The Impact of Changing TOEFL Cut-Scores on University Admissions

Laura Michelle Decker
Brigham Young University
The Impact of Changing TOEFL Cut-Scores on University Admissions

Laura Michelle Decker

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Troy L. Cox, Chair
Teresa R. Bell
K. James Hartshorn
Matthew P. Wilcox

Department of Linguistics and English Language
Brigham Young University

Copyright © 2017 Laura Michelle Decker
All Rights Reserved
ABSTRACT

The Impact of Changing TOEFL Cut Scores on University Admissions

Laura Michelle Decker
Department of Linguistics and English Language, BYU
Master of Arts

As the Test of English as a Foreign Language (TOEFL) is often used as a determiner for university admissions, this study observes the effect on the international student population at a large private university through the examination of the international student admissions data including TOEFL and first-year GPA from 2005-2015. With the anonymous data of 9,837 students, researchers analyzed the result of a cut-score change at the university.

Results indicated that the number of international students decreased at the university. As expected, the TOEFL data revealed a normal distribution for the overall (combined) score and subsection scores, while the GPA data did not. The ANOVA for the TOEFL revealed that the change in cut-scores was not completely implemented in 2010. The GPA results from the ANOVA did not appear to be increasing. Correlation analysis reflected a decrease in the correlation coefficient when comparing results from before and after the cut-score change. Correlations of the subsection TOEFL score presented interesting findings. Multiple regression analysis indicated similar conclusions.

Keywords: predictive validity, TOEFL, cut-scores, admissions, international student success
ACKNOWLEDGEMENTS

I would like to express my gratitude to my thesis advisor, Dr. Troy Cox. He was extremely helpful from the very start, and his continued support was invaluable to me. I sincerely feel that no other advisor would have helped me as much. Also, I would like to thank my committee who jumped in near the end, but definitely had a meaningful impact on the outcome of this thesis. Their fresh eyes provided valuable insight into this work.

Another thanks goes to my amazing cohort who through writing groups and classes provided encouragement and support. And finally, a special thanks to my friends and family who supported me by attending presentations, reading manuscripts, and being listening ears.
# TABLE OF CONTENTS

ABSTRACT .................................................................................................................................... ii

ACKNOWLEDGEMENTS ........................................................................................................... iii

TABLE OF CONTENTS ............................................................................................................... iv

LIST OF TABLES .......................................................................................................................... vi

LIST OF FIGURES ..................................................................................................................... vii

Introduction ..................................................................................................................................... 1

Literature Review ............................................................................................................................ 1

International Student Population ................................................................................................. 1

International student perceptions ............................................................................................... 2

International Student Admissions ............................................................................................... 3

History of the TOEFL .................................................................................................................. 4

Interpreting TOEFL scores ......................................................................................................... 5

Predictive validity of the TOEFL on GPA .................................................................................... 5

International Student Success ..................................................................................................... 7

Definition of academic success .................................................................................................... 7

Use of GPA .................................................................................................................................. 7

Case Study ................................................................................................................................... 8

Methods ......................................................................................................................................... 10

Methods Question 1—Linguistic Diversity .............................................................................. 12

Method Question 2—TOEFL and GPA Relationship .............................................................. 13

Results ........................................................................................................................................... 15
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question 1—Linguistic Diversity</td>
<td>15</td>
</tr>
<tr>
<td>Research Question 2—TOEFL GPA Relationship</td>
<td>17</td>
</tr>
<tr>
<td>Average TOEFL scores by year.</td>
<td>17</td>
</tr>
<tr>
<td>Average GPA by Year.</td>
<td>21</td>
</tr>
<tr>
<td>Correlations between TOEFL and GPA.</td>
<td>23</td>
</tr>
<tr>
<td>Multiple Regression Model.</td>
<td>25</td>
</tr>
<tr>
<td>Limitations</td>
<td>26</td>
</tr>
<tr>
<td>Discussion</td>
<td>26</td>
</tr>
<tr>
<td>Linguistic Diversity</td>
<td>27</td>
</tr>
<tr>
<td>TOEFL and GPA relationship</td>
<td>28</td>
</tr>
<tr>
<td>Conclusion</td>
<td>29</td>
</tr>
<tr>
<td>References</td>
<td>30</td>
</tr>
<tr>
<td>Appendix A</td>
<td>33</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1—Description of Data Sets ............................................................................................................. 12

Table 2—Analyses Run .............................................................................................................................. 14

Table 3—Linguistic Composition: Percentage of International Students .................................................. 16

Table 4—Descriptive Statistic TOEFL(iBT) overall .................................................................................. 17

Table 5—Descriptive Statistics First-Year GPA ........................................................................................ 23

Table 6—Correlation Summaries ............................................................................................................... 24

Table 7—Multiple Regression Model ........................................................................................................ 25
LIST OF FIGURES

Figure 1—Collaboration and Analysis Procedure ................................................................. 11
Figure 2—Box Plot Anomalies in TOEFL Scores ................................................................ 11
Figure 3—Population of Admitted Students Before and After Change ............................... 15
Figure 4—Mean TOEFL Scores by Year, Box Plot ............................................................. 19
Figure 5—Mean TOEFL Scores by Year, Trend Line .......................................................... 20
Figure 6—Confidence Intervals (95%) of TOEFL Score Means by Year .............................. 20
Figure 7—Mean GPA Scores by Year, Box Plot ................................................................. 21
Figure 8—Mean GPA Scores by Year, Trend Line ............................................................. 22
Figure 9—Confidence Intervals (95%) of Mean GPA by Year ........................................... 22
Figure 10—Effect of TOEFL Overall Score on GPA Before and After Change ................. 33
Figure 11—Effect of TOEFL Reading Score on GPA Before and After Change ............... 33
Figure 12—Effect of TOEFL Writing Score on GPA Before and After Change ................ 34
Figure 13—Effect of TOEFL Listening Score on GPA Before and After Change .............. 34
Figure 14—Effect of TOEFL Speaking Score on GPA Before and After Change .............. 35
Introduction

This study examines the relationship between the Test of English as a Foreign Language (TOEFL) and first-year GPA on a change made to TOEFL cut-scores at a private university. The results of this study can be used to help institutions make informed decisions about cut-scores. This research details a method that can be implemented by other universities to increase understanding of TOEFL scores, to encourage collaboration between admissions offices and other university personnel, and to improve the impact of TOEFL score use. Institutions can employ the scientific methods taught within their doors to understand and evaluate incoming international applicants.

Literature Review

Central to both the TOEFL and GPA are the students, and understanding these students plays a vital role not only to admissions decisions but also to the relationship between these two variables. The following sections describe the international student population in the United States (student perceptions of a university degree), international student admission (description and use of the TOEFL), and international student success (definition of academic success and use of GPA).

International Student Population

For many, the value of an education arises from leaving one’s home, perhaps for the first time, and interacting with people from different backgrounds. The introduction to a new world of ideas and individuals from different backgrounds is seen by many to have intrinsic value for the individual learner (Lambert, 2008). International students broaden this intellectual interchange by bringing people from around the world together. Many internationals desire the benefit of a degree from an American university, as it is highly valued in their native country and in the
United States (Bound, Demirici, Gaurav, & Turner 2015). Jobs in the United States are also higher paying (Clemens, Montenegro, & Pritchett, 2008) than jobs in less developed countries (McKenzie, Stillman, & Gibson, 2010), and the pathway to these higher paying jobs can include having a degree from the United States.

In the United States context, value is reciprocated when international students acquire the lingua franca of the business world. The increased diversity opens the doors to increased positive international relations (Peterson, Briggs, Dreasher, Horner, & Nelson, 2002) and can encourage the spreading of ideas. With state-sponsored institutes, international admissions offices might be incentivized to admit foreign students because the out-of-state tuition is financially advantageous. Therefore, admissions offices may face two competing interests: the desire for academically prepared students who can succeed while simultaneously attempting to have a diverse campus.

*International student perceptions.* Maintaining a diverse campus requires understanding the people applying, and this means understanding international students. Even though some governmental barriers exist, and there are some perceptions of discrimination about American universities, universities can make intentional efforts to become internationally friendly (Hser, 2005). Solutions to these obstacles include providing sufficient resources for international students and using objective measures to regulate admission.

When universities admit international students, these students naturally will have interactions with both faculty and classmates. These interactions can bring meaningful information to admission offices to develop resources. Andrade (2010) sought out the faculty opinions on international students in order to improve the teaching of international students. Faculty preferred methods of support for students that did not place more responsibility on them
as professors. Methods faculty tended to support included tutoring, study groups, and technology assistance. These kinds of resources would potentially provide one-on-one help to international students, thus increasing positive interactions.

Additionally, the findings suggest mixed reviews on whether English proficiency was limiting students at the university (Andrade, 2010). While this could indicate problems with proficiency measures, there are several underlying confounding variables to this issue. First, human perceptions are inherently biased. Namely, professors with more exposure to international students may be more generous about English abilities. Second, the requirements and expectations for international students were not explicitly stated in the survey or study and as such it would be desirable to have more information before drawing any conclusions, especially data using more unbiased information.

Regarding the biased nature of perceptions, learner beliefs about their needs may not converge with their actual needs (Liu, Chang, Yang, & Sun, 2011). However, student perceptions do play a role in student approval. Sturman (1996) noticed perception differences when giving students placement exams at an intensive English program. In investigating these differences, he found that students' beliefs about a measure influenced their later satisfaction with the school using that measure through multiple survey questions. Therefore, it behooves a university to seek a needs analysis not only from a wide variety of sources, which include both student and faculty perspectives, but also an analysis that includes more objective measures such as the TOEFL. The TOEFL can provide an equalizing effect as an impartial objective measure for students to prepare.

*International Student Admissions*

In recent years, some research suggests that the number of international students
continues to increase within the United States (Ananyeva, 2014; Bound, Demirici, Gaurav, & Turner, 2015; Chow, & Marcus, 2015; Madge, Raghuram, & Noxolo, 2015). Admission offices are the filter not only for international students, but for American students as well. The belief that students must be sufficiently prepared in order to succeed works as a balance to the increasing desire for an American education. Therefore, admissions offices must make two important determinations: whether students are ready for an environment where English is spoken exclusively and whether or not they will succeed academically. To make these determinations, the TOEFL is one of the many measures used by universities.

**History of the TOEFL.** The TOEFL was developed with the purpose of testing non-native English Speakers who are working on getting into a predominantly English speaking university (Alderson, 2009). There have been three formats of this test: the first—the paper based test (PBT), the second—the computer based test (CBT), and the final—the internet based test (iBT). (Abunawas, 2014; Alderson, 2009). Each version of the test has continued working towards solutions that reduce problems from previous versions and has continued adapting to the changing need of modern technology and education. For countries without the technological infrastructure, the TOEFL PBT is still administered.

Initially, the TOEFL used more of a discrete-point approach, but as the field moved towards a more communicative approach, integrated assessments followed. For instance, Waters (1996), in an in-depth survey analysis, pointed out a need for more integrated tasks which were more authentic and required more crossover between skill areas. Understanding the context and use of TOEFL scores over the past few decades can provide direction for studies and allow for better analysis and decision making. Benefits include improved interpretations of scores, better comparisons between tests, and a better platform on which to compare other tests.
Interpreting TOEFL scores. As a basic outline, the TOEFL iBT score is comprised of four section scores: reading, writing, listening, and speaking. Each of these sections has an emphasis on the communicative functions that are required of academic work (Abunawas, 2014). Additionally, the TOEFL iBT is a norm-referenced test, which tells school admissions a student's ability level in comparison to their peers and thereby allows institutions to vary the score required for entry. This is important, as sometimes cut-scores give the perception that a person must obtain a certain skill level, but this is not the case. At the same time, the Educational Testing Service does provide a general comparability chart to a student's score and their potential proficiency level.

Universities must also consider what will be best for the university. For example, schools that want increased admission rates from international students might lower their cut-scores but also include mentoring programs for those students. Other schools could create a partial acceptance decision dependent on completion of an intensive English Program, which then would lead to full acceptance at the university. As with any assessment, administrators making a high-stakes decision based on an exam need to have a clear understanding of what a particular score means as well as what expectations the university has for students.

Predictive validity of the TOEFL on GPA. Many researchers support the conclusion that proficiency has a relationship with academic achievement. Messer (1984) conducted a study using the paper-based TOEFL and GPA scores, which revealed a significant difference in students with high TOEFL scores and students with low TOEFL scores. Cho and Bridgemen (2006) used the internet based TOEFL score and came to a similar conclusion, finding that there was a "meaningful relationship" between the TOEFL and GPA, despite low correlation values. Additional qualitative research (Ren, Bryan, Min, & Wei, 2007) has supported students' views of
the TOEFL assisting in academic success. Ren et al. (2007) looked into aspects affecting academic success in six students from Eastern Asia. Although the premise of their research was focused on several different factors indicating success, one semi-consistent factor from participants was the impression that studying for the TOEFL helped them improve their English even if it did not indicate academic success—the exception being participants who were in engineering and scientific degrees, who did not feel the same way.

However, some research found opposite or mixed results where TOEFL and GPA did not have a strong relationship. Graham (1987) points out various studies with mixed results: some found no correlation, unsure correlation, and positive correlation between TOEFL and GPA. In his review of previous research literature, Graham questions the use of proficiency tests as a valid measure of language abilities. This suggests the need for replication to ensure that universities are using proficiency test score results effectively.

For subsections, Ginther and Yan (2016) found a positive correlation between GPA and TOEFL writing and speaking sub scores, but a negative correlation with the reading and listening sub scores. A cluster analysis was used to investigate that unexpected finding and there emerged three distinct student profiles: high, low, and discrepant. The negative correlation came from the discrepant profile that consisted of students who did well in reading and listening but not writing and speaking. Furthermore, the students with discrepant profiles were not succeeding academically and their low GPAs contorted the correlational study. To prevent this from happening in the future, they recommended that institutions not only have an overall cut score requirement, but more specific subsection score requirements. For example, universities that require a general cut-score of 80 on the TOEFL may also require that no subsection score is below 20. Overall, these studies reflect a solid interest in the language assessment field in the
relationships between the TOEFL and GPA.

**International Student Success**

Even with careful interpretation of the TOEFL scores, admitting students who will not succeed will not help diversity and may in fact harm the university. For example, peers may resent the international students for holding back the class because of perceived language difficulties. Therefore, the importance of using TOEFL scores lies in the desire for successful students, and the actual use of what "success" in a university means. Promoting a clear set of objectives both long term and short term would provide admissions offices with a clear purpose and direction for what to look for in potential students. In this respect, using the TOEFL scores could be important to a clear understanding of success.

**Definition of academic success.** Academic success is defined differently by administrations, departments, teachers and students, and is often used interchangeably with student success (York, Gibson, & Rankin, 2015). Indeed, the purpose of York et al. was to identify and operationalize this very term. They identified multiple components to academic success within literature, which includes career success, attainment of learning outcomes, persistence, acquisition of skills and competencies, satisfactions, and academic achievement. Unfortunately, these constructs are difficult to measure and the data are rarely collected on a regular basis. However, one measure that is consistently collected is GPA.

**Use of GPA.** There are some challenges associated with using GPA as a measure of academic success. For example, GPA does not account for the difficult classes taken by students, and some classes are more difficult than others. Abunawas (2014), in a meta-analytic dissertation, notes several additional difficulties in using GPA, including grade inflation. A well-defined assessment program can lead to a more accurate grading scale, and departments and
schools often seek a well-defined assessment program that will satisfy accrediting agencies (Walvoord & Anderson, 1998). However, even with a more accurate grading scale, using GPA as a measure of academic success has limitations.

On the other hand, GPA is still commonly used as a measurement of academic success, and positive factors of GPA use are that it is standardized, convenient, easy to interpret, and universally recognized. Additionally, Bacon and Bean (2006) found a large degree of internal reliability with GPA scores. Therefore, even with limitations, GPA has some benefits. For the purposes of this study, academic success was defined as students who received proficient grades in their university classes as measured by first-year GPA.

Case Study

This study explores the impact of establishing and subsequently changing the admissions’ cut score of the TOEFL iBT at a large, private university. In 2005, when the university adopted the iBT, a conversion chart was used to determine that a PBT score of 500 was roughly equivalent to an iBT score of 66. After the initial cut score had been in place for a few years, some concerns arose. For instance, there were complaints from professors that international students were struggling and were unprepared to meet course-work demands. In response to this anecdotal evidence, the Linguistics Department created a post admission ESL test that would require students to take ESL service courses. In addition, an ESL student writing lab was created to serve the unique needs of second language writers.

In 2010, an administrator at the university's Intensive English Program, a non-matriculated program that focused on preparing the English ability of students for university studies, noticed that students with inadequate proficiency were being admitted into the university. He first consulted ETS technical reports and found that a TOEFL score of 66
reflected intermediate language ability. He then conducted a review of *US News and World Report*’s top 100 universities and surveyed the cut-scores of those universities and found that, while ranked 75th, the university had the lowest cut score. After consulting with International Admissions, he received the TOEFL admissions data at the university for 2009-2010 as well as the first-year GPA. Based off these findings and feedback from other entities at the university, the cut-score was changed from 66 to 80 on the TOEFL iBT, with minimum subsection requirements of 20 in each skill. Ultimately, the university went from one of the lowest accepted TOEFL cut-scores to the median of TOEFL cut scores (Cox, 2010).

To examine the effects of the change (i.e., the students were more successful), International Admissions requested further examination of the impact of this change. The expectation at this university was for students to have sufficient English proficiency to meet academic demands. The researchers for the current study received admissions data from 2005 to 2015.

A few decisions for this study were made from the review of literature. First, the review of the top 100 university cut-scores was redone in 2015. The 2015 replication of this original review made note of a couple of new trends in the use of TOEFL scores in the United States top 100 universities. For example, many of the top Ivy League schools no longer provided cut scores, and in some cases, the TOEFL was not even required. Some schools even suggested that students should have proficiency comparable to a native speaker. By reviewing this information, researchers were able to evaluate whether or not there were any additional changes to trends in TOEFL cut-scores.

While other ESL admissions tests exist, the most widely used at this university was the TOEFL-iBT. For this reason, the study focused on this exam. Also, for the purposes of this
study, academic success was defined by first year GPA. While noted earlier that this measure can be problematic, it was the only data available, and albeit imperfect, does provide valuable insight.

With these assumptions in mind, two main points can be extracted: First, the evidence from current literature warrants continual examination of the relationship between TOEFL and GPA as it affects a university. Second, defining and understanding the relationship between TOEFL and academic success will help inform admission committees in the use and understanding of TOEFL scores. Therefore, this study will investigate the impact of the cut-score change at the private university, and specifically work to answer the following questions.

1. How have the new TOEFL requirements influenced the composition and linguistic diversity and linguistic diversity of the student population?
2. To what extent has the change in TOEFL requirements influenced GPA of first year international students? Is there a net gain in GPA scores?

Methods

The configuration of this study required close collaboration with the international admissions office and used a multi-step process (Figure 1). After collecting demographic and academic data, basic descriptive statistics were calculated. When the number of admitted students with TOEFL scores was checked for consistency by year, inconsistencies were found between the newly provided data and the Cox (2010) report (see Figure 2). Namely, the entire population in the 2015 data was smaller than a sample of that same data set taken by Cox five years earlier. After consulting with the international admissions office, it was discovered that data older than five years had had some database fields lost or deleted, and as this was historical
data, it did not negatively impact any currently enrolled students. However, it prevented researchers from doing a complete 10-year longitudinal study on TOEFL and GPA because of the lost TOEFL data.

Figure 1—Collaboration and Analysis Procedure

Collect Data
- Contacted university admissions office

Basic Descriptive Statistics
- Discovered anomalies
- Contacted admission office
- Recovered data from Cox 2010

RQ1
Linguistic Diversity
- Performed analyses
- Collected Percentages
- Chi-Squared Test

RQ2
TOEFL & GPA Relationship
- Performed Analyses
- ANOVA
- Linear Regression
- Multiple Linear Regression

Figure 2—Box Plot Anomalies in TOEFL Scores
Fortunately, the missing TOEFL data did not affect the first research question and all ten years could be analyzed using the 2015 data. However, for the second research question, the years with the usable TOEFL data (2011-2015) were combined with the Cox data set (2009-2010). For more information, see Table 1.

**Table 1—Description of Data Sets**

<table>
<thead>
<tr>
<th>Data Set Name</th>
<th>Sample (n)</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010 Data</td>
<td>166</td>
<td>2009-2010</td>
</tr>
<tr>
<td>2015 Data</td>
<td>9837</td>
<td>2005-2015</td>
</tr>
</tbody>
</table>

Once correcting for the database inaccuracies, researchers performed multiple analyses to answer both research questions, including descriptive statistics from before and after the change, one-way ANOVA tests, correlation analyses, and a multiple regression analysis. The findings from these analyses were then presented to the University's International Admissions office, so they could use the results in making future decisions.

**Methods Question 1—Linguistic Diversity**

The process for identifying the impact on linguistic diversity was straightforward:
identify language categories, calculate the ratios from before and after the cut score change, analyze the data for discrepancies and abnormalities, continue collaboration with the admissions office, identify confounding variables, and synthesize the information collected. A particular challenge was narrowing the language categories. Due to the 127 different native languages of the applicants, 11 main languages were identified based on a sample size (n>20), and the rest (languages with n<20) were placed in an "other" language category. Then, the percentages of admitted applicants were calculated by native language for the combined 2005-2010 years and the 2010-2015 years. Results were compared using a Chi-squared test.

Method Question 2—TOEFL and GPA Relationship

To examine the relationship between TOEFL and GPA three different statistical tests were used: ANOVA, linear regression and multiple linear regression. The ANOVA test determined whether TOEFL and GPA were affected by year. The linear regression showed the relative strength of the direct relationship between the TOEFL and GPA variables, and the multiple linear regression showed how all variables interacted in each other’s presence. When combining the two data sets, the years 2005-2010 had to be removed from the 2015 data set before adding the Cox 2010 data set to make sure that the samples remained independent. Then data sets were inspected for normality and equal variances. Histograms, boxplots, and descriptive statistics were used to double check the method.

The assumption for normality failed as the data for GPA was non-normally distributed, but an ANOVA test is robust to non-normal distributions. In checking for equal variance, the results of a Levene's test on TOEFL data by year, $F(6, 605.5) = 11.4, p < .001$, showed that there was unequal variance in the TOEFL data, and though the result for the GPA data was borderline, $F(6, 795) = 2.32, p = .04$, the result still indicated unequal variance. To account for the unequal
variance, researchers used the Welch ANOVA test, which is more robust to unequal variance. This specific ANOVA test was conducted for both variables to determine the statistical significance of a year on the TOEFL and on GPA. These ANOVA tests ultimately checked for changes in these variables—TOEFL and GPA, and provided a rationale for continuing with the correlation analysis.

The linear regression analysis required adjustments, and the Spearman correlations procedure was chosen over Pearson's correlation model. A limitation in using Spearman's correlation is that there were a good number of ties that could affect the result, but it did help adjust for the non-normal distribution. GPA was calculated as a result of several variables for both before and after the change. Independent variables included TOEFL overall and subsection scores. Table 2 summarizes the tests used for the analysis of the TOEFL GPA relationship. There are limitations with using the Spearman's $r_s$ as opposed to using Pearson's $r$. Pearson's $r$ would be affected by the non-normal distribution of GPA, and Spearmen's $r_s$ is affected by a large number of ranked ties in the data.

**Table 2—Analyses Run**

<table>
<thead>
<tr>
<th>Test</th>
<th>X Variable</th>
<th>Y Variable</th>
<th>Test Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Year</td>
<td>Overall TOEFL</td>
<td>One Way ANOVA</td>
</tr>
<tr>
<td>2</td>
<td>Year</td>
<td>GPA</td>
<td>One Way ANOVA</td>
</tr>
<tr>
<td>3</td>
<td>Overall TOEFL</td>
<td>GPA</td>
<td>Correlation</td>
</tr>
<tr>
<td>4</td>
<td>TOEFL Reading (Before, After)</td>
<td>GPA</td>
<td>Correlation</td>
</tr>
<tr>
<td>5</td>
<td>TOEFL Listening (Before, After)</td>
<td>GPA</td>
<td>Correlation</td>
</tr>
<tr>
<td>6</td>
<td>TOEFL Writing (Before, After)</td>
<td>GPA</td>
<td>Correlation</td>
</tr>
<tr>
<td>7</td>
<td>TOEFL Speaking (Before, After)</td>
<td>GPA</td>
<td>Correlation</td>
</tr>
<tr>
<td>8</td>
<td>Year, TOEFL (Overall and Subsections)</td>
<td>GPA</td>
<td>Multiple-Regression</td>
</tr>
</tbody>
</table>

Moving beyond these comparisons, these simple linear models were followed by a multiple regression model to see how the variables acted in each other's presence.

Specifically, researchers explored the extent to which the variables of the TOEFL overall score,
and the subscores for reading, writing, listening, and speaking, in the presence of each other, have on GPA.

Results

*Research Question 1—Linguistic Diversity*

How have the new TOEFL requirements influenced the composition and linguistic diversity of the student population? For international students, there was an average yearly decrease of 29% in admitted students. The majority of native languages experienced decreases as well (see Figure 3).

*Figure 3—Population of Admitted Students Before and After Change*

For example, native Spanish speakers had an average yearly admission of 113.8 students before the change and an average admission of 73.4 after the change. This change also showed large decreases for more than half of the languages. Nepalese speakers had almost a 100% decrease, but a meeting with the admissions office revealed an expired exchanged program for these students. However, not all native languages saw decreases in admissions to the university.
German speaking students maintained the same acceptance. Some of the languages in the others' category saw small increases: Tagalog speakers had an average of 2 students admitted yearly before the change, and 2.6 students after. Also, international students who were native English speakers did not experience large changes in admission numbers as well.

Additional information was needed to determine how the linguistic composition at the university changed. When the population totals for each of these most common languages were sorted into percentages for each of the five year spans (see Table 3), the result indicated that almost a third of the international student composition were native English speakers, which included international students from Australia, United Kingdom, and Canada. Also there were declines in percentage admitted for international students whose native language were not English, Portuguese, and Chinese.

**Table 3—Linguistic Composition: Percentage of International Students**

<table>
<thead>
<tr>
<th>Native Language</th>
<th>Before</th>
<th>After</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>18%</td>
<td>31%</td>
<td>13%</td>
</tr>
<tr>
<td>Chinese</td>
<td>8%</td>
<td>12%</td>
<td>4%</td>
</tr>
<tr>
<td>German</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Korean</td>
<td>11%</td>
<td>12%</td>
<td>1%</td>
</tr>
<tr>
<td>Portuguese</td>
<td>6%</td>
<td>7%</td>
<td>1%</td>
</tr>
<tr>
<td>French</td>
<td>2%</td>
<td>1%</td>
<td>-1%</td>
</tr>
<tr>
<td>Japanese</td>
<td>3%</td>
<td>2%</td>
<td>-1%</td>
</tr>
<tr>
<td>Russian</td>
<td>2%</td>
<td>1%</td>
<td>-1%</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>3%</td>
<td>2%</td>
<td>-1%</td>
</tr>
<tr>
<td>Spanish</td>
<td>23%</td>
<td>21%</td>
<td>-2%</td>
</tr>
<tr>
<td>Other</td>
<td>16%</td>
<td>10%</td>
<td>-6%</td>
</tr>
<tr>
<td>Nepali</td>
<td>7%</td>
<td>0%</td>
<td>-7%</td>
</tr>
</tbody>
</table>

Also, before 2010, the highest population of international student marked Spanish as their native language, but after 2010 the highest population of international students listed English as their native language. This means that there was a shift in the population being admitted to the University. A chi-squared test reveal that these compositional changes were statistically
significant, $X^2 (11) = 270.24, p < 0.0001$, indicating that the difference between the years before and after the change exhibited more differences than could be explained by chance.

Research Question 2—TOEFL GPA Relationship

To what extent has the change in TOEFL requirements influenced GPA of first year international students? Is there a net gain in GPA scores? To answer this research question the response was divided into sections by the type of test: Average TOEFL scores by year, Average GPA by year, correlations, and multiple regression.

Average TOEFL scores by year. By running a one-way ANOVA of year on a TOEFL scores, researchers verified that a change in TOEFL scores occurred. Additionally, it was noted that the actual change appeared later than 2010, and the admissions office verified that transfer students were not required to meet the new cut score until 2013. Additionally, initial examination of TOEFL scores revealed minimum accepted scores below the cut score requirement every year from 2011-2015 (see Table 4).

<table>
<thead>
<tr>
<th>Year</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>71</td>
<td>91.20</td>
<td>13.88</td>
<td>66</td>
<td>116</td>
</tr>
<tr>
<td>2010</td>
<td>94</td>
<td>88.96</td>
<td>11.28</td>
<td>67</td>
<td>116</td>
</tr>
<tr>
<td>2011</td>
<td>110</td>
<td>89.79</td>
<td>12.86</td>
<td>40</td>
<td>115</td>
</tr>
<tr>
<td>2012</td>
<td>104</td>
<td>89.72</td>
<td>13.38</td>
<td>46</td>
<td>116</td>
</tr>
<tr>
<td>2013</td>
<td>166</td>
<td>95.87</td>
<td>10.52</td>
<td>57</td>
<td>119</td>
</tr>
<tr>
<td>2014</td>
<td>182</td>
<td>96.17</td>
<td>8.66</td>
<td>77</td>
<td>117</td>
</tr>
<tr>
<td>2015</td>
<td>247</td>
<td>96.30</td>
<td>10.21</td>
<td>63</td>
<td>117</td>
</tr>
<tr>
<td>Total</td>
<td>974</td>
<td>93.68</td>
<td>11.50</td>
<td>40</td>
<td>119</td>
</tr>
</tbody>
</table>

Furthermore, the distribution was checked to see whether scores below 80 were random exceptions or if they were more common. Figure 4 shows that there were a number of outlier scores. Though these outliers may not meet the assumptions of normality, Figure 4 also appears to have a normal distribution. Additionally, an ANOVA is robust and can generally handle non-
normal data. Also, outliers that were relevant to the data stayed, while two mistakes were removed through consulting the international admissions office.
Despite looking somewhat unvaried in Figure 4, the examination of the trend line of the means does appear to reflect a positive pattern, and this same pattern is shown in Figure 5, with a large jump from 2012 to 2013. A visual representation of the ninety-five percent confidence intervals, Figure 6, shows an increase in the average of the TOEFL scores as well, but instead of seeing a jump in TOEFL scores after 2010, we again see that the averages of admitted scores appear to jump in 2013. The admissions office supported this finding by explaining that transfer students were not required to meet the minimum cut-score.
Figure 5—Mean TOEFL Scores by Year, Trend Line

Figure 6—Confidence Intervals (95%) of TOEFL Score Means by Year
The results of a one-way ANOVA test revealed that there was a statistically significant difference of year on TOEFL scores admitted to the university, \( F(6, 336.85) = 11.86, p < 0.001 \), which suggests that there is a difference between the average of TOEFL scores admitted to the university. Following the results of the one-way ANOVA, more information was collected on which years appeared to have the greatest and most significant difference in means. A Tukey's pairwise test was conducted to show differences between all year combinations, and the results show the most concentrated differences in the years 2009 and 2013. However, other years did not reveal these strong differences.

**Average GPA by Year.** After checking for normality, equal variance, and independent samples, the Welch's ANOVA was selected. With this particular data, the boxplots, Figure 7, showed the skewness of GPA data with larger tails reaching into the lower GPA. This result is expected for an achievement score, and the ANOVA test is robust again in this distribution. Additionally, the boxplots in Figure 7 do not appear to be moving much.

![Figure 7—Mean GPA Scores by Year, Box Plot](image-url)
deal of overlap. This pattern, or lack of consistent pattern is confirmed in Table 5.

*Figure 8—Mean GPA Scores by Year, Trend Line*

*Figure 9—Confidence Intervals (95%) of Mean GPA by Year*
Table 5—Descriptive Statistics First-Year GPA

<table>
<thead>
<tr>
<th>Year</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>71</td>
<td>3.38</td>
<td>0.54</td>
</tr>
<tr>
<td>2010</td>
<td>94</td>
<td>3.23</td>
<td>0.69</td>
</tr>
<tr>
<td>2011</td>
<td>105</td>
<td>3.25</td>
<td>0.67</td>
</tr>
<tr>
<td>2012</td>
<td>102</td>
<td>3.21</td>
<td>0.64</td>
</tr>
<tr>
<td>2013</td>
<td>154</td>
<td>3.10</td>
<td>0.80</td>
</tr>
<tr>
<td>2014</td>
<td>157</td>
<td>3.31</td>
<td>0.60</td>
</tr>
<tr>
<td>2015</td>
<td>119</td>
<td>3.30</td>
<td>0.65</td>
</tr>
<tr>
<td>Total</td>
<td>802</td>
<td>3.24</td>
<td>0.67</td>
</tr>
</tbody>
</table>

For GPA, the ANOVA test did not reveal a significant difference of year on GPA \( F(6, 325.41) = 2.02, p = .063 \), and in examining the box plot, Figure 7, and the plotted confidence interval, Figure 9, there is not a clear trend. The only years which appear to have any difference in mean were 2009 and 2013 with 95% CIs [3.25, 3.51], [2.97, 3.223], respectively.

Additional information on which particular years appeared to have the most significant differences allowed for the use of Tukey's method. Significant differences were only found when other years were compared to 2013. The reason for the peculiarities founded in 2013 is unknown, but it may be a result of the large number of outliers. No indication was given that 2013 was not representative of the year or that there were any abnormalities about the outliers. Therefore, the results actually show 2013 as being significantly lower than other years. Based on these unexpected results, the relationship between the TOEFL and GPA is unclear. There does not appear to be any net change in GPA after the cut score change. However, running a correlation analysis provides further information on this relationship.

Correlations between TOEFL and GPA. When running a linear regression of GPA based on the TOEFL (overall and subsection) scores, the correlation coefficients are affected by non-normal distributions, unequal variance, independent samples and range restriction. To adjust for the non-normal distribution, the Spearman method was used. In the graphs of this relationship,
Figures 10-14 (see Appendix A), the data appears to be congregated in the top right corner. This distribution may be indicative of the restricted range in that not many students with low proficiencies are admitted to the university. Additionally, we see many students who do not appear to fit the expected model. For example, there were students who received lower TOEFL scores, but achieved higher GPAs, and there were even more students who received higher TOEFL scores, but instead received lower GPAs.

However, despite the unique distribution and violations of several assumptions, the resulting change in correlations before and after the cut score change is summarized in Table 6, and patterns were found. For GPA by TOEFL overall scores, the strength in correlation coefficient ($r_s$) decreased. This same pattern was seen for all other independent variables except for the TOEFL speaking section, which did not have a significant p-value both before and after the cut-score change. Ultimately, this pattern is an indication of range restriction, but patterns found in varying levels of subsection score may suggest relative importance for first year skills.

Table 6—Correlation Summaries

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_s$</td>
<td>$p$</td>
</tr>
<tr>
<td>Reading</td>
<td>.36</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Writing</td>
<td>.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Listening</td>
<td>.23</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Speaking</td>
<td>.01</td>
<td>N.S.</td>
</tr>
<tr>
<td>Overall</td>
<td>.33</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

At the same time, Table 6 reveals another interesting trend in that there appears to be a different strength relevant to subsection skill areas. In other words, some subsections of the TOEFL appear to be of more value in predicting first year GPA than others, with the order—
strongest to weakest—being reading, writing, listening, and then speaking. In contemplating requirements for first year students typically taking general education courses, this is interesting, as many first year courses at this particular university require a great deal of reading. Additionally, in large lecture hall format, common to first year courses at this university, speaking is not as needed for these first year students. These results may have implications for administration and college departments looking to fine-tune the TOEFL requirements based on skills needed in a particular program.

*Multiple Regression Model.* The data was then run to see how GPA was affected based on the TOEFL overall, reading, writing, listening, speaking and year. The regression equation was found to be significant $F(10, 791) = 6.124, p < .0001$. An individual’s predicted GPA is $0.02(\text{Reading}) + <.01(\text{Listening}) - 0.02(\text{Speaking}) + 0.03(\text{Writing})$ where year is recorded as a dummy variable. The results, Table 7, indicated statistical significance for reading and writing, as expected, but the significance for the listening and speaking sections were switched when compared to the correlation analysis. Also, the only year with statistical significance is 2013. Thus, the overall change in GPA is not seen, but the evident necessity of various skill areas is still present.

*Table 7—Multiple Regression Model*

<table>
<thead>
<tr>
<th>Coefficients:</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.63</td>
<td>0.22</td>
<td>12.13</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TOEFLReading</td>
<td>0.02</td>
<td>0.01</td>
<td>3.84</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TOEFLListening</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>0.67</td>
<td>.50</td>
</tr>
<tr>
<td>TOEFLSpeaking</td>
<td>-0.02</td>
<td>0.01</td>
<td>-2.67</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>TOEFLWriting</td>
<td>0.03</td>
<td>0.01</td>
<td>2.75</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Year 2010</td>
<td>-0.12</td>
<td>0.10</td>
<td>-1.17</td>
<td>.24</td>
</tr>
<tr>
<td>Year 2011</td>
<td>-0.09</td>
<td>0.10</td>
<td>-0.94</td>
<td>.35</td>
</tr>
<tr>
<td>Year 2012</td>
<td>-0.14</td>
<td>0.10</td>
<td>-1.37</td>
<td>.17</td>
</tr>
<tr>
<td>Year 2013</td>
<td>-0.29</td>
<td>0.10</td>
<td>-3.04</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Year 2014</td>
<td>-0.08</td>
<td>0.10</td>
<td>-0.79</td>
<td>.43</td>
</tr>
<tr>
<td>Year 2015</td>
<td>-0.09</td>
<td>0.10</td>
<td>-0.86</td>
<td>.39</td>
</tr>
</tbody>
</table>
Limitations

Flaws with this study include the missing data, the required use of an additional data set, the range restriction after the cut-score, the limited information on the classes taken by students, and a non-randomly selected population. Adjustments for these limitations could be made in future research. Additionally, what this study does not reveal is the GPAs of students who were not admitted. The problem with the lack of lower level students may contribute to the lower correlations, and lower level English speakers may not be applying to the university. Ultimately, the universities can still use the results from analyses such as these to make informed decisions. A recommended course of action is more research on the students that are accepted into the university. What appears to contribute to the greatest success for international students? And more specifically, how can administrators measure this ability?

It is important to note that there were several assumptions that were violated. The non-normality of the data reduced some of the correlations, and the range-restrictions limited the ability to draw some conclusions. However, despite these limitations, there were still several meaningful pieces of information for each of the research questions that can be used to inform cut-score decisions and can validate claims for future research of sub-section score analysis.

Discussion

Though other universities may not exhibit the exact same difficulties presented in this case study, this study provides a resource for universities working to investigate their own admissions decisions and use of TOEFL scores. For example, while the university specific to this study chose to set the cut score at 80, other universities may find that this cut-score does not meet the needs of their university. However, as pointed out in the result section, the differences in correlations for the subsection skills may have implications for university decisions. When
strengthened by Ginther and Yan's (2016) study, these findings provide encouragement for examining cut scores at both at a university and department level. In other words, the different levels of importance for TOEFL subsection skills may vary depending on the department.

Within the methods of this case study, two important features are exposed: the increased collaborations between the linguistic and admission offices at the university, which led to better information on cut-score decisions, and the process for analyzing the data, which can be used by future researchers. The crossover between these departments provided the admissions office with the historical context of previous decisions and with empirical evidence. For the Linguistic department, this study provided more insight into the meaning of TOEFL scores. Additionally, this method provided more than anecdotal evidence about student success and a solid rationale for the cut-score decisions chosen by the university. Beyond this, some of the challenges with using historical data were exposed.

**Linguistic Diversity**

Admission decisions can have a wide range of influence on the student body. As the demand for an international education increases or decreases, changes in the cut-score can potentially act a filter for the number of students admitted. As seen with this study, the new requirements did not alter the population of international students applying, rather it shifted the composition of the international student population. The differences represented in Table 3 reflect both a change in the composition of the student body and a potential decrease in linguistic diversity of the university examined. Larger decreases in the non-native English speaking populations, however, did not greatly impact the average GPA. Administrators should be aware of the changes to the student composition as they make changes to policy.

Other adjustments were made as a result of these changes. Many ESL service courses at
the University were no longer offered due to insufficient student enrollment. The only service course that continued to be carried was the pronunciation class. Universities seeking to raise the bar should consider the impact the change may have on the courses offered at the university.

On the other hand, university and colleges with large English as a Second Language (ESL) and Intensive English Programs (IEP) could consider using lower TOEFL scores as an opportunity for expansion. These scores would provide an opportunity to expand resources and classes offered at the university.

**TOEFL and GPA relationship**

This study validates the claims made by previous research that subsection scores provide nuanced meanings for student success. Data revealed a hierarchy for the subsection skills with reading being the strongest indicator of potential first year success. This could be a reflection of typical first year classes as general education classes do require larger amounts of reading.

Information regarding different majors could be used not only to benefit a program's admission, but also to help students better prepare for their intended program of study. It would provide an excellent supplement to admissions decisions. Additionally, more information could be gathered to see whether other subsections are better indicators for second, third, or fourth year students at the university.

At the same time, the values of the ANOVA analysis did not suggest much of a practical difference in the mean GPA scores: basically the average international student stayed in the B to B+ range. Possibilities for this finding may indicate that high cut-scores could lead to less variance accounted for by TOEFL. Additionally, it supports the belief that academic success and GPA are effected by multiple variables, and TOEFL is merely one of those variables. With the minimum benefit provided from GPA, a solution might be to investigate whether students who
obtain high TOEFL scores complete their degrees more quickly. There is a possibility that students with lower abilities may take longer to navigate through university requirements and classes.

As can be seen with the large amounts of variation, one way a university can soften a cut-score would be to use the standard error measurement (SEM) as a window of reasonable doubt, which would be helpful for students who potentially are at a higher ability level than indicated by their raw score. With more than twenty-five percent of the top 100 schools using cut-off scores of 80 or higher, it is possible that potentially successful international students cannot be accepted into desired universities.

Conclusion

With the information that the TOEFL provides, valuable insight into the university cut-score decisions can assist admission offices in the pursuit for academically prepared students who contribute to their university. Additionally, the TOEFL can potentially be fine-tuned and adapted to meet the specific needs of departments and colleges. The results, though unique to the university from which data were derived, can be replicated by other universities to inform decision makers of the impact of TOEFL cut-scores on student success in other contexts.
References


Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT scores to academic performance: Some


Appendix A

Figure 10—Effect of TOEFL Overall Score on GPA Before and After Change

Figure 11—Effect of TOEFL Reading Score on GPA Before and After Change
Figure 12—Effect of TOEFL Writing Score on GPA Before and After Change

Figure 13—Effect of TOEFL Listening Score on GPA Before and After Change
Figure 14—Effect of TOEFL Speaking Score on GPA Before and After Change