

## **Is the TOEFL Score a Reliable Indicator of International Graduate Students' Academic Achievement in American Higher Education?**

**Lan Thi Vu**

Doctoral student in Rhetoric and Composition,  
English Department at Southern Illinois  
University, Carbondale, USA.  
Email address: vulan@siu.edu

**Phu Hoang Vu**

Executive manager of E-center for Professional  
Development  
PhD (ABD) in Curriculum and Instruction,  
Department of Curriculum and Instruction  
Southern Illinois University, Carbondale, USA.

---

**Abstract:** *This study examines whether there is any significant correlation between the TOEFL scores and international graduate students' academic achievement in terms of their GPA. The data were collected via a survey and analyzed through descriptive statistics in order to identify patterns of agreement and disagreement among the group of 464 international graduate students at a Midwestern public university in the U.S. According to the correlation analysis of objective data, TOEFL scores were not found to be an accurate and effective predictor of academic performance as measured by GPA. However, according to students' opinions, their TOEFL scores were a reliable predictor of their academic performance. The findings were interpreted in view of their implications for future international graduate admissions and testing design, and directions for further research were discussed.*

**Keywords:** *Education Testing Service, Test of English as Foreign Language, International English Language Testing System, Traditional Examination, Academic Performance.*

---

### **1. INTRODUCTION**

Higher education at an American university is still luring international students. According to the Open Door report annually published by the Institute of International Education (2012), the number of international students at colleges and universities in the U.S. increased by 5.7% during 2011/2012 academic year to a record high of 764,495 students. As part of the application admission to a U.S. college or university, international students, whose English is not their primary language, must prove their English language proficiency. This requirement has contributed to a boom in the language testing industry in which the Test of English as a Foreign Language (TOEFL), administered by the Educational Testing Service (ETS), is almost the only player. ETS (2010) claimed that TOEFL has been taken by more than 23 million students, and that more than 8,500 colleges, universities and agencies in more than 130 countries have relied on its scores to make admissions.

Due to the significant role of the TOEFL scores in entrance to U.S. universities, there are many concerns about the reliability of TOEFL scores. Whether are the scores considered students' language competences or one of the predictors of their academic success? Are the TOEFL scores really reliable enough for admissions? Are international students satisfied with their language performance in academic settings and with their academic achievement, compared with the TOEFL scores they receive when they are admitted to the university? Reliable answers to these questions are clearly of great importance for admissions of international students. ETS (2012) asserted that its testing instrument has the following advantages: (1) being a highly reliable measure of English-language proficiency, (2) being based on more than 40 years of world-class, scholarly research, (3) having the most objective scoring methods, (4) being built to the highest psychometric standards. However, except ETS' claims, few objective research-based evidences have been found to affirm the reliability of this instrument.

To that end, this study was designed to examine international students' perception of their academic performance in relation with their TOEFL scores and the correlation between the TOEFL scores and international students' academic performance. The findings of the study will contribute to the body of knowledge in the field of testing and evaluation. They also provide international students with concrete information about the validity of this testing instrument. The university's enrollment offices will have

more appropriate decisions on international enrollment and instructors will be informed of international students' language competences to have better instruction methods.

## **2. LITERATURE REVIEW**

The TOEFL has been around for nearly half a century, taking several different forms before arriving at the version test-takers see today. From 1964 until 1998, test-takers took the TOEFL pBT, which was given like a traditional exam. This test was originally based on discrete-point testing, in which each test question was meant to measure one distinct content point and then a score was extrapolated to give an overall view of the test-taker's level. The test format was changed for the first time by using an integrative testing design in which each question required the test-taker to use more than one skill or piece of knowledge at a time. Each test consists of three separately timed sections with multiple-choice questions, which are listening comprehension, structure and written expressions, and reading comprehension. The time duration for the TOEFL pBT test is about three and one-half hours, included time for arrangement at the testing room. The TOEFL pBT is scored from 310 to 680. Currently, this format continues to be administered and is in use in more than 70 countries in the world including the United States, the United Kingdom, many Middle East and Asian countries where Internet is not accessible (ETS, 2012).

In 1998, the Computer-Based TOEFL Test was introduced. This new version consists of four sections: Listening, Structure, Reading, and Writing with multiple-choice questions and a 30-minute essay. It worked on an adaptive model, applied for two sections of the test – Listening and Structure: that is, the question the test-takers saw depended on their answer to the first question in a given section. If they answered the question incorrectly, they would receive easier questions and thus a lower score. In brief, the test was programmed to continuously find questions appropriate for the test-taker of all levels of performance. The time duration of the test was about three to four hours. Its scores were on a scale of 0 to 300 points, and valid for two years. Nowadays, this version is no longer used.

In 2005, ETS launched another version of TOEFL named TOEFL Internet-Based Test. The test, scoring on a scale of 0 to 120 and lasting up to 4 hours, contains four sections - Speaking, Listening, Reading and Writing. The speaking section consists of independent and integrated tasks. Test-takers are allowed to take notes during the test. The Listening section measures the understanding of attitude, degree of certainty, and purpose of spoken texts. In the *Reading section*, test takers are to answer questions to categorize information and fill in a chart or complete a summary. In the *Writing section* test takers are to type on computer, which also involves an integrated task in addition to independent tasks. The grammar skills are tested on questions and tasks in each section. Overall, according to ETS (2012), the TOEFL iBT has such main advantages as measuring the communication skills by integrating different language skills, measuring the English performance in academic settings with the academic focus of 100%, and providing valid and reliable information to support score users in considering a person's English language proficiency.

## **3. THE TOEFL SCORES: A PREDICTOR OF ACADEMIC PERFORMANCE?**

Beside what ETS accredited about its testing instrument, researchers expressed different concerns about the TOEFL as a reliable predictor of academic performance. In 1995, Simmer made a comparison in both the academic performances and the graduation rates between international students having the TOEFL scores in the 580–677 range and those having scores in the 550–579 range. The findings showed that undergraduate students with the TOEFL scores in the 550–579 range not only performed as well as students with scores in the 580–677 range but also as well as their Canadian English native counterparts. Similarly, by examining their TOEFL scores and GPA of 90 first- and second-year students' majoring in English and taking Education as a minor at Bahrain University, Al-Musawi (2001) also found out that the TOEFL scores were not a reliable predictor of students' academic success. Also taking the same approach but with 433 international students at a community college in California, USA, Ng (2008) reported that no significant relationships were found for the TOEFL scores and number of ESL courses in relation to GPA, course completion and retention rates.

Differently, in her study, Salinas (2008) concluded that the TOEFL score was a reliable predictor for international graduate students' academic success. This study focused specifically on Latin American

students with a sample of 36 graduate students enrolled at Texas A&M University-Kingsville during the period of the 2003 fall to the 2005 spring semester. In the same vein, by comparing 6516 students' TOEFL scores in their admission application and their academic records, Wait and Gressel (2009) stated that a positive statistically significant relationship was confirmed between students' TOEFL score and GPA. In addition, the TOEFL scores were also statistically significant in logistic regressions of the Comprehensive Assessment Examination pass rate and graduation rate, indicating increasing probability of success with increasing TOEFL scores.

In conclusion, there have been several studies with controversial stances on the relationship between the TOEFL scores and academic success. The controversial conclusions on this relation probably result at least from the complexity of concept of language proficiency (Canale, 1983) and partly the difficulty in measuring English language skills needed for academic success (Cummins, 1983). However, common things among these studies are that they all focused on the old versions of TOEFL, either the TOEFL pBT or TOEFL cBT. No updated studies on the same issue have been found, especially with the consideration of the latest version of TOEFL, the TOEFL iBT. In addition, previous studies only looked at the data-based relationship between TOEFL scores and GPA but not students' viewpoints to gain an insight into this relationship.

#### **4. METHODOLOGY**

Research questions:

This study attempted to answer the following questions:

1. Is there any correlation between English language proficiency levels as measured by the TOEFL scores of international graduate students and their academic performance, as measured by GPA, in the first semester at their university?
2. How do international graduate students perceive their TOEFL scores in relation to their academic performance in their first semester?

#### **Participants:**

Seven hundred and seventy six international graduate students enrolling at a Midwestern public state university in the Fall semester 2010 were invited to participate in the online survey. Six hundred and ten, accounting for 78.6%, responded to the survey. However, one hundred and forty six participants were removed from the study because they fell into one of the following categories: 1) Their responses were incomplete; 2) They were not required to submit TOEFL scores for university admissions; 3) They stayed in the U.S more than three years before getting admitted to the university. Thus, the exact number of samples selected for the study is 464, accounting for 49% of the total international students at the university in the Fall semester 2010. Of the 464 participants, 77.2% of them took the TOEFL iBT, 15.4% took the pBT version and 7.4% took the cBT. All the TOEFL pBT and the TOEFL cBT scores were converted to the TOEFL iBT scores for the purpose of the study, based on the ETS's score band.

#### **Instrument**

The instrument was a version of an online survey which consisted of two parts. The first part of which was about demographic information such as age, gender and nationality, major, GPA, TOEFL score, years of living in the United States, etc. Part Two was composed of 10 items covering participants' perceptions of their TOEFL scores in relation to their academic performance in their first semester.

#### **4.1 Findings**

**Question 1** attempted to explore whether there is a correlation between English language proficiency levels as measured by the TOEFL scores of international graduate students and their academic performance in the first semester, as measured by GPA. To answer this question, two questions were added into the survey. The first one asked about the TOEFL scores of the participants when admitted to the university and the second one looked at their GPA in their first graduate semester at the university. Because some graduate programs at the university required the TOEFL scores of at least 100 as one of the admission criteria, we decided to divide two cohorts of the sample: one cohort with the TOEFL score

*Is the TOEFL Score a Reliable Indicator of International Graduate Students' Academic Achievement in American Higher Education?*

requirement at least from 80 and above (Cohort 1) and one cohort with the TOEFL scores requirement from and above 100 (Cohort 2). The purpose of this division is to examine whether there is any significant difference in correlation between the two cohorts. If there is a significant difference in correlation between the two cohorts, Question 2 will continue to observe the two cohorts distinctively. If there is no or very little difference in correlation, Question 2 will combine two cohorts into one. Table 1 and 2 summarize Cohort 1's descriptive statistics and paired samples correlations analyzed through SPSS 17.0 program.

**Table 1.** Descriptive statistics for Cohort 1

	Mean	Std. Deviation	N
TOEFL_SCORE	88.1575	8.31722	419
GPA	3.5752	.43447	419

Both the descriptive statistics and correlation analysis revealed that there was a lack of correlation between the TOEFL scores and GPA of the cohort 1. In terms of achievement, the percentage of the TOEFL scores was 73.46%, and that of GPA was 89.38%, which is approximately 16% difference. The correlation analysis shared the lack of relationship with the values  $r = .005$ , and  $p = .916 > .05$ , which is based on how correlation coefficients are usually interpreted by Hinkle, Wiersma, & Jurs (1998). This showed that some participants had high TOEFL scores and low GPA while others had low TOEFL scores but high GPA. There were also some who achieved both high TOEFL scores and high GPA, or who got both low TOEFL scores and low GPA. Overall, there was no linear relationship between the two measures.

**Table 2.** Correlation analysis of Cohort 1

Correlations			
		TOEFL_Score	GPA_Score
TOEFL_Score	Pearson Correlation	1	.005
	Sig. (2-tailed)		.916
	Sum of Squares and Cross-products	28915.604	7.838
	Covariance	69.176	.019
	N	419	419
GPA_Score	Pearson Correlation	.005	1
	Sig. (2-tailed)	.916	
	Sum of Squares and Cross-products	7.838	78.902
	Covariance	.019	.189
	N	419	419

**Table 3.** Descriptive statistics for Cohort 2

	Mean	Std. Deviation	N
TOEFL_Score	110.2889	6.87427	45
GPA_Score	3.9278	.15219	45

**Table 4.** Correlation analysis of Cohort 2

		<b>TOEFL_Score</b>	<b>GPA_Score</b>
<b>TOEFL_Score</b>	Pearson Correlation	1	-.272
	Sig. (2-tailed)		.070
	Sum of Squares and Cross-products	2079.244	-12.541
	Covariance	47.256	-.285
	N	45	45
<b>GPA_Score</b>	Pearson Correlation	-.272	1
	Sig. (2-tailed)	.070	
	Sum of Squares and Cross-products	-12.541	1.019
	Covariance	-.285	.023
	N	45	45

As can be seen in Table 3, the descriptive statistics showed about 92% of achievement in view of TOEFL scores and nearly 98.2% of achievement in view of GPA. According to Hinkle, Wiersma, & Jurs (1998), the correlation analysis in Table 4 revealed a very small negative correlation,  $r = -.272$ ; However, this correlation was not significant,  $p = .07 > .05$ . The negative trend indicated that some participants had a high TOEFL scores but not a perfect GPA of four, and that some had a perfect GPA of four but not high TOEFL scores.

In conclusion, statistics of both cohorts revealed that there was no or very little correlation between the TOEFL scores participants acquired to get admitted to the university and GPA they achieved in the first semester. These results were further proved when the two cohorts were combined for the correlation analysis. Table 5 and Table 6 demonstrate a significant correlation between the TOEFL scores and GPA,  $r = .117$ ,  $p = .012 < .05$ . Although significant, this correlation is very low with the R-squared = .0136 ( $.117 \times .117 = .0136$ ). This signifies that only approximately 1.3% of GPA scores can be explained by the TOEFL scores.

**Table 5.** Descriptive Statistics for both cohorts 1 and 2

	<b>Mean</b>	<b>Std. Deviation</b>	<b>N</b>
<b>TOEFL_Score</b>	92.3297	11.84863	464
<b>GPA_Score</b>	3.6094	.42840	464

**Table 6.** Correlations for both cohorts 1 and 2

		TOEFL_Score	GPA_Score
<b>TOEFL_Score</b>	Pearson Correlation	1	.117
	Sig. (2-tailed)		.012
	N	464	464
<b>GPA_Score</b>	Pearson Correlation	.117	1
	Sig. (2-tailed)	.012	
	N	464	464
*. Correlation is significant at the 0.05 level (2-tailed).			

In summary, as stated above, both cohort 1 and 2 were combined into one sample to allow more variability within the TOEFL scores and the GPA. Even when put together, the correlation analysis confirmed a very low linear relationship between the TOEFL scores and the GPA of 1.3%. Therefore, in the next question, there was no distinction between Cohort 1 and 2.

**Question 2:**

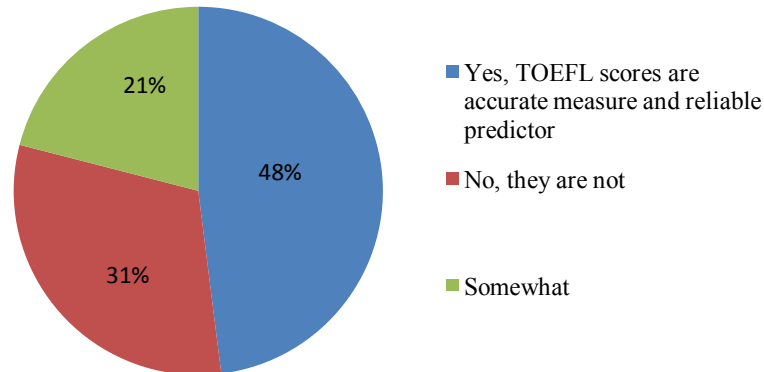
Question 2 focused on examining how international graduate students perceived their TOEFL scores in relation to their English ability through their academic performance in their first semester. To answer this guiding question, nine Likert scale statements and an open-ended question were used to elicit the information. Responses to the nine Likert scale statements were analyzed through SPSS 17.0 while responses to the open-ended question were read one by one to highlight sections relevant to the purpose of the study. The first half of the responses were tallied into three categories “ yes”, “ no”, and “ somewhat ” as showed in Figure 1, and the second half of the responses were noted in the margin and codified.

**Table 7.** Summary of quantitative survey questions

Statements	N	Min	Max	Mean	S.D.
The TOEFL scores when I was admitted to the university truly reflected my English speaking ability in my first semester at the university.	464	1	5	3.5	1.0
The TOEFL scores when I was admitted to the university truly reflected my English listening ability in my first semester at the university.	464	1	5	3.0	0.9
The TOEFL scores when I was admitted to the university truly reflected my English reading ability in my first semester at the university.	464	1	5	4.5	1.0
The TOEFL scores when I was admitted to the university truly reflected my English writing ability in my first semester at the university.	464	1	5	3.0	1.0
Please rank your confidence in your reading skill in your first semester at the university.	464	1	5	3.8	1.0
Please rank your confidence in your writing skill in your first semester at the university.	464	1	5	4.2	1.0
Please rank your confidence in your speaking skill in your first semester at the university.	464	1	5	4.2	0.9
Please rank your confidence in your listening skill in your first semester at the university.	464	1	5	4.0	0.8

These data indicated the following conclusions:

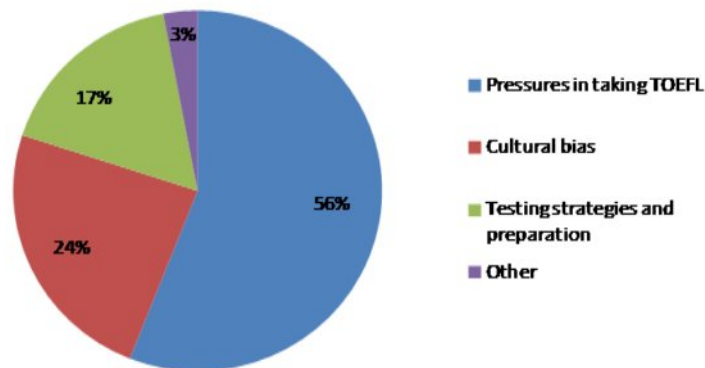
According to the participants, their TOEFL scores when they were admitted to the university truly reflected their English competences in their first semester at the university. International graduate students were confident of their English abilities in their first semester at the university.



**Figure 1.** Participants opinions on whether TOEFL scores are a reliable predictor of their academic success in the university

Combined the results of the Likert scale statements and the first part of the open ended question, it is revealed that international graduate students considered their TOEFL scores a reliable predictor for their academic success.

Regarding the second part of the responses to the open-ended question, reasons to the “Disagreement” responses were categorized into four themes while reasons to “Agreement” and “Somewhat” responses were not categorized because no specific theme could be identified. Chart 2 below summarizes reasons categorized into themes that participants provided.



**Figure 2.** Reasons why TOEFL scores are not an accurate measure and reliable predictor

As observed in the chart above, pressures in taking TOEFL which include the time pressure, making decisions in a short time, and watching the computer screens in a long period of time was considered as the main reason why its scores cannot be viewed as a reliable predictor of their academic success in the university. The following quotation of a participant provides more details how testing pressure affected his or her TOEFL score:

TOEFL is an exam where you have to be four hours, with just 10 minutes break, with all the pressure that the results imply, trying to do your best. I do not think it reflects my English skills. I was so stressed in

one point that was just impossible think, try to speak or understand the audios. Four hours in front of a computer, my sight was so tired I could not read. I pretty sure I could get a better score with a better environment. The best way for me, to improve my English skills was just to live here, talk with people and practice. In my point of view, TOEFL, as well as GRE, are good businesses.

(Female 97, France)

The second factor that makes TOEFL score not reliable is cultural bias, which according to participants involves the inclusion of many terms and concepts in the test that were foreign to their culture. A participant stated:

The content-based questions might need some background knowledge about the U.S. to be comprehended in order to get what the questions want to ask.

(Female 58, China)

Finally, seventeen percent of participants with “Disagreement” responses said that testing strategies and preparation made the TOEFL scores invalid, as one participant said:

My first TOEFL score was low because I did not know how to do it, but later I took a tutor course and learned the strategies to do it. That is why my TOEFL score was much higher than the first one. I don't think my English ability was higher in the second test but just because I know how to do it with strategies.

(Male 236, India)

## **5. DISCUSSIONS AND IMPLICATIONS**

There is no or very low significant correlation between the TOEFL scores and GPA according to the objective data of international graduates' input, so TOEFL scores cannot be regarded as an effective predictor of academic success. This finding is similar to what Al-Musawi, (2001), and Ng (2008) found. However, these previous studies focused on the TOEFL pBT and/or the TOEFL cBT while this present study mainly focused on the TOEFL iBT. Therefore, it helps make a further conclusion that irrespective of different versions of the TOEFL tests, there is not a strong predictive relationship between the TOEFL scores and academic performance as measured by GPA. The plausible interpretation of this is that TOEFL test is not an aptitude and ability test, which is designed to measure the test-takers' ability to acquire a set of skills and predict future achievements; however, it is just simply a language test used as a prerequisite for academic performance. With the finding that the TOEFL scores are not an effective and accurate predictor of international graduate students' academic performance as measured by GPA, the results could then support certain graduate admissions policies. One policy is TOEFL scores should not be used as a predictor or measurement of potential international graduate students' academic success by university admission offices, even though TOEFL scores can be a good tool to provide the admissions offices with a general idea of international students' English proficiency. Instead, prior educational records, work experience or extracurricular experience, personal essays can be used to predict their academic readiness and performance. Another policy is to gradually lower the TOEFL score requirement for graduate admissions until the quality of academic performance is significantly affected. This can help increase the number of international students enrolling in the university, which is one of the top concerns of any university. However, the later policy needs to be considered carefully because that can affect the quality of education and involve gate-keeping issues. On the one hand, the TOEFL scores may be a good gate-keeper; on the other hand, they cannot be used as a good measure of future success, just as a measure of who to let in. And if there were a direct correlation between the TOEFL scores and GPA, the admissions requirements would probably be too low – that is, it may seem that students with low TOEFL scores were being let in and they were not prepared for their academic success. In addition to these two policies, universities are recommended to study the threshold level of English abilities applied for international graduate admissions. The universities might set minimum requirements of each of English skills – reading, listening, speaking and writing skills, as well as have some other forms of assessments on international graduate students' English language rather than depend on TOEFL scores.



Different from the objective data based on TOEFL scores and GPA, subjective data based on students' opinions revealed an interesting finding. More international graduate students believed that their TOEFL scores were a reliable predictor for their academic achievement than those who did not. This remarkable finding cannot be compared to any of the findings of previous studies since no related research was conducted with the purpose of examining the perceptions of international graduate students' TOEFL scores in relation with academic performance. While there can be more than one possible interpretation of these findings, one plausible interpretation is that international graduate students believed in the power of the TOEFL as a predictor because it is a high stakes test. Another interpretation is international graduate students might assume that they were at a proficiency level which helped them perform academically. This really makes a point for a threshold level of English ability covered by the university requirements, not only as a competence at the survival level but also as the ability to perform in academic settings. Last but not the least, they viewed their academic performance as something else other than GPA. These are the issues that future research will look further at to have satisfactory explanations.

Finally, three issues that international graduate students mentioned as challenges to the reliability of the TOEFL scores are also worthy of notice. How a testing instrument can help the test takers deal with pressure, cultural bias and strategic testing preparations will be a question that test designers should take into account to make sure that the TOEFL scores really one of the important components of the admission.

#### REFERENCES

- Al-Musawi, N. (2001) *The validity of scores on TOEFL and FCE for predicting students' success at the university*. Dirasat: Educational Sciences , 28(1)
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. Richards % R. Schmidt (Eds.), *Language and communication* (2-27). New York: Longman.
- Cummins, J. (1983). Language proficiency and academic achievement. In J. Oller (Ed.), *Issues in language testing research* (108-129). Rowley, MA: Newbury House.
- Educational Testing Service. (2012). Retrieved December 10, 2012, from <http://www.ets.org/toefl/>
- Harvey, M.J. (1979). Academic achievement as predicted by the Test of English as a Foreign Language. Unpublished master's thesis, Portland State University, Portland, OR.
- Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1998). *Applied statistics for the behavioral sciences* (4<sup>th</sup> edition). New York: Houghton Mifflin Company.
- Ng, J. N. K (2008). Test of English as a Foreign Language (TOEFL): Good indicator for student success at community colleges? *Dissertation Abstracts International Section A: Humanities and Social Sciences*, Vol 69(1-A), 2008. pp. 82.
- Salinas, A. (2008). A study of TOEFL scores and first year grade point average of Latin graduate international students at Texas A&M University-Kingsville from 2003--2005 as a predictor of academic success in English. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, Vol 68(7-A), 2008. pp. 27-73.
- Simner, M.L. (1995). *Interim report to the ad hoc subcommittee on English Language Proficiency for Admission*. Report submitted to the Senate, University of Western Ontario, December 7, 1995, Exhibit II, Item 3, Appendix 1