EXAMINATION OF THE APPROPRIATENESS OF USING STANDARDIZED TEST SCORES FOR ENGLISH AS A SECOND LANGUAGE (ESL) PLACEMENT

BY
KATERYNA KOKHAN

DISSERTATION
Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics in the Graduate College of the University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:
Professor Fred Davidson, Chair
Associate Professor Peter Mortensen
Associate Professor Peter Golato, Texas State University
Associate Professor Paula Winke, Michigan State University
Abstract

The primary purpose of this multi-paper dissertation was to examine the appropriateness of using standardized test scores for making ESL placement decisions within the context of the University of Illinois at Urbana-Champaign (UIUC). The dissertation consists of an introductory chapter, four manuscript papers, and a concluding chapter. In the first paper, a survey of the current English proficiency requirements in the U.S. universities with very high research activity (Research Universities / Very High [research activity] -- RU/VH) was performed in order to get a general idea about standardized test requirements that international applicants must satisfy in order to be admitted to the university. The results of the survey showed that the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS) are most widely accepted tests. However, the RU/VHs significantly differed in terms of score requirements for admission. A substantial difference was also observed between test score requirements for admission to undergraduate and graduate programs. The conditions for waiver from the English proficiency requirement differed greatly from university to university. In this author’s view, “marketization” of higher education might be one of the strongest reasons triggering variety and complexity of the English proficiency policies at the RU/VHs.

The results of the subsequent three independent research projects (Kokhan, 2012; Kokhan, 2013; Kokhan & Lin, submitted) that were included into this dissertation provided solid research evidence to conclude that standardized language proficiency tests would not be suitable for ESL placement. The results from the second paper (Kokhan, 2012) suggested that there was no particular set of either total or section TOEFL iBT scores which can be used as a reliable criterion for placing students into ESL classes without significant misplacement. This finding was further corroborated in the third paper (Kokhan, 2013) that examined the appropriateness of using standardized tests for placement of undergraduate international students into the ESL writing courses. A distinct pattern of fluctuations of correlation between the TOEFL iBT and the English Placement Test (EPT) over time was also observed (Kokhan, 2012). The correlation was stronger when the time gap between these tests was short but it dramatically
decreased and even became slightly negative around Week 50; however, starting from the 50th week it increased. Language attrition was highlighted in the paper as the main factor contributing to this pattern of correlation.

Kokhan’s (2013) study showed that using language sections of such tests as the SAT and ACT would not be effective for ESL placement of undergraduate students either. Exploratory data analysis demonstrated that only the lowest SAT reading and ACT English scores, and the highest TOEFL iBT total and writing section scores could separate the students between the two ESL placement levels. However, the number of undergraduate ESL students, who scored at the lowest and highest ends of each of these test scales, was less than 5% over the last six years. Thus, setting cutoff scores for such a small fraction of the ESL population may be impractical. For the majority of the undergraduate ESL population, there was about a 40% chance that they would be misplaced if the ESL placement decision were made solely based on the standardized test scores.

Finally, the main objective of the fourth paper (Kokhan and Lin, submitted) was to examine which TOEFL iBT scores (official highest, most recent, average, or self-reported scores) were the best predictors of ESL placement. The findings showed that the self-reported and the highest TOEFL iBT total scores had the strongest association with the ESL placement results. The self-reported and the highest scores also demonstrated the highest classification efficiency in predicting ESL placement of TOEFL iBT repeaters. However, an additional analysis of the ESL placement trends in relation to the total highest, average, most recent and self-reported TOEFL iBT scores indicated that none of the four types of scores were suitable for making accurate ESL placement decisions since, in some cases, students with very low TOEFL iBT scores could be placed into the highest ESL level and students with very high TOEFL iBT scores could be placed into the lowest ESL level.

The key findings, methods of analysis, limitations and suggestions for future research and changes of practice were discussed in the concluding chapter of the dissertation.
To Nicholas
Acknowledgements

I am very grateful to my research advisor, Fred Davidson, for his guidance and support which made this work possible. I greatly appreciate the research opportunities that he gave me. I would also like to thank the rest of the committee – Prof. Peter Golato, Prof. Paula Winke, and Prof. Mortensen – whose thoughtful contributions helped strengthen my dissertation. Special thanks to Mary Ellen Fryer for helping me with paperwork and making my life in graduate school so much easier. Finally, I am truly indebted and thankful to my mother and my husband for their love, support, and encouragement throughout the years.
# Table of Contents

CHAPTER 1: INTRODUCTION

Organization of This Dissertation ................................................................. 1
Significance of This Research ........................................................................... 1
References .......................................................................................................... 2

CHAPTER 2: CURRENT TRENDS IN THE ENGLISH PROFICIENCY REQUIREMENTS FOR
ADMISSION OF INTERNATIONAL STUDENTS TO THE U.S. UNIVERSITIES WITH VERY
HIGH RESEARCH ACTIVITY (Paper One) ............................................................ 4

Abstract ............................................................................................................. 4

Introduction ......................................................................................................... 4

U.S. Universities with Very High Research Activity ............................................. 6

How can the English Proficiency Requirement Be Satisfied at RU/VHs? ............. 7

English Proficiency Tests Accepted by RU/VHs ................................................ 10

Important Terminology ..................................................................................... 15

Subsection Scores Also Matter ......................................................................... 20

Conclusion ......................................................................................................... 22

References .......................................................................................................... 25

CHAPTER 3: INVESTIGATING THE POSSIBILITY OF USING TOEFL SCORES FOR
UNIVERSITY ESL DECISION MAKING: PLACEMENT TRENDS AND EFFECT OF TIME
LAG (Paper Two) .................................................................................................. 44

Abstract ............................................................................................................. 44

Background ........................................................................................................ 45

The Present Study ............................................................................................... 51

Methodology ...................................................................................................... 52

Results ................................................................................................................ 57

Discussion .......................................................................................................... 64

Conclusion ......................................................................................................... 65

References .......................................................................................................... 67

CHAPTER 4: AN ARGUMENT AGAINST USING STANDARDIZED TEST SCORES FOR
PLACEMENT OF INTERNATIONAL UNDERGRADUATE STUDENTS IN ENGLISH
AS A SECOND LANGUAGE (ESL) COURSES (Paper Three) .............................. 70

Abstract ............................................................................................................. 70

Introduction ........................................................................................................ 71
Use of Pre-arrival Test Scores for ESL Placement ............................................................. 72
The Present Study ........................................................................................................... 76
Methodology .................................................................................................................. 79
Results ............................................................................................................................ 85
Discussion ....................................................................................................................... 92
Conclusion ....................................................................................................................... 94
References ....................................................................................................................... 97

CHAPTER 5: TOEFL IBT REPEATERS: HOW TO INTERPRET MULTIPLE SCORE REPORTS
FOR ESL PLACEMENT? (Paper Four) ................................................................................. 102
  Abstract .......................................................................................................................... 102
  Background .................................................................................................................... 102
  Current Trends in the ESL Placement Policy with Respect to TOEFL iBT Test Repeaters .... 106
  The Present Study ......................................................................................................... 106
  Methodology ................................................................................................................ 107
  Results ........................................................................................................................... 110
  Discussion ..................................................................................................................... 119
  References .................................................................................................................... 123

CHAPTER 6: DISCUSSION AND CONCLUSIONS .................................................................. 128
  Discussion of the Results .............................................................................................. 128
  Reflections about Methods of Analysis ......................................................................... 131
  Limitations and Suggestions for Future Research ....................................................... 132
  Suggestions for Changes of ESL Placement Practice ................................................ 133
  Afterword: Changes in the ESL Placement Practice after Fall of 2011 ......................... 134
  References .................................................................................................................... 135

Appendix A: Language proficiency tests accepted by the U.S. universities ..................... 136
Appendix B: Distribution of EPT test-takers by semester in chronological order .............. 139
Appendix C: Overview of the ESL placement policies and TOEFL score requirements .......... 140
Appendix D: Distribution of TOEFL iBT repeaters and non-repeaters according to the native country and the EPT placement levels ....................................................................................................................... 142
Appendix E: Cross classification of the number of TOEFL iBT repeated by levels of English proficiency and ESL placement levels ....................................................................................................................... 144
Chapter 1

Introduction

Organization of This Dissertation

Despite the widespread practice of using standardized language proficiency tests for making admission and ESL placement decisions, the use of such tests for placement purposes is significantly under-researched. The primary objective of this dissertation is to explore the appropriateness of using standardized tests for English as a Second Language (ESL) placement within the context of the University of Illinois at Urbana-Champaign (UIUC). The dissertation consists of an introductory chapter discussing the organization of the dissertation and significance of this research, four separate manuscripts each exploring different aspects of this topic, and a concluding chapter summarizing the main findings and discussing the methods of analysis used in this research as well as limitations and suggestions for future research and changes of practice. Chapter 2 (Paper One) is a lead-in to the main theme of the dissertation. The purpose of this study is to get a general picture about the current trends in the English proficiency requirements for admission of international students to the U.S. universities with Very High Research Activity (Research Universities / Very High [research activity] -- RU/VH). This chapter also offers some critical analysis of the most pronounced trends. Chapter 3 (Paper Two) is a published manuscript examining the changes in correlation between the written part of the UIUC English Placement Test (EPT) and Test of English as a Foreign Language - Internet-based Test (TOEFL iBT) results. In this chapter, I discuss the fluctuation of correlation between these two tests over time and provide some possible interpretations of the observed picture. Chapter 4 (Paper Three) is also a published manuscript and is a follow-up to the research project discussed in Chapter 3. In Chapter 4, I examine the appropriateness of using SAT, ACT and TOEFL iBT scores for ESL placement. Chapter 5 (Paper Four) is a research project performed in collaboration with Chih-Kai (Cary) Lin, a PhD candidate in Educational Psychology at UIUC. The purpose of this study was to examine which scores of TOEFL iBT repeaters (the highest, most
recent, average or self-reported scores) would be the best pre-arrival measure of English language proficiency.

In the concluding chapter, I discuss the main findings of this research and provide some interpretations of the results given the report of other chapters. I also reflect on the methods of analysis used for this research and discuss some limitations and suggestion for future research.

In the final section of this chapter, I offer some suggestions for changes of ESL placement practice at UIUC and generally.

In this dissertation, I am focusing only on the written EPT since the oral EPT assesses ESL candidates’ pronunciation. The latter is outside the scope of this research and my main area of expertise. Additionally, ESL writing is considered a major factor in university language use. For undergraduate students at UIUC, in particular, completion of ESL writing courses fulfills their Composition I requirement.

**Significance of This Research**

The key findings discussed in the dissertation are based on the analysis of a large data set collected over a multi-year period. The data set for Paper One was compiled during Fall 2012 through Spring 2013. The data analyzed in Paper Two were collected over the period of Summer 2006 through Spring 2010. The data set analyzed in Papers Three and Four contained the score information collected over the period of Fall 2006 through Fall 2011.

Increased enrollment of international students to the U.S. universities requires many U.S. universities to establish specific English proficiency polices. To my knowledge, there is no published research examining the complexity and diversity of the English proficiency requirements that are currently being observed in many universities. Paper One in this dissertation provides a comprehensive overview of the English proficiency requirements in all 108 Universities with very high research Activity.

The main contribution of Paper Two (Kokhan, 2012) is a discovery of a time lag effect in the relationship of the TOEFL and the EPT. The graph depicting this effect provided very important evidence
of two things: first, the correlation between the TOEFL and the EPT is never strong, which confirms the need to conduct an institutional English placement test; second, it discounts any suggestions of various people from Campus to simply use the TOEFL as a placement measure.

In Paper Three (Kokhan, 2013), I successfully used graphical illustration for examination of the effectiveness of using combination of test scores for ESL placement. This approach allowed me to obtain the percentages of placement accuracy in relation to various combinations of the TOEFL iBT and ACT/SAT scores.

Graphical approaches to data analysis which were used in Papers Two and Three have never been described in the language testing literature before.

A significant number of test takers choose to repeat the TOEFL iBT test hoping to get higher results. Unfortunately, there is no published research examining students’ multiple TOEFL iBT score reports. Paper Four addresses important issues of interpretation and use of multiple scores of TOEFL iBT repeaters for ESL placement in the context of the University of Illinois at Urbana-Champaign (UIUC). The significance of this study is also in the fact that it is based on a comprehensive campus-wide data set collected over a multi-year period and provides a detailed methodological approach to analyzing the relationship between ordinal ESL placement levels and continuous standardized language test scores, from which other ESL programs could potentially benefit.

References


Chapter 2

Current trends in the English proficiency requirements for admission of international students to the U.S. universities with very high research activity

(Paper One)

Abstract

The number of international students studying at U.S. colleges and universities has increased enormously over the last decade (Institute of International Education [IIE], 2013). This paper examines current trends in the English proficiency requirements for admission to the U.S. universities with very high research activity (Research Universities (very high research activity) -- RU/VH) and informs educators, policy makers, researchers as well as prospective applicants about the variety of the English proficiency policies in the higher education in the United States. The data analyzed in this paper is publicly available information posted on the official websites of the RU/VHs. The results suggest that the RU/VHs differ greatly in terms of standardized language proficiency score requirements for admission of international students. There is also an observed difference between test score requirements for admission to undergraduate and graduate programs. Significant variety is also seen regarding waiver from the English proficiency requirement. In this author’s view, “marketization” of higher education is one of the strongest reasons triggering complexity of the English proficiency policies at the RU/VHs.

Keywords: English proficiency requirement, universities with very high research activity, university admissions

Introduction

Admission policies to American universities are becoming more complex with the number of international applicants increasing each year. During the 2012/13 academic year, the international student enrollment in the USA increased by seven per cent, to 819,644 individuals (IIE, 2013). Standardized language proficiency testing is the easiest way to assess international applicants’ level of English before
their arrival on campus. That is why standardized language proficiency tests gained so much importance and power these days. Many English proficiency tests are advertised nowadays. Some of the tests have a very long validity history (e.g., Test of English as a Foreign Language (TOEFL) first administered in 1964); other tests are new but quite successful (e.g., Pearson Test of English Academic (PTE Academic) first launched globally in 2009 (Ying Zheng, 2011).

If a university accepts scores for more than two language proficiency tests, deciding what language proficiency test to take may sometimes turn into a multiple-choice dilemma for some international students. Moreover, as every rule has an exception, there may also be exceptions to the English proficiency requirement; and for some students, this requirement may be waived entirely. Complexity of admission policies triggers complexity of English as a Second Language (ESL) support to international students, too. International applicants can be granted full or conditional admission with a required ESL classes. In some universities, they can be asked to complete the ESL requirement prior to starting the main course work and in others – simultaneously with the main course work.

Increased competition in the U.S. higher education market (Dill, 2003) encourages colleges and universities to significantly improve their websites in order to provide more information to their students and clients (Klassen, 2002). Prospective applicants can get all the information they need about the admission requirements and apply for entrance online. For this study, I made a comprehensive overview of the published English proficiency requirements for admission of international students to the U. S. universities with very high research activity (RU/VH). The data analyzed in the paper were collected over the period of Fall 2012 through Spring 2013. The findings highlighted in this paper may potentially be interesting for admission policy makers, educators, researchers and international applicants willing to study in the USA.
U.S. Universities with Very High Research Activity

When examining certain aspects of admission and assessment of international students, some researchers focus on Big Ten Universities (e.g., Crusan, 2002; Sato & Hodge, 2009). There are also some studies focusing on the universities that are participating in other major athletic conferences: Pacific Ten, Big Eight, Ivy League or Atlantic Coast Conferences (e.g., McCormick & Tinsley, 1987; Ory & Parker, 1989). Undoubtedly, athletic success helps universities to advertise themselves and to increase the number of applicants (McCormick & Tinsley, 1987; Murphy & Trandel, 1994), which increases tuition revenue, which can be spent on funding scholarships, important research projects, faculty hires, etc. (McEvoy, 2005). However, the formation of these conferences has nothing to do with scholarly or academic success of the universities. There are many smaller universities which are not on the list of the major athletic conferences but are quite successful academically and scientifically. Another option could be to focus on the universities that are members of the Association of American Universities (AAU). One potential downside of choosing AAU for this study could be that AAU is an organization requiring membership rather than an objectively established taxonomy of the universities with a strong system of academic research and education.

Because of the reasons mentioned above, I decided to focus on the universities with very high research activity included in the Carnegie Classification of Institutions of Higher Education (http://classifications.carnegiefoundation.org/). The Carnegie Classification of Institutions of Higher Education is the leading framework for classifying colleges and universities in the United States of America. It was developed by the Carnegie Commission on Higher Education in 1970. This classification allows researchers and educators to make comparisons among U.S. degree-granting colleges and universities. It consists of six classifications: Basic Classification (the traditional Carnegie Classification Framework, it includes the classification of RU/VHs), Undergraduate and Graduate Instructional Program classifications, Enrollment Profile and Undergraduate Profile classifications, and Size and Setting classification. The classification is regularly updated based on the national data from the U.S. Office of
Postsecondary Education, the National Center for Education Statistics’ Integrated Postsecondary Education Data System (IPEDS), the National Science Foundation, and the College Board.

How Can the English Proficiency Requirement Be Satisfied at RU/VHs?

I collected publicly available information about the English proficiency requirements for admission to 108 RU/VHs. The findings of the survey indicate that all the universities from this list do have an English proficiency policy. However, English proficiency requirement for admission may be satisfied in many ways. The three most common ones throughout almost all the RU/VHs are: 1) prior academic experience in the USA, 2) satisfactory results for some standardized English proficiency tests or 3) successful completion of an institutional English language program prior to applying to the university.

The definition of ‘previous academic experience’ that can satisfy the English proficiency requirement differs from university to university. The two most commonly mentioned definitions of ‘previous academic experience’ are:

1. A bachelor’s degree from a regionally accredited U.S. university or college (e.g., Arizona State University, Columbia University in the City of New York, Louisiana State University and Agricultural & Mechanical College, Stanford University, Texas A&M University, the University of Arkansas, the University of Washington – Seattle Campus, the University of Illinois at Urbana-Champaign, Wayne State University, and several other universities);

2. A U.S. high school diploma, usually in combination with satisfactory SAT Verbal and ACT English scores (e.g., Arizona State University, Rensselaer Polytechnic Institute, the University of Washington – Seattle Campus, the University of Arkansas, the University of Nebraska – Lincoln, University of Chicago, the University of Hawaii at Manoa, the University of Oklahoma Norman Campus, the University of Pittsburg - Pittsburg Campus, Brandeis University).

A degree of detail in the descriptions of the English proficiency requirements differ immensely. For example, Arizona State University established the policy according to which applicants who had
attended a regionally accredited college/university in the United States and met one of the following conditions could be exempt from the English proficiency requirement: 1) They earned a bachelor’s degree or higher; 2) They completed at least 12 semester hours of graduate coursework with a cumulative GPA of 3.0 on a 4.0 scale or higher; 3) They completed at least 90 semester hours of undergraduate coursework with a cumulative GPA of 3.0 on a 4.0 scale or higher (all 90 hours must be earned in the United States). By contrast, the California Institute of Technology has very concise exemption criteria for undergraduate and graduate applicants: 1) Applicants who have studied in the US for two or more years; 2) Applicants with a degree from a school whose primary instruction is in English. There are quite a few universities that do not set any specific conditions for exemption from the English proficiency requirement. In a sentence or two, they only state the requirement to take certain English proficiency tests. One of the universities with the least detailed information about the English proficiency requirement is Rockefeller University: there are only two short statements on the official website of the university, “Applicants whose native language is not English must submit evidence of their proficiency in the English language. An official TOEFL score is required” (https://graduateapplication.rockefeller.edu/app/faq/#3_5). According to the Admissions Website of Boston University, “If your first language or the primary language spoken in your home is not English, you must take either the Test of English as a Foreign Language (TOEFL) or the International English Language Testing System (IELTS)” (http://www.bu.edu/admissions/apply/transfer/). On the Undergraduate Admissions Website of the George Washington University, it is stated, “If English is not your first language, you are required to take the TOEFL, regardless of citizenship” (http://undergraduate.admissions.gwu.edu/application-requirements).

It is noteworthy that U.S. citizenship is not generally considered a strong reason for exemption from the English proficiency requirement. A very common phrase that applicants can find in the description of the English proficiency requirement is “We highly recommend that you take the TOEFL if English is not your native language, regardless of your citizenship” (Brown University, http://www.brown.edu/admission/undergraduate/apply-brown/frequently-asked-questions). However,
there are a number of U.S. universities and programs that exempt students from taking English proficiency tests if they have U.S. citizenship. These universities are: Cornell University, the Pennsylvania State University, the Ohio State University, the University of California – Irvine, the University of Tennessee, Virginia Polytechnic Institute and State University. The policy descriptions are provided in Table 2.1.

Table 2.1

*Universities where U.S. Citizenship Grants Exemption from the English Proficiency Requirement*

<table>
<thead>
<tr>
<th>University</th>
<th>Source</th>
<th>Exemption based on U.S. citizenship</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of California – Irvine</td>
<td>UCIrvine Graduate Division Website</td>
<td>“US citizens are not required to take the TOEFL exam”</td>
</tr>
<tr>
<td>The University of Tennessee</td>
<td>The University of Tennessee: Undergraduate Admissions Website</td>
<td>“If your native language is not English, you must demonstrate English proficiency unless you are a US citizen or a Permanent Resident.”</td>
</tr>
<tr>
<td>Virginia Polytechnic Institute and State University</td>
<td>Virginia Tech: Graduate School Website</td>
<td>“TOEFL scores are not required of US permanent residents and US naturalized citizens.”</td>
</tr>
<tr>
<td>Cornell University</td>
<td>Cornell University: Graduate School Website</td>
<td>“The TOEFL requirement may be waived if the applicant meets at least one of these criteria: you are a native citizen of the United States…”</td>
</tr>
<tr>
<td>The Ohio State University</td>
<td>Undergraduate Admissions Website</td>
<td>“Applicants who are citizens of or who have received a bachelor’s degree from one of the following countries are exempt from the English proficiency requirement: …, the United States…”</td>
</tr>
<tr>
<td>The Ohio State University</td>
<td>Graduate Admissions Website</td>
<td>“Applicants who are citizens of or who have received a bachelor's degree or higher in one of the following countries are exempt from the English proficiency requirement: …, the United States …”</td>
</tr>
<tr>
<td>Pennsylvania State University</td>
<td>Undergraduate Admissions Website</td>
<td>“US citizens and permanent residents are not required to provide proof of English language proficiency, regardless of native language.”</td>
</tr>
</tbody>
</table>

Another very common reason for exemption from TOEFL and similar language proficiency tests is completion of an institutional English language program after being admitted conditionally or even prior to applying to the university. The latter option is available at New York University through the American Language Institute, at Washington State University through their Intensive American English Center, at North Carolina State University at Raleigh through their Intensive English Program.
**English Proficiency Tests Accepted by RU/VHs**

All RU/VHs accept students’ TOEFL (Paper-based and Internet-based versions) scores. The TOEFL (Test of English as a Foreign Language) measures the ability of non-native speakers of English to use and understand English at the university level (http://www.ets.org/toefl/ibt/about?WT.ac=toeflhome_ibtabout2_121127). It consists of four sections: Reading, Listening, Speaking (to prompts for a computer recording), and Writing. 94% of the RU/VH universities accept the International English Language Testing System (IELTS) scores. Similarly to the TOEFL, the IELTS measures candidates’ English language proficiency needed for higher education environment (http://www.ielts.org/test_takers_information/what_is_ielts.aspx). It also consists of four components: listening, reading, writing, and speaking (in a face-to-face interview).

Among less widely used tests to satisfy the English Language proficiency requirement are Michigan State University English Language Test (MSU ELT), Advanced Placement English Language, Michigan English Language Assessment Battery (MELAB), Michigan State University Certificate of English Language Proficiency (CELP), Graduate Record Examination (GRE) Analytical Writing, Graduate Management Admission Test (GMAT) Analytical Writing and some sections of SAT and ACT for undergraduate students. Detailed information about the English proficiency tests accepted by 108 RU/VH universities is presented in Appendix A.

It is also emphasized on the admissions websites of many universities that TOEFL and IELTS scores should not be older than 2 years. The only exception that I came across was the policy of the University of Georgia according to which “scores must not be over 5 years old”. However, TOEFL and IETLS scores are not reported by test publishers if they are older than 2 years. Thus, it is very likely that this statement was posted on the website by mistake or it was not deleted after the section to which it belonged had been removed. Although a two-year rule is strictly enforced by test publishers, I was not able to find any research basis for it.

For undergraduate international admissions, it is also a very common practice to waive the TOEFL and IELTS requirement if students’ SAT (the SAT Reasoning Test, formerly called the
Scholastic Aptitude Test, and later the Scholastic Assessment Test) or ACT (American College Testing) scores for certain sections are relatively high. SAT Critical Reading scores that can grant a waiver from a language proficiency test (TOEFL or IELTS) typically range from 430 to 600 and up; and ACT English scores range from 17 to 29. The details are summarized in Table 2.2.

Table 2.2

*SAT Critical Reading and ACT English Scores that Grant Waiver from the English Proficiency Requirement*

<table>
<thead>
<tr>
<th>University</th>
<th>SAT Critical Reading</th>
<th>ACT English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Washington University in St Louis</td>
<td>600</td>
<td>29</td>
</tr>
<tr>
<td>University of Pittsburg – Pittsburg Campus, University of Southern California</td>
<td>600</td>
<td>27</td>
</tr>
<tr>
<td>University of Illinois at Urbana-Champaign</td>
<td>560</td>
<td>25</td>
</tr>
<tr>
<td>Arizona State University, Case Western Reserve University</td>
<td>550</td>
<td>23</td>
</tr>
<tr>
<td>University of Washington – Seattle Campus</td>
<td>550</td>
<td>22</td>
</tr>
<tr>
<td>University of Iowa</td>
<td>540</td>
<td>21</td>
</tr>
<tr>
<td>Pennsylvania State University</td>
<td>530</td>
<td>24</td>
</tr>
<tr>
<td>University of Arizona</td>
<td>530</td>
<td>22</td>
</tr>
<tr>
<td>University of Tennessee</td>
<td>510</td>
<td>21</td>
</tr>
<tr>
<td>Ohio State University, Oregon State University, Texas A&amp;M University</td>
<td>500</td>
<td>21</td>
</tr>
<tr>
<td>Louisiana State University and Agricultural &amp; Mechanical college, Purdue University, University of Cincinnati</td>
<td>480</td>
<td>20</td>
</tr>
<tr>
<td>Mississippi State University</td>
<td>480</td>
<td>19</td>
</tr>
<tr>
<td>Michigan State University</td>
<td>480</td>
<td>18</td>
</tr>
<tr>
<td>University of Oregon</td>
<td>480</td>
<td>n/a</td>
</tr>
<tr>
<td>University of Nebraska - Lincoln</td>
<td>470</td>
<td>20</td>
</tr>
<tr>
<td>University of Arkansas</td>
<td>460</td>
<td>19</td>
</tr>
<tr>
<td>Georgia State University, University of Georgia</td>
<td>430</td>
<td>17</td>
</tr>
<tr>
<td>University of South Florida</td>
<td>“Positive results on the SAT (Critical Reading and Writing) or ACT (English/Writing and Reading) may be used to waive the TOEFL/IELTS requirement.”</td>
<td></td>
</tr>
</tbody>
</table>

Besides the TOEFL and IELTS, international applicants to graduate programs are often asked to provide additional test results. The Graduate Record Examinations (GRE) is a standardized test required for application to many graduate schools in the USA. It measures verbal reasoning, quantitative reasoning, analytical writing, and critical thinking skills. The Graduate Management Admission Test
(GMAT) is a requirement for admission to many MBA programs. It measures candidate’s analytical, writing, quantitative, verbal, and reading skills in standard written English. Both tests contain a writing component; however, as the survey of the official admission requirements suggests, it is not very common to use GRE or GMAT scores for waiver from the English proficiency requirement. The only exception that I was able to find is the University of Arkansas that waives the English proficiency requirement if students achieved 4.5 for GRE Analytical Writing or 4.5 for GMAT Analytical Writing.

The minimum English proficiency score requirements for general admission to the RU/VH universities differ significantly (see Figures 2.1 and 2.2). 89 undergraduate programs and 69 graduate programs set specific score requirements for general admission. The universities for which I was not able to find specific score requirements are those where:

- Language proficiency test scores are not required (e.g., Harvard and Stanford universities – undergraduate admissions);
- Emphasis is only on the subsection scores and there is no total score requirement (e.g., Boston University, the University of Arkansas);
- Only the averages of the total scores for generally admitted students are provided (e.g., Northwestern University, the University of Wisconsin – Madison);
- The universities do not have general campus-wide score requirements and the score values are set by the departments (usually for graduate admissions) (e.g., North Dakota State University, Northwestern University, Tufts University);
- The information about the specific score requirements is missing or cannot be found on the university’s website (e.g., Rockefeller University).

The lowest minimum TOEFL iBT score requirement (a score of 61) for undergraduate admission is set by Arizona State University, the University of Alabama at Birmingham and the University of Arizona (starting from Fall 2013, a minimum score increased to 79). The highest minimum score requirements for both undergraduate and graduate admissions (a score of 104) is set by the University of
Chicago. The lowest score for graduate admissions (a score of 53) is required by the Mississippi State University. The majority of the universities require the scores in the range of 79-80 for both graduate and undergraduate admissions. It is also very common that some departments set higher score requirements than the campus-wide minimums.

Figure 2.1. Minimum total TOEFL iBT scores for general admission to undergraduate programs

Figure 2.2. Minimum total TOEFL iBT scores for general admission to graduate programs
Interestingly, English proficiency requirements for admission to undergraduate and graduate programs may differ within the same university. Undergraduate students may be required to meet higher scores requirements than graduate students. For example, Brown University and Rice University set a TOEFL iBT score of 100 for undergraduate and 90 for graduate students. In other universities, undergraduate students may have lower score requirements than graduate students. For example, Colorado State University sets a total TOEFL iBT score of 71 for undergraduate students and a score of 80 for graduate students. The University of Nebraska - Lincoln sets a TOEFL iBT score of 70 for undergraduate admissions and a score of 79 for graduate admissions. The University of Iowa sets a score of 80 for undergraduate students and 81 for graduate students. The score difference between the two levels in the universities mentioned above is only 9 to 10 points; however, there are a few universities where this difference is more than 10 points, for example, Arizona State University and the University of Oregon established a TOEFL iBT score of 61 for undergraduate students and a score of 80 for graduate students.

Another noteworthy trend is observed regarding the English proficiency test score requirements for students applying to STEM (Science, Technology, Engineering and Mathematics) and non-STEM programs. According to the undergraduate admissions website of Purdue University, the minimum TOEFL iBT score for admission to first year Engineering and College of Science undergraduate programs is 88; whereas students applying to all other colleges have to submit at least a 79. The University of California, Santa Cruz (UCSC) established a TOEFL iBT score of 89 for applicants to graduate programs in Chemistry and Biochemistry, Computer Engineering, Computer Science and Electrical Engineering; whereas the minimum TOEFL iBT score for general admission to UCSC is 83. The minimum TOEFL iBT score for admission to undergraduate program at the University of Cincinnati is 66; however, the TOEFL requirement for Engineering programs is 76 (with minimum 20 in speaking). Undergraduate students majoring in STEM have also to satisfy higher English proficiency requirements at the University of Colorado, Boulder (CU, Boulder). International candidates applying to engineering majors at CU – Boulder are required to submit a TOEFL iBT score of 83 or higher but applicants to all other majors need
to earn at least a 75 on the TOEFL iBT in order to be admitted. International students applying to the graduate Chemistry, Technology, Media & Society, Electrical Engineering, Applied Math programs at CU - Boulder need to get a 100 or higher on the TOEFL iBT. Applicants to the graduate programs in Linguistics, Communications, Dance majors need to get at least a 79 on the TOEFL iBT). A similar trend is observed at the University of Illinois at Urbana-Champaign regarding undergraduate admissions to the College of Business, Engineering and Media (a TOEFL iBT score of 100 versus a score of 79 for general admission) and graduate admissions to Mathematics, Civil Engineering, Computer Science (a TOEFL iBT score of 86 for Math majors and a score of 82 for other STEM versus a TOEFL iBT score of 77 for all non-STEM majors. Different language proficiency score requirements for STEM and Humanities majors might be linked to the fact that these majors differ greatly in payoffs. Many STEM professions are generally the closest to the job market and offer graduates the greatest returns (Davies & Guppy, 1997; Walters, 2002).

**Important Terminology**

It is important to differentiate between the English proficiency requirements for general (“clear,” full or regular) and conditional (provisional or limited status) admission of international students. The majority of the RU/VHs set only general admission requirements and all the test scores that fall below those requirements are considered for conditional admission. However, there are some universities that specifically set minimum score requirements for both general and conditional admissions (see Table 2.3).

Upon being admitted conditionally, students may be required to complete a remedial ESL course (e.g., the Pennsylvania State University, Main Campus, Texas A & M University) or enroll into a local Intensive English program (e.g., Colorado State University, Georgia State University, Iowa State University). Some universities set specific time frames for completing the conditional admission requirements. For example, the University of Arizona asks conditionally admitted students to achieve the required level of English proficiency within one year. At Georgia State University, a student may remain on conditional admission status for no more than three semesters.
Table 2.3

**TOEFL and IELTS Scores for General and Conditional Admission**

<table>
<thead>
<tr>
<th>University</th>
<th>Level (G/UG)</th>
<th>General admission</th>
<th>Conditional admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorado State University</td>
<td>G</td>
<td>TOEFL iBT 80, TOEFL PBT 550, IELTS 6.5</td>
<td>TOEFL iBT 50, TOEFL PBT 475, IELTS 5.5</td>
</tr>
<tr>
<td></td>
<td>UG</td>
<td>TOEFL iBT 79, TOEFL PBT 550, IELTS 6.5</td>
<td>TOEFL iBT 45, TOEFL PBT 450, IELTS 5.0</td>
</tr>
<tr>
<td>Georgia State University</td>
<td>UG</td>
<td>TOEFL iBT 79-80, TOEFL PBT 550, IELTS 5.5</td>
<td>TOEFL iBT 54-55, TOEFL PBT 480, IELTS 4.5</td>
</tr>
<tr>
<td>Iowa State University</td>
<td>UG</td>
<td>TOEFL iBT 71 (min 17 for Speaking and Writing), IELTS 6.0</td>
<td>TOEFL iBT 55, IELTS 5.5</td>
</tr>
<tr>
<td>Michigan State University</td>
<td>UG</td>
<td>TOEFL iBT 79 (no subsection score below 17), TOEFL PBT 550 (no subscore below 52), IELTS 6.5</td>
<td>TOEFL iBT 60-78, TOEFL PBT 500-549, IELTS 6.0</td>
</tr>
<tr>
<td>North Carolina State University at Raleigh</td>
<td>UG</td>
<td>TOEFL iBT 80, TOEFL PBT 550, IELTS 6.5</td>
<td>TOEFL iBT 65, TOEFL PBT 513, IELTS 6.0</td>
</tr>
<tr>
<td>North Dakota State University</td>
<td>UG</td>
<td>TOEFL iBT 71, TOEFL PBT 525, IELTS 6.0</td>
<td>TOEFL iBT 48, TOEFL PBT 460, IELTS 4.5</td>
</tr>
<tr>
<td>Oregon State University</td>
<td>UG</td>
<td>TOEFL iBT 80, TOEFL PBT 550, IELTS 6.5</td>
<td>TOEFL iBT 60, TOEFL PBT 500, IELTS 5.5</td>
</tr>
<tr>
<td>University of Kansas</td>
<td>G</td>
<td>TOEFL iBT 80, TOEFL PBT 550, IELTS 6.5, (TOEFL CBT 213)</td>
<td>TOEFL iBT 61-79, TOEFL PBT 500-547, IELTS 6.0 (TOEFL CBT 173-210)</td>
</tr>
<tr>
<td>University of Maryland – College Park</td>
<td>G</td>
<td>Waiver from on-campus testing: TOEFL iBT (23 for R, L &amp; W); IELTS total score of 6.5 (6.0 for L, R &amp; W)</td>
<td>TOEFL iBT (R, L &amp; W 18-19) IELTS total score of 5.5 (no subsection score below 5.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No waiver from on-campus testing: TOEFL iBT 20 for R, L &amp; W; IELTS total score of 6.0 (no subsection score below 5.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TOEFL iBT 100 (S 22, L 24, R 26, W 24), IELTS 7.0 (L 7, R 7, W 7, S 6.5)</td>
<td>TOEFL iBT 84-99 (S 19, L 18-23, R 21-25, W 21-23), IELTS 6.5 (L 6.5, R 6.5, W 6.5, S 6.5)</td>
</tr>
</tbody>
</table>
A more complex conditional admission policy is established at the University at Buffalo. Here they differentiate between ‘Conditional Admission’ when applicants’ TOEFL/IELTS scores are slightly below the standard minimum and ‘Conditional Bridge Admission’ when applicants’ test scores are significantly below the standard minimum. Each case is reviewed individually and other criteria such as grades and academic performance are usually taken into consideration.

There are a number of universities that state on the official websites that they do not conditionally admit students who have not met the English proficiency requirement. Among such universities are Brown University (undergraduate admissions), the Ohio State University, Main Campus (undergraduate admissions), and the University of Washington, Seattle Campus (undergraduate admissions).

Some universities do not have any minimum English proficiency score requirements for admission, for example, the University of Southern California (graduate admissions) and the California Institute of Technology (graduate admissions). There are also those that cite only the average scores of the admitted international students. For instance, “most competitive applicants” are accepted to New York University with the scores above 100 for the TOEFL iBT and 7.5 for the IELTS. Princeton University provided the average TOEFL iBT scores for admitted students over the last two seasons: total 107, Listening 28, Reading 29, Speaking 24 and Writing 26. “Competitive applicants” to Northwestern University report scores 105 or above for the TOEFL iBT and 7.5 or above for the IELTS. The students admitted to undergraduate programs of the University of Wisconsin, Madison generally have TOEFL iBT scores in the 95-105 range.

It is also necessary to make a distinction between minimum proficiency test scores for admission and cutoff scores for ESL placement. Minimum scores for admission are typically lower than cutoff scores for ESL requirement. Cutoff scores for ESL are usually established at the universities that have local ESL support to international students. Examples of the universities with the ESL cutoff score policy are the following:
• The George Washington University: newly admitted students with TOEFL iBT score of 100 (600 for PBT) and IELTS score of 7.0 (no individual score below 6.0) are exempt from taking the English for Academic purposes placement test.

• The University of Arkansas: students who score 28 or above on the TOEFL iBT (Writing), 5 or above on the TWE, 6.5 or above on the IELTS (Writing), 77 on the PTE-A, or 81 on the ELPT are exempt from additional language instruction.

• The University of Illinois at Urbana-Champaign: students with TOEFL iBT scores below 102, PBT score below 610 and IELTS scores below 6.5 are required to take the English Placement Tests upon arrival at UIUC.

• The University of Iowa: students with TOEFL iBT total score of 100 are exempt from an on-campus English Proficiency Evaluation.

• The University of Southern California: students with TOEFL iBT score of 100 (min 20 for subsections), IELTS score of 7.0 (min 6.0 for subsections) are exempt from taking the International Student Examination (ISE) upon arrival at USC.

• The University of Wisconsin – Madison: students whose TOEFL iBT scores are below 92, PBT score are below 580, IELTS scores are below 7 or MELAB scores below 82 must take an English assessment test upon arrival.

English proficiency requirements turned out to be much more complex than general test requirements for admission to undergraduate and graduate programs. Almost all undergraduate programs that I reviewed require all undergraduate applicants to take ACT or ACT; however, the majority of them do not generally set any specific cutoff scores for these tests, only very few do have some specific SAT and ACT requirements. For example, the ACT requirement for freshmen admission to Louisiana State University and Agricultural & Mechanical College consists of a 22 composite score, with an 18 English subscore and a 19 math subscore. The SAT requirement consists of a 1030 critical reading and math composite score, with a 450 critical reading score and a 460 math score. At North Dakota State
University, an ACT composite score of 21 or higher or SAT score (math and reading) of 970 or higher is recommended for general undergraduate admissions.

Table 2.4

*Universities Accepting TOEFL CBT Scores*

<table>
<thead>
<tr>
<th>TOEFL CBT Score</th>
<th>Level (G/UG)</th>
<th>University</th>
</tr>
</thead>
<tbody>
<tr>
<td>153</td>
<td>G</td>
<td>Mississippi State University</td>
</tr>
<tr>
<td>173</td>
<td>UG</td>
<td>Arizona State University, University of Alabama in Huntsville</td>
</tr>
<tr>
<td>190</td>
<td>UG</td>
<td>University of New Mexico – Main Campus</td>
</tr>
<tr>
<td>193</td>
<td>UG</td>
<td>University of Tennessee</td>
</tr>
<tr>
<td>197</td>
<td>UG</td>
<td>Mississippi State University, University of Kentucky</td>
</tr>
<tr>
<td>213</td>
<td>UG &amp; G</td>
<td>University of Texas at Austin, Washington State University, SUNY at Albany, University of Illinois at Chicago, University of Louisville, Louisiana State University and Agricultural &amp; Mechanical College, Montana State University, University at Buffalo, University of Illinois at Urbana-Champaign, University of Kentucky, University of New Mexico, University of South Florida, University of Tennessee, Virginia Polytechnic Institute and State University</td>
</tr>
<tr>
<td>220</td>
<td>UG</td>
<td>Yeshiva University</td>
</tr>
<tr>
<td>230</td>
<td>UG</td>
<td>Vanderbilt University</td>
</tr>
<tr>
<td>233</td>
<td>UG</td>
<td>Rensselaer Polytechnic Institute</td>
</tr>
<tr>
<td>250</td>
<td>UG</td>
<td>Case Western Reserve University, Virginia Polytechnic Institute and State University</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>Dartmouth College, Tufts University, Yale University</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>Brandeis University, Rice University, Tulane University of Louisiana, Washington University in St Louis</td>
</tr>
</tbody>
</table>

I also came across several universities that publish admission statistics of their admitted undergraduate students. Among such universities are Carnegie Mellon University, Georgia State University, Indiana University, Bloomington, Rutgers University, New Brunswick, and Princeton University. As far as general graduate admissions are concerned, the GRE is required by the majority of
graduate programs; however, most programs do not enforce any strict cut off scores for this test. For example, in the description of the general GRE requirements for entry to Iowa State University, there is a statement, “Most programs do not enforce strict cut-off scores. Rather, they review an applicant’s complete application package and credentials before making a decision” (http://www.admissions.iastate.edu/apply/pdf/us_gr_instr.pdf).

Another surprising finding is that 38 universities still have TOEFL CBT (Computer-based Test) score requirements even though this format of the TOEFL tests was discontinued in 2006 and the CBT scores are no longer valid (see Table 2.4).

**Subsection Scores Also Matter**

Another important trend is that some RU/VHs set not only total score requirements but also subsection score requirements. There are also those that set only subsection score requirements, e.g., Boston University – “We look closely at the score for each section rather than the total” (http://www.bu.edu/admissions/apply/international/application-instructions/). The summary of the subsection score requirements in the RU/VH universities is presented in Table 2.5.

Although I did not find any policy that contained a direct statement saying that students with low standardized test scores could also be admitted, I came across several policy descriptions that can certainly give hope to applicants with low scores and that can potentially alter and possibly improve the enrollment rates of a university. For example, at the University at Buffalo, “each case is reviewed individually and in conjunction with other criteria (grades, academic performance, etc.)” (http://www.buffalo.edu/internationaladmissions/get-ready-to-apply/can-i-get-in/admissions-criteria/bachelor-degree-study-with-english-language-support.html).
### Table 2.5

**RU/VH universities that consider subsection scores for admission**

<table>
<thead>
<tr>
<th>University</th>
<th>Level (UG/G)</th>
<th>Test scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston University</td>
<td>UG+G</td>
<td>College of Communication, College of General Studies and School of Management: TOEFL BT (22 for Writing, 23 for Speaking, 25 Reading, 21 for Listening); Other colleges: TOEFL iBT (22 for Writing, 23 for Speaking, 21 for Reading, 18 for Listening)</td>
</tr>
<tr>
<td>Iowa State University</td>
<td>UG</td>
<td>TOEFL iBT 71 (min score of 17 for Speaking and Writing)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IELTS 6.0 (min score of 5.5 for subsections)</td>
</tr>
<tr>
<td>Johns Hopkins University</td>
<td>UG</td>
<td>TOEFL iBT 26 for Reading, 26 for Listening, 22 for Writing and 25 for Speaking</td>
</tr>
<tr>
<td>Michigan State University</td>
<td>UG</td>
<td>TOEFL iBT 79 (no subsection score below of 17)</td>
</tr>
<tr>
<td>Pennsylvania State University</td>
<td>G</td>
<td>TOEFL iBT 80 (min score of 19 for Speaking)</td>
</tr>
<tr>
<td>Rutgers University</td>
<td>G</td>
<td>TOEFL iBT: 22 for Writing, 23 for Speaking, 21 Reading and 17 for Listening</td>
</tr>
<tr>
<td>University of Alabama in Huntsville</td>
<td>UG</td>
<td>TOEFL iBT 62 (15 for Reading, 16 for Listening, 18 for Speaking, and 13 for Writing)</td>
</tr>
<tr>
<td>University of Chicago</td>
<td>G</td>
<td>TOEFL iBT 104 (min 26 for subsections)</td>
</tr>
<tr>
<td>University of Illinois at Chicago</td>
<td>UG+G</td>
<td>TOEFL iBT 80 (21 for Writing, 20 for Speaking, 17 for Listening, 19 for Reading)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IELTS 6.5 (min 6.0 for subsections)</td>
</tr>
<tr>
<td>University of Iowa</td>
<td>UG</td>
<td>TOEFL iBT 80 (min 17 for subsections)</td>
</tr>
<tr>
<td>University of Kansas</td>
<td>G</td>
<td>TOEFL iBT: min 23 for Reading, Listening and Writing</td>
</tr>
<tr>
<td>University of Michigan – Ann Arbor</td>
<td>UG</td>
<td>TOEFL iBT 88 (23 for Reading and Listening, 21 for Speaking and Writing)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IELTS 6.5 (6.5 for subsections)</td>
</tr>
<tr>
<td>University of Minnesota – Twin Cities</td>
<td>G</td>
<td>TOEFL iBT 79 (21 for Writing, 19 for Reading)</td>
</tr>
<tr>
<td>University of Missouri - Columbia</td>
<td>UG</td>
<td>TOEFL iBT 61 (with no min score below 15)</td>
</tr>
<tr>
<td>University of Nebraska - Lincoln</td>
<td>UG</td>
<td>TOEFL iBT 70 (with min 20 for Writing)</td>
</tr>
<tr>
<td>University of Notre Dame</td>
<td>G</td>
<td>TOEFL iBT 80 (with min 23 for Speaking)</td>
</tr>
<tr>
<td>University of Rochester</td>
<td>UG</td>
<td>TOEFL iBT 100 (with min 25 for Writing, 24 for Speaking, 25 for Reading and 25 for Listening)</td>
</tr>
</tbody>
</table>
Conclusion

The results of this survey suggest that almost all RU/VHs require international applicants for whom English is their second language to demonstrate a certain level of English proficiency. Some universities have very concise English proficiency requirements; others set more detailed and strict English proficiency policies. All the universities accept Paper-based and Internet-based versions of the TOEFL test. The majority also accept the IELTS. Among less widely used tests to satisfy English Language proficiency requirement are Michigan State University English Language Test (MSUELT), Advanced Placement English Language, Michigan English Language Assessment Battery (MELAB), Michigan State University Certificate of English Language Proficiency (CELP), Graduate Record Examination (GRE) Analytical Writing, Graduate Management Admission Test (GMAT) Analytical Writing and some sections of SAT and ACT for undergraduate students.

There is a significant difference among the universities regarding the English proficiency waiver policy. Previous academic experience in the USA is the most common reason for the waiver from the English proficiency requirement. Many universities generally exempt newly admitted students from the English requirement if they have a high school diploma or Bachelor’s degree from a U.S. educational institution or completed an institutional English language program prior to admission.

Very drastic differences are observed in the TOEFL iBT score requirements for admission. However, a significant number of universities require TOEFL iBT scores of 79 or 80. It is also very common that some departments set higher standardized English proficiency score requirements than the campus-wide minimums. In such cases, prospective applicants are encouraged to double check the score requirements with the department to which they are planning to apply. A large number of universities exercise some flexibility in admitting students with lower test scores and additionally set score requirements for conditional (a.k.a. provisional) admission. Conditionally admitted students must be aware of the fact that some programs set specific time frames for completing the English proficiency requirements.
Many universities set TOEFL iBT and IELTS section score requirements in addition to or instead of the total score requirements. There is also an observed difference in score requirements between graduate and undergraduate programs. Another trend is that the English proficiency requirements for students majoring in STEM are generally higher than those in the humanities or arts.

Comparison of the general admission requirements with the English proficiency requirements suggests that the latter are generally more complex and specific. Many programs do not set specific ACT or SAT score requirements for undergraduate admissions and GRE score requirements for graduate admissions. In some case, the GRE is not required at all for general admission.

The observed complexity of the English proficiency policy at RU/VHs prompts an important question: What causes such variety of standardized English proficiency requirements for admission? One good explanation for this may be that the U.S. universities with historically very strong Education and/or TESOL programs that hire teaching assistants to help in providing ESL support to newly-admitted international students can be less strict and more flexible about the admission requirements. This may be the case, for example, for Purdue University, the University of Illinois at Urbana-Champaign, and the University of Southern California. Solid ESL support might also explain the fact why these three universities are the top three leaders in enrollment of international students (IIE, 2013). By contrast, those universities that do not have TESOL programs or those that never experienced shortage of applicants can set higher English proficiency score requirement. This might especially be true for the University of Chicago (a TOEFL iBT score of 104 for general admission).

It is also very likely that the complexity and diversity of the English proficiency requirements is driven by the higher education market. Due to the sensitivity and novelty of this topic, there is very little literature on this issue. Messner and Liu (1995) made an attempt to raise this topic when discussing an arbitrary nature of the cutoff score requirements for admission to graduate programs:

Inquiry into the method by which TOEFL cut-off scores were established by graduate school and admission committees has shown that, in many cases, cut-off scores [for admission purposes] have been arbitrary set with little, if any, scientific rationale for these decisions.
In my view, “marketization” (Williams, 1995) of higher education is the strongest reason for increased stratification between institutions of higher education in student admission scores. Language proficiency scores can serve as a very convenient tool for regulating enrollment of international students. Campuses that wish to attract more international applicants can simply adjust their English proficiency score requirements; whereas those that do not experience shortage of applications can afford to have higher score requirements than other universities. Hoxby (1997) suggested that vigorous market competition for able students and the loss of monopoly power due to geographic integration increased the between-college variance and decreased the within-college variance in students’ admissions test scores over time.

My belief is that there exists strong tension between local ‘ecological’ freedom in interpreting test results according to local needs and broader general unified policies that respect what other educational institutions do and the variability of resources across campuses, for example, presence or absence of a MATESL program. On the one hand, it is good that universities have the right to determine their admission score requirements. On the other hand, some universities, that set low scores in order to attract more applicants but cannot provide adequate ESL support, may significantly lower the quality education.

The bottom line of this study is that prospective international applicants must attentively review all requirements for admission to the universities they wish to apply. The RU/VHs differ dramatically in terms of the English proficiency requirement. It is obvious that students with higher scores have higher chances to be accepted to the RU/VHs than those that scored low on the TOEFL or IELTS. However, due to the growing competition for international applicants there are many universities that are quite flexible on the standardized score requirements and conditionally admit international students with relatively low standardized English proficiency test scores.
References


Institute of International Education. (November 11, 2013). Open Doors 2013: International Students in the United States and Study Abroad by American Students are at All-Time


Louisiana State University. (n.d.). Graduate School: Prospective Student Information. Retrieved October 14, 2012, from 
http://gradlsu.gs.lsu.edu/Prospective%20Students/Prospective%20Student%20Information/item12177.html


Stony Brook University. (n.d.). Applying to the Graduate School – FAQS. Retrieved October 18, 2012, from https://www.grad.stonybrook.edu/ProspectiveStudents/faq.shtml#scores


http://www.miami.edu/coe/index.php/coe/prospective_graduate_student/graduate_admission/graduate_admission_application_checklist/


Chapter 3

Investigating the possibility of using TOEFL scores for university ESL decision-making:

Placement trends and effect of time lag¹

(Paper Two)

Abstract

The English Placement Test (EPT) at the University of Illinois at Urbana-Champaign (UIUC) is designed to provide an accurate placement (or exemption from placement) of international students into the ESL writing and pronunciation classes. Over the last five years, UIUC has experienced an increase in the number of international students taking the EPT. Because of the costs of the EPT, various stakeholders have suggested using TOEFL as a substitute placement tool. The purpose of this research is to find out whether the TOEFL iBT can serve as a suitable pre-screening tool for the placement of international students into ESL writing courses at UIUC.

While there is a general, statistically significant trend that students with higher TOEFL iBT scores are placed into higher levels of ESL classes, there is no particular set of either total or section scores which can be used as a reliable criterion for dividing students into ESL classes without significant misplacement. Furthermore, the correlation between TOEFL iBT and the EPT shows a distinct pattern of time dependence: the correlation is stronger when the time gap between these tests is short but it dramatically decreases and even becomes slightly negative around Week 50; however, starting from the 50th week it increases. Some possible interpretations of the findings as well as recommendations for score users are discussed.

Keywords

TOEFL, time lag, ESL placement, correlation.

¹ http://ltj.sagepub.com/content/29/2/291
The final, definitive version of this paper has been published in Language Testing, 29/2, April/2012 by SAGE Publications Ltd, All rights reserved. © [Kokhan, K].
Background

Placement tests vs. proficiency tests. “Placement tests are designed to assess students’ level of language ability so that they can be placed in the appropriate course or class. Such tests can be used on aspects of the syllabus taught at the institution concerned, or may be based on unrelated material” (Alderson et al., 1995, p. 11). According to Wall, Clapham and Alderson (1994), placement tests are subdivided into two types. The first type has a proficiency orientation and may be considered an institutional version of a proficiency test such as TOEFL or the Michigan Test. The content of such tests usually has no direct relationship to the content of the language courses into which students are placed. The second type is developed to specifically reflect the nature of language courses. Universities may give such tests to international students at the beginning of their first semester at the university in order to find out if the students need to take any courses targeting the language skills which are necessary to study in a particular academic environment.

Conversely, proficiency tests are developed to show whether students have managed to reach a certain level of their language abilities (Alderson et al., 1995) but they may fail to fully meet all the specific language needs of a certain program or educational institution. According to Ascher (1990), English proficiency tests do not measure students’ active use of English and may inadequately assess how well students will be able to do in an English-speaking academic environment.

Use of post-admission tests for ESL placement. According to Fulcher (1997), fair and accurate assessment of students’ abilities and their placement into appropriate language courses is essential to all academic departments. The main objective of placement tests is reduction to an absolute minimum the number of students who may be at risk of failing some courses or not getting their academic degrees because of their poor language skills.

Wall, Clapham and Alderson (1994) reported the research findings obtained by the Language Testing Research Group (LTRG) about the nature and validation of an institutional placement test at Lancaster University. The researchers confirmed the importance of administering post-admission tests since it was generally hard to make conclusions about the level of English judging by that little evidence
reported by international students when applying to the university. They also warned against self-assessment as a tool for gathering the information about the language need of international students since the placement test populations were generally heterogeneous and it would be unreasonable to expect students coming from different educational and cultural environment to accurately estimate the level of their English.

The paper by Wall, Clapham and Alderson (1994) was published almost 17 years ago and the admission policies of many universities have undergone some changes since then. For example, many North American universities are now setting quite strict English language requirements, according to which international students have to take some “recognized” proficiency tests before applying to the university (cf. according to Wall, Clapham and Alderson (1994), the challenges of using the information reported by students about the level of their English were related to the interpretation of the unknown test scores, incomplete or outdated score reports and English course grades). The availability of such homogeneous databases should establish a new direction in language testing research since now it may be much easier to explore the effectiveness of using language test scores for pre-screening international students before their arrival on campus.

**Use of standardized language proficiency tests for ESL placement.** The practice of using external proficiency for ESL (or EAP) placement is quite widespread. Since more and more international students choose to study in English-speaking countries, universities try to find the most optimal pre-screening tools in order to further decide if their admitted international students need to take some ESL courses. For example, TOEFL scores are widely used for making admission decisions and ESL placement due to a high reliability and accessibility of this test. The purpose of TOEFL is to evaluate how the applicants whose native language is not English can understand and use English in academic settings. In a survey study of 78 public and private colleges and universities in the United States, Williams (1995) reported that 26% of the institutions relied on TOEFL scores for placement into the ESL writing courses.

Some studies show that standardized tests of English can be quite effective for predicting EAP (English for academic purposes) placement. A study by Fox (2004) on the use of the Canadian Academic
English Language (CAEL) Assessment provided some evidence that this test was generally very effective in identifying EAP support classes (only 10 (3%) students were misplaced in Fall 2000). The effectiveness of CAEL might be explained by the fact that it “differs from other standardized tests in that its specific purpose is to identify test-takers who have the ability to use English for academic purposes in university classrooms” (Fox, 2004, p. 438). However, Fox (2004) warned against premature conclusions by explaining that this test happened to be effective for making appropriate decisions within one university program and that the study was just the beginning of what should be a continuing line of research.

Five years later, Fox (2009) published a paper examining some challenges of implementing a new admissions policy according to which standardized proficiency tests (TOEFL, IELTS) were used for placement into the levels of the EAP program at a Canadian university. The implementation of this policy resulted into an extreme amount of variability within EAP class levels which was a direct indication of misplacement. A big discrepancy between the students at the top and the students at the bottom of the class inevitably had some negative impact on teaching. The researcher suggested that the problems of misplacement could be overcome with the help of some systematic diagnostic assessment which could help teachers to adjust their instruction to individual student needs.

The issues related to the use of proficiency tests for placement were also discussed by Green and Weir (2004) who suggested that not all proficiency tests were suitable for placement into language classes. The authors explored the effectiveness of using a trial version of the Global Placement Test (GPT), paper-based proficiency test of grammatical knowledge developed to reflect course content and intended to be used as a practical and efficient instrument for placing students into relevant courses at language schools. The researchers expressed their doubt concerning the effectiveness of quick grammar-based placement tests, which traditionally were viewed as efficient and practical, in providing sufficient information for diagnostic mediation or sequencing of instructional materials. Overall, they concluded that such proficiency tests might yield some crude indications of students’ abilities but, unless supported by “more extensive procedures”, they were unlikely to provide some detailed diagnostic information that
would be useful for instructors and learners. This conclusion goes in line with the opinion of Mullen (2008) who points out that an increasing number of educational institutions are using proficiency tests for ESL placement but proficiency tests do not test the incremental developmental changes that define learning within an EAP program.

**Validity evidence supporting the use of TOEFL iBT scores for ESL placement.** Important study evidence in favor of using TOEFL iBT scores for ESL placement was obtained by Wang, Eignor, and Enright (2008). Participants (N=981) who were studying in English speaking countries were asked to indicate whether they were placed into (a) an IEP (Intensive English Program) but not courses in other subjects (e.g., biology or business), (b) ESL courses and courses in other subjects, or (c) courses in other subjects but not ESL courses. A correlation between academic placement and TOEFL iBT scores was analyzed using one-way ANOVA. For the total score the effect of academic placement was large ($\eta^2=.17$) and significant, $F(2,978) = 101.8$, $p < .0001$. Effects for academic placement were also significant for four skills: listening, $F = 83.9$, df = 2,978, $p < .0001$, reading, $F = 86.3$, df = 2,978, $p < .0001$, speaking, $F = 59.6$, df = 2,978, $p < .0001$, and writing, $F = 75.3$, df = 2,978, $p < .0001$. The differences in test scores between students who were enrolled in the ESL language development courses or IEPs, and those enrolled in only content courses were large and statistically significant. However, differences in scores between students enrolled in IEPs and students placed into ESL courses and subject matter classes were significant only in writing. All these findings indicate that TOEFL iBT scores could be suitable for ESL placement decisions.

Additional evidence in favor of using TOEFL scores for ESL placement is provided in the most recent report on the reliability of TOEFL iBT scores is “Reliability and Comparability of TOEFL iBT Scores” (2008). According to ETS, the reliability estimates for the reading (0.85), listening (0.85), speaking (0.88), and total (0.94) TOEFL iBT scores are relatively high, whereas the reliability of the writing score is somewhat lower (0.74). This is a typical result for writing measures consisting of only two tasks (Breland, Brigdeman, & Fowles, 1999, as cited in ETS, 2008). As pointed out in the report, estimates for measures comprising a small number of time-consuming tasks (cf. two essays in the writing
section of TOEFL) are often lower than estimates for measures composed of many short and less time-consuming tasks. Based on these results, ETS recommends using total scores for making high-stakes decisions (e.g., admissions to college or graduate school) because total TOEFL iBT scores reflect four language skills and because they are more reliable. Subsection score reports may be used if there are special demands of the curriculum or a need for additional language training (2008). However, decision makers should also take into consideration such factors as the number of applications for admission and the information about how candidates, collectively, have performed on the test (ETS, 2005a). Even though TOEFL iBT scores are considered to be reliable, “score users are encouraged to consider other evidence of English proficiency in addition to the TOEFL score” (www.ets.org).

**Notion of language attrition.** According to Des Brisay (1994), it is unreasonable to compare the scores from the tests written six months apart. It may be that test-takers made no use of their English over the period of six months; that is why it is quite possible that the scores on the second test will be lower. Alternatively, examinees may have had intensive English language training over this interval of time and the scores on the second test can reflect the impact of this training. This process of disintegration or losing the structure of the language learned after the native language is known as attrition (Hansen, 1999). Overall, there seems to be no agreement among the researchers about the factors causing second language attrition. Most of the hypotheses in second language studies were originally developed from the studies on L1 attrition. According to Bardovi-Harlig and Stringer (2010), the most discussed hypothesis of L2 attrition is the regression hypothesis which “predicts that the path of attrition is the reverse of the path of acquisition (last in, first out)” (Bardovi-Harlig & Stringer, 2010, p. 15). Other researchers believe in the hypothesis which is “the more you know, the less you lose” (Hansen, 1999, p. 151). This approach suggests that high second language proficiency provides some temporary immunity from L2 loss and better long-term outcomes for learners. Some other studies show the evidence in favor of retrieval failure hypothesis which refers to the problem of language retrieval due to lack of access to linguistic knowledge rather than its total loss (Clark & Jorden, 1984; Russell, 1999). A number of researchers also suggest that L2 skills may demonstrate different vulnerability to attrition. Production skills, such as speaking and
writing, are found to be more vulnerable to attrition than receptive skills (listening and reading) (Bahrick, 1984; Scherer, 1957; Snow et al., 1988).

To my knowledge, there is no research examining the effect of attrition on the validity period of standardized language proficiency test scores, in particular, TOEFL scores. The validity period of TOEFL scores, as recommended by ETS, is two years after the test date; however, there is no research evidence provided by the test developers why the scores are valid for two years. ETS Global BV, one of ETS’ subsidiaries, provides a short explanation for the validity period of scores; however, their recommendations are not supported by any cited research evidence either, “ETS test scores are valid as long as the level of English proficiency of the examinee remains the same. This proficiency can, however, improve or decline over time. For this reason, the more recent the score, the more likely it is to be a valid indication of the examinee’s proficiency. ETS Global recommends that a test score be considered valid for up to 2 years” (ETS Global: FAQ). The validity period of the score reports for another major standardized English proficiency test, the International English Language Testing System (IELTS), is also two years. According to the test developers, “The IELTS test provides an accurate picture of a candidate’s language skills at a given moment. For this reason, the validity of a score as a precise representation of a candidate’s abilities will inevitably diminish in time” (IELTS: A guide for agents). However, they suggest that the IELTS scores older than 2 years may also be accepted by score users but only if candidates show some proof that they have maintained or tried to improve their English. Again, no references are provided to support a two-year score validity period for this test, either.

Since there have been no studies examining the effect of time on candidates’ English proficiency as measured by standardized language tests, the present study will be an attempt to fill this gap in the language testing field. Moreover, this research is a response to institutional pressure to re-evaluate ESL placement practices.
The Present Study

The major motivation for this research is the necessity to reconsider the existing ESL placement policies due to a growing number of international students who chose to study at the University of Illinois and changes in TOEFL format over the last decade. The cutoff score which determines who is required to take the EPT was established at UIUC at the time when the paper-based version was generally accepted. The current cutoff score of 102 was set as a result of the conversion of 610 on the TOEFL PBT scale to iBT TOEFL score using the ETS conversion table (see ETS, 2005b). A question arises: Is this cutoff score still valid? In my understanding, the cutoff score should reflect the level of English proficiency sufficient to receive with a high probability the highest possible (Level 3) score on the writing part of the EPT.

In this relation, a surprising conclusion can be drawn from the data shown in the paper by Lee and Anderson (2007) “Validity and topic generality of a writing performance test”. The study was conducted using the data from the English Placement Test (EPT) administered at the University of Illinois at Urbana-Champaign in 1998-2003. The authors used TOEFL PBT and CBT scores (rescaled to PBT). As it turned out, a student whose total TOEFL PBT score was, for example, 610 (the exemption score), had only about 20% chance to be placed at the highest Level 3 (Figure 2 in Lee and Anderson (2007), p. 323). A student with the highest possible score on TOEFL still had about a 45% chance to be placed into “Level 2”, which required a completion of one mandatory ESL writing class, and even a 10% chance to be placed into “Level 1” with two required classes. All this indicates the absence of a clear TOEFL cutoff score which can be used in pre-screening. In the view of these results, the approach taken by MIT and Indiana University, where all non-native speakers of English are required to take ESL placement test regardless of their TOEFL scores, seems quite reasonable.

The relatively high probability of incorrect ESL placement seen in the study based on TOEFL PBT scores and cutoff values could even further change with the evolution of TOEFL and introduction of TOEFL iBT. On the one hand, the total TOEFL score, which now includes speaking section scores, could be even less reliable for predicting writing skills than the older TOEFL formats. On the other hand, with
more and more focus on the writing assignments in TOEFL, test takers spend more time on preparation for typical writing activities akin to those found in the EPT tests, better understand the required form and grading criteria, thus, may have better chances to meet benchmarks for exemption or placement at higher ESL levels.

This study explores the effectiveness of using TOEFL iBT scores for predicting placement into the ESL writing courses at UIUC. The following questions were addressed in this research:

1. Do TOEFL iBT total and subsection scores accurately predict placement into different levels of the ESL writing classes at the University of Illinois?

2. Does the correlation between TOEFL iBT and the EPT depend on the time between taking these two tests? If so, can this dependence affect the accuracy of ESL placement?

Methodology

Participants. Overall, over the period of Summer 2006 through Spring 2010, there was a significant increase in number of international students taking the EPT at the University of Illinois (see Appendix B). The majority of students took the EPT at the beginning of Fall semester. The demographic analysis of the data used for this research showed that there were 15% more male students who have taken EPT than female ones. There was a significant difference in the age of EPT test-takers. The median age of students at the moment of taking EPT was 22.6 years. The youngest test-takers were about 16 years old and the oldest ones were almost 53. The analysis of the EPT test-takers’ degrees expected after the graduation from the University showed that the majority were graduate students. In particular, the students in Master’s programs made up to 41.6% out of the total percentage of the EPT test-takers. Doctorate students made up to 19.1%. Roughly one fourth out of the total number of students (25.6%) were undergraduate students. Non-degree students were the least numerous category (13.5%).

During the 2006 through 2010 academic years, students from 77 countries took the EPT. The students from China make up to 29.5%, followed by South Korean (18.1%) and Taiwanese (15.6%) students. The fourth largest group of students (however, significantly less numerous than the above
mentioned ones) are the students from Turkey (3.6%), followed by the students from India (2.7%), Japan (2.4%), Thailand (2.3%), France (2%), and Malaysia (2%). The percentage of students from other countries taken separately is below 2%.

For this analysis, I will use only the TOEFL iBT data since it is the newest form of the test and the largest number of students (60.9%), whose score information was available in the obtained data sets, took this format of the test. Moreover, the paper-based format of the test is gradually phasing out and this fraction of the TOEFL iBT test-takers should only increase in the future.

**Instruments.** The TOEFL iBT scores and the written EPT results were analyzed in this research. The TOEFL iBT test measures the ability of students to use and understand English at the university level. It evaluates how well they combine listening, reading, speaking and writing skills to perform academic tasks (www.ets.org). The University of Illinois at Urbana-Champaign is one of the universities that use this test for admission decisions. It is the campus-wide requirement that all international students with the TOEFL iBT score 102 and below have to take the English Placement Test (EPT) before the beginning of their first semester at the University. The EPT consists of two parts: an oral interview and a written test. Completion of the ESL requirement is one of the conditions for graduation from the University of Illinois.

The written EPT is designed to assess students’ skills to write an academic essay. During this test, students read an article on a certain global issue, listen to a short lecture delivered by an EPT proctor, after that they discuss in groups a set of questions related to their essay question and then they write an essay referring to both the reading passage and the lecture. After the test, the essays are scored by trained EPT raters. In case of disagreement between two raters, an essay is given for rating to a third party. Based on the scores obtained, students are placed into certain EPT levels (Level 1, 2 or 3). Level 1 is the lowest placement level and Level 3 is the highest. The latter corresponds to a required writing course for undergraduate students, equivalent to Composition I requirement for native English speakers, or an optional advanced writing course for graduate students. The inter-rater reliability on the EPT is reported
in Lee and Anderson’s (2007) article. The reliability between two raters grading an essay is quite high and ranges from 0.75 to 0.95. More recent research work by Jang (2010) showed similar reliability estimates.

In this paper, I will present the findings based on the analysis of two types of data: self-reported and official. The self-reported information was collected at the same time as the EPT was administered. It included TOEFL scores of the EPT test-takers, date of taking TOEFL, major, visa status, and country of origin and individual university student ID numbers (UINs). These data were gathered since summer of 2006 till spring of 2010 by the EPT graduate assistants in the Linguistics Department which had the right of the legal custody of these data.

The official data were collected by the UIUC Office of Admissions and Records during the students’ application process. This data set contained TOEFL and other language proficiency test scores, native country, major, and department. All matching of the official data with UINs was done by DMI (Division of Management Information) at UIUC. After matching and sanitizing the data, the DMI staff sent the official data to the Department of Linguistics via FERPA-compliant data transfer.

**Procedures.** Before starting the data analysis, I had to solve an important methodological issue: What scores should be used for the analysis – official or self-reported? Dual data sources (self-reported and official) permitted cross-data reliability comparisons which made it possible to find out how honestly students were reporting their TOEFL scores. The first step was to exclude from the analysis the self-reported total TOEFL scores which were equal to a zero or which were impossible (e.g., TOEFL iBT total scores of more than 120). The scores of the students from the Economics Department and MA in Teaching ESL program were also excluded because, regardless of their TOEFL scores, they were required to take courses which were specifically designed to meet their needs.

Using the available self-reported and official data, I compiled a graph which illustrated how honest the EPT test-takers were when reporting their TOEFL scores (see Figure 1). I took students’ self-reported TOEFL scores and tried to match them with their official TOEFL scores. If a student took TOEFL more than once, I matched a self-reported score with an official score which was the closest to
the self-reported score of this student. I was able to match self-reported and official scores for the vast majority of the students (91.9%).

As Figure 3.1 shows, there is a small bell-shaped distribution observed around the main peak which probably indicates honest mistakes, since some students reported a lower score. Overall, the tendency is skewed in positive direction which means that some students exaggerate their TOEFL scores; however, it is a very small fraction of students (less than 3%) which significantly misrepresent their scores. Thus, I can conclude that self-reported scores can be used with confidence for the research.

![Figure 3.1](image)

**Figure 3.1.** Score difference between self-reported and official TOEFL scores. To build this graph, I matched the self-reported TOEFL scores with the official TOEFL scores of the EPT test-takers. If a student took TOEFL more than once, a self-reported score was matched with that official score which was the closest to the self-reported one for this student.

Another challenge was that 318 students (16%) in the EPT database took TOEFL more than once and I had several score reports for each of them. According to ETS, students can retake the test as many times as they want (www.ets.org). The question arises: What TOEFL scores to use for the analysis – all, the highest, or the most recent scores? University of Illinois at Urbana-Champaign does not have any clear policy regarding multiple TOEFL scores. When registering for the EPT, students do not get any
instructions about which TOEFL scores (all, highest, latest) they have to report. The variety of scores complicates the analysis since it is hard to determine which scores to use – all (including the scores of test repeaters), the most recent ones, or only the highest scores and which scores are the students typically report in the absence of clear instructions.

I ran several independent sample t-tests in order to compare the reliability of self-reported TOEFL scores in relation to the official ones. The findings of this analysis are presented in Table 3.1. The analysis shows that the average of self-reported TOEFL scores fall between “latest” and “highest” scores but the differences are not statistically significant (p=0.48 and p=0.60, respectively).

Table 3.1

Reliability of self-reported and official total TOEFL scores

<table>
<thead>
<tr>
<th>Type of data</th>
<th>N</th>
<th>Mean ±SD</th>
<th>p (t-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported</td>
<td>Self-reported</td>
<td>2363</td>
<td>93.21±10.31</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>2408</td>
<td>91.88±10.98</td>
</tr>
<tr>
<td>Official</td>
<td>Highest</td>
<td>1988</td>
<td>93.38±10.06</td>
</tr>
<tr>
<td></td>
<td>Latest</td>
<td>1988</td>
<td>92.98±10.31</td>
</tr>
</tbody>
</table>

Note. N = number of students. “Highest” – if a student took TOEFL more than once, only the highest available score was used in the analysis. “Latest” – only the last available TOEFL results were used. “All” – all available scores were included in the analysis.

All the subsequent data analysis was performed in two steps. All the figures presented in the paper were built using the self-reported TOEFL scores; however, the repetition of the procedures with the highest official TOEFL scores yielded identical results and they were omitted for clarity and brevity.

Step 1: Probability of placement into the ESL writing courses. The graphs showing the probability of placement into different levels of ESL writing courses at the UIUC were built using total and subsection scores of TOEFL iBT. To build this graph, I converted all available self-reported total TOEFL scores (N=2363) into the iBT scale. I sorted out the scores according to a bin size of 4 (a bin size of 2 is used for TOEFL separate section scores). For each bin, I calculated percentages of placement into each
writing EPT level. After that, I represented the values graphically. Using the available self-reported TOEFL iBT scores, I also ran one-way ANOVA and the post-hoc Scheffe test to explore the differences in the TOEFL iBT scores between the placement levels.

**Step2: Correlation between the Time of Taking TOEFL iBT and the English Placement Test.** To compile a graph showing the correlation between the times of taking these two tests, I went through a number of stages. First of all, I calculated the periods (in weeks) between TOEFL iBT and the EPT for each test-taker and sorted out students based on this parameter. Most international students who were required to take the EPT took it between 40th and 60th week since the date of taking TOEFL iBT. My next step was to divide the test-takers into the groups containing 20% of students (~281 test-takers). After that, I took the first group (from Student 1 to Student 281) and calculated correlation coefficient between their EPT and TOEFL scores for this group. Then I plotted this value as a line on a graph. The beginning of the line corresponded to the EPT-to-TOEFL time for Student 1 and the line ended with the EPT-to-TOEFL time for Student 281. Then I repeated the same procedure for the next group starting with Student 2 and ending with Student 282. I was doing this procedure until I reached the last member (1406) in the whole list. The graphs showing a distinct patterns of correlation between TOEFL iBT (total and subsection scores) and the EPT are discussed in detail in the Results section.

**Results**

**Probability of Placement into ESL Classes.** Differences in the TOEFL iBT scores between the placement levels were tested by one-way ANOVA and the post-hoc Scheffe test using SPSS 17. The data analysis shows that the average total TOEFL iBT scores are significantly different among all three placement levels (F (2, 1403) =33.48, p<0.001). The differences between all three placement groups in the writing section scores are also significant (F (2, 1403) = 30.33, p<0.001). The significance of differences in the total and writing sections scores is also confirmed by the Scheffe test (see Table 3.2). The differences in the speaking scores are more significant between the two highest levels - Level 2 and 3- (p<0.001); however, they are less significant between Level 1 and 2 (p=0.04). The differences in the
listening section scores are significant only between Level 1 and 2 \((p=0.002)\). The differences in reading are the least significant for the students placed at Level 2 and Level 3 \((p = 0.895)\). Based on the findings of this data analysis, I can conclude that total and writing section scores may be the most suitable for predicting ESL placement. Moreover, these two scores may also be a good reference for EPT raters in case of some disagreement between them concerning the level of placement. If Rater 1 thinks that a student should be exempt but Rater 2 believes that this student should be placed into Level 2, a TOEFL total or writing score can be used to finalize the placement level.

Table 3.2
\textit{Comparison of means of various TOEFL iBT scores for students grouped on the basis of their EPT placement level}

<table>
<thead>
<tr>
<th></th>
<th>Level 1 Mean ±SD</th>
<th>Level 2 Mean ±SD</th>
<th>Level 3 Mean ±SD</th>
<th>Level 1/2 (p)</th>
<th>Level 2/3 (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>89.1±12.9</td>
<td>92.4±8.2</td>
<td>95.4±8.5</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Writing</td>
<td>21.8±3.8</td>
<td>22.8±3.0</td>
<td>23.8±2.8</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Speaking</td>
<td>19.3±3.3</td>
<td>19.8±2.9</td>
<td>20.8±2.9</td>
<td>.040</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Listening</td>
<td>23.5±4.8</td>
<td>24.4±3.8</td>
<td>25.2±3.4</td>
<td>.002</td>
<td>.016</td>
</tr>
<tr>
<td>Reading</td>
<td>24.5±5.2</td>
<td>25.4±3.5</td>
<td>25.6±3.7</td>
<td>.001</td>
<td>.895</td>
</tr>
</tbody>
</table>

Level 1 \(N=346\); Level 2 \(N=769\); Level 3 \(N=291\); incomplete score reports were excluded; \(\alpha = 0.01\)

Figure 3.2 illustrates the ESL placement based the students’ total TOEFL iBT scores. A student who scored 102 (TOEFL cutoff score), has only about a 20% chance to be exempt from ESL classes, almost a 25% chance to get into Level 1 and about a 25% chance to be placed into Level 2. A surprising fact is that the students with low total TOEFL scores have quite a high chance to get into more advanced ESL classes (Level 3). For example, a student with a score of 80 (the established minimum for admission is 79, unless an exception is made) has a 10% chance to be exempt, a 40% chance to be placed into Level 1, and surprisingly almost 50% chance to be placed into Level 2. Furthermore, there are no total TOEFL cutoff scores which can be used for a reliable placement into any of the three ESL levels.
Figure 3.2. Probability of ESL placement in relation to total TOEFL scores. The frequency of total TOEFL scores was calculated with bin size of 4 for each level of ESL placement and later normalized. To build this graph, I converted all available self-reported total TOEFL scores into the iBT scale. I sorted out the scores according to a bin size of 4 (a bin size of 2 was used for TOEFL separate section scores). For each bin, I calculated percentages of placement into each writing EPT level.

Figure 3.3 illustrates the placement probability of students into different ESL levels based on the TOEFL iBT reading, listening, writing and speaking section scores. Similarly to the total TOEFL scores, high section scores do not necessarily result in the placement of students into the highest ESL level. For example, students with a score of 27-28 for reading have a 27% chance to be exempt, about a 30% chance to be placed to Level 2 of ESL classes, and about 43% to be placed to Level 1.
Figure 3.3. Probability of ESL placement in relation to TOEFL reading, listening, writing and speaking subscores. ▲ - Level 1, ● – Level 2, ■ – Level 3. The frequency of TOEFL section subscores was calculated with bin size of 2 for each level of ESL placement and later normalized.

Scores 25 to 26 are the most representative for listening section. An average person with such a score has about a 21% chance to be exempt, approximately a 51% chance to be placed to Level 2 classes, and a 28% chance to be placed to Level 1, and none of these three frequency values seem to be negligible.

Almost 400 students scored 22 out of 30 for writing which makes this score the most representative in the distribution of scores for this section. Students with this score have a 17% chance to be exempt, a 58% chance to be placed to Level 2, and about a 25% chance to be placed to Level 1.

A person with the most common score for speaking (20 out of 30) has a 20% chance to be exempt and a 20% chance to be placed to Level 1. This indicates that scores for speaking section cannot be a very accurate criterion for predicting ESL placement.

In summary, while the students with higher TOEFL iBT scores are typically placed into higher levels of ESL classes, there is no distinct pattern which would give me the evidence to claim that either
total or subsection scores can be used for reliably predicting ESL placement. There is no cutoff score which can be used as a reliable criterion for dividing students into three groups without significant misplacement. However, the results of one-way ANOVA and the post-hoc Scheffe indicate the total and writing TOEFL iBT scores may be used by the EPT raters as an additional reference when deciding which ESL level a student should be placed into.

**Correlation between the Time of Taking TOEFL iBT and the English Placement Test.** I also explored the effect of time on the correlation between TOEFL iBT and the EPT. Using the R statistical package 2.10.0, I calculated Pearson's correlation coefficients for groups of 281 students. Figure 3.4 illustrates a distinct tendency in the change of correlation over the whole period of time. The correlation is stronger if the time gap between taking TOEFL iBT and the EPT is short; however, the more time passes, the weaker the correlation becomes. It dramatically goes down until it reaches the lowest point around Week 50 and it even becomes negative between the 40th and 60th weeks. After Week 50, it, surprisingly, goes up. Based on this, I can conclude that overall TOEFL iBT scores which are about a year old cannot accurately predict English language proficiency demonstrated by the EPT results. Moreover, if we take into consideration the fact that most international students, who are required to take the EPT, take it exactly around Week 50, it will mean that TOEFL scores do not predict English writing proficiency for the majority of the EPT test-takers at the University of Illinois.
I also decided to check whether the same correlation pattern would be observed if we took section scores separately. It turns out that the pattern is preserved and the figures below (see Figure 3.5) demonstrate that no matter whether we use total TOEFL iBT scores or separate section scores for the analysis of the changes of the correlation over time, the pattern remains almost the same: the correlation dramatically decreases from Week 1 to Week 40 and is consistently low (even negative) around Week 50 but after Week 50 it goes up.
The least dramatic changes of the correlation are observed with speaking section scores: the changes of the correlation are in the range of 0.05 to 0.25. It is noteworthy that the correlation for speaking scores reaches its initial level (0.25) by the end of the second year; however, the correlation patterns of other three sections do not show signs of such a recovery. Out of all four sections, reading scores seem to be the only TOEFL scores which get old already after a year. Even though after the 50\textsuperscript{th} week correlation seems to go up (from negative to zero) it does not manage to get higher than that.

Writing and listening section scores seem to have the strongest correlation with the EPT when the TOEFL iBT scores are either newer than about 30 to 40 weeks or older than 50 weeks.

All in all, the most surprising finding is that the TOEFL iBT scores (with exception of reading section scores) older than a year better correlate with the EPT. Possibly, it is so because those who submitted their TOEFL scores older than 1 year had already studied in some English-speaking countries for a year and in a year for some reason they decided to change a program. They could have simply

---

**Figure 3.5.** Comparison of correlation patterns between time of taking TOEFL iBT and the EPT. L - a graph built using the listening section scores; R – reading section scores, S – speaking section scores, W – writing section scores.
resubmitted their old TOEFL scores which satisfied the requirement of their new program. The fact that they might have studied, for example, in the United States, helped them to maintain a good level of English and they also became familiar with the requirements to writing in the academic settings - that is why they did relatively well on the EPT.

Discussion

This study was motivated by the necessity to examine the possibilities of reevaluation of the current EPT policies due to the increased number of international students entering the University of Illinois and changes in the TOEFL format over the last 10 years.

Before starting my research, I needed first to solve an important methodological problem: Can self-reported TOEFL scores be used with confidence for the analysis? The comparison of the averages and individually matched scores from the official and self-reported data showed that the self-reported data could be used with confidence for the research. The fact that students were not given any specific instructions about what scores (highest or latest) to report in the EPT background questionnaire did not seem to affect any qualitative patterns.

The first research question addressed in this paper was: “Do TOEFL iBT total and subsection scores accurately predict placement into different levels of the ESL writing classes at the University of Illinois?”

The research evidence suggests that the students placed into higher ESL levels generally have higher TOEFL scores. The comparison of the means of TOEFL iBT total and subsection scores for all three placement levels showed that there were significant differences between all placement levels in the TOEFL writing and total scores ($p > 0.001$), and between Levels 2 and 3 in the speaking scores. The least significant differences were observed between Levels 2 and 3 in the reading section scores ($p = 0.895$).

Additionally, it seems that for any score that I choose as a cut off score, there will always be a significant fraction of misplaced students. I compiled several graphs illustrating the probability of placement into the ESL classes (see Figures 3.2 and 3.3). The graphs demonstrate that there is no
particular set of either total or section scores which could accurately predict the ESL placement since there is a significant number of cases when the students with very low TOEFL scores were exempt or vice versa – the students with relatively high scores were placed into the lower levels of the ESL writing classes.

The second research question was: “Does the correlation between TOEFL iBT and the EPT depend on the time between taking these two tests? If so, can this dependence affect the accuracy of ESL placement?” The findings suggest that there is a distinct correlation pattern between the time of taking TOEFL and the writing EPT: the correlation is the strongest when the time gap between the tests is short but later it dramatically decreases and reaches the lowest point (below zero) around Week 50. However, starting from the 50th week, it increases. The same trend is observed for total and sections TOEFL iBT scores (see Figures 3.4 and 3.5). My assumption is that a gradual drop of the correlation may be caused by Type 3 attrition - loss of L2 in an L1 environment (de Bot & Bert Weltens, 1985). It is possible that before taking TOEFL students have intensive preparation for the test but during the time between the application to the university and the time of taking the EPT they do not refresh their knowledge of the essay writing strategies and gradually lose their writing skills – that is why their scores do not correlate with the EPT. The increase of the correlation after Week 50 may be due to the reuse of the TOEFL scores after about year of the exposure to English in an English-speaking country. However, there may be additional factors explaining such correlation pattern, for example, country of origin or prior education of the EPT candidates, their major, and the level of studies (graduate, undergraduate or non-degree). The impact of these factors will be explored in future research.

Conclusion

Such factors as the increase of international students who choose to study in the United States (Institute of International Education, 2010) and current financial constraints may force some U.S. educational institutions that have ESL programs to consider using the results of standardized language proficiency tests for predicting ESL placement. This study is examining the possibility of using TOEFL
scores for ESL placement at a large mid-western university. The findings show that TOEFL may not be fully suitable for placement into the ESL writing courses. Due to absence of distinct patterns of probability of ESL placement, neither the ranges of TOEFL total and subsection scores, nor a specific cutoff (exemption) score could be established.

An unexpected finding is a strong effect of time lag. It turns out that a year is quite enough for students’ TOEFL scores to lose their relevance. However, surprisingly, correlation somewhat recovers after a year. On the one hand, this pattern gives us some insight into the changes of language skills over time. On the other hand, this finding may raise a number of provocative questions, for instance, should score users consider the age of pre-arrival test scores when they interpret them? Or should they set policies specifying a certain time period during which international students have to take TOEFL before arriving to the USA?

Based on the findings of this research, it is recommended for TOEFL and EPT score users:

a) to use the TOEFL iBT total or/and writing section scores as an additional reference when deciding which level a student should be placed into in case of some disagreement between the EPT raters. The results of one-way ANOVA and the post-hoc Scheffe show that the averages of the TOEFL iBT total and writing sections scores are significantly different among all three placement levels.

b) to avoid making conclusions about English proficiency of international students based on their TOEFL iBT scores that are about a year old. According to the pattern of correlation between the time of taking TOEFL iBT and the EPT, the most recent TOEFL iBT total and section scores seem to correlate better with the EPT than the scores which are about 50 weeks old.

Acknowledgements

I would like to thank Prof. Davidson from the University of Illinois at Urbana-Champaign for his help with getting the official TOEFL data. I am also grateful to Sun Joo Chung for providing the self-reported TOEFL data for my research. Also, I would like to thank two anonymous reviewers for their valuable comments.
References


http://web.mit.edu/fll/www/languages/ESLOptions.shtml#EET


Chapter 4

An argument against using standardized test scores for placement of international undergraduate students in English as a Second Language (ESL) courses

(Paper Three)

Abstract

Development and administration of institutional ESL placement tests require a great deal of financial and human resources. Due to a steady increase in the number of international students studying in the United States, some U.S. universities have started to consider using standardized test scores for ESL placement. The English Placement Test (EPT) is a locally administered ESL placement test at the University of Illinois at Urbana-Champaign (UIUC). This study examines the appropriateness of using pre-arrival SAT, ACT, and TOEFL iBT test scores as an alternative to the EPT for placement of international undergraduate students into one of the two levels of ESL writing courses at UIUC. Exploratory analysis shows that only the lowest SAT Reading and ACT English scores, and the highest TOEFL iBT total and Writing section scores can separate the students between the two placement courses. However, the number of undergraduate ESL students, who scored at the lowest and highest ends of each of these test scales, has been very low over the last six years (less than 5%). Thus, setting cutoff scores for such a small fraction of the ESL population may not be very practical. As far as the majority of the undergraduate ESL population is concerned, there is about a 40% chance that they may be misplaced if the placement decision is made solely on the standardized test scores.

Keywords

ESL placement, test purpose, test misuse, validity, impact

2 http://ltj.sagepub.com/content/30/4/467
The final, definitive version of this paper has been published in Language Testing, 30/4, October/2013 by SAGE Publications Ltd, All rights reserved. © [Kokhan, K].
Introduction

Each year increasing numbers of international undergraduate students choose to pursue their college degrees in the USA. This is especially true for students from China and South Korea since the economies of these countries are characterized by relative stability and education overseas is considered one of the ways to make an investment in their future (Heckman, 2003; Lemke, 2011; Varghese, 2008). During the 2010/11 academic year, the international student enrollment in the USA increased by five percent to 723,277 (Institute of International Education [IIE], 2011). According to the U.S. Department of Commerce, international students and their accompanying family members contributed more than $21 billion to the U.S. economy through spending money on tuition and living expenses (IIE, 2011). Many public schools are reaping financial benefits from growing numbers of international students who pay out-of-state tuition rates. However, they may also face a number of challenges which include, but are not limited to, a lack of teaching resources, state budget cuts and uncertainty in funding levels, and the development of resources, policies and regulations which would take into account the educational background, needs, and expectations of international undergraduate students. In some university contexts, the increase in the university-wide tuition revenue may not be directly connected to the funding of ESL programs.

Many schools in the United States are starting to look at how they can maintain the quality and attractiveness of education for their international students in the most economical way. One area in which it is possible for a school to save money is in its ESL placement services. Some universities have already set fixed fees for their institutional ESL placement tests (e.g., Indiana University – Purdue Indianapolis (IUPUI) charges $27 for its IUPUI ESL Screening test, and the University of California, Los Angeles (UCLA) charges $20 for its English as a Second Language Placement Examination). Another tempting option is to completely discontinue ESL placement tests and use pre-arrival test scores as a measure of students’ language proficiency and as criteria for placement into ESL courses. This paper presents a discussion of the possible challenges of such an option and provides some evidence that common college
admission tests, which are widely used for course placement of domestic undergraduate students, are not suitable for ESL placement.

In this paper, I will use the phrase “international undergraduate students” to refer to undergraduate ESL students, since they constitute the majority of the international student population at UIUC (more than 95%, according to International Student & Scholar Services at UIUC (2010)).

Use of Pre-arrival Test Scores for ESL Placement

Bachman and Palmer (1996) identified six major qualities that can be used to evaluate all aspects of test development and use. These qualities are reliability, validity, impact, practicality, authenticity, and interactiveness. All of these qualities are important and should be kept in balance; however, because of a number of factors, primarily financial and logistical, some score users may accord a disproportionate amount of weight to practicality. Practicality is defined as “the relationship between the resources that will be required in the design, development, and use of the test and the resources that will be available for these activities” (Bachman & Palmer, 1996, p. 36). The development and administration of ESL placement tests require a great deal of human and financial resources and it is not surprising that some score users may opt for a less costly and less time-consuming alternative.

There are a number of pre-arrival variables recorded under the profile of each undergraduate international student that could potentially be used for ESL placement instead of a placement test. Among such variables are students’ high school transcripts, which contain information about the courses taken in high school and the grades earned for each course. This information may definitely be helpful in making ESL placement decisions, but there are at least two reasons why, if used alone, this information may be insufficient. Firstly, the transcripts of international students may be difficult to interpret. For example, the content of some courses may not correspond to the content of similar courses taken by U.S. students, and the transcript course names may be inaccurately translated. In addition, different foreign educational systems may have different grading scales and the concept of GPA may not exist in some countries at all. Even if a large group of international students comes from the same country (e.g., China or South Korea),
their grades may also be difficult to compare because they studied in different schools, had different teachers and even possibly were taught according to different school programs. Secondly, there may be some cases in which international undergraduate students apply to U.S. universities several years after finishing high school or after completing part of a BA or BS degree in their home country. In both cases, high school transcripts alone would not provide an adequate picture of students’ ability in order to make ESL placement decisions.

Some students can also self-report their exit exam scores from previous institutions. This information may be useful for getting a general idea about the preparedness of these students for college studies, but it may not be very reliable for ESL placement since the information about the content of such exams and their level of difficulty may be hard to obtain and evaluate.

Using the list of the top 20 universities with the largest number of international students provided in a report of the Institute of International Education (November, 2011). Table 4.1 provides a short overview of the officially published admission and English proficiency requirements for international undergraduate admissions (see Table 1).

Almost all of the universities from the list explicitly require SAT or ACT scores for admission of undergraduate international students, with the exception of Penn State University, University Park, Northeastern University, and the University of Minnesota, Twin Cities. These universities consider those scores recommended for international admissions rather than required. Another option is to use TOEFL iBT scores, which are also widely accepted in many U.S. universities.
<table>
<thead>
<tr>
<th>University</th>
<th>Required tests: SAT/ACT</th>
<th>Minimum English proficiency requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. University of Southern California</td>
<td>SAT/ACT</td>
<td>TOEFL iBT(100, 20 in each section), or SAT Critical Reading (600), or ACT English (27)</td>
</tr>
<tr>
<td>2. University of Illinois – Urbana-Champaign</td>
<td>SAT/ACT</td>
<td>ACT English (25), or SAT Critical Reading (550), or TOEFL iBT (79)</td>
</tr>
<tr>
<td>3. New York University</td>
<td>SAT/ACT with Writing</td>
<td>TOEFL iBT (100) or instruction in English for 3 years + SAT/ACT scores</td>
</tr>
<tr>
<td>4. Purdue University – West Lafayette</td>
<td>SAT/ACT</td>
<td>TOEFL iBT (79; or 88 for Freshmen Engineering), or SAT Critical Reading (480), or ACT English (20)</td>
</tr>
<tr>
<td>5. Columbia University</td>
<td>SAT/ACT and 2 SAT Subject Tests</td>
<td>TOEFL iBT (100), or SAT Critical Reading or Writing (650)</td>
</tr>
<tr>
<td>6. University of California – LA</td>
<td>SAT/ACT with Writing</td>
<td>TOEFL iBT (83)</td>
</tr>
<tr>
<td>7. Ohio State University</td>
<td>ACT/SAT with Writing</td>
<td>TOEFL iBT(79), or SAT Critical Reading and Verbal score (500), or ACT English score (21)</td>
</tr>
<tr>
<td>8. University of Michigan – Ann Arbor</td>
<td>SAT/ACT with Writing</td>
<td>TOEFL iBT (88; with 23+ in L. &amp; R. and 21+ in S. &amp; W.), or SAT Critical Reading (600)</td>
</tr>
<tr>
<td>9. Michigan State University</td>
<td>Composite ACT/Composite SAT</td>
<td>TOEFL iBT (79, with subscores 17+), or SAT Critical Reading (480), or ACT English (18)</td>
</tr>
<tr>
<td>10. Harvard University</td>
<td>SAT/ACT with Writing and 2 Subject Tests</td>
<td>The SAT I and two SAT II Subject Tests. TOEFL and IELTS are not required</td>
</tr>
<tr>
<td>11. Indiana University – Bloomington</td>
<td>SAT/ACT</td>
<td>TOEFL iBT (79); SAT Critical Reading (500) or ACT English (21)</td>
</tr>
<tr>
<td>12. Boston University</td>
<td>SAT/ACT with Writing</td>
<td>TOEFL iBT; College of Communication, College of General Studies, School of Management: W.22; S.23; R.25; L. 21; Other Colleges: W.22; S. 23; R. 21; L. 18 TOEFL iBT (80)</td>
</tr>
<tr>
<td>13. University of Florida</td>
<td>SAT/ACT with Writing</td>
<td>SAT Writing (600-800) and ACT Writing (26 or higher) = credit for Freshman English</td>
</tr>
<tr>
<td>14. University of Texas – Austin</td>
<td>SAT/ACT with Writing</td>
<td>SAT Writing (600-800) and ACT Writing (26 or higher) = credit for Freshman English</td>
</tr>
<tr>
<td>15. Penn State University – University Park</td>
<td>Optional: SAT/ACT</td>
<td>TOEFL iBT (84 with min. subscores of R. 19, L. 19, S. 22 and W. 20), or SAT Critical Reading/ ACT equivalent (550)</td>
</tr>
<tr>
<td>16. Northeastern University</td>
<td>Optional: SAT/ACT</td>
<td>TOEFL iBT (79 or 70-78 + English course)</td>
</tr>
<tr>
<td>17. SUNY University at Buffalo</td>
<td>SAT (Critical Reading and Math)/ACT with Writing</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1
*Overview of the officially published test requirements for undergraduate admissions*
English proficiency requirements may be satisfied not only by total TOEFL scores but also by the score reports for certain sections of the SAT and ACT. The University of Texas – Austin even gives credit for Freshman English, if an international undergraduate student earned 600-800 on the SAT Writing or 26 and higher on the ACT Writing. The idea of using the SAT, ACT, or TOEFL iBT for the placement of international students in ESL courses may seem attractive since all international students provide such score reports; however, a question arises: How suitable are these tests for this purpose?

Tests are designed for certain purposes. When formulating a statement of test purpose, test developers usually determine the target population, set the target domains of language use, the range of knowledge, skills and abilities (Fulcher & Davidson, 2009). According to Alderson, Clapham and Wall (1995), using a test for purposes for which it was not intended is “one of the commonest problems in testing” (p.170). The reason for test misuse may lie in “the fundamental tension that is bound to exist within testing organizations” (Fulcher & Davidson, 2009, p. 125). Rather than restricting score interpretation to those test takers for which the test was developed, test providers may be tempted to gain extra revenue by increasing testing volume.

The practicality of standardized tests and their strong face validity is not sufficient for using these tests for new purposes. Alderson, Clapham and Wall (1995) suggest that a test may be valid for more than one purpose but “if it is to be used for any purpose, the validity of the use for that purpose needs to be established and demonstrated” (p. 170). Unfortunately, the problem of using standardized admission test scores for purposes other than originally intended is under-researched. There are very few studies that discuss the misuse of standardized tests for placement, and those that are available mainly focus on the analysis of the reflections of faculty regarding the consequences of using test scores for placement. For
example, Morante (1987) suggested that admission tests could not substitute for placement tests since they pursued different goals. According to him, admissions tests are designed to differentiate among better, more competent students, whereas the purpose of a placement test is to differentiate among less proficient students. Morante (1987) also pointed out that the biggest concern of the faculty was that admissions tests did not measure what they wanted their students to know or what they taught to their students.

Fox (2009) analyzed major trends regarding the perceptions and actions of teachers and student performance in response to a newly imposed policy which allowed L2 students to use external proficiency test scores (e.g., TOEFL and IELTS) for placement in English for Academic Purposes (EAP) courses in a Canadian university. The findings indicated that there were a number of negative consequences resulting from the use of standardized proficiency test scores. To begin with, students placed in the same level had very diverse abilities, and there was a huge difference between students at the top and students at the bottom of the EAP classes. In addition, the teachers reported that the new policy had a negative impact on teaching since it was harder to maintain a good teaching pace and meet the needs of all the students — stronger students were demotivated, and weaker ones were struggling. The results presented by Fox (2009) may not necessarily mean that using proficiency test scores for placement purposes in other contexts is a bad idea. It is possible that the test scores were not working for EAP placement because the curriculum for the EAP program required a different set of abilities, and getting a certain score on the TOEFL did not necessarily mean that a student with such a score could perform well in a particular EAP class.

The Present Study

The major motivation for this research is a necessity to explore the appropriateness of using standardized test scores for ESL placement at the University of Illinois at Urbana-Champaign (UIUC). A significant increase in the number of international students admitted to undergraduate programs at UIUC in combination with some state budget constraints pose a number of challenges to the University
administration as well as to the Department of Linguistics, which is in charge of the English Placement Test (EPT) and the ESL courses. The University administration is currently encouraging research that explores the possibility of using pre-arrival variables such as external test scores for placement in the ESL writing courses.

The majority of EPT candidates take this test before the beginning of each fall semester. Table 4.2 illustrates the increase of undergraduate EPT candidates with regard to the semester when the test was taken. A comparison of the number of students in Fall 2006 with that in Fall 2011 is the most striking: 49 undergraduate EPT test takers in Fall 2006 versus 398 test takers in Fall 2011.

According to the University policy, all international undergraduate students, except those who are from countries where English is the official language and medium of instruction\(^1\), are required to take the EPT no matter what score they have received on a language proficiency test. Based on the EPT results, undergraduate students are placed in either ESL 114 or ESL 115. Completion of the sequence ESL 114/ESL 115 satisfies the Composition I requirement (i.e., the University requirement of satisfactory proficiency in the use of written English) at UIUC. ESL 115 is the equivalent of RHET 105 “Principles of Composition” (the required course for English L1 students). English L2 students placed into this level can complete either of the two courses in order to fulfill their Composition I requirement.

Table 4.2

<table>
<thead>
<tr>
<th>Semester</th>
<th>N</th>
<th>Semester</th>
<th>N</th>
<th>Semester</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring’11</td>
<td>42</td>
<td>Summer’11</td>
<td>15</td>
<td>Fall’11</td>
<td>398</td>
</tr>
<tr>
<td>Spring’10</td>
<td>38</td>
<td>Summer’10</td>
<td>3</td>
<td>Fall’10</td>
<td>379</td>
</tr>
<tr>
<td>Spring’09</td>
<td>28</td>
<td>Summer’09</td>
<td>1</td>
<td>Fall’09</td>
<td>258</td>
</tr>
<tr>
<td>Spring’08</td>
<td>13</td>
<td>Summer’08</td>
<td>1</td>
<td>Fall’08</td>
<td>241</td>
</tr>
<tr>
<td>Spring’07</td>
<td>6</td>
<td>Summer’07</td>
<td>1</td>
<td>Fall’07</td>
<td>30</td>
</tr>
<tr>
<td>Spring’06</td>
<td>n/a</td>
<td>Summer’06</td>
<td>n/a</td>
<td>Fall’06</td>
<td>49</td>
</tr>
<tr>
<td>Total</td>
<td>1503</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: EPT date information is missing for 247 international undergraduate students.
A different practice of placement in English composition courses is established for undergraduate English L1 students. They also have to fulfill the Composition I requirement during their freshman year; however, they do not have to take an English Placement Test, as they are simply placed into the Rhetoric courses based on their highest achieved ACT English or SAT Critical Reading scores (see Table 4.3). For example, if students score between 21 and 31 on the ACT English, they can take only one four-credit course (RHET 105), which fulfills their Composition I requirement.

Table 4.3

*Rhetoric placement cutoff scores for undergraduate English L1 students (2011-2012)*

<table>
<thead>
<tr>
<th>ACT English score</th>
<th>Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-19</td>
<td>Fall RHET 101 and Spring RHET 102; Take RHET 100 both semesters</td>
</tr>
<tr>
<td>20</td>
<td>Fall RHET 103 and Spring RHET 104</td>
</tr>
<tr>
<td>21-31</td>
<td>RHET 105, or Fall CMN 111 and Spring CMN 112</td>
</tr>
<tr>
<td>32-36</td>
<td>Earned 4 hours of credit for RHET 105; Fulfilled Composition I requirement</td>
</tr>
</tbody>
</table>

Note: RHET 100 (Rhetoric Tutorial), RHET 101 (College Writing I), RHET 102 (College Writing II), RHET 103 (College Composition I), RHET 104 (College Composition II), RHET 105 (Principles of Composition), CMN 111 (Oral and Written Communication I), CMN 112 (Oral and Written Communication II).

The placement policy described above may have a number of advantages. To begin with, there is no need to hire and train proctors and raters for a placement test. In addition, course administration can know beforehand how many sections they need to open and how many instructors/teaching assistants they need to hire. Finally, they can early inform their newly admitted students about the specific composition courses that they will be required to take. The idea of such a placement policy seems to be very budget-friendly. The main question that has to be answered is “Will such a policy work for international undergraduate students?”

The external standardized test scores directly reported for each international student by test publishers seem to be the most suitable instrument for ESL placement since they are considered to be more objective and are proven to be quite reliable measures of the applicants’ skills and, most
importantly, they are easy to interpret. Since the profiles of almost all international undergraduate students at UIUC contain the information about their SAT, ACT, or TOEFL iBT scores, this paper addresses the following research questions:

1. Which of the three tests – the ACT, SAT, or TOEFL iBT – is the most accurate pre-arrival measure for placement of international undergraduate students into the ESL writing courses?

2. Can paired combinations of ACT, SAT, and TOEFL iBT scores be used for placement of international undergraduate students into the ESL writing courses?

Methodology

**Data.** The longitudinal data analyzed in this paper represent operational results from the period of Fall 2006 - Fall 2011. The participants (i.e., international undergraduate students) were not physically involved in the data collection process since the information about their test scores and other admission variables is the property of the University, and the data were provided for the research with the permission of the Institutional Review Board. The researcher received de-identified data.

The data were collected by the UIUC Office of Admissions and Records during the students’ application process. The data set contains the information about students’ SAT, ACT, and TOEFL iBT scores as well as other variables of interest (native country, major, department, etc.). Matching and sanitizing of the data were performed by the Division of Management Information (DMI) at UIUC. The information about students’ ESL placement results was provided by the Department of Linguistics and was further matched with the official score reports by the researcher using students’ University ID numbers (UINs), which were available in both data sets.

The data set contains the score reports for the following number of the international undergraduate students: the SAT (N= 1051), the ACT (N =1047), and the TOEFL iBT (N = 1257). The ratio of female to male EPT test takers is about 42% to 58%. The test takers from China constitute more than half of the international undergraduate student population (62.4%). They are followed by students from South Korea (15.7%), Taiwan (5.3%), Malaysia (4.4%), and other countries (12.2%).
For repeat test takers, the highest SAT, ACT and TOEFL iBT scores of the international undergraduate students were analyzed in this research. The reason for choosing the highest test scores rather than the most recent or initial ones is that the University of Illinois at Urbana-Champaign as well as a number of other U.S. universities (e.g., Columbia University, Ohio State University, Indiana University – Bloomington, and others) considers their applicants’ highest admission test scores for making undergraduate admission decisions.

**Instruments.** The SAT (the SAT Reasoning Test, formerly called the Scholastic Aptitude Test, and later the Scholastic Assessment Test) is a standardized college admission test. It is owned and developed by the College Board, a non-profit organization in the Unites States. The SAT consists of three sections: Critical Reading, Mathematics and Writing. For each section, a student can receive a score on a scale of 200-800. The total score for the test is calculated by summing the scores for the three sections. The Critical Reading (or formerly Verbal) section contains different types of questions including sentence completion and questions about short and long reading passages. The Mathematics (or Quantitative) section consists of three scored sections with multiple choice and grid-in questions (test takers write their answers inside a grid on the answer sheet). The Writing section consists of multiple choice questions (error identification, sentence improvement and paragraph improvement questions) and a short essay. The maximum essay score is 12 and it contributes 28% towards the total Writing section score (http://sat.collegeboard.org/home).

The ACT (American College Testing) is a standardized multiple-choice test for high school achievement and college admissions in the USA produced by ACT, Inc. It measures high school students’ capability to complete college-level course work. It consists of four subject tests: English, Mathematics, Reading, and Science. It also has an optional Writing Test, which measures students’ skills in planning and writing a short essay. Subject test scores range from 1 to 36. The composite ACT score is the average of all four sections. Students who take the Writing Test also get a separate writing score ranging from 2 to 12 and a combined English/Writing score ranging from 1 to 36. The Writing Test score does not affect the composite ACT score (http://www.actstudent.org/).
The TOEFL iBT (Test of English as a Foreign Language, Internet-based Test) measures the ability of non-native speakers of English to use and understand English at the university level. It consists of four sections: Reading, Listening, Speaking, and Writing. The scores for each section range from 0-30. A total score is formed by adding the scores from each section, and ranges from 0-120. The Reading section consists of 3-4 reading passages and questions about them. The Listening section consists of six passages followed by a number of questions about them. The Speaking section contains two independent and four integrated speaking tasks. In the independent tasks, the test takers are asked to express their opinion on familiar topic; in the integrated speaking task, they are asked to read a passage and listen to an audio recording before providing responses to questions. The Writing section consists of one independent essay task (an opinion essay) and one integrated essay task (the test takers have to read a short passage and listen to a short lecture and write a summary about the important points in the listening passage and explain how these relate to the key points of the reading passage) (www.ets.org).

The English Placement Test (EPT) administered at the University of Illinois at Urbana-Champaign consists of two parts: an oral interview and a written test. In this study, I analyzed the written EPT results. The written test is specifically developed to assess the ability of newly admitted international students to write an academic essay using the information from a reading passage and a lecture. The inter-rater reliability of the written EPT was reported in Lee and Anderson’s (2007) article. The reliability between two raters grading an essay was quite high and ranged from 0.75 to 0.95. Jang (2010) reported similar reliability estimates in her doctoral dissertation. The percentage of misplaced students based on instructors’ evaluations was relatively small - 10.71% (Cho, 2001). More information about the test is available in the EPT Bulletin (http://www.linguistics.illinois.edu/students/placement/documents/eptbulletin_jan2011.pdf) and in Kokhan (2012). Based on the EPT results, students can be placed into one of three ESL levels: ESL 113, 114, and 115. ESL 113 is considered an entry level course to the ESL 114/115 sequence. The number of the students placed in ESL 113 is very low in the database (less than 4% of students over the period Fall 2006 through 2011), and ESL 113 was not offered in 2010.
and 2011. Students who should be placed in this level are routinely admitted to ESL 114. Due to a very small sample size of the students placed in ESL 113, their test scores were not analyzed in this research.

**Procedures.** To explore the probability of placement of international undergraduate students in levels of the ESL writing courses, I analyzed ACT, SAT, and TOEFL iBT scores and the English Placement Test results using an exploratory data analysis approach. The justification for my choice of this particular set of variables is presented below.

According to Sawyer (1989), researchers and policy-makers have traditionally relied on correlation coefficients and associated tests of statistical significance in validating tests for use in college admission and placement. The most common approach is to use a prediction equation. However, the appropriateness of this method depends on the following assumption: “The course grades being predicted are valid measure of the academic skills learned in the course, rather than measures of random or irrelevant factors” (Sawyer, 1989, p. 11). I share the opinion of the author that if this assumption is not met, then it would be inappropriate to establish any placement policy on grade predictions and the results demonstrating the statistical relationship between placement variables (e.g., ACT/SAT scores and high school grades) and course grades (e.g., ESL course grades and GPAs) would be irrelevant to the validity argument. The reason why I chose the SAT, ACT, TOEFL iBT and ESL placement results as principal variables for the analysis rather than high school grades or college GPAs is because the above mentioned assumption does not seem to be fully true when applied to the ESL writing courses.

To begin with, student achievement in the ESL writing courses at UIUC is not measured by standardized tests or assignments as it is usually done in the sciences. In math classes, for example, it is common practice to give students a test (or multiple tests), and the grade for the test is based on whether a student provided the correct answer or not. However, when it comes to writing assignments, there is no correct answer expected. What often matters for ESL instructors in this context is whether their students are able to write well-structured and coherent papers and whether they can properly develop their arguments. Grammatical correctness (which is admittedly easier to measure) may not be the primary goal
of such courses. Additional aspects that are commonly reflected in the final grade of an ESL class are class attendance, participation, timely assignment submission, and so on.

Another reason why ESL course grades may not be appropriate variables for this analysis is that the ESL course administration at UIUC allows relative freedom to teaching assistants regarding the use of the instructional materials and the choice of topics. ESL teaching assistants are often full-time graduate students pursuing their MA in TESL degrees and are encouraged to develop their own lesson plans or adapt previously used teaching materials depending on the needs of their students. Moreover, different teaching assistants may differ by their teaching styles: Some may be strict in grading whereas others may add extra credit just for students’ effort in performing the task. All this may significantly affect the reliability of course grades. Sawyer (1989) suggested that if the reliability of course grades was low (e.g., 0.4) and the reliability of ACT scores was high (e.g., 0.9), one could not expect the correlation between ACT scores and course grades to exceed 0.6.

A university GPA as a variable commonly used in research on the predictive validity of standardized test scores (Aleamoni & Oboler, 1978; Bridgeman, McCamley-Jenkins & Ervin, 2000; Coyle & Pillow, 2008; Lenning & Maxey, 1973; Lins, Abell & Hutchins, 1966) may not be a suitable variable for the present research either. The reason for this is that the ESL students at UIUC are not typically separated according to their major or department into the ESL sections. GPAs of students from different departments may not be comparable: Some departments may have more relaxed grading guidelines, whereas others may have more stringent grading. Duke (1983) suggests that a GPA is made up of components that are not equivalent among students from different departments since the findings indicate that grades in some departments can also measure characteristics unrelated to students’ academic achievement. Moreover, researchers may overlook another important factor: Students need to complete a course in order to receive a grade for it. However, there are some students who choose to withdraw from a course or drop out of the university. Thus, using a GPA for this kind of research may not be appropriate.

My primary objective was to explore the ESL placement trends of international undergraduate students in relation to their SAT (Writing, Reading, Math, and Essay), ACT (Composite, English, Math
and combined English/Writing) and TOEFL iBT (total and section) scores and to check whether it would be possible to draw any definitive conclusions about the probability of ESL placement based on the observed patterns. I first ran a two-sample t-test to explore the differences between the means of the SAT, ACT, and TOEFL iBT scores of the students placed into the two ESL levels. Then, I built a series of graphs using OriginPro 8.5 software. The scores were sorted out according to the following bin sizes:

- The TOEFL iBT: a bin size of 2 for section scores and a bin size of 4 for total scores;
- The SAT: a bin size of 20 for Writing, Reading, and Math scores and a bin size of 1 for essay scores;
- The ACT: a bin size of 2 for section and composite scores.

In this research, a bin size is defined as a range of numerical test score values according to which students are grouped and according to which the probability of their placement into certain levels of ESL classes is calculated. Dots on the graph represent the centers of the bins.

![Graph](image)

**Figure 4.1.** Graphic illustration of a hypothetical ideal case of ESL placement using total TOEFL iBT scores.

I aimed to perform the analysis in the most practical, simple and illustrative way. An ideal picture of distribution of students in relation to their test scores is illustrated by Figure 4.1. My
assumption was that if certain test scores were better predictors of students’ ESL placement, there would be a clear divide between the placement levels in relation to their test scores. Also, students who scored very high on a certain standardized test would not be placed in the lower level ESL classes and vice versa.

To get a clearer idea about the differences in the actual and expected ESL placement results and to find out whether a combination of any pair of test scores would help in predicting ESL placement, I also performed a composite analysis of all available standardized test scores using MATLAB 7.11.0 (R2010b). For each possible pair of cutoff scores (e.g., TOEFL iBT total and TOEFL iBT Writing section scores), I compared the expected placement of undergraduate students with their actual placement in the ESL levels \(N = 800-1000\). Then I calculated the fractions of students who were correctly placed in relation to their paired test scores and plotted the percentages on a graph using different color shades. To avoid any conflation of the scores, I also subtracted the Writing TOEFL iBT score from the total TOEFL iBT score and built a graph demonstrating the percentage of correctly placed students in relation to TOEFL iBT total (excluding the Writing score) and TOEFL iBT Writing section scores. The findings of this analysis and their interpretations will be discussed in the Results section.

**Results**

The results of the two-sample \(t\) tests are summarized in Table 4.4. The findings indicate that the mean SAT Reading, ACT English, TOEFL iBT total, TOEFL iBT Speaking, and TOEFL iBT Writing section scores of the two groups of students (ESL 114 and ESL 115) are significantly different at the \(p < 0.01\) level. As for other sections, it is not surprising that the mean SAT Math and ACT Math scores of the two groups are not significantly different. There is no statistically significant difference between the mean ACT Composite scores of the ESL 114 and ESL 115 students, so these scores may not be suitable for predicting ESL placement either. The results of this analysis can be useful to get a general idea about the score differences between the two placement levels; however, an additional data analysis is needed to determine whether there is a test score or a range of scores that would separate the students into the two ESL levels.
Figure 4.2 illustrates the probability of placement into the two levels of the undergraduate ESL courses in relation to SAT scores. SAT Reading scores that range from about 350 to 400 seem to best separate all of students placed into the two levels: About 85% of students with such scores are placed in the lower placement level; however, if students score more than 400 on the SAT Reading, they have on average a 50% chance of placement into either of the two levels. There is no clear trend of ESL placement observed for other section scores either: Students with very high SAT scores can be placed in the lower ESL level, and students with very low SAT scores can be placed in the higher ESL level. For example, a student who scores 700 on the SAT Writing (note: according to the College Board report “SAT Percentile Rank,” 2011, 96% of SAT test takers earned a score below 700) has about a 40% chance of being placed in ESL 115 and a 60% chance of being placed in the lower level, ESL 114.

Table 4.4

<table>
<thead>
<tr>
<th>Test</th>
<th>ESL 114 (Mean ± SD)</th>
<th>ESL 115 (Mean ± SD)</th>
<th>p (t-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT Writing</td>
<td>571.11 ± 66.52</td>
<td>581.06 ± 67.70</td>
<td>0.021</td>
</tr>
<tr>
<td>SAT Reading</td>
<td>496.35 ± 51.43</td>
<td>507.51 ± 47.80</td>
<td>3.9·10^{-4}*</td>
</tr>
<tr>
<td>SAT Math</td>
<td>753.47 ± 49.00</td>
<td>749.29 ± 52.59</td>
<td>0.20</td>
</tr>
<tr>
<td>SAT Essay</td>
<td>8.02 ± 1.37</td>
<td>8.19 ± 1.40</td>
<td>0.076</td>
</tr>
<tr>
<td>ACT Composite</td>
<td>28.19 ± 1.97</td>
<td>28.29 ± 2.04</td>
<td>0.46</td>
</tr>
<tr>
<td>ACT Math</td>
<td>34.53 ± 1.75</td>
<td>34.32 ± 2.06</td>
<td>0.079</td>
</tr>
<tr>
<td>ACT English/Writing</td>
<td>25.99 ± 3.09</td>
<td>26.45 ± 3.13</td>
<td>0.024</td>
</tr>
<tr>
<td>ACT English</td>
<td>21.93 ± 2.69</td>
<td>22.44 ± 2.49</td>
<td>0.0017*</td>
</tr>
<tr>
<td>TOEFL iBT total</td>
<td>92.17 ± 7.75</td>
<td>94.43 ± 7.90</td>
<td>5.8·10^{-7}*</td>
</tr>
<tr>
<td>TOEFL iBT Reading</td>
<td>24.79 ± 3.40</td>
<td>25.03 ± 3.60</td>
<td>0.28</td>
</tr>
<tr>
<td>TOEFL iBT Listening</td>
<td>24.09 ± 3.69</td>
<td>24.55 ± 3.46</td>
<td>0.043</td>
</tr>
<tr>
<td>TOEFL iBT Speaking</td>
<td>20.59 ± 2.61</td>
<td>21.21 ± 2.50</td>
<td>1.4·10^{-4}*</td>
</tr>
<tr>
<td>TOEFL iBT Writing</td>
<td>23.04 ± 2.74</td>
<td>24.02 ± 2.77</td>
<td>3.0·10^{-8}*</td>
</tr>
</tbody>
</table>

Note: SAT/ESL 114 N = 493-600, SAT/ESL 115 N = 355-421; ACT/ESL 114 N = 576-611, ACT/ESL 115 N = 400-436; TOEFL iBT total/ESL 114 N = 732, TOEFL iBT total/ESL 115 N = 525; TOEFL iBT section/ESL 114 N = 579, TOEFL iBT section/ESL 115 N = 427-428. The number of score observations for different sections varies because the score reports of some students are incomplete. α = 0.01
Interestingly, very low ACT English scores (in the range of 15-17) similarly to very low SAT Reading scores (see Figure 4.2) seem to separate the students quite well between the two levels (only a 10% chance of placement in ESL 115); however, the number of students with such low scores is less than 2% in the database, and exemption of such a small fraction of students from local placement tests may not result in a significant reduction of expenses associated with the administration of these tests.

The lines illustrating the placement of students in relation to their ACT scores seem to be smoother and straighter (see Figure 4.3); however, there is no ACT score which would yield a clear divide between students placed into the two ESL levels. Let us take, for example, a Composite ACT score of 33 (according to the ACT report on the national ranks of test scores, 99% of students scored below 33): About 40% of undergraduate international students with this score are placed in ESL 114 and about 60% are placed in ESL 115 which is the equivalent of RHET 105 “Principles of Composition.”
Figure 4.3. Probability of ESL placement in relation to ACT scores.

Figure 3 also illustrates that the policy of placement into the Rhetoric courses that is currently applied to undergraduate English L1 students (see Table 4.3) may not work for international students. According to the Campus policy, undergraduate English L1 students who score 21-31 on the ACT English are required to take only one course (RHET 105), and this course will fulfill their Composition I requirement. This policy will be problematic if applied to international undergraduate students because, according to Figure 3, they have almost a 50/50 chance of being placed into either of the two ESL levels.

The placement distribution in relation to total TOEFL iBT scores (see Figure 4.4) looks more promising. Students who score 105 and higher on the TOEFL iBT are generally placed in the higher ESL level: 80% of the students with this score were placed in ESL 115, and only 20% were placed in ESL 114. A similar pattern is observed for TOEFL iBT Writing section scores (see Figure 4.5): 80% of the students who scored in the range of 20-30 were placed in ESL 115, and only 20% were placed in ESL 114. The comparison of the placement patterns in relation to all four section scores shows that students who score low on any of these sections are generally placed in the lowest ESL level. The patterns only differ for

88
very high section scores. For students who score high on Listening and Reading, the separation between the levels is not very clear. For example, a student, who scores 30 on Listening, has a 45% chance of being placed into ESL 115 and a 55% chance of being placed into ESL 114. However, when students get very high scores on Writing and Speaking, they may have a much lower chance of being placed into ESL 114. Unfortunately, all these observations may not be very helpful for formulation of new ESL policy since there are very few students who get extreme TOEFL iBT scores.

Figure 4.4. Probability of ESL placement in relation to total TOEFL iBT scores.

Figure 4.5. Probability of ESL placement in relation to TOEFL iBT section scores.
The results of the composite score analysis demonstrate that combinations of any possible pairs of test scores, if required for ESL placement, would not be effective either. Figure 4.6 illustrates the percentages of correctly placed students \((N = 1007)\) in relation to various pairs of TOEFL iBT total and section cutoff scores. The percentage of correct placement is coded according to the color bar to the right of the figure. Let us assume that the University wants to require a combination of these test scores for placing undergraduate students into the ESL writing courses. If we pick, for example, a TOEFL iBT score of 95 and a TOEFL iBT Writing section score of 23 as hypothetical cutoff scores for ESL placement, the color at their intersection will show that only about 56% of students would be correctly placed. Higher TOEFL iBT scores in combination with higher TOEFL iBT Writing section scores (e.g., TOEFL iBT total = 102 and TOEFL iBT Writing = 25) seem to yield the highest percentage of correctly placed students (about 62%), but the fraction of misplaced students is still significant (about 38%).

Figure 6 (left panel) also demonstrates that taking TOEFL iBT total scores alone (left edge of the graph) will result in the same placement as taking a combination of TOEFL iBT total and TOEFL iBT Writing section scores; for example, a TOEFL iBT total score of 102 is as good as a TOEFL iBT total score of 102 + a TOEFL iBT Writing score of 25. Using only TOEFL iBT Writing scores (bottom edge of the graph) would produce a slightly worse placement accuracy (about 60%). To avoid any conflation of the scores, I subtracted the TOEFL iBT Writing score from the total TOEFL iBT score and built a similar graph demonstrating the percentage of correctly placed students in relation to various pairs of TOEFL iBT Listening + Reading + Speaking and Writing section scores (Figure 6, right panel). The highest placement accuracy demonstrated in the graph is 62% (a combination of the range of scores 75-77 for Listening, Reading and Speaking and the range of scores 25-27 for Writing).
Figure 4.6. Percentage of correctly placed undergraduate students in relation to various pairs of TOEFL iBT total and Writing section cutoff scores.

A combination of ACT English and TOEFL iBT total scores and a combination of TOEFL iBT total and SAT Writing scores would be even less effective for ESL placement (see Figure 4.7) than paired TOEFL iBT total and TOEFL iBT Writing section scores. The percentage of correct placement is coded according to the color bar to the right of the figure. TOEFL iBT total $N = 1007$, ACT English $N = 912$, and SAT Writing $N = 894$. What the graphs show is that pairs of cutoff values (low ACT English scores (10-23) or relatively low SAT Writing scores (300-550) in combination with relatively high TOEFL iBT total scores (100-105)) result in about 62% of correctly placed students. Figure 4.7 also suggests that taking TOEFL iBT total scores alone will be as effective as taking them in combination with certain SAT Writing or ACT English scores.
Figure 4.7. Percentage of correctly placed undergraduate students in relation to various pairs of TOEFL iBT total and ACT English cutoff scores and in relation to TOEFL iBT total and SAT Writing scores.

The figures for other possible combinations of scores (e.g., TOEFL iBT total and ACT Composite, TOEFL iBT total and SAT Essay, TOEFL iBT total (excluding Writing) and SAT Writing, and so on) are very similar to Figure 4.7, so they are omitted for reasons of space. All in all, the findings of the composite score analysis show that setting pairs of cutoff scores would be ineffective for ESL placement.

Discussion

Despite the obvious practicality of using external standardized tests for ESL/EAP placement, some cautionary messages about the negative consequences of such practice are currently being seen in the research literature (di Gennaro, 2006; Fox, 2009). The present study is an attempt to enrich the ongoing discussion on the use of pre-arrival test scores for ESL placement by providing research-based evidence obtained within the context of the University of Illinois at Urbana-Champaign. The first research question addressed in this paper was “Which of the three tests – the ACT, SAT, or TOEFL iBT – is the most accurate pre-arrival measure for placement of international undergraduate students into the ESL writing courses?”
The following official score reports were available for the analysis: SAT (Writing, Reading, Math, and Essay), ACT (Composite, English, Math, and combined English/Writing) and TOEFL iBT (total and section) scores. Overall, none of the standardized tests, when used as a variable for analyzing placement probabilities, yielded the ideal pattern of ESL placement. A possible reason for this is that all of them are designed to serve different purposes. The ACT is an achievement test (i.e., a curriculum and standards-based test) measuring what a student has learned in school (www.actstudent.org). The SAT is more of an aptitude test, testing reasoning and verbal abilities and is designed to predict the performance of students in college (www.collegeboard.org). TOEFL iBT is an English proficiency test, measuring the ability of international students to use and understand English at the university level on a very broad spectrum (www.ets.org). All of these tests are primarily designed for making admission decisions, whereas placement tests “are designed to assess students’ level of language ability so that they can be placed in the appropriate course or class” (Alderson et al., 1995, p. 11). Such tests are commonly aligned with the syllabus of the program at a particular institution.

The results of the two-sample t-tests showed that the mean SAT Reading, ACT English, TOEFL iBT, Speaking and Writing section scores of the two levels were significantly different. However, $p$-values could not provide any specific information about the placement trends, as they simply indicated that the two placement groups differed in their mean test scores. Thus, an exploratory data analysis approach was used to examine the trends of ESL placement over the period of Fall 2006 - Fall 2011.

The data analysis revealed two interesting observations. Firstly, very low SAT Reading and ACT English scores seem to discriminate better between the two placement levels: Students who scored only 350-400 on the SAT Reading (less than 2% of the population), and students who scored 15-17 on the ACT English (also less than 2% of the population) have only about a 10% chance of being placed in the higher ESL level. Secondly, a good separation between the placement levels is observed in relation to high TOEFL iBT total and TOEFL iBT Writing section scores. The graph for TOEFL iBT total scores (see Figure 4.4) indicates that over the last six years, there has been a trend that students with a score of 105 and higher (3.2% of the EPT candidates who took the TOEFL iBT) are generally placed in the higher
ESL level, and only about 20% of students with these scores are placed in the lower level. A very similar pattern of ESL placement is observed in relation to high TOEFL iBT Writing section scores. The finding that there is a contrast between the ESL students with very low and very high standardized test scores seems to be logically expected, but establishing a placement policy based on this observation will probably not help the university to save a substantial amount of financial resources due to a very small number of students with such extreme scores.

The second research question was “Can paired combinations of ACT, SAT, and TOEFL iBT scores be used for placement of international undergraduate students into the ESL writing courses?” The findings of the composite score analysis essentially corroborate the results of the analysis of the ESL placement trends. Using any possible pairs of TOEFL iBT, ACT, and SAT scores for placing undergraduate students into the ESL classes is not appropriate for placement. As illustrated by Figures 4.6 and 4.7, even the most favorable combination of scores (where the fraction of correctly placed students is the highest – about 62%) will still produce a significant percentage of misplaced students (about 38%). Overall, only a marginal improvement is observed (<5%) in the percentage of correctly placed students over placing them all in the lower ESL level.

**Conclusion**

The results of this research indicate that only students’ lowest ACT English and SAT Reading scores and their highest TOEFL iBT total and TOEFL iBT Writing section scores would work for placement of international undergraduate students into ESL classes at the University considered in this study. This finding prompts two very important questions that universities need to answer before establishing any ESL placement policy based on standardized test scores. First, do they consider 80% to be acceptable placement accuracy? The results show that there is about a 20% chance that students with low TOEFL iBT total and TOEFL iBT Writing scores may be placed into the highest ESL level. Additional on-site testing may still be needed to prevent misplacement. Second, will there be any financial gains from setting cutoff scores for such a small fraction of incoming international
undergraduate students? The number of students with extreme standardized test scores has been very low (less than 5%) over the period of Fall 2006 - Fall 2011. Reduction of the EPT test taker population by 5% will not significantly save human and financial resources spent on the administration of this test.

A decision to rely on standardized proficiency tests for ESL placement may largely depend on the situation, for example, enrollment rates, availability of human and financial resources, and so on. However, I believe that the quality of the education should be the top priority for any educational institution. Score users should realize that neither language proficiency tests, nor language sections of aptitude and achievement tests are designed for ESL placement purposes. None of them is based on any particular ESL program. If these tests are used for ESL placement, the impact of the decisions that are based on them may be quite adverse (Fox, 2009). Crusan (2002) pointed out that universities must ask themselves whether ease of administration and low cost are “enough to justify financial and academic hardships to students” (p. 26). Because some U.S. universities charge for ESL courses, a requirement to take an additional ESL course will result in extra financial expenses for students. Moreover, if misplaced, students may fail to significantly improve their English skills. In a class of less proficient ESL students, they will get bored and frustrated about low progress in English. In a class of more proficient ones, students will most likely lag behind. Instructors and teaching assistants may feel uncomfortable teaching in an environment of dissatisfied students who do not have any motivation for attending their classes and performing all tasks and homework assignments. Finally, being completely unfamiliar with conventions of academic writing, international undergraduate students, who do need some help with their English but are exempt from ESL courses based on their standardized test scores, may face many language-related challenges at the university.

Some U.S. universities set specific requirements for combinations of scores to prevent a compensatory approach, according to which high scores for some subsections can compensate on low scores for other subsections. For example, the University of Southern California requires a minimum TOEFL iBT total score of 100 and a minimum TOEFL iBT section score of 20). The results of the composite score analysis indicate that such an approach will not be better than using a single test score for
ESL placement purposes since the fraction of misplaced students is also quite high when score combinations are used.

In conclusion, this study suggests standardized test scores do not work for ESL placement for the majority of international undergraduate students at the University considered here: Even though the mean test scores of the students from the two placement levels are statistically different in some cases, there is about a 40% probability that the majority of them may be placed in the wrong level if this decision is made solely on the standardized test scores. This observation adds weight to the importance of an institutionally developed and administered placement instrument such as the ESL placement test, which is aligned with the ESL curriculum so that it can more efficiently distribute ESL students between the placement levels.

Acknowledgements

I would like to thank Prof. Davidson from the University of Illinois at Urbana-Champaign for his help with getting the official data for my research. I am also grateful to anonymous reviewers for their valuable comments.

Notes

1. The countries where English is the official language and medium of instruction - Australia, Bahamas, Barbados, Belize, Botswana, Canada (except Québec), Fiji, Gambia, Ghana, Guyana, Ireland, Jamaica, Kenya, Lesotho, Liberia, Malta, Mauritius, New Zealand, Nigeria, Papua New Guinea, Philippines, South Africa, Swaziland, Tanzania, Trinidad & Tobago, Uganda, United Kingdom, United States (except Puerto Rico), Zambia, Zimbabwe (Prof. F. Davidson, personal communication, March 7, 2012)

2. IELTS scores are also accepted by UIUC but the number of the undergraduate international students who took this test is not substantial in the data set, so they are not analyzed in this research.
References


Arizona State University. (n.d.) Freshman admission requirements and timelines. Retrieved from https://students.asu.edu/freshman/requirements


Indiana University - Purdue University Indianapolis. (n.d.). Student information: Placement testing at IUPUI. Retrieved from http://tc.iupui.edu/testing/students/


Chapter 5

TOEFL iBT repeaters: How to interpret multiple score reports for ESL placement?

(Paper Four)

Kateryna Kokhan
University of Illinois at Urbana-Champaign, USA

Cary Lin
University of Illinois at Urbana-Champaign, USA

Abstract

The vast majority of U.S. universities accept TOEFL scores for admission and placement into ESL classes. A significant number of candidates choose to repeat the test hoping to get higher results. Unfortunately, there is no published research examining students’ multiple TOEFL score reports. This paper addresses the issue of interpretation and use of multiple scores of TOEFL iBT repeaters for making ESL placement decisions in the context of the University of Illinois at Urbana-Champaign (UIUC). The main research question considered in our study was: Which TOEFL iBT scores (official highest, most recent, average or self-reported scores) are the best predictors of ESL placement? The findings indicate that the self-reported and the highest TOEFL iBT total scores have the strongest association with the ESL placement results. The self-reported and the highest scores also demonstrate the highest classification efficiency in predicting ESL placement of TOEFL iBT repeaters. The results and implications of the study are discussed.

Keywords
Test repeaters, ESL placement, TOEFL, predictive validity

Background

Previous research on test repeaters. Repeaters are examinees who repeated the exam at least once, regardless of the time interval between the tests and/or the number of repeated tests (Yang, Bontya
There were several publications examining the interpretation of repeaters’ scores (Boldt, 1986; Wightman, 1990; Yang, Bontya, & Moses, 2011; Zhao, Oppler, Dunleavy, & Kroopnick, 2010). Although most of these studies were conducted to explore multiple test scores other than TOEFL, some of the findings are relevant to the present research and will be discussed below.

Boldt, Centra and Courtney (1986) were interested in finding the best way for treating multiple SAT scores of individuals in college admission. They investigated five methods of treating multiple test scores: simple average, weighted average (giving all the scores, except the latest, the same weight, and giving all the scores, except the highest, the same weight), latest and highest. The combinations of variables and the methods of treating multiple scores were evaluated in regression equations developed using one-time testers. The results showed that the simple arithmetical average was as valid as the weighted average. Overall, all treatments of multiple scores resulted in underprediction of actual college grades; however, the highest scores were slightly better predictors. For reasons of validity and simplicity, the authors recommended using an average of scores because the arithmetic average had the highest correlation with college GPA. However, as far as the effects on prediction are concerned, the average turned out to be the worst.

Wightman (1990) conducted a study to determine what Law School Admission Test (LSAT) scores (initial, latest or highest) could most accurately predict the subsequent performance of test takers in law schools. The author used a least-squares regression analysis of different LSAT scores of test repeaters on first-year average in law school. The researcher found that, in general, repeaters tended to get lower LSAT scores than one-time test takers regardless of whether initial, latest or highest scores were considered. Another finding was that repeaters and one-time test takers performed comparably in their undergraduate academic work; however, one-time test takers earned generally higher GPAs in law school than LSAT repeaters. The author concluded that the average LSAT score was the best predictor for the majority of law schools. This finding was later confirmed by Thornton, Stilwell and Reese (2006) and Thornton Sweeney, Marcus and Reese (2010), who examined the validity of the most recent, highest and average LSAT scores in terms of predicting first-year law school grades. Wightman (1990) also pointed
out that the second best predictor was the initial test score. However, this finding contradicted the results of Pitcher (1977), who found that the initial score produced “the most deviant predicted means” and suggested that the use of initial test scores would be unfair to applicants. Wightman (1990) assumed that the differences in the findings between the results of Pitcher (1977) and her study may be due to intensive test preparation or coaching among most recent repeaters, which could have inflated the most recent scores of test repeaters, making them less predictive of law school performance.

Zhao, Oppler, Dunleavy and Kroopnick (2010) investigated the validity of four approaches (average, most recent, highest-within-administration, and highest-across-administration) to using repeaters’ Medical College Admission Test (MCAT) scores to predict Step 1 the United States Medical Licensing Examination (USMLE) scores. The findings suggested that the best approach for computing repeaters’ score for admission purposes was to take the average across all administrations. The authors pointed out that: “Students with the same MCAT average score are expected to perform the same on Step 1 exam regardless of the number of attempts to achieve that average. As such, admissions committees can have confidence that MCAT total scores computed by averaging across all administrations are comparable regardless of the number of times a student took the MCAT exam” (Zhao et al., 2010, p. S67). The authors warned against using the most recent, highest-within-administration and highest-across-administration score approaches because they were associated with a higher prediction error. These approaches can result in a so-called ‘overprediction’ of repeaters’ Step 1 total scores as repeaters were expected to perform less well on Step 1 than non-repeaters and repeaters who were tested fewer times but had the same MCAT total score computed using one of these approaches.

Yang, Bontya and Moses (2011) investigated repeater effects on score equating. They analyzed a significant number of self-reported test scores across a wide range of administrations from a graduate admissions examination that was administered in a non-English language. The authors reported that repeater scores across testing occasions were fairly stable. However, there were large scale score gains/losses for a broad range of administrations. The test-retest correlation was 0.74 for Verbal and 0.72 for Quantitative section for the overall repeater group. The results also suggested that high-performing
test takers might not improve their scores by retesting as would the low-performing examinees. The
researchers admitted that further research would be needed in order to better understand repeater
performance patterns while taking into consideration the demographic information.

**Previous research on TOEFL iBT repeaters.** To our knowledge, an ETS research report by
Zhang (2008) is the only published paper analyzing the scores of TOEFL iBT repeaters. Zhang’s
assumption was that under normal circumstances, the scores of test takers may vary insignificantly on the
condition that the tests were repeatedly taken within a very short period of time and no intensive training
occurred during this time frame. The repeater sample (N=12,300) analyzed in the study was a self-
selected sample of candidates who took one TOEFL iBT test in a month after taking the other. The author
did not mention though how she knew that no TOEFL preparation had occurred prior to the second test
session.

Zhang (2008) found that there was a good correlation between individual sections of TOEFL iBT
(0.78 for reading, 0.77 for listening, 0.84 for speaking, and 0.77 for writing), and between the total scores
of two tests (0.91). She concluded that small changes were observed in the test scores between the first
and the second test of the repeaters. The distribution of score changes resembled a symmetrical bell-
shaped distribution and the total scores only slightly increased (Δ=3.74).

Even though the paper by Zhang (2008) was not investigating the relationship between repeaters’
test scores and their ESL placement results, there are some data reported by the author that have drawn
our attention. The standard deviation of the average score change as reported in the paper was quite high
(9.50). Most universities usually have a narrow range of scores (about 20 points, see Appendices A & B)
between the minimum for admission and the minimum for exemption from additional on-campus testing,
so a large uncertainty can matter and ESL students may be misplaced.

Unlike the repeaters in Zhang (2008), the TOEFL iBT repeater population considered in the
present study represents a typical international student population in a large public university in the
United States and is not limited to short-term repeaters only.
Current Trends in the ESL Placement Policy with Respect to TOEFL iBT Test Repeaters

Each year increasing numbers of international students choose to study in the USA. According to the Institute of International Education (November, 2011), the international student enrollment in the United States increased by five per cent over the period of 2010-2011 academic years. Using the list of top 20 universities with the largest number of international students provided in the report, we made a short overview of the officially published ESL placement policies (see Appendix C). Overall, the universities differ in terms of TOEFL iBT cutoff score requirements for ESL placement: some campuses set higher cutoff score requirements (e.g., the Ohio State University), others set the scores considerably lower (e.g., Michigan State University), and there are those that do not have any specific cutoff score policies at all. The majority of the universities with the largest number of international students have local ESL placement tests and all of them offer ESL courses or their equivalents. A short overview of the TOEFL iBT score requirements for admission as published on the official websites of these universities (see Appendix C) shows that the universities are very specific about the TOEFL iBT score requirements for admission but the issue of interpreting multiple test scores does not seem to have been addressed by them. Some universities do not have any published guidelines for interpreting the scores of TOEFL iBT repeaters but they do have a policy regarding the scores of ACT and SAT repeaters: only the highest scores are considered when application is reviewed (e.g., the University of Texas – Austin, and the University of Illinois at Urbana-Champaign). Despite an abundance of the institutionally established means to help new international students who meet the admission requirements but still struggle to meet the English language demands of their programs of study, there are no published policies at most universities which would help to interpret multiple score reports of TOEFL iBT repeaters for ESL placement.

The Present Study

This study is motivated by the current need to find the most reliable and cost-effective ESL placement policy at the University of Illinois at Urbana-Champaign due to a significant increase in the
number of international students. The main research question addressed in the paper is: Which scores (official highest, most recent, average or self-reported scores) are the best predictors of ESL placement? Additional research questions considered in the study include:

1. Who are TOEFL iBT repeaters at the University of Illinois at Urbana-Champaign?
2. Is there a significant correlation between the number of TOEFL iBT tests repeated and the level of achievement on the TOEFL iBT and the English Placement Test (EPT)?
3. Is there a significant difference among the highest, most recent, average and self-reported scores of TOEFL iBT repeaters?

Methodology

Data. The data analyzed in this paper were collected over the period of Fall 2006 - Fall 2011. The findings reported in the paper are based on the analysis of two types of data: self-reported and official. The self-reported information was collected at the time when students were registering for the EPT. The self-reported data include the TOEFL iBT scores of the EPT test takers, their expected degree, gender, country of origin, and individual university student ID numbers (UINs).

The official data were collected by the UIUC Office of Admissions and Records during the students’ application process. The data set contains information about students’ TOEFL iBT scores, their native country, major, and department. Matching of the official data with UINs was performed by DMI (Division of Management Information) at UIUC. After matching and sanitizing the data, the DMI staff sent the official data to the Department of Linguistics via FERPA-compliant data transfer.

Overall, the data set contains the TOEFL iBT scores of 3032 students wherein the number of TOEFL iBT repeaters is 474 (15.6%). The score reports of TOEFL iBT repeaters that were incomplete and the TOEFL iBT scores of the students who were placed into specifically developed ESL courses regardless of their placement level (students from the Economics Department and MA in Teaching ESL program) were excluded from the analysis. Thus, the total number of TOEFL iBT repeaters analyzed in
this study was 396. A detailed demographic analysis of TOEFL iBT repeaters will be presented in the Results section of the paper.

**Instruments.** TOEFL iBT (Test of English as a Foreign Language, Internet-based Test) measures the ability of non-native speakers of English to use and understand English at the university level. It consists of four sections: Reading, Listening, Speaking, and Writing. The scaled scores of each section range from 0-30. A total score is formed by adding the scores of each section, and ranges from 0-120. The Reading section consists of 3-4 reading passages and questions about them. The Listening section consists of six passages followed by questions about them. The Speaking section contains two independent and four integrated speaking tasks. The Writing section consists of one independent essay task and one integrated essay task (for details, see http://www.ets.org/toefl/ibt/about).

The English Placement Test (EPT) consists of two parts: an oral interview and a written test. It is usually administered several days before the beginning of students’ first semester at UIUC. Completion of the ESL requirement is one of the conditions for non-native speakers of English for graduation from the University. Based on the results of the written and oral parts of the EPT, students are placed into the writing and speaking ESL courses respectively. In this study, we are focusing only on students’ written EPT results.

The written EPT test is a paper-and-pencil placement test specifically developed to assess students’ ability to write an academic essay using a combination of skills (reading, listening, speaking and writing). On a test day, EPT test takers are asked to write an argumentative essay using the information from both a lecture delivered by a trained EPT proctor and a reading passage from a journal article. The structure of the written EPT is very close to the integrated writing assignments in the TOEFL iBT. The main difference though is the presence of a group discussion in which EPT test takers participate right after they finish the first draft of an essay. Based on the scores on this test, students are placed into three levels of the ESL writing classes (Level 1, 2 or 3). Level 1 is the lowest placement level and Level 3 is the highest. The latter corresponds to a required writing course for undergraduate students (equivalent to the
Composition I requirement for domestic undergraduate students), or to an optional advanced writing course for graduate students.

The EPT has a long history of validity research reflected in the master’s theses, doctoral dissertations and publications of the graduate students from UIUC who helped in developing, exploring and improving this test. Once a year, the EPT administration organizes rater trainings for new EPT raters. In order to avoid any rater effects acquired over time - so called “rater drift” (Hoskens & Wilson, 2001; Wilson & Case, 2000), experienced raters regularly participate in recalibration sessions. The inter-rater reliability on the EPT was reported in Lee and Anderson’s (2007) article. The reliability between two raters grading an essay was quite high and ranged from 0.75 to 0.95. Jang (2010) reported similar reliability estimates for the EPT. The percentage of misplaced students based on instructors’ evaluations is reported to be relatively small - 10.71% (Cho, 2001). In addition, the majority of test takers believe that the EPT adequately reflects their English academic writing ability in that source-based writing is required in most academic courses (Cho, 2001). Perhaps one of the most appealing content-relevance features is the tight resemblance between the EPT tasks (e.g., information gathering, multiple drafts, group discussion, etc.) and classroom activities typically expected in the ESL curriculum on which the EPT is based. More information about the test can be found in the EPT Bulletin (http://www.linguistics.illinois.edu/students/placement/) and in Kokhan (2012).

**Procedures.**

*Description of the TOEFL iBT repeater population at UIUC.* The TOEFL iBT repeater population was described using the following variables: gender, degree, major, and native country. The results of these analyses were also compared to those of non-repeaters.

*Relationship between the number of tests repeated and the level of achievement on the TOEFL iBT and on the EPT.* The Cochran-Mantel-Haenszel statistics was used for the analysis. It takes into account the order information of the number of tests repeated and the levels of achievement on the test in performing hypothesis testing about the correlation between the two ordinal variables.
Analysis of the differences among the highest, most recent and average scores of TOEFL iBT repeaters. A within-subject ANOVA test was conducted to see whether the highest, most recent, average and self-reported TOEFL iBT scores differ from each other. A follow-up Tukey test was used to locate sources of pair-wise differences among the four types of TOEFL iBT scores of test repeaters.

Relationship between ESL writing placement levels and TOEFL iBT scores (highest, most recent, average and self-reported). We examined the placement trends of TOEFL iBT repeaters over the period of Fall 2006 – Fall 2011. We first sorted out the total highest, most recent, average, and self-reported TOEFL iBT scores according to a bin size of 4. For each bin, we calculated percentages of placement into each ESL level. After that, we represented the values graphically. In addition, the empirical placement trends in relation to TOEFL iBT scores were compared with model-based placement trends to see whether the observed trends triangulate with the model-based approach. We employed a regression model for the data analysis since this approach was widely used in previous research on test repeaters (Boldt, 1986; Thornton et al., 2006; Thornton et al., 2010; Zhao, 2010). More specifically, we performed the cumulative logistic regression in analyzing the relationship between ordinal responses and continuous predictors. In this study, it is used to model the probability of test takers being placed in the three ESL writing placement levels with respect to their TOEFL iBT scores. It should be noted that the association between the EPT results and TOEFL iBT scores would be masked by the limited range of ESL placement levels (i.e., three levels only) if the analyses were done in ordinary linear regression. Nonetheless, due to the ordinal nature of ESL placement levels and small number of ESL levels (only three), the cumulative logistic regression is appropriate for analyzing the association between ESL placement and TOEFL iBT scores as long as adequate variability in TOEFL iBT scores is observed (see Appendix D) for each ESL placement level.

Results

Description of the TOEFL iBT repeater population at UIUC. TOEFL iBT repeaters constitute 15.6 % of the entire EPT test taker population over the Fall’06-Fall’11 period. The scores of 396 TOEFL
iBT repeaters were analyzed using SAS ver. 9.3 and R software ver. 2.13.2. The majority of TOEFL iBT repeaters at UIUC took the test twice (81.06%) and a sizeable fraction of candidates took it three times (14.90%). Among the test repeaters, 46.21% are undergraduate students, 34.60% are students pursuing a Master’s degree, 17.42% are doctoral students, and 1.77% are exchange (non-degree) students. The ratio of female to male is 44.44% to 55.56%. Test takers from China constitute more than half of test repeaters population (53.54%). They are followed by students from Taiwan (12.63%), South Korea (10.61%), Indonesia (3.18%), Turkey (2.53%), Thailand (2.02%) and the remaining nationalities are under 2% each.

In order to determine how representative TOEFL iBT repeaters are of the entire EPT test taker population, we examined the differences between repeaters and non-repeaters (test takers who took TOEFL iBT only once) with respect to their gender, expected university degree, native country and placement results. The proportions of male and female candidates are very similar in both groups: female repeaters (44.4%) versus female non-repeaters (45.1%) and male repeaters (55.6%) versus male non-repeaters (54.7%). The comparison of TOEFL iBT repeaters and non-repeaters according to the expected degree shows that non-degree students are least likely to repeat the test: the fraction of non-degree students in the repeater population is only 1.8% whereas their fraction in the non-repeaters group is 8.7%. The fractions of PhD and undergraduate (UG) students in the repeater group are somewhat higher than those in the non-repeaters which means that they are more likely to repeat the test: UG - 46.2% versus 43.3% and PhD – 17.1% versus 12.0%; however, the fraction of Master’s students in the repeater group is lower than in the non-repeaters group: 34.6% versus 36%. The analysis of the distribution of TOEFL iBT repeaters and non-repeaters according to their native countries shows that the students from India are least likely to repeat TOEFL iBT (only 3 repeaters out of the total 89); they are followed by the students from Japan (3 repeaters out of 42) (see Appendix D for further information). Although the fraction of the students from Indonesia in the entire population of test takers is not high (1%), their percentage in the repeaters group is much higher than that in the non-repeaters group (3.28% versus 0.6%). A tendency for repeating TOEFL iBT is also observed among students from China, Turkey, Iran,
Columbia and Kazakhstan. It is difficult to draw any conclusions about the differences in the fractions of the students from other countries due to their low number in the database.

More than half of the test repeaters in this study were placed into Level 2 which corresponds to an advanced academic ESL writing course. 23.23% of test repeaters were placed into Level 1 (an introductory academic writing course) and 20.96% were placed into Level 3 (exemption and an optional writing course for graduate students or a required course equivalent of the Composition I requirement for undergraduate students). It is noteworthy that the repeaters placed into the three levels differ by their average TOEFL iBT scores: the students placed to higher levels have generally higher average TOEFL iBT scores (see Appendix D). The comparison of the average TOEFL iBT scores of repeaters and non-repeaters (see Appendix D) suggests that the students with higher scores seem to be less likely to repeat the test.

Relationship between the number of TOEFL iBT tests repeated and the level of achievement on the TOEFL iBT. The next step was to examine the correlation between the number of TOEFL iBT taken repeatedly and the levels of achievement on the test. Using R software ver. 2.13.2, we determined the highest scores of TOEFL iBT repeaters out of all the score reports available. Then we grouped the test repeaters’ TOEFL iBT highest total scores into three categories according to the tertiles of ranked scores. Scores 90 and below constituted the first third of the ranked scores; scores between 91 and 97 were the second; scores 98 and above were the third. TOEFL iBT repeaters were further grouped into three categories according to the number of times they took the test: those who took the test twice, those who took it three times and those who took the test four times and more. Thus, the analysis was based on a 3-by-3 contingency table (see Appendix E), which was used to test for non-zero correlation between the number of test repeated and the levels of achievement on the test. The Cochran-Mantel-Haenszel statistics was used in this analysis. Results ($M^2 = 0.41, p = 0.52$) show that the null hypothesis of zero correlation is retained at .05 level, suggesting the number of TOEFL iBT taken repeatedly and the levels of achievement on the test are independent of each other for test repeaters at UIUC. Our decision to group TOEFL iBT scores by tertiles was to ensure that each group had sufficient frequency counts in the
analysis; nonetheless, another grouping scheme based on the UIUC admission minimum TOEFL iBT score of 79 and the EPT exemption criterion of 102 was used as well. \(^5\) Again, results \((M^2 = 0.12, p = 0.73)\) indicate that the number of TOEFL iBT test repeated is not related to the levels of achievement on the test.

Using the same statistical method, we examined whether the number of TOEFL iBT tests repeated was related to the levels of ESL placement. Again, the test repeaters were grouped into three groups according to the number of the tests repeated. The analysis was based on a 3-by-3 contingency table (see Appendix E) with three groups of test repeaters and three levels of ESL placement. The results \((M^2 = 1.39, p = 0.24)\) suggest that the number of times TOEFL iBT is taken and the levels of ESL placement are independent of each other for test repeaters at UIUC.

**Analysis of the differences among highest, most recent, average and self-reported scores of TOEFL iBT repeaters.** For this part of the analysis, we first calculated the average of TOEFL iBT scores of test repeaters. As shown in Table 5.1, self-reported total TOEFL iBT scores have the highest average (92.35), followed by the highest (92.19), most recent (90.72) and average (88.05) scores.

<table>
<thead>
<tr>
<th>Score</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest</td>
<td>396</td>
<td>92.19</td>
<td>8.65</td>
</tr>
<tr>
<td>Most recent</td>
<td>396</td>
<td>90.72</td>
<td>9.49</td>
</tr>
<tr>
<td>Average</td>
<td>396</td>
<td>88.05</td>
<td>9.26</td>
</tr>
<tr>
<td>Self-reported</td>
<td>396</td>
<td>92.35</td>
<td>8.52</td>
</tr>
</tbody>
</table>

In order to explore the significance of the differences among all scores, we ran a within-subjects ANOVA. The results indicate that there is a significant difference among the highest, most recent, average and self-reported TOEFL iBT total scores \((F = 220.57, df = 3, 1185 \text{ and } p < .0001)\), suggesting that at least one of the TOEFL iBT scores differs from the others.
A follow-up analysis using Tukey pair-wise method shows that the average TOEFL iBT total score is the main source of pair-wise differences among the four types of scores from test repeaters (see Table 5.2). The average TOEFL iBT total score is the lowest among the four types of scores. In addition, self-reported TOEFL iBT scores do not differ significantly from both the highest and most recent scores. The difference between the highest and most recent scores is also insignificant. These pair-wise results seem to suggest that the test repeaters are inclined to report their TOEFL iBT total scores that are very close to either their highest or most recent scores. It also suggests that the test takers are very aware of their own scores and are strategizing about which ones to submit. In an earlier study, Kokhan (2012) matched self-reported and official scores for 91.9% of students in a similar data set and found that less than 3% of students significantly exaggerated their TOEFL scores.

Table 5.2

<table>
<thead>
<tr>
<th>Score</th>
<th>Difference</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>Adj. p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest-Average</td>
<td>4.14</td>
<td>2.49</td>
<td>5.78</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Most recent-Average</td>
<td>2.67</td>
<td>1.03</td>
<td>4.31</td>
<td>.0002</td>
</tr>
<tr>
<td>Self-reported-Average</td>
<td>4.29</td>
<td>2.65</td>
<td>5.49</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Most recent-Highest</td>
<td>-1.47</td>
<td>-3.11</td>
<td>.18</td>
<td>.099</td>
</tr>
<tr>
<td>Self-reported-Highest</td>
<td>.16</td>
<td>-1.49</td>
<td>1.80</td>
<td>.99</td>
</tr>
<tr>
<td>Self-reported-Most recent</td>
<td>1.62</td>
<td>-.02</td>
<td>3.27</td>
<td>.054</td>
</tr>
</tbody>
</table>

**Relationship between ESL writing placement levels and TOEFL iBT scores (highest, most recent, average and self-reported).** Figure 5.1 illustrates an ideal scenario of ESL placement. In order to be usable for placement purposes, TOEFL iBT scores should significantly differ between the placement levels. For example, all students with low TOEFL iBT scores (e.g., 1-80) should be placed in Level 1 (the lowest ESL level), all students with higher TOEFL iBT scores (e.g., 81-100) should be placed in Level 2, and all students with the highest TOEFL iBT scores (e.g., 101-120) should be placed in Level 3.
To check whether the actual ESL placement is any close to the ideal scenario of placement, we built a series of graphs illustrating the placement trends in relation to the highest, most recent, average and self-reported total TOEFL iBT scores (see Figure 5.2). The results suggest that, no matter what scores we take (highest, most recent, average or self-reported), there is no distinct pattern of placement which would help us to determine as to what scores of TOEFL iBT repeaters are better predictors of ESL placement. According to all four graphs, students with TOEFL iBT scores ranging from 70 to 105 have on average a 60% chance to be placed into Level 2 of the ESL writing classes. The repetition of the procedures with the highest, average, most recent and self-reported section scores yields very similar results, so they are omitted for reasons of space.

Figure 5.1. The ideal scenario of ESL placement based on the hypothetical TOEFL iBT score ranges for the three placement levels.
Figure 5.2. Probability of ESL placement in relation to the total highest, most recent, average and self-reported scores of TOEFL iBT repeaters.

To further explore the predictive capacity of the TOEFL iBT scores of test repeaters for ESL placement purposes, we conducted four cumulative logistic regression analyses based on the four types of TOEFL iBT total scores (highest, most recent, average and self-reported) respectively. The cumulative logistic model helps to collect two important pieces of information. First, it examines the extent to which the ESL placement levels are associated with test repeaters’ TOEFL iBT scores. Second, the analysis can estimate the classification efficiency in predicting the ESL placement levels based on test takers’ TOEFL iBT scores.

Inherent in the cumulative logistic regression model is the proportional odds assumption, according to which the logistic regression coefficient is the same for all cumulative logistic curves. In the present study, this implies that the strength of association between TOEFL iBT scores and the odds of ESL placement is identical across all ESL placement levels (i.e. Level 1, 2 and 3). Table 5.3 shows the statistics of score test for the proportional odds assumption according to each TOEFL iBT score from test
repeaters. This statistics can also be used as an alternative to goodness-of-fit statistics, such as deviance and Pearson chi-square.

Table 5.3

*Score test for the proportional odds assumption*

<table>
<thead>
<tr>
<th>Score</th>
<th>Proportional Odds Assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chi-square</td>
</tr>
<tr>
<td>Highest</td>
<td>.28</td>
</tr>
<tr>
<td>Most recent</td>
<td>.30</td>
</tr>
<tr>
<td>Average</td>
<td>.38</td>
</tr>
<tr>
<td>Self-reported</td>
<td>.07</td>
</tr>
</tbody>
</table>

For each TOEFL iBT total score from test repeaters, the data appear to satisfy the assumption of proportional odds. For example, the cumulative regression analysis based on the highest TOEFL iBT total scores shows that the proportional odds assumption is retained (p = .60), suggesting that there is no strong evidence of lack-of-fit. These results also indicate that the association between the TOEFL iBT total scores and ESL placement does not differ significantly at the higher and the lower ESL placement levels.

Table 5.4 shows an excerpt of logistic regression output from SAS ver. 9.3. The TOEFL iBT scores in this analysis are standardized according to the standard deviation of each score so that the results from the four cumulative logistic regression analyses are comparable.

Table 5.4

*Summary table of cumulative logistic regression of ESL placement on four types of TOEFL iBT total scores respectively (n=396)*

<table>
<thead>
<tr>
<th>Score</th>
<th>Logistic regression coefficient</th>
<th>Standard error</th>
<th>Pr &gt; ChiSq</th>
<th>Odds ratio estimates</th>
<th>Somers’ D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest</td>
<td>.4115</td>
<td>.1011</td>
<td>&lt;.0001</td>
<td>1.509</td>
<td>.180</td>
</tr>
<tr>
<td>Most recent</td>
<td>.3498</td>
<td>.0994</td>
<td>.0004</td>
<td>1.419</td>
<td>.171</td>
</tr>
<tr>
<td>Average</td>
<td>.3919</td>
<td>.1003</td>
<td>&lt;.0001</td>
<td>1.480</td>
<td>.169</td>
</tr>
<tr>
<td>Self-reported</td>
<td>.4542</td>
<td>.1022</td>
<td>&lt;.0001</td>
<td>1.575</td>
<td>.200</td>
</tr>
</tbody>
</table>
The logistic regression coefficient represents the degree to which ESL placement is associated with each TOEFL iBT score. Results show that for each TOEFL iBT total score, the logistic regression coefficient is significant at the .01 level, suggesting that there is a strongly significant association between the ESL placement and each TOEFL iBT total score. For example, for one standard unit increase in the highest TOEFL iBT total score, the odds of a test taker being placed at a higher ESL level are 1.509 times (i.e., odds ratio estimate) the odds of the test taker being placed at a lower ESL level. Overall, the values of the estimated odds ratio indicate that the self-reported TOEFL iBT total scores from test repeaters have the strongest association with ESL placement results (1.575), followed by the highest (1.509), average (1.480) and most recent scores (1.419).

Somers’ D is probably the most commonly used statistic in estimating the classification efficiency of cumulative logistic regression models (O’Connell, 2006). In this study, it is used as a rank order correlation between the predicted probability of ESL placement given the TOEFL iBT total scores and the observed ESL placement trends. Using different TOEFL iBT total scores in predicting the ESL placement levels, we found that self-reported TOEFL iBT total scores had the highest classification efficiency, followed by the highest, most recent and average total scores.

However, we must emphasize that there is a substantive difference between association and classification efficiency in an analysis such as this. The significant association between TOEFL iBT scores and ESL placement is to be interpreted in a probabilistic sense (i.e. a person with a higher TOEFL iBT score is more likely to be placed at a higher placement level) whereas classification efficiency is to be interpreted as the absolute predictive power of TOEFL iBT scores with respect to ESL placement results. Although the four types of TOEFL iBT scores are significantly associated with ESL placement, results show that the predictive power is low for any TOEFL iBT scores, ranging from .169 to .200. Furthermore, classification error rates based on the cumulative logistic model are above 57% regardless of which TOEFL iBT score is used.
Given that the current study focused on the EPT written test only and that the EPT required examinees to incorporate information from a lecture and an article, we also conducted separate analyses that looked into the relationship between the ESL placement levels and TOEFL iBT Writing scores only, and the relationship between the ESL placement levels and TOEFL iBT combined scores based on the Writing, Listening and Reading sections. Results based on the TOEFL iBT Writing scores showed that the odds ratio estimates ranged from 1.314 to 1.448 and the classification efficiency estimates were from 0.135 to 0.171; results based on the TOEFL iBT combined scores showed that the odds ratio estimates were between 1.236 and 1.470 whereas the Somers’ D were between 0.103 and 0.160. Similarly, a strongly significant association was observed between the ESL placement levels and TOEFL iBT scores; however, the classification efficiency was low for any TOEFL iBT scores from test repeaters based on either the TOEFL iBT Writing section scores only or the combined scores for Writing, Listening, and Reading.

Discussion

Overall, the TOEFL iBT repeaters are representative of the entire population of EPT test takers. Gender does not seem to be a factor which determines the tendency for repeating the test. The expected university degrees do not explain test repetition either; however, we found that non-degree students are least likely to repeat the TOEFL iBT. The latter observation may be due to the fact that these students arrive in the USA as exchange students and the English language proficiency requirements may not be strictly applied to them. Countries of origin do not seem to be closely related to the tendency to repeat the TOEFL iBT; however, there are two interesting observations in this respect. The students from India are least likely to repeat the TOEFL iBT, which may be due to the fact that English is the second official language in India and this category of candidates does not experience much difficulty in reaching the score which would satisfy the English proficiency requirements for admission. The students from Indonesia are most likely to repeat the TOEFL iBT as their fraction in the repeater group is much higher than that in the non-repeater group (3.28% versus 0.6%), which indicates that they may experience much more difficulty in reaching
the minimum English proficiency requirement for admission than the students from other countries. These observations are consistent with the ETS’ Test and Score Data Summary (2011): TOEFL iBT total and section score means of students from India are significantly higher than those of students from Indonesia.6

It is necessary to point out that the TOEFL iBT test repeaters in our study have an average highest total score of 92.19. According to the TOEFL score guide, test takers who score higher than 90 can be considered high-proficient English learners (http://www.ets.org/toefl/ibt/scores/understand/). Our results also indicate that the number of the TOEFL iBT tests repeated and the levels of achievement on the test are not related to each other. Our results are in line with Yang et al.’s (2011) findings that high-performing test repeaters are less likely to have significant changes by repeating the test. In addition, our results showed that the number of the TOEFL iBT tests repeated and the levels of ESL placement are independent of each other for the TOEFL iBT test repeaters at UIUC.

Pair-wise comparisons show that the average TOEFL iBT scores are significantly lower than the other three TOEFL iBT scores. In addition, the differences between the highest and most recent scores, between the highest and self-reported scores, and between the most recent and self-reported scores are not significant. In fact, the highest, most recent and self-reported TOEFL iBT scores from test repeaters at UIUC do not differ by more than 1.63 points. The test repeaters at UIUC tend to self-report either their highest or most recent TOEFL iBT scores.

The analysis of ESL placement trends for the period of Fall 2006 – Fall 2011 in relation to the total highest, average, most recent and self-reported TOEFL iBT scores indicates that none of the scores are reliable for making accurate ESL placement decisions since, in some cases, students with very low TOEFL iBT scores can be placed into the highest ESL level and students with very high TOEFL iBT scores can be placed into the lowest ESL level. The results of the cumulative logistic regression analyses demonstrate that the absolute predictive power of the four types of TOEFL iBT scores is generally low. However, the self-reported TOEFL iBT scores, followed by the highest TOEFL iBT scores, have the strongest association with the ESL placement results and they also demonstrate the highest classification
efficiency in predicting ESL placement. We believe that the significant association between the TOEFL iBT scores of test repeaters and their ESL placement results could be explained by the fact that both tests deal with academic English language proficiency. The poor predictive capacity of TOEFL iBT scores may be due to the difference between the intended purposes of the TOEFL iBT and the EPT. TOEFL iBT scores are primarily used for admission purposes as a pre-arrival measure of students’ ability to use and understand English in the academic environment. On the other hand, the EPT is a local placement test administered several days before students’ first semester at UIUC. It simulates a real classroom environment and is better aligned with the goals of the University’s ESL writing curriculum. The low predictive power of TOEFL iBT scores may be due to its lack of sensitivity to the specific objectives embedded in the ESL curriculum, from which the test specification of the EPT is derived. Under the circumstances with limited financial resources and in the situation where it is impractical to require all students to take ESL placement test, the self-reported or the highest TOEFL iBT scores might be the best option for considering when making ESL placement decisions.

Although clearly we must be cautious in generalizing our results to other educational institutions, we need to point out that our study is based on a comprehensive campus-wide data set collected over a multi-year period and provides a detailed methodological approach to analyzing the relationship between ordinal ESL placement levels and continuous standardized language test scores, from which other ESL placement programs could benefit.

The results of this study prompt the question: Can TOEFL iBT scores of any TOEFL test taker (repeater and non-repeater) be used for ESL placement? This question is beyond the scope of this study and we do not attempt to answer it in this paper. By concluding that TOEFL iBT repeaters’ scores would not appropriate for ESL placement at the University considered here, we are not in any way implying that the TOEFL iBT does not work for placement of the majority of ESL population in other universities. However, we do want to emphasize that, because of the under-researched nature of this topic, score users must cautious when making ESL or EAP placement decisions based solely on standardized language proficiency scores.
Notes

1. Highest-within-administration score is a score from the administration in which students received the highest total score (Zhao et al., 2010).

2. Highest-across-administration score is a total score computed by summing the highest section scores across administrations (Zhao et al., 2010).

3. Yang, Bontya and Moses (2011) did not explicitly mention the name of the test.

4. A cutoff score (or a cut-point) is “that score at or above which students will be classified one way and below which students will be classified differently” (Brown, 1996, p. 249). In case of ESL placement of international students, such a cut-point separates students who need some remedial ESL courses from those whose language skills satisfy the academic requirements set by the university.

5. Note, that 20 TOEFL iBT test repeaters who scored below 79 in our sample were offered conditional admission to UIUC.

References


Purdue University. (n.d.) How to apply. Retrieved from http://www.gradschool.purdue.edu/admissions/apply_toefl.cfm


The University of Texas at Austin. (n.d.). About ESL. Retrieved from http://world.utexas.edu/esl/about


Zhao, X., Oppler, S., Dunleavy D., & Kroopnick, M. (2010). Validity of four approaches of using repeaters’ MCAT scores in medical school admissions to predict USMLE Step 1 total scores. *Academic Medicine, 85,* S64–S67
Chapter 6
Discussion and Conclusions

Discussion of the Results

The overview of the English proficiency requirements in the RU/VHs confirmed that standardized language proficiency tests are widely used to satisfy the English proficiency requirement for admission. The TOEFL and IELTS are the most widely accepted tests for university admission purposes. However, the most striking trend in the admission policies is that the score requirements established by the universities significantly differ. The results of this overview and review of the relevant literature prompted me to reach the conclusion that such variations in test scores may be due to mere “marketization” of higher education prevalent in the United States. Because of the prevailing “marketization” some schools can lower their entry scores to raise enrollment while other schools may raise their admission scores if they wish to be more selective in choosing their international applicants. Besides standardized test scores, other reasons for exemption from the English proficiency requirement include: a bachelor’s degree or higher from regionally accredited U.S. college or university; completion of certain number of years of high school in a U.S. high school; completion of certain number of hours of graduate coursework with a satisfactory cumulative GPA; a degree from a school where primary instruction is in English. The TOEFL requirement may be waived if undergraduate applicants received high scores for SAT Critical Reading or ACT English. American citizenship is not generally considered a reason for exemption from the English proficiency requirement with exception of six campuses that do not require U.S. citizens to fulfill this requirement. Conditional (a.k.a. provisional) admission of students with low language proficiency test scores is very common for RU/VHs. Conditionally admitted students may be required to attend intensive English classes or ESL courses prior to starting their studies at the university or simultaneously with taking main courses. Completion of the institutional intensive English program prior to applying to the university can also satisfy the English proficiency requirement.
Another important observation is that standardized test scores for admission to undergraduate and graduate programs may significantly differ within the same university. In some institutions, undergraduate students may be required to meet higher scores requirements than graduate students; in other universities, undergraduate students may have lower score requirements than graduate students. The observed score difference between the two levels is generally 9-10 points.

The overview of the ESL placement policies in 20 U.S. universities with the largest number of international students (Appendix C) prepared as part of Paper Four confirmed that ESL placement testing and ESL coursework are very common in many universities. Similarly to test score requirements for admission, significant variability is observed regarding TOEFL iBT cutoff scores for ESL placement. Having observed the dominance of the TOEFL in the international admissions and ESL placement requirements, I examined the appropriateness of using this test scores for making ESL placement decisions. Graphic illustration of ESL placement trends was chosen by me as the main method for exploring probability of ESL placement. One of the ways to examine the probability of placement is to closely analyze changes of the correlation between the TOEFL and the EPT over certain period of time; here it is about two years. The correlation pattern obtained is characterized by stronger correlation when the time gap between the tests is short. Around Week 50, correlation dramatically decreases and reaches the lowest point (below zero). Starting from the 50th week, correlation goes up. This trend is observed for TOEFL total and section scores. I attribute a gradual drop of the correlation to English language attrition. Before taking the TOEFL, most candidates have intensive preparation for the test; however, during the time between submitting applications to the university and taking the EPT they do not practice any essay writing strategies and their writing skills gradually deteriorate which explains why their scores do not correlate with the EPT. Reuse of TOEFL scores after about year of the exposure to English in an English-speaking country might explain the increase of the correlation after Week 50. The point to notice is that the majority of the ESL population at UIUC takes the EPT around Week 50, which leads to the conclusion that the TOEFL seems to be ineffective for placement for the majority of the EPT candidates.
The University of Illinois at Urbana-Champaign has experienced rapid growth of the international undergraduate student population. In Paper Three, I closely examined such variables as SAT (Writing, Reading, Math, and Essay), ACT (Composite, English, Math, and combined English/Writing) in addition to the TOEFL iBT scores. The findings indicate that only students’ lowest ACT English and SAT Reading scores and their highest TOEFL iBT total and TOEFL iBT Writing section scores would work for placement of international undergraduate students into ESL classes at the University of Illinois. I would like to emphasize that the number of students with extreme standardized test scores has been very low (less than 5%) over the period of Fall 2006 through Fall 2011. Thus, reduction of the EPT test taker population by 5% will not significantly save human and financial resources spent on the administration of this test.

TOEFL iBT can be taken multiple times. A close look at the data set available for the analysis revealed that 15.6% of the EPT test takers were TOEFL iBT repeaters. To find out which scores predict best of all the ESL placement results (the highest, most recent, average or self-reported), a series of statistical analyses were performed in a paper co-authored with Cary Lin. The results indicate that the self-reported and the highest TOEFL iBT total scores have the strongest association with the EPT results. They also demonstrate the highest classification efficiency in predicting ESL placement. However, the absolute predictive power of the four types of TOEFL iBT scores is generally low; thus, none of them is a good predictor of ESL placement. The significant association between TOEFL iBT scores and ESL placement can be due to the fact that both tests deal with academic English language proficiency. The poor predictive capacity of TOEFL iBT scores can be explained by the difference between the intended purposes of the TOEFL iBT and the EPT. The TOEFL iBT is primarily developed for making admission decisions prior to students’ arrival on campus. The EPT simulates a real classroom environment and is better aligned with the goals of the University’s ESL writing curriculum. This test is usually administered before students’ first semester at the university. The low predictive power of TOEFL iBT scores may be a result of a lack of sensitivity of the test to the specific objectives embedded in the ESL curriculum.
Even though very different research methods were used to explore the appropriateness of using standardized test scores for ESL placement for different groups of the EPT candidates (Paper Two – all EPT candidates, Paper Two – undergraduate EPT candidates, Paper Four – EPT candidates who repeated the TOEFL iBT), the results from all these papers firmly suggest that using test scores for ESL placement as the only measure of students’ ability to use and understand English may result in significant percentage of misplaced students.

It is noteworthy that despite that fact that all the chapters are individual research projects, the findings and their interpretations are interrelated. For example, the survey of the English Proficiency requirements presented in Paper One is an extended version of the survey of test score requirements discussed in Papers Three and Four. The surveys in the latter two papers are based on the classification of 20 U.S. universities with the largest number of international students compiled by the Institute of International Education; whereas Paper One is a comprehensive overview of the English proficiency requirements in 108 research universities in which those 20 universities are included.

Reflections about Methods of Analysis

A huge data set created the opportunity for me to apply not only traditional statistical methods of data analysis but also some innovative approaches. One of the absolutely new approaches that I used was the examination of the changes of correlation between the TOEFL and the EPT over time (Paper Two). To get a picture of the correlation between these two tests, I wrote an algorithm that could graphically demonstrate gradual changes of correlation for groups of students over time. In my view, no other method could better capture the pattern of correlation.

Graphical illustration of the probability of placement using various combination of cutoff scores (Paper Three) was used for exploring the appropriateness of using standardized test scores for ESL placement. The algorithm compared the expected placement of undergraduate students with their actual placement in the ESL levels. The fractions of students who were correctly placed in relation to their paired test scores were calculated and plotted on a graph using different color shades.
By contrast, the study on TOEFL iBT repeaters was based entirely on traditional statistical methods of data analysis (Chapter 5). Because of the ordinal nature of ESL placement levels (Levels 1, 2 & 3), cumulative logistic regression was chosen by us as the most appropriate approach for analyzing the association between ESL placement and TOEFL iBT scores. This approach allowed us to compare precise values of logistic regression coefficients which represented the degree to which ESL placement is associated with each of the four types of TOEFL iBT scores (highest, most recent, average, and self-reported). We would not be able to achieve such precise comparison if only a graphical approach to data analysis and interpretation were used.

**Limitations and Suggestions for Future Research**

All the standardized tests discussed in this dissertation were examined in the context of the U.S. higher education. Although TOEFL is also widely used outside the USA, the investigation of that is outside the scope of this research work and is not considered in this dissertation.

One of the potential limitations of this research is that the researcher could not conduct a follow-up qualitative study with the EPT candidates whose scores were available for the analysis. The data were collected over an extended period of time which in combination with privacy concerns made it extremely difficult to track the EPT test takers from earlier years. Examination of additional variables could potentially strengthen this research and provide some valuable insights, in particular regarding the fluctuation of correlation between TOEFL and EPT over time. It would be interesting to know if the EPT candidates took any intensive writing courses upon arrival to the University or had any intensive language instruction after taking the TOEFL in the candidate’s home country. Follow-up interviews with the EPT test-takers could amplify the findings. Qualitative research could also explain some variations in TOEFL iBT repeaters’ scores.

Another issue that needs further investigation is the effect of essay format (typed versus handwritten) on candidates results. The preliminary results suggest that higher-level students seem disadvantaged when writing by hand (Prof. Paula Winke, personal communication, March 8, 2013).
Language attrition is one of the hypotheses that could explain the distinct correlation pattern between the EPT results and the TOEFL iBT scores (Kokhan, 2012). However, there may be additional factors explaining such correlation pattern but they were not considered in this study, for example, country of origin or prior education of the EPT candidates, their major, and the level of studies (graduate, undergraduate or non-degree). A sudden drop in the correlation pattern around Week 50, when the majority of students take the EPT, might also be related to such highly unexplored and rarely discussed issue as cheating on either the TOEFL or the local placement test. The impact of all these factors will be explored in future research.

**Suggestions for Changes of ESL Placement Practice**

Ideally all English L2 students should take the English Placement Test before the beginning of their first semester at the University. Based on the analysis of placement trends (Papers Two through Four), there is no particular set of either total or section scores which could accurately predict the ESL placement since there is a significant number of cases when the students with very low TOEFL scores were exempt or vice versa – the students with relatively high scores were placed into the lower levels of the ESL writing classes.

Introduction of a policy according to which all newly admitted students should take ESL placement tests may be the most objective approach. Local ESL placement tests are specifically designed to place the students in the appropriate ESL levels; whereas standardized language proficiency tests are primarily designed for making admission decisions and cannot adequately assess how students are likely to perform in the ESL classes. Undoubtedly, the biggest downside of this policy is increased financial expenses on administering ESL placement tests. Moreover, there are some universities that do not have any local ESL placement instruments at all.

If the first recommendation cannot be realized due to some reasons, for example, lack of financial and teaching resources, score users are advised to follow these recommendations:
1. Based on the results from Paper Two, it is recommended to use TOEFL iBT total or/and writing section scores as an additional reference when deciding which level a student should be placed into in case of disagreement between the EPT raters. The results of one-way ANOVA and the post-hoc Scheffe performed suggest that the averages of the TOEFL iBT total and writing sections scores are statistically significantly different among all three placement levels.

2. Based on the results from Paper Two, it is also recommended to avoid making any ESL placement decisions and conclusions regarding the English proficiency of the EPT candidates based on their TOEFL iBT scores that are about a year old. According to the pattern of correlation between the time of taking TOEFL iBT and the EPT, the most recent TOEFL iBT total and section scores seem to correlate better with the EPT than the scores which are about 50 weeks old.

3. It is recommended to avoid setting any combinations of cutoff scores for making ESL placement decisions. Based on the results from Paper Three, using combination of test scores for ESL placement will not be effective either.

4. As far as TOEFL iBT repeaters are concerned, their self-reported or the highest TOEFL iBT scores might be the best option for considering when making ESL placement decisions under the circumstances with limited financial resources and in the situation where it is impractical to require all students to take ESL placement test.

Afterword: Changes in the ESL Placement Practice after Fall of 2011

The major change in the ESL placement since Fall of 2011 is the introduction of an online version of the English Placement Test to complement the traditional paper-and-pencil format. The reason for this change is that the University wants to have the ESL placement results prior to students’ arrival on campus which would make course planning much easier. The online EPT was offered in Summer of 2012 and 2013. This test format is available for only two days a week during the summer semester and is open only for undergraduate candidates. The main difference in the structure of the online test from the traditional version is that students are asked to participated in the online Moodle forum; whereas in the
paper-and-pencil format, they are asked to take part in a group discussion regarding the main points of a lecture and a reading passage which are used to write an argumentative essay. The EPT team, a group of experience EPT proctors and raters, have developed around 35-40 versions of the test. Several hundreds of the ESL candidates have already taken this test. The availability of early EPT test results made it possible for students to get early ESL advising and course selection. The main concern that remains regarding the online test format is the possibility of cheating on the test. The new version of the EPT is currently being explored by the doctoral students. Sun Joo Chung has found that there exists some problem with communication during this test which results in very low percentage of students participating in the online forum that is considered a required part of the test (Prof. Fred Davidson, personal communication, April 18, 2014).

References

## Appendix A

### Language proficiency tests accepted by the U.S. universities

<table>
<thead>
<tr>
<th>RU/VH</th>
<th>Tests</th>
<th>RU/VH</th>
<th>Tests</th>
<th>RU/VH</th>
<th>Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona State University, Tempe, AZ</td>
<td>TOEFL PBT &amp; iBT, IELTS, PTE</td>
<td>Purdue University – Main Campus, West Lafayette, IN</td>
<td>TOEFL all versions, PTE Academic, IELTS</td>
<td>University of Illinois at Chicago, Chicago, IL</td>
<td>TOEFL all versions, IELTS</td>
</tr>
<tr>
<td>Boston University, Boston, MA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>Rensselaer Polytechnic Institute, Troy, NY</td>
<td>TOEFL all versions, IELTS</td>
<td>University of Illinois at Urbana-Champaign, Champaign, IL</td>
<td>TOEFL all versions, IELTS</td>
</tr>
<tr>
<td>Brandeis University, Waltham, MA</td>
<td>TOEFL PBT, CBT &amp; iBT; IELTS</td>
<td>Rice University, Houston, TX</td>
<td>TOEFL all versions, IELTS</td>
<td>University of Iowa, Iowa City, IA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Brown University, Providence, RI</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>Rockefeller University, New York, NY</td>
<td>TOEFL (no info about the versions)</td>
<td>University of Kansas, Lawrence, KS</td>
<td>TOEFL iBT, IELTS</td>
</tr>
<tr>
<td>California Institute of Technology, Pasadena, CA</td>
<td>TOEFL all versions, PTE Academic, IELTS</td>
<td>Rutgers University – New Brunswick, New Brunswick, NJ</td>
<td>TOEFL PBT &amp; iBT, IELTS, MELAB, PTE Academic</td>
<td>University of Kentucky, Lexington, KY</td>
<td>TOEFL all versions, IELTS</td>
</tr>
<tr>
<td>Carnegie Mellon University, Pittsburgh PA</td>
<td>TOEFL iBT, IELTS</td>
<td>Stanford University, Stanford, CA</td>
<td>TOEFL, but some programs may also accept IELTS, PTE</td>
<td>University of Louisville, Louisville, KY</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Case Western Reserve University, Cleveland, OH</td>
<td>TOEFL all versions, IELTS</td>
<td>Stony Brook University, Stony Brook, NY</td>
<td>TOEFL all versions, IELTS</td>
<td>University of Maryland – College Park, College Park, MD</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Colorado State University, Fort Collins, CO</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>SUNY at Albany, Albany, NY</td>
<td>TOEFL all versions, IELTS</td>
<td>University of Massachusetts Amherst, Amherst, MA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Columbia University in the City of New York, New York, NY</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>Texas A&amp;M University, College Station, TX</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>University of Miami, Coral Gables, FL</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Cornell University, Ithaca, NY</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>The University of Tennessee, Knoxville, TN</td>
<td>TOEFL all versions, IELTS</td>
<td>University of Michigan – Ann Arbor, Ann Arbor, MI</td>
<td>TOEFL PBT &amp; iBT, IELTS, MELAB, ECPE, CPE, CAE</td>
</tr>
<tr>
<td>CUNY Graduate School and University Center, New York, NY</td>
<td>TOEFL</td>
<td>The University of Texas at Austin, Austin, TX</td>
<td>TOEFL all versions, IELTS</td>
<td>University of Minnesota – Twin Cities, Minneapolis, MN</td>
<td>TOEFL PBT &amp; iBT, IELTS, MELAB, ECPE, CPE, CAE</td>
</tr>
<tr>
<td>Dartmouth College, Hanover, NH</td>
<td>TOEFL all versions, IELTS</td>
<td>Tufts University, Medford, MA</td>
<td>TOEFL all versions, IELTS</td>
<td>University of Missouri – Columbia, Columbia, MO</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Duke University, Durham, NC</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>Tulane University of Louisiana, New Orleans, LA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>University of Nebraska – Lincoln, Lincoln, NE</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Emory University, Atlanta, GA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>University at Buffalo, Buffalo, NY</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>University of New Mexico – Main Campus, Albuquerque, NM</td>
<td>TOEFL all versions, IELTS, CPE, CAE</td>
</tr>
<tr>
<td>Institution</td>
<td>TOEFL, IELTS, MELAB</td>
<td>University</td>
<td>University</td>
<td>TOEFL, IELTS, PTE Academic</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Florida State University, Tallahassee, FL</td>
<td>TOEFL PBT &amp; iBT, IELTS, MELAB</td>
<td>University of Alabama at Birmingham, Birmingham, AL</td>
<td>University of North Carolina at Chapel Hill, Chapel Hill, NC</td>
<td>TOEFL PBT &amp; iBT, MELAB, IELTS, PTE Academic</td>
<td></td>
</tr>
<tr>
<td>George Washington University, Washington, DC</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>University of Alabama at Huntsville, Huntsville, AL</td>
<td>University of Notre Dame, Notre Dame, Indiana</td>
<td>TOEFL IBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Georgetown University, Washington, DC</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>University of Arizona, Tuscon, AZ</td>
<td>University of Oklahoma Norman Campus, Norman, OK</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Georgia Institute of Technology – Main Campus, Atlanta, GA</td>
<td>TOEFL</td>
<td>University of Arkansas, Fayetteville, AR</td>
<td>University of Oregon, Eugene, OR</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Georgia State University, Atlanta, GA</td>
<td>TOEFL PBT &amp; iBT, IELTS, GSTEP</td>
<td>University of California – Berkeley, Berkeley, CA</td>
<td>University of Pennsylvania, Philadelphia, PA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Harvard University, Cambridge, MA</td>
<td>TOEFL iBT</td>
<td>University of California – Davis, Davis, CA</td>
<td>University of Pittsburgh - Pittsburgh Campus, Pittsburgh, PA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Indiana University – Bloomington, Bloomington, IN</td>
<td>TOEFL PBT &amp; iBT, IELTS, MELAB</td>
<td>University of California – Irvine, Irvine, CA</td>
<td>University of Rochester, Rochester, NY</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Iowa State University, Ames, IA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>University of California – Los Angeles, Los Angeles, CA</td>
<td>University of South Carolina – Columbia, Columbia, SC</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Johns Hopkins University, Baltimore, MD</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>University of California – Riverside, Riverside, CA</td>
<td>University of South Florida – Tampa, Tampa, FL</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Louisiana State University and Agricultural &amp; Mechanical College, Baton Rouge, LA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>University of California – San Diego, La Jolla, CA</td>
<td>University of Southern California, Los Angeles, CA</td>
<td>TOEFL IBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Massachusetts Institute of Technology, Cambridge, MA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td>University of California – Santa Barbara, Santa Barbara, CA</td>
<td>University of Utah, Salt Lake City, UT</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Michigan State University, East Lansing, MI</td>
<td>TOEFL PBT &amp; iBT, IELTS, MSUETL, MELAB, Celp</td>
<td>University of California – Santa Cruz, Santa Cruz, CA</td>
<td>University of Virginia, Main Campus, Charlottesville, VA</td>
<td>TOEFL PBT &amp; iBT, IELTS, PTE Academic</td>
<td></td>
</tr>
<tr>
<td>Mississippi State University, Mississippi State, MS</td>
<td>TOEFL all versions, IELTS</td>
<td>University of Central Florida, Orlando, FL</td>
<td>University of Washington – Seattle Campus, Seattle, WA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>Montana State University, Bozeman, MT</td>
<td>TOEFL all versions, IELTS</td>
<td>University of Chicago, Chicago, IL</td>
<td>University of Wisconsin – Madison, Madison, WI</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
<tr>
<td>New York University, New York, NY</td>
<td>TOEFL iBT, IELTS</td>
<td>University of Cincinnati – Main Campus, Cincinnati, OH</td>
<td>Vanderbilt University, Nashville, TN</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
<td></td>
</tr>
</tbody>
</table>
Table A.1. (Continued)

<table>
<thead>
<tr>
<th>University</th>
<th>Tests Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Carolina State University at Raleigh, Raleigh, NC</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>University of Colorado at Boulder, Boulder, CO</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Virginia Commonwealth University, Richmond, VA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>North Dakota State University – Main Campus, Fargo, ND</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>University of Connecticut, Storrs, CT</td>
<td>TOEFL PBT &amp; iBT, IELTS + TOEFL CBT for UG</td>
</tr>
<tr>
<td>Virginia Polytechnic Institute and State University, Blacksburg, VA</td>
<td>TOEFL all versions, IELTS</td>
</tr>
<tr>
<td>Northwestern University, Evanston, IL</td>
<td>TOEFL all versions, IELTS</td>
</tr>
<tr>
<td>University of Delaware, Newark, DE</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Washington State University, Pullman, WA</td>
<td>TOEFL all versions, IELTS, MELAB</td>
</tr>
<tr>
<td>Ohio State University – Main Campus, Columbus, OH</td>
<td>TOEFL PBT &amp; iBT, IELTS, MELAB</td>
</tr>
<tr>
<td>University of Florida, Gainesville, FL</td>
<td>TOEFL PBT &amp; iBT, IELTS, MELAB</td>
</tr>
<tr>
<td>Washington University in St Louis, St Louis, MO</td>
<td>TOEFL all versions, IELTS</td>
</tr>
<tr>
<td>Oregon State University, Corvallis, OR, OR</td>
<td>TOEFL PBT &amp; iBT, IELTS, AP English L &amp; C</td>
</tr>
<tr>
<td>University of Georgia, Athens, GA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Wayne State University, Detroit, MI</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Pennsylvania State University – Main Campus, University Park, PA</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>University of Hawai’i at Manoa, Honolulu, HI</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Yale University, New Haven, CN</td>
<td>TOEFL all versions, IELTS, PTE Academic</td>
</tr>
<tr>
<td>Princeton University, Princeton, NJ</td>
<td>TOEFL iBT, IELTS</td>
</tr>
<tr>
<td>University of Houston, Houston, TX</td>
<td>TOEFL PBT &amp; iBT, IELTS</td>
</tr>
<tr>
<td>Yeshiva University, New York, NY</td>
<td>TOEFL all versions</td>
</tr>
</tbody>
</table>
### Appendix B

**Distribution of EPT test-takers by semester in chronological order**

Table B.1. *Distribution of EPT test-takers by semester in chronological order*

<table>
<thead>
<tr>
<th>Term</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer'06</td>
<td>57</td>
<td>2.2</td>
</tr>
<tr>
<td>Fall'06</td>
<td>340</td>
<td>13.1</td>
</tr>
<tr>
<td>Spring'07</td>
<td>56</td>
<td>2.2</td>
</tr>
<tr>
<td>Summer'07</td>
<td>59</td>
<td>2.3</td>
</tr>
<tr>
<td>Fall'07</td>
<td>371</td>
<td>14.3</td>
</tr>
<tr>
<td>Spring'08</td>
<td>58</td>
<td>2.2</td>
</tr>
<tr>
<td>Summer'08</td>
<td>71</td>
<td>2.7</td>
</tr>
<tr>
<td>Fall'08</td>
<td>593</td>
<td>22.8</td>
</tr>
<tr>
<td>Spring'09</td>
<td>91</td>
<td>3.5</td>
</tr>
<tr>
<td>Summer'09</td>
<td>116</td>
<td>4.5</td>
</tr>
<tr>
<td>Fall'09</td>
<td>681</td>
<td>26.2</td>
</tr>
<tr>
<td>Spring'10</td>
<td>104</td>
<td>4.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2597</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>
### Appendix C

**Overview of the ESL placement policies and TOEFL score requirements**

Table C.1. *Overview of the ESL placement policies as published on the official university websites*

<table>
<thead>
<tr>
<th>University</th>
<th>Cutoff score</th>
<th>ESL placement test</th>
<th>ESL coursework</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. University of Southern California</td>
<td>100 (min 20 for sections)</td>
<td>The International Student English Examination</td>
<td>Courses at American Language Institute</td>
</tr>
<tr>
<td>2. University of Illinois – Urbana-Champaign</td>
<td>102</td>
<td>The English Placement Test</td>
<td>ESL service courses</td>
</tr>
<tr>
<td>3. New York University</td>
<td>No info</td>
<td>All international students take a placement test at the American English Institute</td>
<td>Non-credit English courses</td>
</tr>
<tr>
<td>4. Purdue University – Main Campus</td>
<td>No info</td>
<td>Oral English Proficiency Exam for international TAs</td>
<td>An academic writing course and a course on classroom communication</td>
</tr>
<tr>
<td>5. Columbia University</td>
<td>No info</td>
<td>The American Language Program (ALP) Essay Exam</td>
<td>ALP classes</td>
</tr>
<tr>
<td>6. University of California – Los Angeles</td>
<td>No info</td>
<td>All international students take the UCLA English as-a Second Language Placement Exam</td>
<td>English as a second language courses</td>
</tr>
<tr>
<td>7. Ohio State University – Main Campus</td>
<td>114</td>
<td>A written English composition placement test</td>
<td>Courses in the ESL Composition Program, The English for Academic Purposes (EAP) courses</td>
</tr>
<tr>
<td>8. University of Michigan – Ann Arbor</td>
<td>No info</td>
<td>The Academic English Evaluation placement assessment</td>
<td></td>
</tr>
<tr>
<td>9. Michigan State University</td>
<td>79</td>
<td>The MSU English Language Test</td>
<td>English language courses at the English Language Center Courses at the Institute for English Language Programs Special part-time or full-time English courses Non-credit English courses</td>
</tr>
<tr>
<td>10. Harvard University</td>
<td>Depends on the program</td>
<td>The Harvard English language placement test</td>
<td></td>
</tr>
<tr>
<td>11. Indiana University – Bloomington</td>
<td>105</td>
<td>The Indiana English Proficiency Exam</td>
<td></td>
</tr>
<tr>
<td>12. Boston University</td>
<td>No info</td>
<td>All admitted students take an English placement examination</td>
<td></td>
</tr>
<tr>
<td>13. University of Florida</td>
<td>No info</td>
<td>No information</td>
<td>Courses at the English Language Institute Courses through ESL Services Specified remedial English courses</td>
</tr>
<tr>
<td>14. University of Texas – Austin</td>
<td>No info</td>
<td>No information</td>
<td></td>
</tr>
<tr>
<td>15. Penn State University – University Park</td>
<td>Total 80, S 19</td>
<td>No information</td>
<td></td>
</tr>
<tr>
<td>16. Northeastern University</td>
<td>Min score of 52</td>
<td>A placement test through American Classroom Program Placement evaluation at the English Language Institute</td>
<td>Courses through the American Classroom Program UG – at least one English course; G – a summer Intensive English Program Courses through Minnesota English Language Program ESL courses at Georgia Tech’s Language Institute English courses through AECP</td>
</tr>
<tr>
<td>17. SUNY University at Buffalo</td>
<td>Conditionally admitted</td>
<td>Placement examination</td>
<td></td>
</tr>
<tr>
<td>19. Georgia Institute of Technology</td>
<td>No info</td>
<td>American English and Culture program placement exam</td>
<td></td>
</tr>
<tr>
<td>20. Arizona State University</td>
<td>Conditionally admitted</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table C.2. Overview of the published TOEFL iBT score requirements in the US universities with the largest number of international students

<table>
<thead>
<tr>
<th>University</th>
<th>Minimum TOEFL iBT score required for admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. University of Southern California</td>
<td>At least 20 on for each section</td>
</tr>
<tr>
<td>2. University of Illinois – Urbana-Champaign</td>
<td>79</td>
</tr>
<tr>
<td>3. New York University</td>
<td>UG: no minimum score, but recommended 100; G: requirements differ, range 95-100</td>
</tr>
<tr>
<td>4. Purdue University – Main Campus</td>
<td>UG: 79; G 77: W – 18; S – 18; L – 14; R - 19</td>
</tr>
<tr>
<td>5. Columbia University</td>
<td>UG:100; G: depends on a program</td>
</tr>
<tr>
<td>6. University of California – Los Angeles</td>
<td>UG 83; G 87: W- 25; S – 24; R – 21; L - 17</td>
</tr>
<tr>
<td>7. Ohio State University – Main Campus</td>
<td>79; conditional admission 61-79</td>
</tr>
<tr>
<td>8. University of Michigan – Ann Arbor</td>
<td>88-100, section scores at least 23 in L &amp; R; 21 in S&amp;W</td>
</tr>
<tr>
<td>9. Michigan State University</td>
<td>UG: 79, at least 17 for each section; provisional admission 45-78; G: no score below 19 for R, L, and S; no W score below 22; minimum average score of 80</td>
</tr>
<tr>
<td>10. Harvard University</td>
<td>No general requirements</td>
</tr>
<tr>
<td>11. Indiana University – Bloomington</td>
<td>79</td>
</tr>
<tr>
<td>13. University of Florida</td>
<td>80</td>
</tr>
<tr>
<td>14. University of Texas – Austin</td>
<td>79</td>
</tr>
<tr>
<td>15. Penn State University – University Park</td>
<td>UG: 80; G: 80 (min for S 19)</td>
</tr>
<tr>
<td>16. Northeastern University</td>
<td>UG total 84; G 79-80 but 100 for Architecture, Biology, Journalism, Regulatory Science; and 93 for Chemistry</td>
</tr>
<tr>
<td>17. SUNY University at Buffalo</td>
<td>79, some programs permit conditional admission</td>
</tr>
<tr>
<td>18. University of Minnesota – Twin Cities</td>
<td>79 with minimum W – 21 and R - 19</td>
</tr>
<tr>
<td>19. Georgia Institute of Technology</td>
<td>Differ by program, range 79-100</td>
</tr>
<tr>
<td>20. Arizona State University</td>
<td>UG 61 (some departments set higher scores); G 80</td>
</tr>
</tbody>
</table>
Appendix D

Distribution of TOEFL iBT repeaters and non-repeaters according to the native country and the EPT placement levels

Table D.1. Distribution of TOEFL iBT repeaters and non-repeaters according to the native country

<table>
<thead>
<tr>
<th>Native country</th>
<th>Total population</th>
<th>Non-repeaters</th>
<th>Repeaters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency (%)</td>
<td>Frequency (%)</td>
<td>Frequency (%)</td>
</tr>
<tr>
<td>China</td>
<td>1464 (49.6)</td>
<td>1252 (48.9)</td>
<td>212 (53.54)</td>
</tr>
<tr>
<td>South Korea</td>
<td>387 (13.1)</td>
<td>345 (13.5)</td>
<td>42 (10.61)</td>
</tr>
<tr>
<td>Taiwan</td>
<td>379 (12.8)</td>
<td>329 (12.9)</td>
<td>50 (12.63)</td>
</tr>
<tr>
<td>India</td>
<td>89 (3.0)</td>
<td>86 (3.4)</td>
<td>3 (0.76)</td>
</tr>
<tr>
<td>Thailand</td>
<td>51 (1.7)</td>
<td>43 (1.7)</td>
<td>8 (2.02)</td>
</tr>
<tr>
<td>Turkey</td>
<td>58 (2.0)</td>
<td>48 (1.9)</td>
<td>10 (2.53)</td>
</tr>
<tr>
<td>Japan</td>
<td>42 (1.4)</td>
<td>39 (1.5)</td>
<td>3 (0.76)</td>
</tr>
<tr>
<td>Malaysia</td>
<td>37 (1.3)</td>
<td>32 (1.3)</td>
<td>5 (1.26)</td>
</tr>
<tr>
<td>France</td>
<td>35 (1.2)</td>
<td>32 (1.3)</td>
<td>3 (0.76)</td>
</tr>
<tr>
<td>Brazil</td>
<td>29 (1.0)</td>
<td>28 (1.1)</td>
<td>1 (0.25)</td>
</tr>
<tr>
<td>Indonesia</td>
<td>29 (1.0)</td>
<td>16 (0.6)</td>
<td>13 (3.28)</td>
</tr>
<tr>
<td>Iran</td>
<td>24 (0.8)</td>
<td>19 (0.7)</td>
<td>5 (1.26)</td>
</tr>
<tr>
<td>Colombia</td>
<td>20 (0.7)</td>
<td>15 (0.6)</td>
<td>5 (1.26)</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>19 (0.6)</td>
<td>14 (0.5)</td>
<td>5 (1.26)</td>
</tr>
<tr>
<td>Other countries</td>
<td>291 (9.8)</td>
<td>260 (10.1)</td>
<td>31 (7.82)</td>
</tr>
<tr>
<td>Total</td>
<td>2954 (100)</td>
<td>2558 (100)</td>
<td>396 (100)</td>
</tr>
</tbody>
</table>

Table D.2. Distribution of TOEFL iBT repeaters according to the EPT placement levels

<table>
<thead>
<tr>
<th>EPT placement</th>
<th>Frequency</th>
<th>Percent</th>
<th>Mean iBT total score</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>92</td>
<td>23.23</td>
<td>84.83</td>
<td>11.24</td>
</tr>
<tr>
<td>Level 2</td>
<td>221</td>
<td>55.81</td>
<td>88.58</td>
<td>8.60</td>
</tr>
<tr>
<td>Level 3</td>
<td>83</td>
<td>20.96</td>
<td>90.22</td>
<td>7.53</td>
</tr>
<tr>
<td>Total</td>
<td>396</td>
<td>100</td>
<td>88.05</td>
<td>9.26</td>
</tr>
</tbody>
</table>
Table D.3. Distribution of TOEFL iBT non-repeaters according to the EPT placement levels

<table>
<thead>
<tr>
<th>EPT placement</th>
<th>Frequency</th>
<th>Percent</th>
<th>Mean iBT total score</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>492</td>
<td>21.3</td>
<td>90.42</td>
<td>11.62</td>
</tr>
<tr>
<td>Level 2</td>
<td>1226</td>
<td>53.1</td>
<td>92.32</td>
<td>8.48</td>
</tr>
<tr>
<td>Level 3</td>
<td>589</td>
<td>25.5</td>
<td>94.62</td>
<td>8.02</td>
</tr>
<tr>
<td>Total</td>
<td>2307</td>
<td>100</td>
<td>92.50</td>
<td>9.24</td>
</tr>
</tbody>
</table>

Note. Students from MA in TESL program and students majoring in Business and Economics (N=251) were excluded from this analysis since they are required to be enrolled into the ESL writing courses specifically developed to meet their language needs.
Appendix E

Cross classification of the number of TOEFL iBT repeated by levels of English proficiency and ESL placement levels

Table E.1. Cross classification of the number of TOEFL iBT repeated by levels of English proficiency

<table>
<thead>
<tr>
<th>Freq Count (Column %)</th>
<th>Proficiency Levels Based on Tertiles</th>
<th>1&lt;sup&gt;st&lt;/sup&gt;</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt;</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt;</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Tests</td>
<td></td>
<td>79.58</td>
<td>81.88</td>
<td>81.90</td>
<td>321</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>113</td>
<td>113</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>22</td>
<td>22</td>
<td>15</td>
<td>59</td>
</tr>
<tr>
<td>4 and more</td>
<td></td>
<td>15.49</td>
<td>15.94</td>
<td>12.93</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>142</td>
<td>138</td>
<td>116</td>
<td>396</td>
</tr>
</tbody>
</table>

Note. Proficiency levels are based on the tertiles of ordered highest score distribution: 1st (1-90), 2nd (91-97) and 3rd (98-120).

Table E.2. Cross classification of the number of TOEFL iBT repeated by ESL placement levels

<table>
<thead>
<tr>
<th>Freq Count (Column %)</th>
<th>ESL Placement Levels</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>75</td>
<td>174</td>
<td>72</td>
<td>321</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(81.52)</td>
<td>(78.73)</td>
<td>(86.75)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>13</td>
<td>35</td>
<td>11</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(14.13)</td>
<td>(15.84)</td>
<td>(13.25)</td>
<td></td>
</tr>
<tr>
<td>4 and more</td>
<td></td>
<td>4</td>
<td>12</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.35)</td>
<td>(5.43)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>92</td>
<td>221</td>
<td>83</td>
<td>396</td>
</tr>
</tbody>
</table>