *English language assessment and the Chinese learner*. New York and London: Routledge

Laborda, J. (November 2008) 'Is the TOEFL exam aimed at everyone? Research considerations in the training and application of the TOEFL exam abroad', 14.

Lee, C. (2016) 'Second language learners' self-perceived roles and participation in face-to-face English (sic) writing consultations', 63, p. 51-64.

Lee, I. and Coniam, D. (2013) 'Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong', *Journal of Second Language Writing*, 22, p. 34-50.

Lei, S. (2015) 'Variation in study patterns among college students; A review of literature', *College Student Journal*.

Lewis, S and Hardy I. (2015) 'Funding, reputation and targets: the discursive logics of high-stakes testing', *Cambridge Journal of Education*, 45(2), pp. 245-264.

Li, Q. (2017) 'Changes in the motivation of Chinese ESL learners: a qualitative investigation', *English Language Teaching*, 10(1), p. 112-122.

Ling, G. et.al. (June 2014) 'Do TOEFL iBT scores reflect improvement in English-language proficiency? Extending the TOEFL iBT validity argument', TOEFL iBT research report series and ETS research report No. 22 RR 14-09. *Educational Testing Service*: Princeton NJ.

Liu, F. and Stapleton, P. (2014) 'Counterargumentation and the cultivation of critical thinking in argumentative writing: investigating washback from a high-stakes test', *System*, 45, p. 117-128.

Liu, O. et.al. (December 2016) 'Pilot testing the chinese version of the ETS proficiency profile critical thinking test', TOEFL iBT research report series and ETS research report No. 22 RR 14-42. *Educational Testing Service*: Princeton NJ.

*Market Line* (February 2017) 'Company profile: New Oriental Education Group, Inc.',

Marketwired (April 1, 2010) "Record number of TOEFL test takers in China."  
<http://www.marketwired.com/press-release/record-number-of-2009-toeflr-test-takers-in-china-1165853.htm> (Accessed 2017/7/1)

Malone, M. and Montee, M. (December 2014) 'Stakeholder's belief's about the TOEFL iBT test (sic) as a measure of academic language ability', TOEFL iBT research report series and ETS research report No. 22 RR 14-42. *Educational Testing Service*: Princeton NJ.

Maxwell, J. A. (2004). 'Casual explanation, qualitative research, and scientific inquiry in education', *Educational Researcher*, 33(1), pp. 3–11.

Mertens, D. (2010) *Research and evaluation in education and psychology*, Los Angeles: Sage.

Messic, S. (May 1996) 'Validity and washback in language testing', *Educational Testing Service*: Princeton NJ.

Metzger, K. et. al. (2018) 'SMASH: A diagnostic tool to monitor student metacognition, affect and study habits in an undergraduate science course', *Journal of College Science Teaching*, 47(3), pp. 88-99.

Moore, A. (2000) *Teaching and learning; Pedagogy, curriculum and culture*, New York: Routledge.

Mujis, D. (2004) *Doing quantitative research in education*. London: Sage Publication.

Neuman, L. (2013) *Social research methods: Quantitative and qualitative approaches*, USA: Pearson.

Nichols, S., et. al. (2012) 'High-Stakes testing and student achievement: u[dated analyses with NAEP data', *Education Policy Analysis Archives*, 20(20).

Nonis, S. and Hudson, G. (2010) 'Performance of college students: Impact of study time and study habit', *Journal of Education for Business*, 85, pp. 229-238.

Oroujlou, N. and Vahedi, M. (2011) 'Motivation, attitude, and language learning', *Social and Behavioral Sciences*, 29, p. 994-1000.

Pan, Y. (2016) 'Traditional and non-traditional test preparation practices: learner performance and perspective.' *Electronic journal of foreign language teaching*, 13(2), p. 170-183.

Pan, Yi-Ching. (2016) 'Traditional and non-traditional test preperation practices: learner performance and perspectives', *Electronic Journal of Foreign Language Teaching*, 13(2), p. 170-183.

Pardis, J. and Jia, R. (2017) 'Bilingual children's long-term outcomes in English as a second language: Language environment factors shape individual differences in catching up with monolinguals', *Developmental Science*, 20:e12422.

Qian, D. (2010). 'From TOEFL pBT to TOEFL iBT: recent trends, research, landscape and Chinese learners.' in Cheng, L. and Curtis A. (ed.) *English language assessment and the Chinese learner*. New York and London: Routledge

Qin, X. (2015) "'I must impress the raters!'" An investigation of Chinese test-takers' strategies to manage rater impressions', *Assessing Writing*, 25, p. 22-37.

Rajaram, K. (2013) 'Followers of Confucianism or a new generation? Learning culture of mainland Chinese: In pursuit of Western-based business education away from mainland China', *International Journal of Teaching and Learning in Higher Education*, 25(3), pp. 369-377.

Rauhala, E. (2015) 'China's big test: cheating allegations dog Chinese students applying to US colleges-but the reality is more complex', *Time*, April 15, 36-41.

Rahimi, F. et.al. (2014) 'Using an argument-based approach to ensure fairness of high stakes tests' score-based consequence', *Procedia Social and Behavioral Sciences*, 98, p. 1461-1468.

Reay, D. and William, D. (1999) "'I'll be a nothing': structure, agency and the construction of identity through assessment", *British Educational Research Journal*, 25(3) p. 343-354.,

Seddigh, R., et. al. (2016) 'Participating in university entrance exam despite repeated failure: a qualitative study of participants experiences', *International Journal of Medical Education*, 7, pp. 345-353.

Sevior, M. (2015) 'Assessing academic writing on a pre-sessional EAP course: designing assessment which supports learning', *Journal of English for Academic Purposes*, 18, p. 84-89..

Shell, D. et. al. (September 1995) 'Self efficacy, attribution and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement level differences', *American Psychological Association*, 87(3), pp. 386-398.

Shohamy, E. (2000) 'The relationship between language testing and second language acquisition, revisited', *System*, 28, p. 541-553.

Slomp, D. (2012) 'Challenges in assessing the development of writing ability: theories, constructs and methods', *Assessing Writing*, 17, p. 81-91.

Strickler, L. and Atalli, Y. (2010) 'Test taker's attitudes about the TOEFL iBT', *Educational Testing Service*: Princeton NJ.

Taylor, C. and Angelis, P. (2008) 'The evolution of the TOEFL' in Chapelle C. Enright, M. and Jamieson, J (ed.) *Building a validity argument for Test of English as a foreign language*, New York: Routledge.

Terwel, J. (2006) 'Curriculum differentiation: Multiple perspectives and developments in education', *Journal of Curriculum Studies*, 37(6), pp. 653-670.

'The need for EAP teacher knowledge in assessment' (2015) *Journal of English for Academic Purposes*, 18, p. 3-8.

'Validity evidence supporting the interpretation and use of TOEFL iBT scores' (2011) *Educational Testing Service*: Princeton NJ.

Vasishth, S. (December 2017) *The replication crisis in science*. Available at: <https://thewire.in/science/replication-crisis-science> (Accessed: 29 May 2018).

Wall, D. (2000) "The impact of high-stakes testing on teaching and learning: can this be predicted or controlled?" *System*, 28, p. 499-509.

Wang, M. et. al. (April 2018) 'Harsh parenting and academic achievement in Chinese: Potential mediating roles of effortful control and classroom engagement', *Journal of School Psychology*, 67, p. 16-30.

Wang, X. and Cheng, L. (2010) "The realities of English language assessment and the Chinese learner in China and beyond." *English Language Assessment and the Chinese Learner* in Cheng, L. and Curtis A. (ed.) *English language assessment and the Chinese learner*, New York and London: Routledge

Wragg, E. (2010) *Assessment and learning in the secondary school*, New York: Routledge.

Xin, Z. (July 2008) 'Practice on assessing grammar and vocabulary: the case of the TOEFL', 5(7), p. 46-57.

Yin, R. (2009) *Case study research*. 4<sup>th</sup> edn. London: Sage.

Zhan, Y. and Wan, Z. (2014) 'Dynamic nature of washback on individual learners: the role of possible selves', *Assessment & Evaluation in Higher Education*, 39(7), p. 821-839.

## Appendices

### Appendix A: Focus Group Questions

Below are the questions which will be asked during the focus group. The blank spaces are included to add follow up questions as necessary.

#### English Use

| Question:                                                                          | Notes: |
|------------------------------------------------------------------------------------|--------|
| What are you interested in studying at university?                                 |        |
|                                                                                    |        |
| Why do you think are the advantages to studying in an English language university? |        |
|                                                                                    |        |
| What do you think is the easiest part of learning English?                         |        |
|                                                                                    |        |
| What do you think is most difficult about learning English?                        |        |
|                                                                                    |        |
|                                                                                    |        |
| How do you use English in your daily life? For                                     |        |

|                                                                                           |  |
|-------------------------------------------------------------------------------------------|--|
| example do you watch English movies or talk to friends in English?                        |  |
|                                                                                           |  |
| How confident do you feel when you are using English?                                     |  |
|                                                                                           |  |
| What concerns you the most about going to university in an English speaking environment?  |  |
|                                                                                           |  |
| What are you most confident about going to university in an English speaking environment? |  |
|                                                                                           |  |
| What do you think university English classes will be like?                                |  |
|                                                                                           |  |
| What do you think your university major classes will be like?                             |  |
|                                                                                           |  |
| What do you think your university general education requirements will be like?            |  |
|                                                                                           |  |

#### Purpose of TOEFL iBT test

| Question                                                                                                                              | Notes |
|---------------------------------------------------------------------------------------------------------------------------------------|-------|
| What do you think the goal of the TOEFL iBT test is?                                                                                  |       |
|                                                                                                                                       |       |
| Do you think the TOEFL measures your ability to use English in daily life? Why or why not?                                            |       |
|                                                                                                                                       |       |
| Do you think the TOEFL measures your ability to use English at school? Why or why not?                                                |       |
|                                                                                                                                       |       |
| Do you think your score reflects your English ability?                                                                                |       |
|                                                                                                                                       |       |
| Describe the experience of taking the test                                                                                            |       |
|                                                                                                                                       |       |
| Why do you think the test is designed the way it is?                                                                                  |       |
|                                                                                                                                       |       |
| What do you think the independent writing tasks are measuring?                                                                        |       |
|                                                                                                                                       |       |
| Why do you think they include tasks where you read and write or listen and write rather than doing the different tasks independently? |       |
|                                                                                                                                       |       |
| Why do you think the listening tasks are                                                                                              |       |

|                                                                            |  |
|----------------------------------------------------------------------------|--|
| organized the way they are?                                                |  |
|                                                                            |  |
| What do you think the purpose of the speaking section is?                  |  |
|                                                                            |  |
| Do you ever discuss the task with classmates or friends?                   |  |
|                                                                            |  |
| Do you ever discuss or read about the test online?                         |  |
|                                                                            |  |
| What do you think the impression of other Chinese students of the test is? |  |
|                                                                            |  |

### Preparation Techniques

| Question                                                                       | Notes |
|--------------------------------------------------------------------------------|-------|
| How long have you studied English?                                             |       |
|                                                                                |       |
| How long have you studied for TOEFL?                                           |       |
|                                                                                |       |
| How important do you think the TOEFL test is?                                  |       |
|                                                                                |       |
| What do you think is the best way to prepare for the TOEFL test?               |       |
|                                                                                |       |
| How do you individually study for the test?                                    |       |
|                                                                                |       |
| Have you taken a test preparation class specifically for TOEFL?                |       |
|                                                                                |       |
| Describe a typical TOEFL test preparation class?                               |       |
|                                                                                |       |
| What do they tell you the best way to prepare for the test are?                |       |
|                                                                                |       |
| How well do you think these classes prepare you for the exam?                  |       |
|                                                                                |       |
| How would you compare these classes to normal English classes?                 |       |
|                                                                                |       |
| Has your overall English ability improved because of test preparation classes? |       |
|                                                                                |       |

## Appendix B: Student Questionnaire

Student Number X

Please answer the following questions to the best of your ability.

If a question is only about the Test of English as a Foreign Language (TOEFL) and you have not taken it, please leave the question blank.

If you cannot or do not want to answer a question please leave it blank.

- 1) How many years have you studied English? \_\_\_\_\_
- 2) What do you plan to study at University? \_\_\_\_\_
- 3) Why do you want to study at an English speaking University?

---

---

---

- 4) Put the following in order of most worried (1) to least worried (6) about studying in an English language university

Understanding Lectures \_\_\_\_\_  
Unfamiliar Accents \_\_\_\_\_  
Speaking to Teachers and Classmates \_\_\_\_\_  
Amount of Reading \_\_\_\_\_  
Writing Papers \_\_\_\_\_  
Taking Tests \_\_\_\_\_

Do you have any other worries about studying in an English Language University? If yes, please explain further.

---

---

- 5) Have you ever taken the TOEFL test? \_\_\_\_\_
- 6) If yes, how many times? \_\_\_\_\_
- 7) If you have taken the test more than once, **circle one** of the following answers that best describes the change.

a) Improve a little                      b) improve a lot                      c) stay the same                      d) decrease

- 8) Have you ever taken classes at an outside training school? \_\_\_\_\_

- 9) Have you ever taken classes specifically for standardized test preparation? \_\_\_\_\_

- 10) Have you taken classes specifically to study for TOEFL?? \_\_\_\_\_

**Circle your level of agreement with the following statements**

- 11) The English used in the TOEFL test felt natural

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

12) The TOEFL test is mostly fair

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

13) My TOEFL score shows my ability to use English

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

14) I understand why I received the score that I received on the TOEFL test

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

15) The TOEFL Reading Section showed my ability to read in English

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

16) The TOEFL Writing Section showed my ability to write in English

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

17) The TOEFL Speaking Section showed my ability to speak in English

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

18) The TOEFL Listening section showed my ability to understand spoken English

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

19) Studying for the TOEFL test has improved my ability to communicate in English

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

20) Studying for the TOEFL test has improved my ability to use English in school

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

21) I feel more confident about studying at an English language university because I have prepared for the TOEFL

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

disagree

22) The TOEFL reading, writing, speaking and reading tasks are similar to what I think the academic experience at university will be like

- a) Strongly Agree      b) Somewhat agree      c) Somewhat disagree      d) Strongly disagree

23) To study for the TOEFL or Other standardized English test I have used the following preparation techniques. Place an X next to all that apply

Taking practice multiple choice questions \_\_\_\_\_

Writing sample essays \_\_\_\_\_

Studying reading questions by dividing into types of questions \_\_\_\_\_

Recording speaking samples \_\_\_\_\_

Studying TOEFL vocabulary book \_\_\_\_\_

Looking up unfamiliar words in a dictionary \_\_\_\_\_

Studying sample English Sentences to learn new Vocabulary \_\_\_\_\_

Writing diagrams to organize ideas \_\_\_\_\_

Making outlines of information \_\_\_\_\_

Looking for key terms in writing questions \_\_\_\_\_

Looking for key terms in reading questions \_\_\_\_\_

Listening for key terms \_\_\_\_\_

Learning speeches \_\_\_\_\_

Listening to English conversations or lectures \_\_\_\_\_

Repeating Listening examples \_\_\_\_\_

Dictating listening samples \_\_\_\_\_

Studying Test Format \_\_\_\_\_

Studying Writing Examples \_\_\_\_\_

Talked to other students about test questions online or in person \_\_\_\_\_

Guess what topics will be on the test \_\_\_\_\_

Looked at how question is set up \_\_\_\_\_

Note taking technique \_\_\_\_\_

How to guess the best choice \_\_\_\_\_

Reading articles \_\_\_\_\_

Have writing and speaking examples judged by others \_\_\_\_\_

Other \_\_\_\_\_

21) What do you think the 3 most important ways to study for the TOEFL or other Standardized tests are? Either from the list above or other's you have used.

---

---

---

## Appendix C: Counselor Interview Questions

### Interview Questions for College Counselors

- 1) When you teach students to study for the test of English as a foreign language what is your usual process?
- 2) What strategies have you have seen or heard of students doing for studying TOEFL that you would categorize as learning English language skills or knowledge?
- 3) What strategies have you seen or heard of students doing to study for TOEFL that you would categorize as 'testwise', that is, strategies which are based on understanding how the test is set up in order to improve scores.
- 4) Which type of strategy do you believe is more popular? Why do you think that is?
- 5) Which type do you believe to be more effective for raising students TOEFL scores? Why?
- 6) Have you observed any difference in study habits for TOEFL between student of high, medium and low English ability? If yes what difference do you see?
- 7) Who do you believe has the most influence on student's beliefs about how to study in general and for TOEFL in particular?
- 8) What is your understanding of what a student means by "Studying TOEFL Vocabulary" or "Studying TOEFL book"?
- 9) What is your understanding of what a student means by "looking for key terms"?
- 10) What do you believe are the most common ways for students to study for each of the four sections of the test, that is writing, speaking, listening and reading?

## Appendix D: Student Interview Questions

### Student Interview Questions

- 1) How long have you been studying English?
- 2) In that time how has your English learning changed?
- 3) Do you think that is a typical experience of students like yourself?
- 4) Why do you think you should take Test of English as a Foreign Language over other English language tests?
- 5) Do you think TOEFL is fair? Why or why not?
- 6) When you get your score, do you understand why you received the score you got?
- 7) How many times have you taken the TOEFL?
- 8) Has your score gone up, down or stayed the same?
- 9) Why do you think this is and what do you think would be the best way to improve?
- 10) Has your study method changed since you started to study for the TOEFL?
- 11) Do you think your English ability has improved as a result of studying for the test?
- 12) Which part of the test is most difficult? Why?
- 13) How do you study vocabulary?
- 14) Have you heard of other methods student use to study vocabulary? If yes, what are those methods?
- 15) How do you study for writing and speaking?
- 16) Have you heard of other methods student use to study writing and speaking? If yes, what are those methods?
- 17) How do you get your writing and speaking practice judged?
- 18) What do you think your fellow students think are the most effective methods for improving TOEFL scores?
- 19) What kind of methods have for note taking do you use?
- 20) Do you know of any methods other students use?
- 21) Do you know what other students do to study the format of the test?

- 22) If you aren't sure of an answer to a question are there any ways to improve the chance of getting the correct answer?
- 23) If you encounter an unfamiliar word while you are studying, what is the best way to learn that word?
- 24) When you are doing reading or listening practice, do you read or listen for key terms? Why or why not?
- 25) Do you think it is a good idea to do practice tests? Why or why not?
- 26) Do you go take other English classes besides the ones at your high school? If yes, how do they compare?
- 27) How would you compare preparing for western universities with studying for Chinese universities?
- 28) Do you know of students going to training centers for the Chinese tests? Why do you think they go or don't go??
- 29) Do you feel like you will be ready to attend an English language university when the time comes? Why or why not?
- 30) If you could improve the TOEFL or any aspect of your English language education, how would you do this?

#### Appendix E: Student Questionnaire Data

##### Descriptive Questions: English Language Experience: Survey Questions 1, 8, 9, 10

| Students    | N= | Average years studied English | Percent Taken Outside Classes | Percent Taken TOEFL Classes |
|-------------|----|-------------------------------|-------------------------------|-----------------------------|
| All         | 75 | 11.5                          | 71(95%)                       | 71 (95%)                    |
| All S3      | 44 | 12.3                          | 40 (91%)                      | 41 (93%)                    |
| All S2      | 31 | 10.4                          | 31 (100%)                     | 30 (97%)                    |
| Top S3      | 15 | 12.3                          | 14 (93%)                      | 14 (93%)                    |
| Mid S3      | 15 | 12.2                          | 13 (87%)                      | 14 (93%)                    |
| Low S3      | 12 | 12.2                          | 11 (92%)                      | 12 (100%)                   |
| S3 No TOEFL | 2  | 14                            | 1 (50%)                       | 1 (50%)                     |

|             |    |       |          |           |
|-------------|----|-------|----------|-----------|
| Top S2      | 9  | 11    | 8(89%)   | 8 (89%)   |
| Mid S2      | 12 | 10.1  | 11 (92%) | 12 (100%) |
| Low S2      | 6  | 9.75  | 6 (100%) | 6 (100%)  |
| S2 No TEOFL | 4  | 11.25 | 3 (75%)  | 4 (100%)  |

Descriptive Questions: TOEFL Experience: Survey Questions 6-7

| Students | N= | Average number of TOEFL tests Taken | Improved a Lot | Improved a Little | Same | Decreased |
|----------|----|-------------------------------------|----------------|-------------------|------|-----------|
| All      | 75 | 3.7                                 | 36             | 24                | 6    | 1         |
| All S3   | 44 | 4.7                                 | 24             | 13                | 4    | 1         |
| All S2   | 31 | 2.5                                 | 12             | 11                | 2    | 0         |
| Top S3   | 15 | 4.3                                 | 10             | 3                 | 1    | 1         |
| Mid S3   | 15 | 4.9                                 | 8              | 5                 | 2    | 0         |
| Low S3   | 12 | 4.75                                | 6              | 5                 | 1    | 0         |
| Top S2   | 9  | 2.2                                 | 6              | 1                 | 0    | 0         |
| Mid S2   | 12 | 2.4                                 | 5              | 4                 | 2    | 0         |
| Low S2   | 6  | 2.85                                | 1              | 6                 | 0    | 0         |

Descriptive Questions: Plan to Study: Survey Question 2

| Students    | N= | STEM | Business and Economics | Humanities, Social Sciences, Arts | Undecided |
|-------------|----|------|------------------------|-----------------------------------|-----------|
| All         | 75 | 28   | 18                     | 20                                | 7         |
| All S3      | 44 | 17   | 7                      | 13                                | 4         |
| All S2      | 31 | 11   | 11                     | 7                                 | 3         |
| Top S3      | 15 | 6    | 1                      | 6                                 | 2         |
| Mid S3      | 15 | 8    | 1                      | 4                                 | 1         |
| Low S3      | 12 | 3    | 4                      | 2                                 | 1         |
| S3 No TOEFL | 2  | 0    | 1                      | 1                                 | 0         |
| Top S2      | 9  | 4    | 4                      | 2                                 | 0         |
| Mid S2      | 12 | 2    | 3                      | 4                                 | 2         |
| Low S2      | 6  | 4    | 2                      | 1                                 | 0         |

|             |   |   |   |   |   |
|-------------|---|---|---|---|---|
| S2 No TEOFL | 4 | 1 | 2 | 0 | 0 |
|-------------|---|---|---|---|---|

Descriptive Questions: Why Study at Western University: Survey Question 3: Part 1

| Students    | N= | Quality of Education | Experience Different Cultures | Flexibility | Unintelligible | Political Freedom | No Answer |
|-------------|----|----------------------|-------------------------------|-------------|----------------|-------------------|-----------|
| All         | 75 | 12                   | 24                            | 16          | 9              | 2                 | 7         |
| All S3      | 44 | 4                    | 13                            | 9           | 5              | 2                 | 2         |
| All S2      | 31 | 8                    | 11                            | 7           | 4              | 0                 | 5         |
| Top S3      | 15 | 3                    | 5                             | 5           | 1              | 0                 | 1         |
| Mid S3      | 15 | 0                    | 4                             | 1           | 0              | 0                 | 0         |
| Low S3      | 12 | 0                    | 3                             | 2           | 4              | 0                 | 0         |
| S3 No TOEFL | 2  | 1                    | 1                             | 1           | 0              | 0                 | 1         |
| Top S2      | 9  | 3                    | 4                             | 3           | 0              | 0                 | 1         |
| Mid S2      | 12 | 3                    | 3                             | 4           | 0              | 0                 | 1         |
| Low S2      | 6  | 2                    | 4                             | 0           | 4              | 0                 | 2         |
| S2 No TEOFL | 4  | 1                    | 0                             | 0           | 0              | 0                 | 1         |

Descriptive Questions: Why Study at Western University: Survey Question 3: Part 2

| Students    | N= | GaoKao | Improve English | Technology | Future Opportunities | Parents |
|-------------|----|--------|-----------------|------------|----------------------|---------|
| All         | 75 | 3      | 24              | 4          | 10                   | 2       |
| All S3      | 44 | 3      | 13              | 4          | 5                    | 1       |
| All S2      | 31 |        | 11              | 0          | 5                    | 1       |
| Top S3      | 15 | 0      | 3               | 1          | 2                    | 0       |
| Mid S3      | 15 | 2      | 8               | 2          | 3                    | 1       |
| Low S3      | 12 | 1      | 2               | 1          | 0                    | 0       |
| S3 No TOEFL | 2  | 0      | 0               | 0          | 0                    | 0       |
| Top S2      | 9  | 0      | 0               | 0          | 3                    | 0       |

|             |    |   |   |   |   |   |
|-------------|----|---|---|---|---|---|
| Mid S2      | 12 | 0 | 7 | 0 | 0 | 0 |
| Low S2      | 6  | 0 | 2 | 0 | 1 | 1 |
| S2 No TEOFL | 4  | 0 | 2 | 0 | 1 | 0 |

Likert Scale Questions: 11-22

|                                                    |           |                |
|----------------------------------------------------|-----------|----------------|
| 1) The English used in the TOEFL test felt natural |           |                |
| Senior 3                                           | Senior 2  | All            |
| 2.143                                              | 1.912     | 2.053          |
| S3 Top                                             | S3 Middle | Senior 3 Lower |
| 2.067                                              | 2.133     | 2.231          |
| S2 Top                                             | S2 Middle | S2 Lower       |
| 2.1                                                | 1.917     | 1.75           |

|                                  |                 |                |
|----------------------------------|-----------------|----------------|
| 2) The TOEFL test is mostly fair |                 |                |
| Senior 3                         | Senior 2        | All            |
| 2.024                            | 2.059           | 2.039          |
| Senior 3 Top                     | Senior 3 Middle | Senior 3 Lower |
| 1.867                            | 2.067           | 2.181          |
| Senior 2 Top                     | Senior 2 Middle | Senior 2 Lower |
| 2                                | 2.25            | 1.75           |

|                                                   |                 |                |
|---------------------------------------------------|-----------------|----------------|
| 3) My TOEFL score shows my ability to use English |                 |                |
| Senior 3                                          | Senior 2        | All            |
| 1.975                                             | 2.382           | 2.162          |
| Senior 3 Top                                      | Senior 3 Middle | Senior 3 Lower |
| 1.867                                             | 2.067           | 2              |
| Senior 2 Top                                      | Senior 2 Middle | Senior 2 Lower |
| 2.4                                               | 2.417           | 2.125          |

|                                                                      |                 |                |
|----------------------------------------------------------------------|-----------------|----------------|
| 4) I understand why I received the score the I received on the TOEFL |                 |                |
| Senior 3                                                             | Senior 2        | All            |
| 1.927                                                                | 1.697           | 1.824          |
| Senior 3 Top                                                         | Senior 3 Middle | Senior 3 Lower |
| 2.067                                                                | 1.786           | 1.909          |
| Senior 2 Top                                                         | Senior 2 Middle | Senior 2 Lower |
| 1.7                                                                  | 2.417           | 1.5            |

|                                                             |                 |                |
|-------------------------------------------------------------|-----------------|----------------|
| 5) The TOEFL read section showed my ability to read English |                 |                |
| Senior 3                                                    | Senior 2        | All            |
| 1.857                                                       | 1.794           | 1.829          |
| Senior 3 Top                                                | Senior 3 Middle | Senior 3 Lower |
| 1.733                                                       | 2.133           | 1.636          |
| Senior 2 Top                                                | Senior 2 Middle | Senior 2 Lower |
| 1.9                                                         | 1.833           | 1.5            |

|                                                                    |          |       |
|--------------------------------------------------------------------|----------|-------|
| 6) The TOEFL writing section showed my ability to write in English |          |       |
| Senior 3                                                           | Senior 2 | All   |
| 2.238                                                              | 1.912    | 2.092 |

|              |                 |                |
|--------------|-----------------|----------------|
| Senior 3 Top | Senior 3 Middle | Senior 3 Lower |
| 2.467        | 2.333           | 1.909          |
| Senior 2 Top | Senior 2 Middle | Senior 2 Lower |
| 2.1          | 1.75            | 1.625          |

|                                                                     |                 |                |
|---------------------------------------------------------------------|-----------------|----------------|
| 7) The TOEFL speaking section showed my ability to speak in English |                 |                |
| Senior 3                                                            | Senior 2        | All            |
| 2.119                                                               | 2.059           | 2.092          |
| Senior 3 Top                                                        | Senior 3 Middle | Senior 3 Lower |
| 2.4                                                                 | 2.133           | 1.818          |
| Senior 2 Top                                                        | Senior 2 Middle | Senior 2 Lower |
| 2                                                                   | 2.167           | 1.75           |

|                                                                        |                 |                |
|------------------------------------------------------------------------|-----------------|----------------|
| 8) The TOEFL listening section showed my ability to understand English |                 |                |
| Senior 3                                                               | Senior 2        | All            |
| 1.833                                                                  | 2               | 1.908          |
| Senior 3 Top                                                           | Senior 3 Middle | Senior 3 Lower |
| 1.8                                                                    | 1.867           | 1.818          |
| Senior 2 Top                                                           | Senior 2 Middle | Senior 2 Lower |
| 2.1                                                                    | 2.083           | 1.5            |

|                                                                                  |                 |                |
|----------------------------------------------------------------------------------|-----------------|----------------|
| 9) Studying for the TOEFL test has improved my ability to communicate in English |                 |                |
| Senior 3                                                                         | Senior 2        | All            |
| 2.139                                                                            | 2.118           | 2.132          |
| Senior 3 Top                                                                     | Senior 3 Middle | Senior 3 Lower |
| 2                                                                                | 2.267           | 2.083          |
| Senior 2 Top                                                                     | Senior 2 Middle | Senior 2 Lower |
| 2.2                                                                              | 2.167           | 1.875          |

|                                                                             |                 |                |
|-----------------------------------------------------------------------------|-----------------|----------------|
| 10) Studying for the TOEFL has improved my ability to use English in school |                 |                |
| Senior 3                                                                    | Senior 2        | All            |
| 1.976                                                                       | 2.029           | 2              |
| Senior 3 Top                                                                | Senior 3 Middle | Senior 3 Lower |
| 1.8                                                                         | 2.133           | 2              |
| Senior 2 Top                                                                | Senior 2 Middle | Senior 2 Lower |
| 2.2                                                                         | 2.167           | 1.625          |

|                                                                                                             |                 |                |
|-------------------------------------------------------------------------------------------------------------|-----------------|----------------|
| 11) I feel more confident about studying at an English language university because I prepared for the TOEFL |                 |                |
| Senior 3                                                                                                    | Senior 2        | All            |
| 2.524                                                                                                       | 2.206           | 2.382          |
| Senior 3 Top                                                                                                | Senior 3 Middle | Senior 3 Lower |
| 2.33                                                                                                        | 2.667           | 2.267          |
| Senior 2 Top                                                                                                | Senior 2 Middle | Senior 2 Lower |
| 2.5                                                                                                         | 2               | 2              |

|                                                                                                                                           |          |     |
|-------------------------------------------------------------------------------------------------------------------------------------------|----------|-----|
| 12) The TOEFL reading, writing, speaking and reading tasks are similar to what I think the academic experience at university will be like |          |     |
| Senior 3                                                                                                                                  | Senior 2 | All |

|              |                 |                |
|--------------|-----------------|----------------|
| 2.357        | 2               | 2.197          |
| Senior 3 Top | Senior 3 Middle | Senior 3 Lower |
| 2.533        | 2.267           | 2.273          |
| Senior 2 Top | Senior 2 Middle | Senior 2 Lower |
| 1.9          | 2               | 2.25           |

Question 4: Ranking Most to Least Worried about Studying in a Western University

| 1: Understanding Lectures |    |         |   |   |    |    |    |    |
|---------------------------|----|---------|---|---|----|----|----|----|
| Students                  | N= | Average | 1 | 2 | 3  | 4  | 5  | 6  |
| All                       | 75 | 3.8     | 8 | 4 | 19 | 20 | 13 | 12 |
| All S3                    | 44 | 3.9     | 4 | 4 | 7  | 14 | 7  | 8  |
| All S2                    | 31 | 3.7     | 4 | 0 | 12 | 6  | 6  | 4  |
| Top S3                    | 15 | 4.5     | 0 | 1 | 3  | 4  | 2  | 5  |
| Mid S3                    | 15 | 3.7     | 2 | 2 | 2  | 4  | 2  | 3  |
| Low S3                    | 12 | 3.4     | 2 | 1 | 2  | 4  | 3  | 0  |
| S3 No TOEFL               | 2  | 4       | 0 | 0 | 0  | 2  | 0  | 0  |
| Top S2                    | 9  | 4       | 0 | 0 | 3  | 4  | 1  | 1  |
| Mid S2                    | 12 | 4.3     | 0 | 0 | 5  | 1  | 4  | 2  |
| Low S2                    | 6  | 3       | 2 | 0 | 2  | 1  | 1  | 1  |
| S2 No TOEFL               | 4  | 2       | 2 | 0 | 2  | 0  | 0  | 0  |

| 2: Unfamiliar Accents |    |         |    |    |    |    |    |    |
|-----------------------|----|---------|----|----|----|----|----|----|
| Students              | N= | Average | 1  | 2  | 3  | 4  | 5  | 6  |
| All                   | 75 | 3.5     | 11 | 12 | 12 | 22 | 12 | 10 |
| All S3                | 44 | 3.5     | 8  | 5  | 8  | 9  | 8  | 6  |
| All S2                | 31 | 3.6     | 3  | 7  | 4  | 13 | 4  | 4  |
| Top S3                | 15 | 3       | 4  | 2  | 4  | 2  | 1  | 2  |
| Mid S3                | 15 | 3.9     | 2  | 0  | 4  | 3  | 4  | 2  |
| Low S3                | 12 | 3.4     | 2  | 3  | 0  | 3  | 3  | 1  |
| S3 No TOEFL           | 2  | 5       | 0  | 0  | 0  | 1  | 0  | 1  |
| Top S2                | 9  | 4.1     | 1  | 1  | 2  | 6  | 1  | 3  |
| Mid S2                | 12 | 3.6     | 1  | 1  | 2  | 6  | 2  | 0  |
| Low S2                | 6  | 3.2     | 1  | 2  | 0  | 0  | 1  | 1  |
| S2 No TOEFL           | 4  | 2.5     | 0  | 3  | 0  | 1  | 0  | 0  |

| 3: Speaking to Teachers and Classmates |    |         |    |    |   |   |    |    |
|----------------------------------------|----|---------|----|----|---|---|----|----|
| Students                               | N= | Average | 1  | 2  | 3 | 4 | 5  | 6  |
| All                                    | 75 | 3.7     | 13 | 9  | 3 | 5 | 11 | 11 |
| All S3                                 | 44 | 3.3     | 10 | 14 | 1 | 3 | 4  | 12 |
| All S2                                 | 31 | 4.4     | 3  | 3  | 2 | 3 | 10 | 9  |
| Top S3                                 | 15 | 2.9     | 3  | 6  | 1 | 2 | 1  | 2  |
| Mid S3                                 | 15 | 3.4     | 4  | 4  | 0 | 1 | 1  | 5  |
| Low S3                                 | 12 | 3.9     | 3  | 2  | 0 | 0 | 2  | 5  |
| S3 No TOEFL                            | 2  | 1.5     | 0  | 2  | 0 | 0 | 0  | 0  |

|                |    |      |   |   |   |   |   |   |
|----------------|----|------|---|---|---|---|---|---|
| Top S2         | 9  | 5.2  | 0 | 0 | 0 | 1 | 5 | 3 |
| Mid S2         | 12 | 4.3  | 1 | 2 | 0 | 2 | 4 | 3 |
| Low S2         | 6  | 2.4  | 2 | 1 | 1 | 0 | 1 | 0 |
| S2 No<br>TEOFL | 4  | 5.25 | 0 | 0 | 1 | 0 | 0 | 3 |

| 4: Amount of Reading |    |         |    |    |    |    |   |    |
|----------------------|----|---------|----|----|----|----|---|----|
| Students             | N= | Average | 1  | 2  | 3  | 4  | 5 | 6  |
| All                  | 75 | 3.2     | 18 | 12 | 15 | 10 | 8 | 12 |
| All S3               | 44 | 3.3     | 11 | 6  | 9  | 4  | 6 | 8  |
| All S2               | 31 | 3.1     | 7  | 6  | 6  | 6  | 2 | 4  |
| Top S3               | 15 | 3.6     | 1  | 4  | 3  | 2  | 3 | 2  |
| Mid S3               | 15 | 3.2     | 4  | 2  | 3  | 1  | 3 | 2  |
| Low S3               | 12 | 3.1     | 6  | 0  | 1  | 1  | 0 | 4  |
| S3 No<br>TOEFL       | 2  | 3       | 0  | 0  | 2  | 0  | 0 | 0  |
| Top S2               | 9  | 2.5     | 3  | 2  | 1  | 1  | 0 | 1  |
| Mid S2               | 12 | 3.3     | 2  | 4  | 2  | 1  | 2 | 2  |
| Low S2               | 6  | 3       | 2  | 0  | 2  | 1  | 0 | 1  |
| S2 No<br>TEOFL       | 4  | 3.75    | 0  | 0  | 1  | 3  | 0 | 0  |

| 5: Writing Papers |    |         |    |    |    |   |    |   |
|-------------------|----|---------|----|----|----|---|----|---|
| Students          | N= | Average | 1  | 2  | 3  | 4 | 5  | 6 |
| All               | 75 | 3.1     | 16 | 18 | 18 | 5 | 13 | 8 |
| All S3            | 44 | 3       | 11 | 9  | 12 | 4 | 6  | 5 |
| All S2            | 31 | 3.2     | 5  | 9  | 6  | 1 | 7  | 3 |
| Top S3            | 15 | 3.2     | 3  | 4  | 6  | 0 | 2  | 3 |
| Mid S3            | 15 | 2.7     | 5  | 4  | 2  | 1 | 1  | 2 |
| Low S3            | 12 | 3.1     | 3  | 0  | 4  | 3 | 2  | 0 |
| S3 No<br>TOEFL    | 2  | 3.5     | 0  | 1  | 0  | 0 | 1  | 0 |
| Top S2            | 9  | 2.9     | 1  | 3  | 3  | 0 | 2  | 0 |
| Mid S2            | 12 | 2.8     | 2  | 5  | 2  | 1 | 0  | 2 |
| Low S2            | 6  | 3.5     | 2  | 0  | 1  | 0 | 2  | 1 |
| S2 No<br>TEOFL    | 4  | 4.25    | 0  | 1  | 0  | 0 | 3  | 0 |

| 6: Taking Tests |    |         |    |    |    |   |   |    |
|-----------------|----|---------|----|----|----|---|---|----|
| Students        | N= | Average | 1  | 2  | 3  | 4 | 5 | 6  |
| All             | 75 | 2.9     | 26 | 11 | 11 | 8 | 6 | 12 |
| All S3          | 44 | 3       | 13 | 7  | 7  | 6 | 4 | 7  |
| All S2          | 31 | 2.7     | 13 | 4  | 4  | 2 | 2 | 5  |
| Top S3          | 15 | 4.3     | 1  | 0  | 3  | 5 | 2 | 4  |
| Mid S3          | 15 | 2.6     | 7  | 2  | 2  | 1 | 1 | 2  |
| Low S3          | 12 | 2       | 5  | 4  | 2  | 0 | 1 | 0  |
| S3 No<br>TOEFL  | 2  | 4       | 0  | 1  | 0  | 0 | 0 | 1  |
| Top S2          | 9  | 3.4     | 2  | 1  | 2  | 1 | 1 | 2  |
| Mid S2          | 12 | 2.3     | 7  | 1  | 2  | 0 | 0 | 2  |
| Low S2          | 6  | 2.6     | 2  | 1  | 0  | 1 | 1 | 0  |

|             |   |     |   |   |   |   |   |   |
|-------------|---|-----|---|---|---|---|---|---|
| S2 No TOEFL | 4 | 2.5 | 2 | 1 | 0 | 0 | 0 | 1 |
|-------------|---|-----|---|---|---|---|---|---|

Question 23: Selected Preparation Techniques: Part 1

| Students    | N= | Taking Practice multiple Choice questions | Writing Sample Essays | Studying reading Questions by dividing into types of questions | Recording Speaking Samples | Studying TOEFL Vocabulary Book | Looking up unfamiliar words in a dictionary |
|-------------|----|-------------------------------------------|-----------------------|----------------------------------------------------------------|----------------------------|--------------------------------|---------------------------------------------|
| All         | 75 | 67                                        | 70                    | 57                                                             | 58                         | 71                             | 57                                          |
| All S3      | 44 | 41                                        | 38                    | 34                                                             | 30                         | 38                             | 29                                          |
| All S2      | 31 | 26                                        | 31                    | 23                                                             | 28                         | 27                             | 28                                          |
| Top S3      | 15 | 15                                        | 14                    | 11                                                             | 12                         | 14                             | 15                                          |
| Mid S3      | 15 | 14                                        | 11                    | 11                                                             | 9                          | 13                             | 9                                           |
| Low S3      | 12 | 11                                        | 12                    | 11                                                             | 9                          | 10                             | 4                                           |
| S3 No TOEFL | 2  | 1                                         | 1                     | 1                                                              | 0                          | 1                              | 1                                           |
| Top S2      | 9  | 9                                         | 9                     | 8                                                              | 8                          | 9                              | 8                                           |
| Mid S2      | 12 | 10                                        | 12                    | 8                                                              | 10                         | 12                             | 10                                          |
| Low S2      | 6  | 5                                         | 6                     | 4                                                              | 6                          | 6                              | 6                                           |
| S2 No TOEFL | 4  | 2                                         | 4                     | 3                                                              | 4                          | 3                              | 4                                           |

Question 23: Selected Preparation Techniques: Part 2

| Students    | N= | Studying Sample English Sentences to Learn new Vocabulary | Writing Diagrams to Organize Ideas | Making Outlines of Information | Looking for Key Terms in Writing | Reading for Key Terms | Listening for Key Terms |
|-------------|----|-----------------------------------------------------------|------------------------------------|--------------------------------|----------------------------------|-----------------------|-------------------------|
| All         | 75 | 42                                                        | 32                                 | 47                             | 48                               | 66                    | 57                      |
| All S3      | 44 | 20                                                        | 16                                 | 27                             | 24                               | 35                    | 29                      |
| All S2      | 31 | 22                                                        | 16                                 | 20                             | 24                               | 31                    | 26                      |
| Top S3      | 15 | 9                                                         | 5                                  | 11                             | 5                                | 10                    | 9                       |
| Mid S3      | 15 | 4                                                         | 7                                  | 7                              | 9                                | 13                    | 11                      |
| Low S3      | 12 | 6                                                         | 4                                  | 8                              | 10                               | 11                    | 9                       |
| S3 No TOEFL | 2  | 1                                                         | 0                                  | 1                              | 0                                | 1                     | 0                       |
| Top S2      | 9  | 7                                                         | 4                                  | 4                              | 6                                | 9                     | 8                       |
| Mid S2      | 12 | 8                                                         | 4                                  | 8                              | 8                                | 11                    | 9                       |

|                |   |   |   |   |   |   |   |
|----------------|---|---|---|---|---|---|---|
| Low S2         | 6 | 5 | 4 | 4 | 5 | 8 | 7 |
| S2 No<br>TEOFL | 4 | 2 | 4 | 4 | 4 | 3 | 4 |

Question 23: Selected Preparation Techniques: Part 3

| Students       | N= | Learning Speeches | Listening to English Conversations or lectures | Repeating Listening Examples | Dictating listening samples | Studying Test Format | Studying Writing Examples |
|----------------|----|-------------------|------------------------------------------------|------------------------------|-----------------------------|----------------------|---------------------------|
| All            | 75 | 44                | 62                                             | 49                           | 41                          | 54                   | 58                        |
| All S3         | 44 | 20                | 33                                             | 27                           | 23                          | 29                   | 29                        |
| All S2         | 31 | 24                | 29                                             | 22                           | 18                          | 25                   | 29                        |
| Top S3         | 15 | 7                 | 12                                             | 9                            | 6                           | 13                   | 13                        |
| Mid S3         | 15 | 5                 | 9                                              | 11                           | 9                           | 9                    | 8                         |
| Low S3         | 12 | 8                 | 11                                             | 6                            | 8                           | 7                    | 7                         |
| S3 No<br>TOEFL | 2  | 0                 | 1                                              | 4                            | 0                           | 0                    | 1                         |
| Top S2         | 9  | 5                 | 6                                              | 6                            | 6                           | 6                    | 8                         |
| Mid S2         | 12 | 9                 | 11                                             | 8                            | 4                           | 10                   | 11                        |
| Low S2         | 6  | 6                 | 8                                              | 6                            | 4                           | 5                    | 6                         |
| S2 No<br>TEOFL | 4  | 4                 | 4                                              | 2                            | 3                           | 4                    | 4                         |

Question 23: Selected Preparation Techniques: Part 4

| Students       | N= | Talked to other students about test questions online or in person | Guess what topics will be on the test | Looked at how question is set up | Note taking Techniques | How to guess the best choice | Reading Articles | Have reading and speaking examples judged by others |
|----------------|----|-------------------------------------------------------------------|---------------------------------------|----------------------------------|------------------------|------------------------------|------------------|-----------------------------------------------------|
| All            | 75 | 38                                                                | 40                                    | 34                               | 56                     | 45                           | 60               | 51                                                  |
| All S3         | 44 | 17                                                                | 19                                    | 20                               | 32                     | 23                           | 33               | 28                                                  |
| All S2         | 31 | 21                                                                | 21                                    | 14                               | 24                     | 22                           | 27               | 23                                                  |
| Top S3         | 15 | 9                                                                 | 8                                     | 9                                | 11                     | 7                            | 11               | 13                                                  |
| Mid S3         | 15 | 6                                                                 | 7                                     | 7                                | 10                     | 7                            | 10               | 9                                                   |
| Low S3         | 12 | 1                                                                 | 4                                     | 4                                | 11                     | 8                            | 12               | 6                                                   |
| S3 No<br>TOEFL | 2  | 1                                                                 | 0                                     | 0                                | 0                      | 1                            | 0                | 0                                                   |
| Top S2         | 9  | 9                                                                 | 7                                     | 4                                | 7                      | 7                            | 8                | 7                                                   |
| Mid S2         | 12 | 9                                                                 | 7                                     | 7                                | 10                     | 9                            | 10               | 11                                                  |
| Low S2         | 6  | 3                                                                 | 3                                     | 1                                | 4                      | 4                            | 6                | 2                                                   |

|             |    |   |   |   |   |   |   |   |   |
|-------------|----|---|---|---|---|---|---|---|---|
| S2<br>TEOFL | No | 4 | 4 | 4 | 3 | 3 | 3 | 3 | 4 |
|-------------|----|---|---|---|---|---|---|---|---|

#### Preparation Techniques: Most Important By Students: Part 1

| Students    | N= | Vocabulary | note taking technique | practicing MCQs | Reading Articles | Studying Writing | Listening to English | Writing Sample Essays | Reading Sample Essays | unintelligible |
|-------------|----|------------|-----------------------|-----------------|------------------|------------------|----------------------|-----------------------|-----------------------|----------------|
| All         | 75 | 34         | 7                     | 31              | 9                | 1                | 10                   | 7                     | 2                     | 21             |
| All S3      | 44 | 18         | 2                     | 24              | 3                | 1                | 5                    | 6                     | 0                     | 7              |
| All S2      | 31 | 16         | 5                     | 7               | 6                | 0                | 5                    | 1                     | 2                     | 14             |
| Top S3      | 15 | 3          | 2                     | 12              | 1                | 1                | 4                    | 3                     | 0                     | 1              |
| Mid S3      | 15 | 8          | 0                     | 8               | 2                | 0                | 1                    | 3                     | 0                     | 5              |
| Low S3      | 12 | 6          | 0                     | 4               | 0                | 0                | 0                    | 0                     | 0                     | 1              |
| S3 No TOEFL | 2  | 1          | 0                     | 0               | 0                | 0                | 0                    | 0                     | 0                     | 0              |
| Top S2      | 9  | 5          | 0                     | 3               | 1                | 0                | 3                    | 1                     | 2                     | 1              |
| Mid S2      | 12 | 8          | 4                     | 3               | 3                | 0                | 1                    | 0                     | 0                     | 4              |
| Low S2      | 6  | 3          | 0                     | 0               | 1                | 0                | 1                    | 0                     | 0                     | 8              |
| S2 No TOEFL | 4  | 0          | 1                     | 1               | 1                | 0                | 0                    | 0                     | 0                     | 1              |

#### Preparation Techniques: Most Important By Students: Part 2

| Students    | N= | Background to test topics | practice speeches | outlining | writing and speaking judged | test format | Looking at writing examples | test strategies | novels | movies |
|-------------|----|---------------------------|-------------------|-----------|-----------------------------|-------------|-----------------------------|-----------------|--------|--------|
| All         | 75 | 1                         | 3                 | 3         | 4                           | 11          | 5                           | 9               | 4      | 6      |
| All S3      | 44 | 0                         | 0                 | 2         | 2                           | 8           | 4                           | 7               | 3      | 4      |
| All S2      | 31 | 1                         | 3                 | 1         | 2                           | 3           | 1                           | 2               | 1      | 2      |
| Top S3      | 15 | 0                         | 0                 | 0         | 1                           | 6           | 1                           | 2               | 0      | 0      |
| Mid S3      | 15 | 0                         | 0                 | 2         | 0                           | 2           | 2                           | 3               | 0      | 1      |
| Low S3      | 12 | 0                         | 0                 | 0         | 1                           | 0           | 1                           | 2               | 3      | 2      |
| S3 No TOEFL | 2  | 0                         | 0                 | 0         | 0                           | 0           | 0                           | 0               | 0      | 1      |
| Top S2      | 9  | 1                         | 0                 | 0         | 0                           | 2           | 0                           | 1               | 0      | 0      |
| Mid S2      | 12 | 0                         | 2                 | 1         | 2                           | 1           | 1                           | 1               | 0      | 1      |
| Low S2      | 6  | 0                         | 1                 | 0         | 0                           | 0           | 0                           | 0               | 1      | 0      |
| S2 No TOEFL | 4  | 0                         | 0                 | 0         | 0                           | 0           | 0                           | 0               | 0      | 1      |

### Preparation Techniques: Most Important By Students: Part 3

| Students    | N= | dictating<br>listening | live in<br>English<br>environment | study<br>schedule | talk to<br>English<br>speakers | use<br>English<br>daily<br>life | key<br>terms | have<br>confidence | grammar | typing | none |
|-------------|----|------------------------|-----------------------------------|-------------------|--------------------------------|---------------------------------|--------------|--------------------|---------|--------|------|
| All         | 75 | 3                      | 2                                 | 1                 | 2                              | 4                               | 2            | 1                  | 1       | 1      | 48   |
| All S3      | 44 | 2                      | 1                                 | 0                 | 2                              | 4                               | 2            | 1                  | 1       | 1      | 26   |
| All S2      | 31 | 1                      | 1                                 | 1                 | 0                              | 0                               | 0            | 0                  | 0       | 0      | 22   |
| Top S3      | 15 | 1                      | 0                                 | 0                 | 0                              | 3                               | 1            | 0                  | 0       | 1      | 6    |
| Mid S3      | 15 | 0                      | 1                                 | 0                 | 1                              | 0                               | 1            | 1                  | 1       | 0      | 3    |
| Low S3      | 12 | 1                      | 0                                 | 0                 | 0                              | 1                               | 0            | 0                  | 0       | 0      | 14   |
| S3 No TOEFL | 2  | 0                      | 0                                 | 0                 | 1                              | 0                               | 0            | 0                  | 0       | 0      | 3    |
| Top S2      | 9  | 0                      | 0                                 | 0                 | 0                              | 0                               | 0            | 0                  | 0       | 0      | 9    |
| Mid S2      | 12 | 1                      | 1                                 | 0                 | 0                              | 0                               | 0            | 0                  | 0       | 0      | 6    |
| Low S2      | 6  | 0                      | 0                                 | 1                 | 0                              | 0                               | 0            | 0                  | 0       | 0      | 6    |
| S2 No TOEFL | 4  | 0                      | 0                                 | 0                 | 0                              | 0                               | 0            | 0                  | 0       | 0      | 1    |

### Number of Mistakes By Students on Questionnaire

| Students    | N= | Mistakes<br>Q 4 | Other<br>Mistakes |
|-------------|----|-----------------|-------------------|
| All         | 75 | 22              | 14                |
| All S3      | 44 | 12              | 8                 |
| All S2      | 31 | 10              | 6                 |
| Top S3      | 15 | 4               | 2                 |
| Mid S3      | 15 | 5               | 3                 |
| Low S3      | 12 | 2               | 3                 |
| S3 No TOEFL | 2  | 1               | 0                 |
| Top S2      | 9  | 1               | 4                 |
| Mid S2      | 12 | 3               | 1                 |
| Low S2      | 6  | 4               | 1                 |
| S2 No TOEFL | 4  | 2               | 0                 |

Appendix F: Student TOEFL and APGPA Data

Students TOEFL and APGPA Scores

| Year 3 |                      |           |
|--------|----------------------|-----------|
|        | Most Recent<br>TOEFL | APGP<br>A |
| 34     | 116                  | 90.3      |
| 26     | 112                  | 92.3      |
| 35     | 112                  | 92.5      |
| 49     | 111                  | 97.3      |
| 24     | 110                  | 95.7      |
| 25     | 110                  | 97.3      |
| 5      | 109                  | 91.7      |
| 41     | 108                  | 90.7      |
| 44     | 108                  | 91.2      |
| 1      | 105                  | 83.8      |
| 4      | 104                  | 85.2      |
| 6      | 104                  | 90.7      |
| 23     | 103                  | 91.2      |
| 36     | 102                  | 79.3      |
| 9      | 101                  | 84.2      |
| 20     | 101                  | 83.2      |
| 21     | 101                  | 86.2      |
| 19     | 100                  | 85        |
| 37     | 100                  | 89.2      |
| 38     | 100                  | 80.5      |
| 39     | 100                  | 91.3      |
| 40     | 100                  | 91.3      |
| 18     | 99                   | 93.3      |
| 48     | 98                   | 86.3      |
| 30     | 95                   | 86.3      |
| 31     | 95                   | 84.7      |
| 32     | 95                   | 78.7      |
| 33     | 91                   | 94.5      |
| 42     | 91                   | 89.8      |
| 27     | 88                   | 73.3      |
| 46     | 87                   | 86.7      |
| 2      | 85                   | 89.5      |
| 47     | 85                   | 81        |
| 13     | 84                   | 75        |
| 14     | 84                   | 85        |
| 45     | 84                   | 85.7      |

|    |    |      |
|----|----|------|
| 43 | 83 | 82.2 |
| 15 | 80 | 75.2 |
| 3  | 79 | 85.8 |
| 29 | 78 | 82.3 |
| 16 | 77 | 68.5 |
| 10 | 75 | 75.7 |
| 17 | 75 | 76.7 |
| 28 | 66 | 77.5 |
| 12 | 37 | 57.5 |
| 11 | 35 | 44.5 |
| 7  |    | 75   |
| 8  |    | 81.2 |
| 22 |    | 89.2 |

### Year 2

|     |     |      |
|-----|-----|------|
| 50  | 106 | 93   |
| 51  | 98  | 85.3 |
| 52  | 74  | 82.2 |
| 53  | 105 | 95.2 |
| 54  | 79  | 84.5 |
| 55  | 91  | 85.3 |
| 56  | 96  | 95.3 |
| 57  | 89  | 82.2 |
| 58  | 50  | 74   |
| 59  |     | 75   |
| 60  | 53  | 69.8 |
| 61  | 71  | 80.8 |
| 62  | 92  | 89.8 |
| 63  | 91  | 85   |
| 64  | 73  | 77.5 |
| 65  | 68  | 83.5 |
| 66  |     | 80.7 |
| 67  | 88  | 92   |
| 68  | 50  | 79   |
| 69  | 81  | 86.7 |
| 70  | 82  | 77   |
| 70b | 70  | 78.5 |
| 71  | 77  | 79.2 |
| 72  | 84  | 81.3 |
| 73  | 41  | 67.8 |
| 74  |     | 72   |

|            |    |      |
|------------|----|------|
| <b>75</b>  | 76 | 77.8 |
| <b>76</b>  | 88 | 79.3 |
| <b>77</b>  | 63 | 69.8 |
| <b>78</b>  | 89 | 90.5 |
| <b>79</b>  | 96 | 97.8 |
| <b>80</b>  | 89 | 62.7 |
| <b>81</b>  | 77 | 78.3 |
| <b>82</b>  | 83 | 81   |
| <b>83</b>  | 61 | 74.2 |
| <b>84</b>  | 96 | 92.3 |
| <b>85</b>  | 97 | 76.2 |
| <b>86</b>  | 80 | 89.8 |
| <b>87</b>  | 62 | 83.7 |
| <b>87b</b> | 91 | 87.5 |
| <b>88</b>  | 83 | 94.5 |

## Appendix G: Counselor Interview Transcripts

Counselor 1:

I: I'm sitting down with (Counselor Number 1)?

C: Uh huh

I: My first question is, when you teach students to study for the test of English as a foreign language, what is your usual process

C: Actually I'm not a teacher who teaches English to the students, I'm a , I give the consulting information to the students, sometimes in discuss TOEFL, SAT, with the students because I will let them know how to arrange their test and what resources they can get online and sometimes, just test their current ability by taking a test through the TPO software, and then I will look at their scores in the four sections and then figure out which section they should put great emphasis on, and the. We can plan or fix a time for their test

I: So what strategies have heard or seen students doing for studying TOEFL that you would categorize as learning English language skills or knowledge

C: ...for the international standardized tests students may put more effort on their performance or how to improve their scores on the test than learning the English knowledge, so the concern is using the software for practice, sometimes they believe practice makes perfect, so they will put much time, put more time, for practice that they can, some of them will try to look back to their performance, figure what kind of types of questions they make mistakes and what kind of abilities they should improve, sometimes just look scores and impression their numbers instead of ability

I: My next question is so what strategies have seen or heard of students doing that you would categorize as test wise, that is based on the test in order to improve scores?

C: Most of them will register or sign up for Astrill, I mean out of school class, and they will divide into three types we call it one is the basic class, the next is the class for 80, 90, I mean the score 80 or 90 maybe the third type is fighting for 100 , they will register for a class, some of them will use the TPO practice at home, do some practice, also they will focus on the vocabulary books, the old real books , try to do some real tests

I: So which type of strategy do you think is more popular, believe is more popular?

C: First two, register for class and TPO software

I: Why?

C: Oh...they will give them the direct impression with a number, with a score, how they performed. With the class they may think the teacher have experience in teaching English or the exam oriented English teaching.

I: Ok, which type do you believe is more effective for raising TOEFL scores?

C: You mean the effect?

I: Well like, learning more English language or the test wise strategies? Which works better?

C: I think vocabulary is the most basic for Chinese students, if they can't read the vocabularies they can't figure out the true meaning of the passage, and second they should establish the ability of understanding of writing English with their own comprehension, some of them they just recite the sample provided by the teacher, they don't know how to use it how to write it with their own voices

I: Have you observed differences in study habits of students of high, medium and low English ability?

C: The high performance I observe is give. By (Student), you must know her, she has not participated in any out of school classes he just get involved in our foreign teachers classes and she has a very good habits of reading habits, books, both English and Chinese, so I believe her English has built up by involving our class and what she has learned is the English knowledge not just the English language. And the mid one I think they can understand and they can use English in the most of their times when they write the research paper they will organize their own ideas they still feel a little afraid of organizing their own language. And their uh, the most the uh, the worst students cannot even read the English passage, for example uh, I can't name the students name, this morning he gave me the paper let me print it out, one of the assignments of the human geography, the first line of his paper is that, what the picture shows the country built the um, as my common sense I think he wants to express the distribute of Brazil's population, I think he, he both cannot comprehend the language and he is writing and he didn't have a real understanding of the human geography

I: So uh, who do you believe has the most influence on the student's beliefs about studying in general but the TOEFL in particular?

C: Uh, I do believe the parents are the first teachers of the children, for (Student), her mother gave her great freedom to explore the knowledge she is interested in so she has the freedom to choose the book she likes, still have much time to read, just read , not focus on her scores. But some students are focused by their parents to do more uh, practice the reading, so I think the most influence is the parents. The second of course is subject teachers or counselors who will give the students suggestions, some of them will agree with our suggestion, some will at first but then their parents... parents suggestion will conquered our ideas

I: So where do the parents get their ideas?

C: Uh, sometimes they will focus on the other agencies, some agencies may provide how to plan the study abroad process, and some will give the suggestion about English training

I: So uh, what is your understanding of what a student means by quote studying TOEFL vocabulary or quote studying TOEFL book?

C: You mean what do they usually do?

I: Like if they say, what are you doing? I said I'm studying TOEFL vocabulary, what do you mean ?

C: For some recite, some will try to figure out which words have a similar meaning with one specific vocabulary or refer to sample sentence and figure out in which Constance should be the word used, but some just see how to write or to recite.

I: So similarly what does a student mean if they say quote looking for key terms?

C: Key terms? Uh I think most of them will recite the definition with understanding use, how to apply in the real context

I: Uh, what do you believe are the most common ways for students to study for each of the sections of the test, writing, speaking, reading listening?

C: Sorry I don't...

I: So what do you believe are the common ways for students to study for each of the four section of the TOEFL test, writing speaking reading

C: For writing they will get some sample from their teacher, try to recite I'm, typical sentence for statement, argument, spreading ideas, for comprehension they will do a practice, they will characterize most of the test, most of the questions into different types, for example the information clues, the word vocabulary question there is a way one student I know she usually do the practice according to the types. But I think the scientific way for students to do the reading section is to read the text totally and then go through the question instead of put them into specific types and omitting the whole picture passage, for the speaking section they will also try to recite some samples, uh form listening students will write down some part of the information they have listened but they can't write down the whole meaning they have listened to

I: Thank you so much

## Counselor 2

I: I'm here with (Counselor 2) my first question is when you teach students for the test of English as a foreign language, TOEFL, what is your usual process

C: Usually I will talk to them about the test, the test is not something you can prepare to recite of words or sample questions answers, speaking writing, outside training centers will provide lots of sample essays so that they can have answers for the speaking and writing part, but usually I will tell them, first of all TOEFL is a language ability test, it's not where you see the same thing so you do the questions and get the high mark, secondly I will suggest the official guide, the official guide you can use to build your vocabulary...sorry not the vocabulary test... first of all is vocabulary, and we have known that you have to prepare for each part, listening , speaking, reading and writing, based on official guide, like what questions you will have, what are the skills to answer these ones, what the test tests you, you have to know that, the target, so you will know like... prepare the test will be much easier

I: So what strategies have you seen or heard of students studying for TOEFL that you would categorize as learning English language skills or knowledge

C: ...learning language skills...how about the for example vocabulary like a at first they will recite the vocabulary book, like the Chinese definition and some they have sentences, but doesn't actually work, the students acknowledge not very, how to say, efficient, so and I suggest them based on the English definition and you can make the English definition like from the dictionary, it's difficult because this is now the speed they usually recite the words, but when they realize it is very helpful, I recite the words and do the reading from the textbook they will make the similar the same vocabulary words and they will double check it and they will not remember it...so they double check in the dictionary and they will read again the same words and they have to memorize that. This is the student she told me is really helpful when she studies vocabularies

I: What are some strategies you have seen or heard of that you would categorize as test wise, that is based on an understanding of how the test is set up in order to improve scores

C: ...test wise ... well I do know they memorize some bridge words for the writing and speaking parts, for example , to conclude, or somehow...some words like this they will somehow and they will use them in their essays, I think that's kind of test wise skills

I: What kind of strategy do you think is more popular? And why?

C:...depends on different centers

I: Why

C: I don't think, I think the students the ability and when they understand things the study is the personal thing think they need some extra help from the ...how you say... the centers ....teachers who are experts with the tests, to teach them some test wise.

I: Which type do you think is more effective for raising TOEFL scores and why?

C: Reading...based on my understanding of students in this age I think they lack of reading , both in Chinese and English, so this is the right input, reading and listening, you will have the output

I: So uh, have you observed any differences between study habits for TOEFL between students of high, medium and low English ability

C: Uh...

I: So top achievers, low achievers, medium achievers... have you seen any differences in what they do?

C: Yeah, of course. So like low medium, how do I say, medium low they don't have the active process of learning I think they, the difference between top and low is the process...the top achievers take care initiative on their own learning, so when they read and listen they are active, their brain is active so they catch the words they don't know and they learn, the medium ones they probably just go through, even they learn probably won't notice they don't learn, they probably won't notice that. And the difference between middle, medium and lower is the medium can do according to the requirement....at least you can, like teacher and parents can do this they probably can do so they will get through. The lowers ones not like they can follow steps they can do according to their pace so that's the difference ok

I: Who do you believe has the most influence on student's beliefs about how to study in general and for then TOEFL in particular?

C: That's a big question, if you trace back it should be their parents, based on their education, , know that they need to learn for themselves, learn everywhere, this is the process, well when they grow up as a teenager, teenagers I think, the will get they have got a lot of information, information from the counselors, information from the training school, and from the enrichment center, their parents, so it's based on the personality, the personal experience. Like for example if it's the top student they believe they need to study for themselves, they need to take initiative when the enrichment center teacher tells them you need to recite the sample question, they won't agree with that, but the medium or lower achieving student they are going to believe, recite this word you're going to get 100, or yes I can do that, so they seem different

I: So what is you're understanding of if a student said, studying TOEFL vocabulary or studying TOEFL book. What do they mean when they say that?

C: In the market they have lots of publishers, they say SAT prep, TOEFL prep, so the first one is always vocabulary, otherwise you won't understand it, so this understanding they have a lot of book and they ...list the words, some they list the words in alphabetical order, some they mix it...and when they say studying with the vocabulary book they mean recite the book itself

I: What is your understanding of what a student means by looking for key terms?

C: Hmm?

I: They say I'm looking for key terms in a listening or reading?

C: Uh...that means when they reading they want to catch the theme they want to catch the meaning, they will point out their...keys, in the first few sentences it's helpful to them for finding main idea

I: Ok, what do you believe are the most common ways or methods to study for each of the four parts of the test, writing, speaking listening and reading?

C: They do practices, that's the main thing, they practice, like with the test online, they have all the mock it's not the real TOEFL questions their the retired questions, I don't know how to say that, old TOEFL test questions, and they just practice with those ones

Counselor 3:

I: I'm sitting here with (Counselor 3), is that right?

C: Yes

I: I'm never sure. About that,, so my first question is when you teach students to study for TOEFL, test of English as a foreign language, what's your usual process?

C: I will try to determine what...first I will try to determine what, which part I would like to focus wi5 students for the period, that is listening, speaking, any of those kinds of skills, which skill I would like to focus on for this period, I would like to find suitable materials for them to use, during my class time I would like to make students participate more, like I want to give them time to practice it's much more than the time I tell them the skills or steps they need to prepare for the exam, because the students actually take extra like classes outside the classroom and I am pretty sure they are aware of the procedure of the test and the uh abilities or skills, that's why I mainly focus on the like practice rather than

I: So what strategy have you seen or heard of students studying for TOEFL that you would categorize as studying English language skills or knowledge?

C: Actually, most of them practice TPO test online, for that actually they categorize the test as four parts ,listening speaking...four parts, when they do practice the test, I believe they actually improve their language ability, listening ability, reading ability, I believe they do need to read more, but like, literature, nonfiction materials because the TPO materials are actually very....limited in the knowledges for students to master their reading abilities.

I: Sorry what strategies have you seen or head of students doing that you would categorize as test wise, that is based on an understanding of how the test is set up in order to improve scores or....

C: Again TPO, but maybe actually there are anticipated tests, there will be some anticipation before the test, the students will try to get those resource questions, in some way it might help

them improve their scores, others there will be some outside classes, teachers will tell students how to eliminate unreasonable or impossible choices in the multi choice questions, there will be the writing samples, the speaking samples which the students will memorize, and they will just put the actual test into those forms, so they will enhance their scores

I: Which strategy do you believe is more popular

C: The writing sample , almost every student uses that, I believe those samples are not really helpful because ETS is cutting down the scores for using samples because they believe everyone is writing the same things

The same beginning the same examples, so if you use the sample not give you high score

I: Why do you think that's so popular

C: It's easy for them to use, just need to memorize some sentences memorize some linking words and they will use it when they actually write something, it's not really help them with their actual abilities

I: So what type of strategies are most useful for raising students TOEFL scores and why?

C: Well for year one students they need to memorize the vocabulary, find their own vocabulary system, and the grammar, most of them just translate their Chinese thoughts to English sentences, but those expressions don't even exist, so I think they need to read more proper English. And learn to write a proper English sentence.

I: And year two?

C: Year two for example , they need to really push the. Out of their comfort zone. I've seen students like they are good at writing, they just keep doing it saying if I'm not good at speaking I won't try, I don't even like to practice, I do believe they need to focus on the things they are not so good at

I: Have you observed any difference in study habits for TOEFL between students of high, medium and low language ability?

C: Can you repeat the question?

I: So have you seen any difference in strategies or what have you for our best students, medium students...

C: The strategies they use

I: Yeah, like are high achievers more likely to use one strategy over another

C: Actually the best students have high scores in every subject they just have a good study system or habit, they tend to read everything thoroughly, make sure nothing is confusing them, make them like, repeat, make sure everything is well in their study. It's more like they, I don't know, study habit, Keep going I don't think, but for the low achieving students most of them have problems with their vocabulary, all of them have problems with their vocabulary, they don't want to recite vocabulary, I have students only take 20 minutes a day to go through their vocabulary, so they couldn't memorize anything

I: So who do you believe has the most influence on students beliefs about study in general, but also about TOEFL in particular?

C: ....Yeah, Teachers, training center?

I: Yeah, like their parents or whoever

C: Well it depends, students who take classes outside, their outside teachers , like the New Oriental, they have the authority because the students, I think once they pay for something that's valuable and teachers have the, I don't know, know more about it

I: So what is your understanding of what a student means by studying vocabulary or studying TOEFL book? What do they mean but that

C: Well vocabularies means reciting vocabulary, the other one?

I: Studying TOEFL book

C: Maybe practice the TPO test, because there is book version and online version

I: So what's your understanding of what a student means by studying for key terms?

C: Looking for key terms, in reading?

I: Reading or listening...

C: Uh, like TOEFL and SAT, some questions are evidence based so I believe they try to find the evidence so they can eliminate the impossible choices.

I: And so finally what do you believe are the most common way for students to study for each section of the test, writing speaking, listening and reading?

C: You mean a common way?

I: What's the most common way?

C: Reading helps a lot in reading, oh no let me change answer listening because uh, except from the writing, all the other parts involve listening, they listening good

I: What's their method, for each of the four?

C: What are they currently using

I: Yeah what's the most common study habit for them to learn each of the four skill?

C: TPO

I: Ok thank you very much

## Appendix H: Student Interview Transcripts

### Student 1

I: Ok so I'm here with (Student 1)

S: Yes

I: Thank you for coming, how long have you been studying English?

S: Since kindergarten

I: So how many years

S: More than 12

I: In that time how has your English learning changed?

S: Changed? You mean?

I: Like how has teaching study....

S: Like when I was in kindergarten i was actually in international kindergarten it's actually bilingual, so half of the time I speak English , it's not about grammar, just speaking English, then in primary school I generally speak less but we do more exercise about grammar, and uh....in middle school I come to TJFLS, so I need to speak, to write, grammar, I need to know the culture background

I: Uh, do you think that is a typical experience for students like yourself

S: I think no, I....because my parents they might need to speak English as a way to communicate, so they let me spend more time than other student son this

I: How many times....so you've taken the TOEFL?

S: Yeah

I: How many times?

S: Three times

I: Has your score gone up, stayed the same or gone down?

S: Up

I: Why do you think that is?

S: Um...because, for me I take the TOEFL courses, they will tell you some skills to help you achieve high scores, at the same time use the book to recite vocabulary, for those reasons I have a better TOEFL grade

I: Do you think the TOEFL is fair?

S: I don't think the TOEFL is fair

I: Why not?

S: For TOEFL you don't really need to improve yourself, just learn the skills to get a higher score

I: So when you get a score do you know why you received that score?

S: Yeah because you have a chart that is very detailed to show you that what your weak point

I: So has your study method changed since you started studying for the TOEFL?

S: Yes I think so, because previously I learn because I want to improve my English, now I just want to improve my score

I: Has your English improved as a result of studying for the test?

S: It has improved but something I have to do I improve myself

I: What is the most difficult part and why?

S: I think the speaking part is most difficult because I think for these students, for students like me it's very hard to speak to a machine in part because you don't have communication

I: How do you study for vocabulary

S: I use the vocabulary book, I recite

I: Have you heard of any other methods students use for studying vocabulary, if yes what are those methods?

S: I think some of my classmates they read books in English everyday, other students watch films, tv series to improve their vocabulary

I: How do you study for writing and speaking?

S: For speaking I practice it with my TOEFL teacher, for hundreds of topics, you do the same thing you will have a natural reaction when you are in the test, for writing if we also practice a lot of topics, I write it than teacher revise it

I: Have you heard of any other methods students use for studying writing and speaking, if yes what are those methods?

S: Some students they practice with some apps, With some native speakers, for writing they write journal for for read a book, journal

I: How do you get you writing and speaking practice judged?

S: Judged..?

I: If you do a practice how do you know how you did?

S: The judgment is from my TOEFL teacher

I: What do your fellow students believe are the most effective methods for improving TOEFL scores?

S: I think my friend my classmates and I the first is know the skills, it's important, time management how when you see some types of topics you have the answer, you don't even need to look at the article, the other is vocabulary

I: What kinds of methods for note taking do you use?

S: Uh, note taking...when I'm listening it's really fast, so I don't have time I just write the key words and write some symbols

I: Do you know of any methods for note taking that other students use?

S: I think most students use this method

I: Do you know what other students do to study the for, at of the test...how it's organized

S: I think it's the teachers, TOEFL teachers, they tell you the format, the methods and uh, what's the best strategy for this

I: If you are not sure of an answer is there a way to improve the chance of getting the right answer?

S: I think so, for example I reading there are some methods teachers told me where if you see this type of question just uh, when have four choice and only three can be correct you can do that

I: If you encounter an unfamiliar word when you are studying what is the best way to learn it?

S: I would first check it in dictionary, I would write it in my notebook and I can see it every day

I: Ok, when you are doing listening practice do you listen for key terms, why or why not?

S: Key terms?

I: Key ideas...main ideas....

S: Oh, yeah I think so, because ....the questions around it so you have to better understand it

I: Do you think it's a good idea to do practice tests? Why or why not?

S: Yes I think it's important, the first reason I think is because when you do the practice test, the mock test you have the kind of experience of taking TOEFL test, you feel like when you take the real test so you won't be nervous, the second reason when you take the test you know exactly what's going on

I: Do you take other English classes besides the ones at your high school?

S: Yes, I also take classes enrichment classes out the school, summer time I will take theater courses

I: TOEFL?

S: Yes definitely and SAT

I: How do they compare to your regular English classes?

S: Compare to our classes....our classes we read the textbook, the teacher help you to understand the ...to inspire you , make you think, help you know, the critical thinking, for TOEFL it's just practice over and over and over again

I: How would you compare preparing for western university with preparing for Chinese university?

S: For western university you study you have to keep the GPA, the standard tests, TOEFL, SAT, but for Chinese you take Gaokao, you just do what the teacher tells you to do and you can achieve an high score

I: Do you know of student going to training schools for Chinese tests, like Gaokao?

S: Yes, I think they go to the center more frequently than us

I: More frequently?

S: Yes

I: Do you feel like you will be ready to attend English language university?

S: I think I will, in high school I attend the AP courses, I think in my opinion they are similar to the university courses

I: If you could improve one aspect of TOEFL or your English education in general what would it be?

S: I think for the standard test they shouldn't just have a very standard format for the teachers outside the school they can find the rubric and they can study for it, help you...doesn't really take English to achieve a higher score

I: Thank you so much

Student 2

Researcher: I'm sitting down with (Student 2)

Student: Yes

R: My first question is how long have you been studying English

S: I have been studying English since grade four,

R: So what would that be?

S: About nine or ten years

R: In that time how has your English learning changed

S: Naturally I studied English in Primary school but it was fairly easy and I didn't focus on that because our English course is quite easy, it's not challenging but since middle school it's more challenging and I kind of falling behind

R: What do you mean easier, what was challenging about it?

S: Well for example primary school students just study vocabularies and has simple conversations and in middle school teachers focus on the grammar questions especially in many schools we focus on the test

R: Is that the Zhongkao?

S: Yes the Zhongkao...In Zhongkao grammar is a very large part so we need to study...even some grammar question is kind of inappropriate...basically nobody speaks like that but the question just perceive if you understand the grammar

R: What about in the last two, three years?

S: The last two years studying English? In?

R: Everything

S: What do you mean?

R: So everything you've done to study English in the last few years, high school and in any outside classes

S: In high school because we are in AP center my English ability improved a lot because...well because in Middle School basically for the writing and reading, but in after TOEFL in SAT I have to study...sorry...I need to improve all my abilities, listening...speaking

R: Do you think this is typical experience for students like you?

S: ....I think this experience is quite nice but I don't think it's for every student in China

R: How about compared to your fellow classmates

S: They also have a great improvement

R: Have you taken TOEFL?

S: Yes I've taken it six or seven times

R: Why did you choose TOEFL over the other ones?

S: I think it's because the TOEFL is more widely used, accepted by more people around me...the most popular choice

R: Do you think the test is fair?

S: Is it fair...I have to say...not very fair

R: Why not?

S: I'm not sure I get the right...uh...and I feel good but the score doesn't come out great sometimes I feel tired and I'm not in a mood but the score seems much better... I'm not sure it can measure my ability perfectly

R: When you get your score do you understand why you got the score you did?

S: I think so....I mean I know why my score came out that way, the TOEFL Score has two parts the reading score and the listening score I have no doubts they're all multiple choice and they have a valid a perfect choice for that. The questionable part is on the...speaking and writing part, they're judged by people, judged by people

R: In the times you've taken the TOEFL test has your score gone up, down, stayed the same?

S: Well my score goes up slowly, that's why I take it so many times

R: Why do you think that is?

S: Well I think my English ability improves slowly

R: What do you think is the best way to improve your score?

S: I think the best way is to take a class

R: What kind of class?

S: You know the class in New Oriental for test taking

R: Has your study method changed for TOEFL?

S: Yes my method has changed because every test has a trick and you...so...focus on what the test measure....like uh...like for example the speaking part of TOEFL the will ask you fifteen questions and you can improve f you practice these kinds of questions even though you won't face these kinds of questions in reality when doing this practice my English ability proves less than my Test ability.

R: So do you think your English ability has improved as a result of studying for the test?

S: Yes in some way

R: Which part is most difficult?

S: Which part is most difficult? For me is reading

R: Why?

S: Because reading requires a long time focus on a passage for a long time...usually make me tired

R: How do you study for vocabulary?

S: Mainly in two methods...one method is vocabulary book

R: What does that mean?

S: You will have a book...a book published by an education institution that puts every word in TOEFL in that book and you recite a chapter...every chapter twenty five or twenty six. Another method is reading, I think that is a much better way. We can read textbooks, novels to learn vocabularies

R: Have you heard of any other methods, things other students might have used to study vocabulary

S: Other methods to study vocabulary?

R: Like on the internet

S: Ah there's another way....there's an A P P on the phone that could helps people recite

R: An app to help recite?

S: Yes

R: How about for writing and speaking?

S: You mean in general or for TOEFL?

R: Specifically for TOEFL test

S: I do TOEFL test practice

R: What does that mean?

S: They have software called PPO...TOEFL practice online they put every published, even unpublished, some questions that might be stolen so they ...some questions are stolen but they put everything there and simulate the test, timing, they even have a screen like it...I mainly use this method

R: How do you grade them? Like how would you grade the writing test?

S: The writing test?

R: Like how do you know if you did good or bad?

S: Mainly through teachers....

R: They send that to a teacher and they grade them?

S: That's one method, you can also find a real TOEFL teacher and they'll grade the,

R: Ok, have you of any other methods that other students use for speaking and writing?

S: Speaking and writing...other than find a teacher? They possibly watch videos on YouTube or Chinese website that will have some tips.

R: Ok

R: : So what do you think students think are the most effective methods for improving TOEFL score?

S: Speaking frankly the English ability is most important because the tips and tricks can only help you for a while, the English abilities are on point...for a student like me for the last year...my final TOEFL score was in the third the last year, while some outstanding students like (Student) like (Student) finish in their first year....so...I think about why I fall behind, so I think it's the English ability

R: Ok, so what about note taking do you ever do note taking?

S: Yes I do take the notes

R: So what method do you use?

S: I will take a record of I don't do well...so can a do example?

R: Yes

S: So I find which part of my writing is ....the weakness of my passage...like which part is not very good so I will recite the teacher's words and put them in a notebook and so when I finally do a TOEFL test I will look through this book and see what kinds of questions I might have.

R: Do you know of other methods students use?

S: Other methods? Not this one

R: Do you know what other students do to study the format of the test?

S: Format?

R: Like, you were saying how you study specifically for a test

S: ...no

R: Ok, If you are unsure about the answer to a question, so if like a multiple choice question, is there a way to improve your chance of guessing the right answer

S: Multiple choice question

R" You don't know the question but you...

S: Can see the right answer? Yes, for example avoid absolute answers...

R: I got yah

S: The tone is too firm

R: If you encounter an unfamiliar word while you are studying what is the best way to learn that word?

S: Do you mean in practice or in test?

R: Either

S: In daily practice I will check the dictionary and ... if I will remember them...in test all you can do is guess...from context before and context after

R: When you are doing listening or reading practice do you listen for key terms?

S: No, because listening to key terms might ... initially I did do such thing but .... when I found out... if I don't focus on the whole context I might mix the relationships between the few key terms

R: Do you think it's a good idea to do practice tests?

S: Yes I think so

R: Why?

S: Because everything is good with practice...you need to familiarize yourself with the time and method as well

R: Ok, so do you go to any other classes besides your high school English classes?

S: Yes

R: How do they compare?

S: Compare with the school.... in classes outside school we focus mainly on the test

R: Ok

S: But in school we mainly improve English ability

R: So how would you compare comparing preparing for a western university to how the students in the main school prepare for a Chinese university?

S: How do I compare them....frankly I think we have some kinds of advantages because we can study more broadly we can pick the subject we want to study... we have more choice more choice...the second one we study advance...because in Chinese school we have Gaokao, in Gaokao one or two points miss a lot so practice a hundred, thousand times just improve one point or two points but when we study for the western university we can save that time and study more subjects...yes like in study history in Chinese improve score 95 to 98 but here I can get a 90 in history and 90 in economics , two subjects

R: Do students go to, like with TOEFL, but for the Gaokao?

S: ...sorry I didn't understand

R: So like you said you went to outside schools to study for TOEFL, do students in other schools do the same for Gaokao

S: Yes, yes, some of these institutions are initially formed for the preparation of Gaokao

R: Interesting, so do you think you're ready for a western university?

S: Yes

R: So if you could improve the TOEFL what would you do?

S: I think I would shorten the time

R: Why?

S: Takes like 4 or 5 hours and like my head start to get hurt

R: Thank you very much (Student 2)

Student 3

I: I am sitting down with (Student 3)

S: Yes

I: So my first question is how long have you been studying English

S: I have been studying English since I was 3, so uh 40, uh 14 years

I: In that time how has your English learning changed

S: You mean in what aspect?

I: Yeah

S: Well uh first at the beginning I just learned words, since then I mostly learn from Social media I uh listen to music, I was uh, I like American culture, especially African American culture, I learned English from that, native accent

I: Do you think that is a typical experience for students like you?

S: I think so.

I: So why, so you have taken the TOEFL right?

S: Yeah

I: How many times?

S: One times.

I: Well that changes my next question, uh, so uh why do you think you should take then TOEFL test?

S: To study abroad , to show my English ability

I: Do you think TOEFL is Fair?

S: No

I: Why not

S: I think I can ace it if I learn in the proper way

I: The proper way?

S: The test skills, your English ability is one aspect but your test skill is another aspect you need to know what is on the test

I: So when you get your score do you feel you understand the score you got

S: Yeah

I: Why

S: They will send you a letter that has every aspect of the test, why I got that,

I: So uh, why, uh what do you think is the best way to improve a TOEFL score?

S: Well the best way to improve TOEFL score is to use English daily, to everyday speak English with the foreign teachers to ....fellow classmates, well uh you know language is all about use, if you don't use it than lose it

I: So has your study method changed since you started to study for TOEFL?

S: Yes

I:How?

S: I particularly focus on the uh, test, what kind of questions the TOEFL will ask me. And I will practice questions many times and I will uh, collect some stories from me, from my daily life, for TOEFL writing, to present for the essay

I: Do you think your English ability has improved as a result of studying for the test?

S: Of course

I: How so?

S: Well I think I knew a lot of new words from a TOEFL, my English becomes more academic from TOEFL, I can write a proper paper from studying TOEFL, and I know how to speak or write in the proper way, native speakers understand

I: hat part of the test is most difficult?

S: Reading I think

I: Why?

S: Lot of new words for me

I: So how do you study vocabulary?

S: Well I will use particular vocabulary to make a sentence and will recite the sentence

I: Have you heard of other methods students use to study for vocabulary?

S: Yeah

I: Like what

S: Well, I mean there's a stereotype write the words a thousand times with the Chinese description until you can remember but I think that's not useful, because eventually you will need to use those words

I: Is that popular?

S: I think it's popular because it's fast a don't require a lot of thinking

I: How do you study for writing and speaking?

S: Writing and speaking... well I practice my writing with my foreign teachers and I will write about one essay per week and I submit to my teacher, well about the speaking part I will speak English with my parents, not everyday, but I will teach my parents how to speak English and also in school I will talk to my foreign teachers

I: Ok, have you heard of other ethos's you have heard of students using to study writing and speaking? If yes what are those methods?

S: Uh I mean, there's methods the main school, normal high school students use they just write an essay a paragraph around 200 words to uh, mainly focus on the test, Gaokao, and they will practice, not that native but use precise grammar

I: Ok, how do you get your writing and speaking practice judged, how do you know if you did well?

S: My teacher will give me the response

I: Which teachers?

S: You know (Counselor) from the counselor office

I: Ok, what do you believe other students think are the most effective methods for improving TOEFL scores?

S: Go out for classes...well the class offered outside will give you specific methods for the test, like a model to modify your language if you put into an essay you will get a high score on a TOEFL test

I: What kind of methods have you used for note taking ?

S: Note Taking? Well I will write if my professor, my teacher, speak very fast I will write the short, simplify words, like human geography I will write HG

I: Ok, do you know of other methods other students use?

S: Uh, I don't know

I: Do you know what other students do to study the format?

S: Format?

I: How it's organized

S: Oh yeah, format, you mean the other students...I don't know I never discuss this

I: If you aren't sure of the answer to a question is there any way to improve the chance of getting the correct answer?

S: Improve the chance of getting the correct answer...

I: So if you don't know the answer can you improve your chance of getting the right one?

S: Oh, i will guess probably

I: If you encounter an unfamiliar word while you are studying, what is the best way to learn that word?

S: If you see the word you don't know in a paragraph or sentence I will learn that sentence, and also I will know how to use that word, this is a method the counselors show us

I: Cool, when you are doing reading or listening practice do you read or listen for key terms?

S: Yeah

I: Why?

S: Efficiency, when you take notes you just take the significant words, you remember

I: Do you think it's a good idea to do practice tests?

S: Yeah

I: Why?

S: When you will get nervous when you TOEFL test, but if you do a lot of TOEFL tests it's normal for you to go for TOEFL test.

I: Do you take other English classes besides the ones at your high school

S: No

I: How would you compare preparing for western university to preparing for Chinese university?

S: Western universities you have to be versatile, you have to learn a lot of things, study for tests, you may have activities and writing a paper, doing a program and collaborating with you classmates

I: So you know other students...oh sorry do you feel like you will be ready to attend English language university when the time comes?

S: Yeah

I: Why?

S: Because uh, was study hard now and if I pass TOEFL test it convince I can handle the university level course

I: If you could improve TOEFL or any aspect of your English education, what would you do?

S: Pardon?

I: If you could improve the TOEFL, or any aspect of your English education what would you improve?

S: Syntax, When I'm writing a sentence according to (American English Teacher), she will tell me I'm not writing like a native speaker, I will improve my syntax by reading essays by the native speakers, and uh, just directly learn the,

I: Ok