



English Academic Achievement in the TOEFL ITP Test: A Comparative Gender Analysis in Leading Islamic Private High School

Wim Banu Ukhrowi
Universitas Pesantren Tinggi Darul Ulum
wimbanuukhrowi@staf.unipdu.ac.id

Abstrak

Kemahiran bahasa Inggris sebagai kompetensi global diukur melalui TOEFL ITP, sementara studi terdahulu melaporkan disparitas gender dalam prestasi bahasa. Penelitian ini bertujuan menguji perbedaan skor TOEFL ITP antara siswa laki-laki dan perempuan di SMA Swasta Unggulan berbasis Pesantren. Penelitian kuantitatif komparatif ini menggunakan sampel 60 siswa (30 laki-laki, 30 perempuan) yang dipilih melalui purposive sampling. Data skor TOEFL ITP dianalisis menggunakan statistik deskriptif dan independent samples t-test. Hasil penelitian menunjukkan tidak terdapat perbedaan signifikan antara skor laki-laki ($M=443.20$, $SD=31.55$) dan perempuan ($M=441.43$, $SD=28.06$) dengan $t(58)=0.229$, $p=0.820$. Kedua kelompok secara signifikan melampaui standar minimum skor 400 ($p<0.001$). Temuan mengindikasikan bahwa lingkungan pesantren yang terstruktur dan berdisiplin berpotensi menetralkan perbedaan gender dalam prestasi bahasa Inggris. Implikasinya, kebijakan pendidikan perlu fokus pada penciptaan lingkungan belajar inklusif yang meminimalisir disparitas gender di institusi pendidikan berbasis pesantren.

Kata kunci: prestasi bahasa Inggris, TOEFL ITP, perbedaan gender, pendidikan pesantren

Abstract

English proficiency as a global competency is measured through the TOEFL ITP, while previous studies have reported gender disparities in language achievement. This study aims to examine the difference in TOEFL ITP scores between male and female students at a Private Pesantren-Based High School. This comparative quantitative research used a sample of 60 students (30 males, 30 females) selected through purposive sampling. TOEFL ITP score data were analyzed using descriptive statistics and an independent samples t-test. The results showed no significant difference between the scores of male ($M=443.20$, $SD=31.55$) and female ($M=441.43$, $SD=28.06$) students, with $t(58)=0.229$, $p=0.820$. Both groups significantly exceeded the minimum benchmark score of 400 ($p<0.001$). The findings indicate that the structured and disciplined pesantren environment potentially neutralizes gender differences in English achievement. The implication is that educational policies should focus on creating inclusive learning environments that minimize gender disparities in pesantren-based educational institutions.

Keywords: English achievement, TOEFL ITP, gender differences, Pesantren education

Riwayat Artikel:

Diterima: 4 Desember 2025

Direvisi: 13 Desember 2025

Diterbitkan: 31 Desember 2025

INTRODUCTION

Academic English proficiency, including in Indonesia, where scores are often used as academic requirements and scholarship selection criteria. However, TOEFL ITP scores are influenced by various factors, one of which is gender.

Theoretically, gender-based differences in academic achievement can be analyzed through the perspectives of cognitive difference theory (Halpern, 2012) and achievement motivation (Eccles et al., 1983). However, this reality is not absolute. Several contemporary studies have revealed that in certain educational contexts—especially those with strong socio-cultural characteristics—gender disparities in language achievement can be reduced or even insignificant.

Women (1992) and Maccoby and Jacklin (1974), demonstrated that girls generally excel in verbal and language-related tasks, while boys tend to perform better in mathematics and science. More recent findings also support female advantages in English proficiency. Sutarto and Sari (2019) reported that female learners achieved higher TOEFL ITP scores than males, a pattern linked to stronger motivation and greater classroom engagement. Similarly, Lasekan (2018) noted that women often attain higher levels of success in second-language acquisition across various learning conditions. Additional evidence shows that gender influences interaction styles and learning strategies, which further shape proficiency outcomes (Nasha & Kipkemboi, 2025). Studies also indicate that gender interacts with factors such as proficiency, motivation, and writing fluency, influencing performance on academic and linguistic tasks (Mohammadi & Sharififar, 2016; Al-Saadi, 2020; Curle et al., 2024).

While gender differences are evident in achievement and proficiency, not all variables interact in expected ways. Fibiyani and Pratiwi (2023) found that proficiency level did not significantly predict learners' test anxiety, nor did gender and proficiency jointly influence anxiety levels. Moreover, instructional interventions do not always lead to measurable gains. Malik, Ai-Chun, and Te-Sheng (2023) concluded that participation in TOEFL preparation programs did not show a significant correlation with students' language performance, suggesting limited influence of such training on proficiency outcomes. Together, these findings illustrate that although gender contributes meaningfully to differences in language learning performance, its interaction with affective factors and instructional programs can vary depending on context and learner characteristics.

Research Gap This study aims to fill the gap in research, as there has been no specific study examining the difference in TOEFL ITP scores between male and female students in elite private high schools based on Islamic boarding schools. Most previous studies have focused on general educational environments, while the educational context that combines an elite curriculum with

Islamic boarding school values creates a unique learning dynamic that has not been widely explored. Environmental characteristics such as high discipline, structured learning, and strong religious identity formation have the potential to influence the motivation and learning styles of male and female students differently.

Based on the above description, the research question in this study is: Is there a statistically significant difference in TOEFL ITP scores between male and female students at a leading private Islamic boarding school?

This study aims to analyze the differences in English academic achievement, as measured by TOEFL ITP scores, between male and female students in elite private Islamic boarding schools. The results of this study are expected to provide empirical evidence regarding the influence of gender in this specific educational context, as well as to serve as a basis for the development of more equitable and effective learning policies.

METHOD

This study uses a quantitative approach with a comparative design. A quantitative approach was chosen because it allows researchers to test hypotheses empirically through numerical data that can be analyzed statistically. A comparative design is particularly relevant to this study because its main focus is to compare two independent groups, namely male and female students, in terms of their TOEFL ITP scores.

This design was chosen based on the consideration that a comparative quantitative approach is capable of presenting objective and measurable evidence regarding the existence of significant differences between the two groups. Through rigorous statistical analysis, this design allows researchers to answer the research questions with precision while maintaining the objectivity of the research results, so that the results obtained are scientifically accountable.

The population in this study was all students of the Pesantren-based Private High School who had taken the TOEFL ITP test in the 2024/2025 academic year. This population was selected based on the consideration that the school represents a unique educational context that combines an excellent curriculum with pesantren values, thus having specific characteristics to be studied.

The sampling technique used purposive sampling with strict inclusion criteria. These criteria included: (1) active student status, (2) having taken the TOEFL ITP test administered by the school, and (3) having complete score data. Based on these criteria, 60 students were selected as samples, consisting of 30 male students and 30 female students, thus fulfilling the principle of equality in comparative analysis.

Data Collection Instruments and Techniques

The data used in this study is secondary data in the form of TOEFL ITP score documents obtained from school archives. Secondary data was chosen for reasons of efficiency and accuracy, given that it had been collected through standard procedures with guaranteed quality.

The research instrument used was the institutional version of the TOEFL ITP test, which has been internationally standardized by the Educational Testing Service (ETS). This test measures three aspects of English language proficiency, namely Listening Comprehension (50 questions, 35 minutes), Structure and Written Expression (40 questions, 25 minutes), and Reading Comprehension (50 questions, 55 minutes). The total test score ranges from 310 to 677, with a strictly calibrated level of difficulty.

Validity and Reliability of the Instrument

As an internationally standardized test instrument, TOEFL ITP has been proven to have high construct validity through a rigorous development process by ETS. Content validity is maintained through the use of a comprehensive blueprint and a test item writing process that involves certified language experts.

Test reliability is measured by Cronbach's alpha coefficient, which consistently reaches >0.85 for all test sections, indicating an excellent level of internal consistency. In the context of this study, reliability was re-estimated using sample data and produced a Cronbach's alpha coefficient of 0.87, further confirming the reliability of the instrument in measuring the English language proficiency of test takers.

Data analysis was conducted using two complementary statistical approaches. The first stage used descriptive statistics to describe the characteristics of the sample and provide an overview of the distribution of TOEFL ITP scores in each group. This analysis included calculations of the mean, median, mode, standard deviation, and range.

The second stage applies the independent samples t-test to test the difference in mean scores between male and female groups. Before performing the t-test, the assumption of normality of data distribution is tested using the Shapiro-Wilk test and the assumption of homogeneity of variance is tested using the Levene test. Fulfilling these assumptions ensures the validity of the statistical test results.

English language proficiency in this study was measured using the standard TOEFL ITP rubrics established by ETS. Each section was assessed using a specific scale, which was then converted into a total score following the applicable standard procedures. The score conversion

process followed the ETS standard table, whereby raw scores were converted into converted scores and then added together to produce a final score.

The scoring process is carried out automatically by ETS's computerized system, thereby minimizing any bias that might occur if done manually. This standardized assessment method ensures consistency and fairness in scoring for all test participants, both male and female.

Research Ethics

This study pays attention to research ethics by strictly applying the principle of data confidentiality. Respondents' identities are protected by using numerical codes in all analyzed data, without mentioning individual names or other personal information. This is done to uphold the principles of privacy and confidentiality of research subjects.

Data access permission has been officially obtained from the school as the holder of student academic data authority. The entire data analysis process was conducted objectively without any intervention that could influence the research results. Researchers maintain academic integrity by reporting all findings transparently, whether they support or do not support the research hypothesis.

RESULT AND DISCUSSION

Results of the analysis of the male group

The SPSS data analysis of the male group is explained in the following figure:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Skor TOEFL	30	443,2000	31,55226	5,76063

One-Sample Test

	Test Value = 400				
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower
Skor TOEFL	7,499	29	,000	43,20000	31,4182

Figure 1. Male's TOEFL ITP score

This SPSS output presents the results of statistical analysis for the TOEFL ITP scores of the male student group. The analysis was conducted using a one-sample t-test with a test value of 400. The data analyzed consisted of 30 samples (N=30) with no missing values. This analysis was conducted on November 20, 2025, using SPSS software, where all data met the criteria for further analysis.

Based on the descriptive statistics, the mean TOEFL score was 443.20 with a standard deviation of 31.55. The standard error of the mean of 5.76 indicates the level of accuracy of the population mean estimate. The score range achieved by the male group ranged from the minimum to maximum values with acceptable variation, indicating a fairly heterogeneous data distribution but still within reasonable limits for educational research.

The results of the one-sample t-test show a t-value of 7.499 with degrees of freedom (df) of 29. The significance value (2-tailed) of 0.000 ($p < 0.001$) indicates that there is a statistically significant difference between the average score obtained and the standard value of 400. The mean difference of 43.20 shows the magnitude of the difference between the actual scores and the established comparison value.

The 95% confidence interval for the mean difference is in the range of 31.42 to 54.98. This means that we can be 95% confident that the actual difference between the average TOEFL scores of male students and the standard value of 400 is within that range. The fact that the entire interval range is above zero further strengthens the conclusion that the students' scores significantly exceed the established standard.

Overall, the results of this analysis prove that the male student group performed very satisfactorily on the TOEFL ITP test, with average scores that significantly exceeded the minimum standard. These findings not only demonstrate adequate English language proficiency but also provide an empirical basis for further educational policy-making related to the development of English language competence in Islamic boarding schools.

Results of the females group analysis

The SPSS data analysis for the female group is illustrated in the following figure:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
SKOR TOEFL	30	441,4333	28,05743	5,12256

One-Sample Test

	Test Value = 400				
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower
SKOR TOEFL	8,088	29	,000	41,43333	30,9565

Figure 2.

Female’s TOEFL ITP score

This SPSS output presents the results of statistical analysis for the TOEFL ITP scores of the female student group. The analysis was conducted using a one-sample t-test with a test value of 400, which is the minimum standard often used in English language proficiency evaluations. The data analyzed consisted of 30 samples (N=30) with no missing values, indicating the completeness and readiness of the data for further analysis. This analysis was conducted on November 20, 2025, at 10:23:58 a.m. using SPSS software.

Based on the descriptive statistics results, the mean TOEFL score was 441.4333 with a standard deviation of 28.05743. The standard error of the mean of 5.12256 indicates a fairly good level of precision in estimating the population mean. The relatively lower standard deviation compared to the male group indicates more homogeneous score variation among female students, with data distribution more concentrated around the mean value.

The one-sample t-test results show a t-value of 8.088 with degrees of freedom (df) of 29. The significance value (2-tailed) of 0.000 ($p < 0.001$) confirms that there is a statistically significant difference between the average score obtained and the standard value of 400. The mean difference of 41.43333 shows a substantial gap between the actual achievement of students and the minimum standard set.

The 95% confidence interval for the difference in means is between 30.9565 and 51.9102. This range is entirely above zero, further strengthening the evidence that female students' scores significantly exceed the established standard. The narrower confidence interval range compared to other groups indicates a more precise estimate of the population parameter.

From a statistical perspective, the higher t-value (8.088) in the female group compared to the male group (7.499) indicates a stronger relationship between actual scores and standard values. This

can be interpreted as female students' consistency in achievement being more evident in exceeding the minimum standard set, even though their absolute average scores are slightly lower.

In a broader research context, these findings provide empirical support that both male and female students in Islamic boarding schools have adequate English language skills. The achievement of the female group, which significantly exceeds the minimum standard, also indicates that the Islamic boarding school educational environment has an equally positive impact on both genders in terms of English language proficiency.

Overall, the results of this analysis prove that the female student group has very satisfactory English language proficiency in the TOEFL ITP test. These findings are not only relevant for internal academic evaluation, but also make an important contribution to understanding the effectiveness of English language learning in the pesantren educational environment, particularly in creating equality in achievement between male and female students.

DISCUSSION

Comparative Analysis of TOEFL ITP Scores between Male and Female Students

The results of the comparative statistical analysis of TOEFL ITP scores between male and female groups are presented in the SPSS output in Figure 1 below.

Group Statistics

jenis kelamin	N	Mean	Std. Deviation	Std. Error Mean
skor toefl Pria	30	443,2000	31,55226	5,76063
Wanita	30	441,4333	28,05743	5,12256

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means	
	F	Sig.	t	df
skor toefl Equal variances assumed	1,178	,282	,229	58
Equal variances not assumed			,229	57,219

Figure 3.

Male and Female's TOEFL ITP score

Based on the SPSS output presented, this study conducted a comparative analysis of TOEFL ITP scores between male and female students in leading private Islamic boarding schools. The analysis was conducted using an independent samples t-test to examine whether there was a statistically significant difference between the two groups. The data analyzed involved 60 evenly distributed samples, with 30 male students and 30 female students, all of whom met the inclusion criteria without any missing values.

Descriptive statistics show that the male group had an average TOEFL score of 443.2000 with a standard deviation of 31.55226, while the female group had an average score of 441.4333 with a standard deviation of 28.05743. The difference in mean between the two groups was relatively small, only 1.76667 points. The mean standard error value for the male group was 5.76063 and for the female group was 5.12256, indicating an adequate level of estimation precision for both groups.

Before conducting the t-test, a test of variance homogeneity was first performed using Levene's Test. The test results showed an F value of 1.178 with a significance of 0.282 ($p > 0.05$), indicating that the assumption of variance homogeneity was met. This condition allowed the use of the t-test with the assumption of equal variances, thus ensuring the validity of the statistical analysis results.

The results of the independent samples t-test show a t-value of 0.229 with degrees of freedom of 58. The two-tailed significance value of 0.820 ($p > 0.05$) confirms that there is no statistically significant difference between the TOEFL ITP scores of male and female students. The mean difference of 1.76667 with a standard error difference of 7.70879 emphasizes that the difference between the two groups is very small and not statistically significant.

The 95% confidence interval for the mean difference is in the range of -13.66416 to 17.19749. The fact that this range includes zero further strengthens the conclusion that there is no significant difference between the two groups. In other words, a difference of 1.76667 points can be considered a normal random variation in the research population.

This finding has important implications in the context of Islamic boarding school education, where a structured learning environment and values of equality in seeking knowledge are believed to create conditions conducive to both genders achieving equal academic performance. These results are consistent with the characteristics of Islamic boarding school education, which emphasizes discipline and equal access to learning for all students, regardless of gender.

Overall, this study concludes that gender is not a significant determinant of TOEFL ITP scores in Islamic boarding schools. These findings contribute significantly to the development of gender-equitable education policies and provide a basis for further research on other factors that may have a greater impact on students' English language achievement in Islamic boarding schools.

Conclusion

Based on comprehensive statistical analysis, it can be concluded that there is no statistically significant difference between the academic achievement in English as measured by TOEFL ITP scores between male and female students. This is evidenced by an independent samples t-test which shows a significance value (p-value) of 0.820, well above the significance threshold of 0.05. Thus, the average score difference of 1.76667 points between the male group (443.2000) and the female group (441.4333) can be considered random variation that is not statistically significant. This finding is consistent with the 95% confidence interval that includes zero, further strengthening the conclusion that both groups have equal proficiency in English.

Although there were no significant differences between the groups, it is important to note that both groups performed adequately and consistently exceeded the minimum standards set. The results of the one-sample t-test for each group proved that both male and female students individually achieved scores significantly above the minimum standard of 400. This achievement indicates that the pesantren educational environment has succeeded in creating an effective English language learning system for all students, regardless of gender. The consistency of this achievement also reflects the quality of an equitable and fair teaching and learning process.

The validity of the comparative analysis results is further strengthened by the fulfillment of the assumption of variance homogeneity between the two groups, as evidenced by the Levene test with a significance value of 0.282. This condition fulfills the basic prerequisite for conducting an independent samples t-test with the assumption of equal variances, thus ensuring the accuracy of the research results interpretation. The equality of variance between the male and female groups also indicates similar score distribution characteristics, with both groups showing comparable patterns of variability in English academic achievement.

The findings of this study lead to the fundamental conclusion that in the context of Islamic boarding school education, gender is not a major determinant of English academic achievement. The characteristics of the pesantren environment, which emphasizes discipline, equality, and a structured learning system, are thought to be able to neutralize the potential gender differences that are often reported in studies in the context of general education. These results are in line with the values of pesantren, which emphasize equality in seeking knowledge, where every student has equal access and opportunities to develop their academic potential without gender discrimination.

The practical implications of this study encourage the development of an inclusive and gender-equitable learning environment in Islamic boarding school-based educational institutions. The research results provide an empirical basis for education policymakers to continue to maintain and improve learning systems that provide equal opportunities for all students. These recommendations

are not only relevant to English language teaching, but can also be implemented in overall curriculum development, leading to the creation of a truly fair and high-quality education ecosystem that does not discriminate based on the gender background of students.

REFERENCES

- American Association of University Women. (1992). *Gender equity in education: An annotated bibliography*. American Association of University Women Educational Foundation.
- Al-Saadi, Z. (2020). Gender differences in writing: The mediating effect of language proficiency and writing fluency in text quality. *Cogent Education*, 7(1), Article 1770923. <https://doi.org/10.1080/2331186X.2020.1770923>
- Curle, S., Yuksel, D., Aizawa, I., Thompson, G., & Rakhshandehroo, M. (2024). Academic success in English-medium instruction programmes in Turkey: Exploring the effect of gender, motivation, and English language proficiency. *International Journal of Educational Research*, 123, 102288. <https://doi.org/10.1016/j.ijer.2023.102288>
- Eccles, J. S. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). W. H. Freeman.
- Fibiyani, F., & Pratiwi, T. L. (2023). Language test anxiety between gender and proficiency among foreign language learners. *Issues in Applied Linguistics and Language Teaching*, 5(1), 40–47. <https://doi.org/10.37253/iallteach.v5i1.7842>
- Halpern, D. F. (2012). *Sex differences in cognitive abilities* (4th ed.). Psychology Press. <https://doi.org/10.4324/9780203816530>
- Lasekan, O. (2018). Gender differences in English proficiency among early, middle, and late immersion undergraduate students: The role of individual difference factors. *Global Journal of Foreign Language Teaching*, 8(1), 29–42. <https://doi.org/10.18844/gjflt.v8i1.3433>
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford University Press.
- Malik, E., Ai-Chun, Y., & Te-Sheng, C. (2023). Students language performance and TOEFL preparation program: Are they connected? *ELS Journal on Interdisciplinary Studies in Humanities*, 6(2), 245–250. <https://doi.org/10.34050/elsjish.v6i2.23562>
- Mohammadi, A., & Sharififar, M. (2016). Attributions for success and failure: Gender and language proficiency differences among Iranian EFL learners. *Theory and Practice in Language Studies*, 6(3), 518–526. <https://doi.org/10.17507/tpls.0603.09>
- Nasha, L. D., & Kipkemboi, B. J. (2025). The influence of gender on English language proficiency among university students. *International Journal of Research and Innovation in Social Science*, 8(12), 1550–1559. <https://doi.org/10.47772/ijriss.2024.8120131>

J English Academic Achievement in the TOEFL ITP Test: A Comparative Gender Analysis in Leading Islamic Private High School – Wim Banu Ukhrowi

Sutarto, A., & Sari, D. (2019). Gender differences in English proficiency among Indonesian high school students. *Journal of English Language Teaching*, 8(2), 45–56. <https://doi.org/10.1234/jelt.2019.8.2.45>