

# Improvement of pre- and post-tests and gender differences on TOEFL scores

Saeun LEE

The TOEFL (Test of English as a Foreign Language) scores are important for international students when applying to colleges in the United States. Hence, international students spend much of their time studying and receiving extra academic support in order to obtain higher TOEFL scores. Additionally, TOEFL scores have been explored through gender; however, there is a lack of study to determine whether there is an improvement difference within gender. The purpose of this study is to analyze whether gender plays a role in higher pre and post TOEFL scores prior to and at the end of taking a TOEFL class. The mixed ANOVA was performed to answer the following research questions; 1) Is there a significant difference between students' pre and post TOEFL tests?, 2) Is there a significant difference of TOEFL scores by genders?, 3) Is there an interaction between male and female students' pre and post TOEFL test scores? The result indicated that there was a significant difference between pre and post TOEFL test scores; however, there was no significant difference between genders. Moreover, the interaction was not shown between male and female students' pre- and post TOEFL test scores.

**Key words:** TOEFL, Pre- and Post test scores, Gender

## Table of Contents

- I Introduction
- II Literature Review
  - 1 TOEFL scores and studying time
  - 2 TOEFL scores and gender differences
  - 3 Research questions
- III Methodology
  - 1 Participants and procedure
  - 2 Statistical analysis

- IV Results
  - V Discussion
  - VI Conclusion
- 

## I Introduction

The Test of English as a Foreign Language (TOEFL) exam is administered by the Educational Testing Services (ETS) and it has serviced over 27 million people worldwide (ETS, 2014). The purpose of the TOEFL exam is to assess and evaluate the individual's overall capability of writing, reading, listening and speaking English in higher education, (ETS, 2014). The TOEFL exam is necessary because it provides higher education institutions a score that can be used to evaluate the individual's level of English proficiency in different categories such as speaking, listening, reading, writing, and grammar.

Cho and Bridgeman (2012) proposed that if students have higher scores they are more likely to perform better and have a higher GPA in higher institutions. As found in this study, TOEFL scores are considered as one of the major criteria in determining the success of a student's academic performance prior to and once they enter a U.S. institution. In this respect, earning higher scores on the TOEFL is crucial for international students.

In order to obtain high scores in the TOEFL exam, there is an ongoing debate about whether or not extra studying time would affect the result of the exam, or whether spending time training testing skills would be more effective. Besides the studying time, many researchers also attempt to explore whether gender would impact students performances in TOEFL test. To clarify these two factors in TOEFL scores, the current study aims to investigate whether there is a statistical significant difference between pre and post TOEFL scores (to explore the impact of the amount of studying time) as well as explore whether there is a statistical significant between gender.

## II Literature Review

### 1 TOEFL scores and studying time

A plethora of researchers have conducted research on TOEFL from diverse aspects. One

aspect that has been popularly investigated is whether there is a relationship between hours of studying for the test and the final scores obtained. For instance, Vinke and Jochems (1993) found that when an individual devotes more hours on studying, they are more likely to earn higher scores. The outcome seems to be understandable that more time spend lead the better outcome; however, other studies represent that there is no significant different between amount of time spent on studying and the language proficiency (e.g. Pichette, 2005; Tanaka & Ellis, 2003). Tanaka and Ellis (2003) conducted a study to see if there was a difference in the TOEFL scores of Japanese students before they left for a short-term study abroad session and after they returned to their country. While studying abroad, students spent the majority of their time studying English. The results in this study did not show that there was a significant difference between the pre and post TOEFL scores. This outcome is valuable; however, there is a flaw. When students were in abroad, they did not specifically have TOEFL preparation class to study for the exam. They were studying English in a variety way. Due to the fact that students were not concentrating on studying for the TOEFL test, although their English skills have improved, it did not cause to perform better in the TOEFL exam. However, to claim the idea as a robust one, more studies should be conducted to investigate the influence of the time spent and the improvement of TOEFL scores through TOEFL preparation courses for clearer understanding of whether there is statistical significant difference.

## 2 TOEFL scores and gender differences

Gender is another variable that is frequently researched; however, there is a lack of solid research to determine whether TOEFL scores are significantly different between male and female language learners. From the perspective of psychology, the many differences occur according to the gender (Maccoby & Jacklin, 1974). Generally speaking, there is an assumption that females are more successful in language learning compared to their male counterparts. To explore the assumption, many researchers in language learning field investigated how gender differences can have an impact on students' language learning proficiency.

Hyde and Linn (1998) conducted a meta-analytical study to investigate the language proficiency differences between males and females. This study demonstrated that there is no significant difference by genders for English vocabulary proficiency. However, for reading comprehension, 15 among 21 studies reported that there is a significant difference. In other words, ten studies revealed that female were superior and five studies claimed that male

students were superior on reading comprehension. Hyde and Linn (1998) concluded that in regards to reading, speaking, and writing females were slightly more advantaged. However, since the significant differences were small, they argued that there is no gender difference in their meta-analysis study. In the same fashion, Lin and Wu (2003) also conducted a quantitative to investigate whether there are gender differences in TOEFL and concluded that there is no significant difference.

In contrast, the Educational Testing Service (ETS) found a different outcome. The study showed that female students are more advanced compared to males. (Cole, 1997). For instance, females were superior in writing and reading, although it was a small significant difference. On the other hand, male students were better in listening and comprehension, in terms of vocabulary proficiency (Brimer, 1969; Boyle, 1987). As a result, there is no salient decision whether females or males are better in language.

As shown in the previous literature, some researchers demonstrated that gender differences influences the language learning proficiency; however, others claimed that there is only little or no gender differences impact on language proficiency including TOEFL scores. However, prior to acknowledging a significant difference of English proficiency and gender differences, new research should be conducted on this topic since many studies have been conducted more than decade ago. Moreover, many studies are subjective by observing the gender difference of English proficiency. It might have caused the inaccuracy of determining the significant differences by gender. The instruments used to measure should be objective such as the use of English test scores to gain a clearer outcome. As a consequence, recent and objective proficiency data such as one of the biggest standardized test, TOEFL, would be appropriate to investigate whether there is a significant difference by gender and the test scores.

### 3 Research Questions

This study will investigate the statistical significant difference between students' pre and post TOEFL scores and gender different groups. The purpose of this study is to analyze if students obtained higher TOEFL scores at the end of an eight-week class by comparing pre and post TOEFL test scores. This study also explores if there is a significant difference by gender difference. The following research questions include:

1. Do students' scores change from the pre-test to the post- test on TOEFL?
2. Do gender differences influence student TOEFL scores?
3. Do the gender differences and pre and post-test performance interact each other on TOEFL scores?

### III Methodology

#### 1 Participants and procedure

The research sample consists of data from 32 international students who are from different countries. Their first languages include: Arabic, Japanese, Spanish, Korean, French, Turkish, and Russian. In total, 32 students took the TOEFL two times. There were 21 male and 11 female students who participated in the study. The data was collected from a language institution in the United States. The program was an 8-week intensive English program and students were taught various English skills including TOEFL preparation class. The instrument for this research is students' TOEFL scores. Students took the TOEFL PBT (paper-based test) at the end the session. Students were in class five hours per day, and were assigned homework related to the TOEFL. The total scores of TOEFL consist of three distinguished skills: listening, grammar/structure, and reading. The TOEFL PBT score ranges from 310 to 677.

#### 2 Statistical analysis

A mixed ANOVA was performed as a statistical analysis tool in this study. ANOVA is an appropriate statistical model because this study is observing significance of group differences, a dependent variable of a continuous level (TOEFL scores) and independent variables of categorical level (pre and post-test and gender). Among various types of ANOVA, a mixed ANOVA is the most appropriate statistical method because this study attempts to observe the interaction between gender differences and the two tests that were taken two different times (Lomax, 2012). The within- subject variables are pre and post TOEFL scores and between-subject variables are male and female. To answer the first research question, a statistical significant mean difference between pre and post TOEFL scores will be investigated. Next, the statistical significant mean difference between male and female TOEFL scores will be examined. Then, the effects will be checked between male and females' TOEFL scores within pre and post-test.

Before looking at the outcome of mixed ANOVA analysis, it is important to check the assumption of homogeneity of variance with Levene's test. The Levene's test table indicate that variances are homogenous for all levels of the repeated measures variables because the significance values are greater than  $p > .05$  (Pre = .484, Post = .696).

This study also needs to check the assumption of Sphericity because this study performs a within-subject test of ANOVA. The Mauchly's test of Sphericity represented a dot for

the significant value. The dot means the p value is not significant. This result meets the assumption of Sphericity since it should be nonsignificant in order to continue the analysis.

In addition, two groups have relatively close to the normal distributions, which meets the assumption of normality (See histograms below).

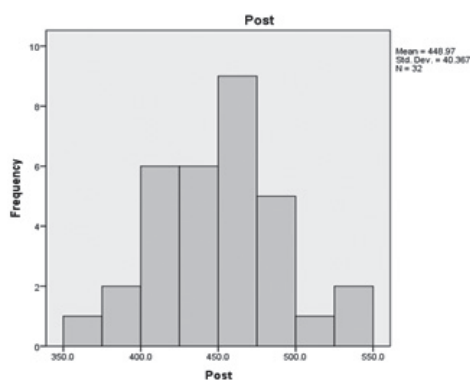
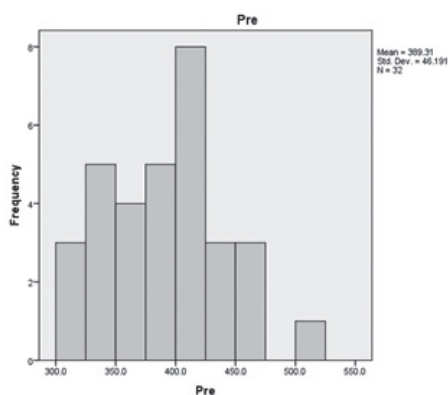


Figure 1. Mean of students' pre TOEFL scores      Figure 2. Mean of students' post TOEFL scores

Lastly, the assumption of independence has been met because the TOEFL scores were collected from the same participants on two different time points. There was an eight week time frame between taking pre and post-test. Also, gender differences yield the independent data points.

Post Hoc test is no need to investigate which groups are significantly different because there are only two groups that are comparing in this study. Post Hoc test is needed when there are more than three groups because ANOVA is an omnibus test which does not tell which groups are statistically significant different or not if there is a difference. The alpha is set to .05 with a 95% confidence interval in this study.

## IV Results

A mixed ANOVA was conducted to compare the effect of pre and post-test and gender differences on TOEFL scores as well as their interaction. Table 1 shows the means, standard deviations, and the sample sizes. For female students (N = 11), the mean of pre TOEFL test

Improvement of pre- and post-tests and gender differences on TOEFL scores (Saeun LEE)

score is 408 and standard deviation (SD) is 51.5. For male students (N = 21), the mean is 379 and SD is 46.19. For the post TOEFL scores, the mean of for females is 464 and SD is 43.5 while the mean for male students is 441 and SD is 37.34.

Table 1

*Descriptive statistics for pre and post TOEFL scores by genders*

	Gender	Mean	Std. Deviation	N
Pre	Female	408.273	51.5055	11
	Male	379.381	40.9811	21
	Total	389.313	46.1914	32
Post	Female	463.727	43.5548	11
	Male	441.238	37.3456	21
	Total	448.969	40.3673	32

The main effect for pre and post TOEFL scores yielded,  $F(1, 30) = 50.575$ ,  $p < .000$  (See table 2 below), indicating a significant difference. This outcome answered the first research question that the means of pre and post TOEFL scores are statistically significant different. The effect size for this analysis, Cohen's  $d$  is 1.9. That is, the pre and post TOEFL scores' means are different by 1.9 standard deviation. This is considered a large effect size according to Cohen's standards. In contrast, the main effect for female's and male's TOEFL scores yielded,  $F(1, 30) = 3.688$ ,  $p > .064$  (see table 3 below), indicating no significant difference. Lastly, Table 2 shows that there was no interaction effect within pre and post TOEFL scores between genders,  $F(1, 30) = .151$ ,  $p > .701$ . The effects of pre and post-test on the TOEFL scores are statistically the same between male and female TOEFL scores. In other words, when the pre and post TOEFL test scores were looked at by genders, the means of TOEFL scores had no statistical significant difference between male and female's pre and post TOEFL test scores.

Table 2

*Test of within-subjects contrast of Pre and post test scores with genders*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
TOEFL	49672.335	1	49672.335	50.575	.000
TOEFL*					
Gender	147.960	1	147.960	.151	.701
Error(TOEFL)	29464.649	30	982.155		

Table 3

*Tests of between-subjects effect of genders*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	10340712.258	1	10340712.258	4002.006	.000
Gender	9528.758	1	9528.758	3.688	.064
Error	77516.476	30	2583.883		

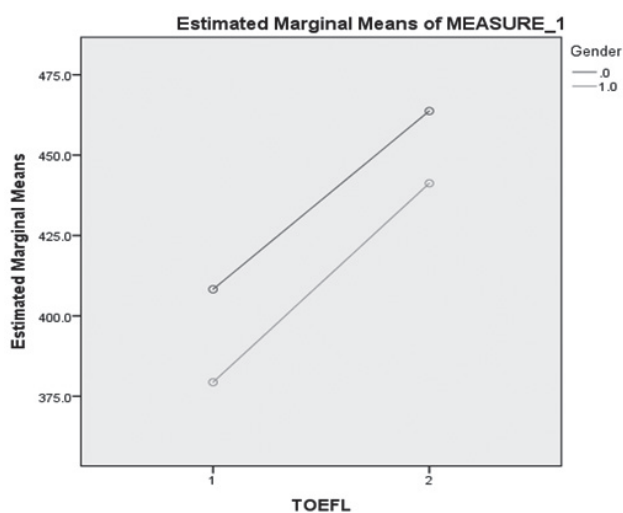


Figure3. Interaction of students' pre and post TOEFL scores between genders



The figure 3 shows that there are parallel lines of improvement from the pre to post- test performances between male and female students.

## V Discussion

Same as any standard tests, obtaining higher scores on TOEFL is crucial for international students who are planning to get admission in the U.S. higher education institutes. In this respect, this study investigated what factors make significant differences to get higher scores on TOEFL. First, the outcome of mixed ANOVA demonstrates that the time spent on studying TOEFL exam influence the students' TOEFL scores. Although there are some studies represented that there is no significant different between amounts of time spent on studying and the language proficiency (e.g. Pichette, 2005; Tanaka & Ellis, 2003; Vinke & Jochems, 1993), in terms of the TOEFL test score, the results in this study demonstrates the significant difference. Similar to the study by Vinke and Jochemes (1993), when students devote more hours to studying, they are more likely to obtain higher scores on the test. The outcome suggests that the more time an individual spends studying for the TOEFL, the better the performance.

On the other hand, the statistical analysis showed that gender differences do not yield a significant difference on students' TOEFL scores. The results of this study echoes the previous studies by Hyde and Linn (1998) insisted that reading, speaking, and writing skills may slightly be advantages for female, although they concluded that there is no gender difference. Additionally, similar to this study, Lin and Wu (2003) found that gender difference was not represented as a significant difference in the TOEFL scores. As shown in previous research, gender difference is not an important factor to explore students TOEFL scores because there is no statistical significant difference.

Another attempt of this investigation demonstrated that there is no interaction within pre and post TOEFL scores between male and female students. This trial was worthwhile to acknowledge that although there is a significant difference between pre and post TOEFL scores, the interaction with the gender difference was not existed. The TOEFL scores from male and female students were improving in a same direction toward the higher scores.

## VI Conclusion

This study examined 32 international students' TOEFL scores from the time spent on studying and gender differences. As expected, the TOEFL scores have improved from pre to the post-test and it has a statistically significant difference. However, there was no significant difference by gender differences. Furthermore, there was no interaction within pre and post TOEFL scores between male and female students. Therefore, the results of the study provide the beneficial information to international students who are planning on studying in the United States. Moreover, the educators of TOEFL can suggest that the more time a student puts into studying, the more likely they will receive a higher TOEFL score. This also helps to inform program developers, in terms of designing classes, by providing an idea of potential outcomes of students preparing for the TOEFL.

There are a few limitations. First of all, there were only 32 participants, considering a robust outcome. Considering there is a large population of international students applying to American universities every year, it is not enough to generalize the results for the TOEFL scores to the general population. Another limitation is the unequal number of male (21) and female (11) students. The equal number of students for both genders may cause different result.

## References

- Boyle, J. P. (1987). Sex Differences in Listening Vocabulary. *Language Learning*, 37: 273-284.
- Brimer, M. A. (1969). Sex differences in listening comprehension. *Journal of Research and Development in Education*, 9, 171-179.
- Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing*, 29(3), 421-442. doi:10.1177/0265532211430368.
- Cole, N. S. (1997). *The ETS gender study: how females and males perform in educational setting*. Princeton, NJ: Educational Testing Service.
- ETS. (2014). Educational Testing Services. Retrieved from <https://www.ets.org/toefl/ibt/about>.
- Halpern, D. F. (1986). *Sex differences in cognitive abilities*. Hillsdale, NJ: Erlbaum.
- Hyde, J. S., Linn, M. C., & Masters, J. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104(1), 53-69.

Improvement of pre- and post-tests and gender differences on TOEFL scores (Saeun LEE)

- Lin, J. & Wu, F. (2004). *Differential performance by gender in foreign language testing*. Paper presented at the annual meeting of the national council on measurement in education (Chicago, IL).
- Lomax, R. G. & Hahs-Vaughn (2012). *An introduction to statistical concepts* (3<sup>rd</sup> ed.). New York, NY: Routledge.
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Pichette, F. (2005). Time spent on reading and reading comprehension in second language learning. *Canadian Modern Language Review*, 62(2), 243
- Tanaka, K. & Ellis, R. (2003). Study-abroad, language Proficiency, and learner beliefs about language learning. *JALT Journal*, 25(1), 63-85.
- Vinke, A., & Jochems, W. (1993). English proficiency and academic success in international postgraduate education. *Higher Education*, 26(3), 275-285. doi: 10.1007/BF01383487.

