

## **ACADEMIC ENGLISH PROFICIENCY: TOEFL ITP COMPARISON BETWEEN UNDERGRADUATES AND GRADUATES AT AN ISLAMIC UNIVERSITY**

**Wim Banu Ukhrowi<sup>1\*</sup>, Binti Qani'ah<sup>2</sup>**

*Universitas Pesantren Tinggi Darul Ulum, Indonesia<sup>1,2</sup>*

\*Corresponding Author: [wimbanuukhrowi@staf.unipdu.ac.id](mailto:wimbanuukhrowi@staf.unipdu.ac.id)

*Article Submission:*  
15 June 2025

*Article Revised:*  
22 June 2025

*Article Accepted:*  
28 June 2025

*Article Published:*  
01 July 2025

### **ABSTRACT**

This research explores the gap in academic English proficiency between undergraduate (S1) and postgraduate (S2) students at a *pesantren*-based university, using TOEFL ITP scores as the primary benchmark. Although academic English has been widely studied in university contexts, limited attention has been given to *pesantren*-based institutions—especially in comparing the proficiency levels of students across educational stages. This study addresses that gap by offering fresh insights from a unique educational setting that combines religious and academic curricula. A quantitative comparative approach applied involved 140 students, divided equally between the undergraduate and postgraduate levels. After testing the data distribution using the Shapiro-Wilk test, results indicated non-normality ( $p < 0.05$ ), leading to the use of the Mann-Whitney U test for statistical comparison. Descriptive analysis revealed that postgraduate students scored higher on average ( $M = 468.9$ ;  $SD = 23.57$ ) than undergraduates ( $M = 447.6$ ;  $SD = 20.32$ ). Furthermore, score ranges differed noticeably: undergraduates commonly scored between 430–460, while postgraduates were more frequently in the 460–500 range. The Mann-Whitney U test confirmed a statistically significant difference between the two groups ( $U = 1890.500$ ,  $p = 0.020$ ). These findings suggest a clear disparity in academic English mastery between undergraduate and postgraduate students in this context. The results highlight the importance of implementing targeted academic interventions at the undergraduate level—such as integrated English instruction, intensive support programs, and context-based language learning. By reinforcing English proficiency early in students' academic journeys, it can prepare their graduates to succeed in global academic and professional arenas.

**Keywords:** Graduate Students, Mann-Whitney U-Test, Non-Parametric Test, TOEFL ITP, Undergraduate Students

## INTRODUCTION

English language skills are currently one of the essential competencies that students in higher education must possess, particularly in responding to the globalization era and the rapid advancement of science. English is not only a tool for international communication, but also a key to accessing scientific literature, international journals, and various learning resources required throughout academic life. Therefore, sufficient English mastery is a crucial requirement for students to compete both academically and professionally at national and international levels.

In the academic setting, the TOEFL ITP (Institutional Testing Program) is commonly used as a standardized measure of students' English proficiency, especially in areas such as reading comprehension, grammar, and listening. The TOEFL ITP offers an objective overview of students' readiness to meet the English language demands of higher education. Furthermore, it often serves as an administrative requirement for academic activities, including graduation, admission into advanced programs, and scholarship applications.

Despite the widespread use of TOEFL ITP, there remains a potential gap in English language proficiency between undergraduate and postgraduate students that requires attention. Postgraduate students are typically expected to have more advanced English proficiency due to the increased complexity of academic tasks at their level. This gap may be influenced by factors such as prior education background, English language learning experiences, and overall academic preparedness. Investigating this difference empirically is essential to confirm whether the proficiency gap is significant and warrants targeted intervention.

This study is situated within a unique educational context: a university based on the *pesantren* (Islamic boarding school) system, where educational traditions strongly emphasize religious teachings. In such institutions, English language instruction often presents distinct challenges, particularly in balancing the religious curriculum with the demands of foreign language acquisition for academic purposes. Research conducted in this setting offers valuable insights into the development of students' English skills and how educational level differences may influence TOEFL ITP outcomes.

The main objective of this study is to identify whether there are significant differences in TOEFL ITP scores between undergraduate and postgraduate students at Islamic boarding school-based universities. By understanding the results, educational institutions can formulate more effective and targeted strategies for English language instruction, particularly to prepare students for academic and professional demands in the future.

Additionally, the findings are expected to serve as a reference for academic leaders in *pesantren*-based universities in designing systematic programs to improve English language quality.

To achieve these objectives, this study adopts a quantitative comparative approach by analyzing the TOEFL ITP scores of 140 students—comprising 70 undergraduates and 70 postgraduates—at a *pesantren*-based university. Data analysis involves descriptive statistics to examine score trends and the Mann-Whitney U Test, a non-parametric statistical method, to identify significant differences between the two independent groups. This methodological approach allows for an objective assessment of the English proficiency gap across different academic levels in a unique institutional context.

## RESEARCH METHOD

This study used a quantitative approach with a comparative design. The comparative design was used because the main purpose of this study was to compare two groups based on their level of education, undergraduate (S1) and postgraduate (S2) students. With this approach, the researcher can test whether there is a significant difference between the two groups in terms of academic English competence.

The population in this study were all active undergraduate and graduate students of an Islamic boarding school-based university in Indonesia. This university has unique characteristics because it integrates a general and Islamic curriculum in its learning process. The research sample consisted of 140 respondents selected using a purposive sampling technique, consisting of 70 undergraduate students and 70 graduate students. The selection of a balanced sample size aims to avoid bias in comparison and increase the validity of the research results.

The main instrument used in this study is the official TOEFL ITP score that participants have obtained through standardized tests organized by official institutions in collaboration with universities. The TOEFL ITP test was chosen because it is objective, internationally standardized, and widely used in academic contexts to measure English language skills. The score used includes three main components of TOEFL ITP, namely Listening Comprehension, Structure and Written Expression, and Reading Comprehension.

Before conducting a comparative analysis, a data normality test was first conducted using the Shapiro-Wilk method, because the number of samples was below 200. This test is used to determine whether the distribution of TOEFL ITP score data in each group meets the assumption of normality which is a requirement in the use of parametric tests. If the test

results show that the data distribution is normal, then parametric tests (such as the Independent Samples t-test) can be used. However, if the data is not normally distributed, then the researcher will use non-parametric tests.

Based on the results of the normality test, if the data showed an abnormal distribution, the analysis was conducted using the Mann-Whitney U Test, which is a non-parametric test to compare two independent groups. This test is suitable for use in situations where the data does not meet the assumption of normality, as found in most foreign language score data in *pesantren*-based educational contexts that have heterogeneous backgrounds. The Mann-Whitney U test will show whether the difference in TOEFL ITP scores between undergraduate and graduate students is statistically significant.

For data analysis, the latest version of SPSS software (Statistical Package for the Social Sciences) was used, as SPSS provides a complete and easy-to-use tool for conducting normality tests, non-parametric tests, and interpretation of statistical results automatically and systematically. All data were analyzed with a significance level of 0.05, meaning that differences were considered significant if the p-value was smaller than 0.05.

There are two main variables in this study. The independent variable is the level of education, which is divided into two categories: undergraduate students and graduate students. Meanwhile, the dependent variable is the TOEFL ITP score, which consists of the total score of the three sub-components of English language skills. With this comparison, the researcher hopes to find out whether an increase in education level correlates with an increase in academic English proficiency.

The selection of these research methods and procedures is based on logical and empirical considerations in accordance with the research objectives. The comparative quantitative design allows the identification of real differences between two groups based on educational levels, which in turn can make an important contribution to the formulation of academic policies, especially in the *pesantren*-based higher education environment. The results of this study are expected to be the basis for the development of English learning programs that are more effective and in accordance with the characteristics of students from various levels.

## **FINDINGS AND DISCUSSION**

As an initial step in data analysis, descriptive statistics were examined to provide an overview of the distribution of TOEFL ITP scores in two groups of students, namely undergraduate (S1) and postgraduate (S2) programs at a *pesantren*-based university. This

stage is important to understand the basic characteristics of the data, such as the distribution of scores, measures of centering, and dispersion, before proceeding to more complex inferential analysis. The following data visualization and SPSS output are presented to strengthen the understanding of the results of the statistical analysis that has been carried out.

**Tabel 1.** analysis of study level

Ranks				
Jenjang pendidikan		N	Mean Rank	Sum of Ranks
Skor TOEFL	Sarjana	70	62,51	4375,50
	pascasarjana	70	78,49	5494,50
	Total	140		

A ranks analysis on the Mann-Whitney U test was conducted to compare TOEFL ITP scores between two groups of students, namely undergraduate (S1) and postgraduate (S2), at *pesantren*-based universities. The table results show that each group consists of 70 respondents. Undergraduate students had a mean rank of 62.51, while graduate students had a higher mean rank of 78.49. The total number of ranks for the undergraduate group is 4375.50, while the postgraduate group is 5494.50. This difference in mean rank indicates that graduate students tend to have higher TOEFL ITP scores than undergraduate students.

To determine whether there is a significant difference in TOEFL ITP scores between undergraduate and postgraduate students in *pesantren*-based universities, an inferential analysis using the Mann-Whitney U non-parametric test was conducted. This test was chosen because the TOEFL ITP score data of the two groups were not normally distributed based on the results of the previous Shapiro-Wilk test. Thus, Mann-Whitney U is considered an appropriate method to compare two independent groups whose data are not normal. The following Mann-Whitney U test output presents the main statistical values that will be the basis for interpreting the differences in English proficiency between the two levels of education.

**Tabel 2.** Mann-Whitney U-test interpretation

Test Statistics <sup>a</sup>	
	Skor TOEFL
Mann-Whitney U	1890,500
Wilcoxon W	4375,500
Z	-2,334
Asymp. Sig. (2-tailed)	,020

a. Grouping Variable: Jenjang pendidikan

Furthermore, inferential statistical tests were conducted using the Mann-Whitney U test, because the data were not normally distributed based on the previous Shapiro-Wilk test ( $p < 0.05$ ). From the test results, the U value = 1890.500, and the Wilcoxon W value = 4375.500, which corresponds to the number of group rankings with lower score values. In addition, the Z value = -2.334 indicates the direction of the difference between the two groups and confirms that there is sufficient deviation from a balanced distribution between the groups.

The significance value, or Asymp. Sig. (2-tailed), obtained is 0.020, which is below the 0.05 significant threshold. Thus, it can be concluded that there is a statistically significant difference between the TOEFL ITP scores of undergraduate and graduate students. This result strengthens the initial assumption that the level of education affects the achievement of TOEFL ITP scores, where postgraduate students show superior results in academic English language skills.

This finding has important implications in the context of curriculum development and foreign language learning in *pesantren*-based universities. The results of statistical tests show that an increase in education level is accompanied by an increase in English language ability, which is reflected in the TOEFL ITP score. Therefore, it is recommended that undergraduate programs pay more attention to strengthening English language skills, either through learning interventions, intensive programs, or integration in the curriculum. Thus, undergraduate students can be prepared more optimally to meet increasingly global academic and professional demands.

The findings of this study, as evidenced by the Mann-Whitney U test, reveal a statistically significant difference between the TOEFL ITP scores of undergraduate (S1) and postgraduate (S2) students ( $p = 0.002$ ). The mean rank scores further support this result, with postgraduate students achieving a higher average (78.49) compared to their undergraduate counterparts (62.51). This disparity underscores the critical role that educational level plays in enhancing academic English proficiency. The improvement observed at the postgraduate level may be attributed to greater exposure to academic English texts, the rigors of thesis writing, and active participation in international scholarly activities. These results suggest that students pursuing advanced degrees are better equipped to perform well in standardized English assessments such as the TOEFL ITP.

Moreover, the elevated TOEFL ITP scores among postgraduate students highlight the strong correlation between educational attainment and the development of academic literacy, particularly in a second language. As students advance in their academic careers,

their engagement with English-language scholarly materials increases, necessitating a deeper comprehension and the ability to produce academic content in English. This is especially evident in *pesantren*-based universities, where postgraduate students are often expected to consult international sources, author scientific manuscripts, and present their research in global academic settings. Thus, the educational level functions not merely as an administrative classification, but as a reflection of cognitive and linguistic advancement fostered through academic engagement.

In *pesantren*-based higher education institutions, these findings carry important pedagogical implications for English language instruction. While Islamic universities traditionally prioritize Arabic and Islamic studies, the current demands of academic globalization necessitate English proficiency as a key competency. Integrating English education with Islamic values and students' academic objectives is essential to prepare graduates for participation in international academic discourse. Accordingly, English instruction should adopt contextualized, task-based approaches that address students' specific academic needs within the *pesantren* environment.

A strategic recommendation derived from these findings is the need to reinforce English for Specific Purposes (ESP) programs at the undergraduate level. Early exposure to academic English is essential for students, particularly in Islamic institutions, to ensure their readiness for postgraduate studies. ESP curricula should focus on essential academic skills such as reading scholarly articles, writing research papers, and delivering presentations within academic contexts. Furthermore, these programs should be discipline-specific—for example, ESP for Islamic Studies, ESP for Nursing, or ESP for Education—so that instruction remains relevant and directly applicable to students' fields of study. This tailored approach can enhance linguistic competencies that align with their academic and professional aspirations.

To bridge the proficiency gap between undergraduate and postgraduate learners, *pesantren*-based institutions should implement comprehensive and sustainable foreign language education policies. Initiatives such as foundational TOEFL ITP training, the establishment of English-speaking communities, and support for student participation in international publication and presentation forums represent practical steps toward cultivating an academic culture of English. By fostering English proficiency from the early stages of higher education, students from *pesantren* backgrounds will be better prepared to compete not only in national academic contexts but also on a global stage. This research provides an empirical foundation for the strategic enhancement of English curricula—

particularly at the undergraduate level—as a vital investment in producing a competitive and globally engaged generation of scholars from Islamic institutions.

The observable difference in TOEFL ITP performance between undergraduate (S1) and postgraduate (S2) students is well-supported by a growing body of research. Melorose et al (2014) found that TOEFL scores are moderately yet meaningfully related to GPA at both levels, suggesting broader academic impacts of English proficiency. Ginther and Yan (2017) confirmed such predictive validity especially among Chinese graduate students. In Indonesian and Omani university contexts, Irudayasamy et al. (2020) showed that ESP courses significantly improve undergraduates' grammar, vocabulary, and specific English skills, while Prawiyata and Barus (2022) measured postgraduate TOEFL low-intermediate proficiency but emphasized its diagnostic utility. Palupi et al. (2020) reported no significant TOEFL reading score differences across learning styles, indicating that academic exposure might outweigh stylistic preferences. Muttaqin and Chuang (2023) found that ESP positively mediates between English proficiency and academic achievement in EMI settings. Noviasmy and Hasanah (2025) identified economics undergraduates' preference for discipline-aligned materials as key to effective language improvement. Fadli Perkasyah et al. (2024) confirmed that ESP-based digital materials boost vocational student engagement and proficiency, and Hartina and Syahrir (2021) emphasized learner-centered ESP needs tailored to student goals in Islamic education contexts. Yu and Champakaew (2024) affirmed that blended ESP teaching markedly improved students' language competence in specific academic fields. Finally, Golubovich et al. (2018) validated TOEFL ITP's value in multiple countries, including Indonesia, as reliable English proficiency indicators. These findings strongly align with your results—S2 students exhibit higher TOEFL ITP mean scores due to sustained exposure to academic English, reading, writing for final projects, and participation in international scholarly contexts. Especially in *pesantren*-based universities, postgraduate students regularly engage with foreign-language literature, research writing, and international discourse, reinforcing the need for richer ESP opportunities at the undergraduate level to close the English proficiency gap

## CONCLUSION

This study concluded that there is a significant difference in TOEFL ITP scores between undergraduate (S1) and postgraduate (S2) students in Islamic universities. The Mann-Whitney U test results According to the TOEFL ITP results, this study finds a statistically significant difference in academic English ability between undergraduate (S1) and



postgraduate (S2) students at an institution with a *pesantren* location. Important educational insights are provided by the findings, especially for Islamic higher education institutions. Curriculum designers and managers of study programs can provide English education that is relevant to students' academic demands from the start by acknowledging this proficiency difference. Students can develop the skills they'll need to handle future academic challenges, such as postgraduate degrees, by integrating English for Academic Purposes (EAP) at the undergraduate level,

The results of this study have high relevance for the development of English language teaching curriculum in higher education, especially those based on Islamic boarding schools. By knowing the ability gap between undergraduate and postgraduate students, study program managers can develop a curriculum that is more responsive to the linguistic needs of students from the beginning of the lecture. English learning is not enough just as a basic course, but needs to be directed to support academic abilities such as understanding scientific articles, compiling papers, and scientific presentations. The needs-based curriculum and the English for Academic Purposes (EAP) approach are highly recommended to be applied to make learning more contextual and targeted.

English learning strategies at the undergraduate level also need to be redesigned to emphasize active, collaborative and academic task-based learning processes. Students need to be given more opportunities to use English in real academic contexts, such as scientific discussions, project presentations, and paper writing. The use of technology, online platforms, and the integration of English literature in cross-course learning can also be effective strategies. With the right learning strategies, undergraduate students can be better prepared to face future academic demands, including if they continue to graduate school.

## REFERENCES

- Ginther, April, and Xun Yan, 'Interpreting the Relationships between TOEFL IBT Scores and GPA: Language Proficiency, Policy, and Profiles', *Language Testing*, 35.2 (2018), pp. 271–95, doi:10.1177/0265532217704010
- Golubovich, Juliya, Florencia Tolentino, and Spiros Papageorgiou, 'Examining the Applications and Opinions of the TOEFL ITP® Assessment Series Test Scores in Three Countries', *ETS Research Report Series*, 2018.1 (2018), pp. 1–30, doi:10.1002/ets2.12231
- Hartina, St., and Syahrir Syahrir, 'The Inappropriateness of English for Specific Purposes (ESP) with Learner's Goals: A Need Analysis on Communication and Islamic Broadcasting Program', *Elsya : Journal of English Language Studies*, 3.3 (2021), pp. 225–35, doi:10.31849/elsya.v3i3.6671

- Irudayasamy, Julius, Nizar Mohammed Souidi, and Carmel Hankins, 'Impact of an Esp Course on English Language Proficiency of Undergraduate Engineering Students: A Case Study at Dhofar University', *International Journal of Higher Education*, 9.2 (2020), pp. 309–20, doi:10.5430/ijhe.v9n2p309
- Melorse, J., R. Perroy, and S. Careas, 'A Meta-Analytic Investigation of the Predictive Validity of the Test of English As a Foreign Language (Toefl) Scores on Gpa', *Statewide Agricultural Land Use Baseline 2015*, 1.December (2015)
- Muttaqin, Syariful, and Hsueh-Hua Chuang, 'English for Specific Purposes, English Proficiency, and Academic Achievement in English-Medium Instruction: A Mediation Analysis', *KnE Social Sciences*, 2023 (2023), pp. 302–17, doi:10.18502/kss.v8i7.13259
- Noviasmy, Yessicka, 'English Language Needs Assessment for Economics Students : An ESP Analysis', 6.1 (2025), pp. 140–55
- Palupi, Asti Diah, Tono Suwartono, and Dewi Sukawati, 'Can Different Learning Styles Impact Efl Learners' Toefl Achievement?', *Premise: Journal of English Education*, 13.1 (2024), p. 51, doi:10.24127/pj.v13i1.8751
- Perkasyah, Musthafa Fadli, and others, 'Elevating English Proficiency : Crafting ESP-Based Digital Multimedia Learning Materials for Vocational Students in Indonesia', 06.02 (2024), pp. 12466–71
- Prawiyata, Yugi Diraga, and Ulian Barus, 'English Language Proficiency through the TOEFL Test Taken by UMNAW Postgraduate Students', *International Journal of English and Applied Linguistics (IJEAL)*, 2.3 (2022), pp. 342–47, doi:10.47709/ijeal.v2i3.1687
- Yu, Ruirui, and Wilawan Champakaew, 'The Impact of Blended-ESP Instructional Model on Fostering English Proficiency among Students Majoring in Traditional Chinese Medicine', 7.2 (2024), pp. 109–23