Open Access | Original Research
Journal of Intercultural Communication, 25(2), 2025 | PP: 88– 100
https://doi.org/10.36923/jicc.v25i2.1061

Do Gender and Major Affect Preparatory Year Students' Performance on English Tests Across Proficiency Levels?

Hamdallah Alhusban[1], Shakul Tewari[2]

Article History:

Received: 24-12-2024
Revision: 09-03-2025
Accepted: 26-03-2025
Publication: 01-06-2025

Cite this article as:

Alhusban, H. A., & Tewari, S. (2025). Do Gender and Major Affect Preparatory Year Students' Performance on English Tests Across Proficiency Levels?. *Journal of Intercultural Communication*, *25*(2), 88-100. doi.org/10.36923/jicc.v25i2.1061

Corresponding Author:

Hamdallah Abdulkarim Alhusban
Department of English Language, Deanship of Preparatory Year & Supporting Studies, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia. Email: halhusban@iau.edu.sa

Abstract: This study investigates the influence of gender and academic major on student performance in two English language assessments—the English Placement Test (EPT), administered by the National Center for Assessment (NCA), and the Preparatory Year Program Tests (PYPT), conducted in-house at a Saudi university. The primary aim was to examine whether these demographic variables significantly affect English language proficiency during the Preparatory Year Program (PYP), a foundational stage for university-level study. Drawing on a sample of 2,913 students, the study employed Bonferroni post-hoc analysis and factorial ANOVA to analyze differences in test scores across gender, major, and proficiency levels. The results revealed that female students consistently outperformed male students across all proficiency levels and test sub-domains. This gender-based difference may reflect broader sociocultural factors, including shifts in national education policy and evolving societal expectations. While students enrolled in health-related majors showed higher average proficiency scores, the academic major overall did not exert a statistically significant effect on general language performance. Entry-level proficiency emerged as a strong predictor of test outcomes. The findings underscore the critical role of gender and initial proficiency in shaping language achievement and suggest that standardized assessments may not fully capture the nuanced learning needs of all student groups. The study concludes with a call for gender-sensitive instructional support, differentiated curriculum design, and inclusive assessment strategies. Future research should explore qualitative dimensions such as student motivation, engagement, and study habits, particularly among male learners, to inform equitable language education practices in the Saudi context and beyond.

Keywords: English Placement Test, English Proficiency Test, English as a Foreign Language (EFL), Gender Differences, Higher Education, Saudi Learners, Proficiency Levels

1. Introduction

In the Kingdom of Saudi Arabia, English is taught as a foreign language (EFL). University students are typically required to participate in a Preparatory Year Program (PYP), designed to bridge the gap between secondary and tertiary education. Since English is the medium of instruction (EMI) in Saudi higher education, the PYP aims to develop students' language skills to enable them to comprehend and use specialized academic discourse relevant to their fields of study.

To ensure students receive instruction appropriate to their language proficiency, the PYP administers placement tests developed by the National Center for Assessment NCA, 2018). These assessments identify students whose English proficiency falls below the B2 level of the Common European Framework of Reference for Languages (CEFR). Based on their results, students are enrolled in level-specific and major-specific courses (e.g., engineering, science, medicine). For instance, engineering students are introduced to the genres and registers typical of technical disciplines, while medical-track students learn to analyze case studies and compose clinical reports. These courses are offered separately to male and female students, with curricula tailored to meet the academic demands of each college.

While these courses are both level-specific and discipline-aligned, factors such as student proficiency, gender, and chosen major significantly influence the delivery and effectiveness of the curriculum. Although the NCA's placement test and in-house formative and summative assessments offer valuable insights into students' overall English proficiency, they do not reveal how these independent variables interact to affect learning outcomes.

Some studies have shown that female students tend to perform slightly better than male students in language tests, often producing syntactically longer and more complex sentences (James, 2010; Lahuerta Martínez, 2018). Other research has found that female students are generally more proficient and more willing to communicate in English than their male counterparts (S. H. Huang & K. J. Tsai, 2024). Additionally, the academic discipline in which students are engaged appears to influence language acquisition; students exposed to content-integrated tasks grounded in real-world

[1] Department of English Language, Deanship of Preparatory Year & Supporting Studies, Imam Abdulrahman Bin Faisal University, Saudi Arabia
[2] Birla Institute of Technology & Science, Pilani (BITS Pilani), Saudi Arabia

scenarios often achieve better language learning outcomes (Juan-Garau & Jacob, 2015).

While these studies have examined the individual effects of gender or major on language proficiency, fewer have investigated how these factors interact, particularly in EFL contexts like Saudi Arabia's. Furthermore, most existing research relies on perception-based surveys or small-scale classroom observations, with limited use of authentic, large-scale institutional assessment data. This study lies in its empirical analysis of real test performance data to explore the combined effects of gender and academic major on English language proficiency across multiple proficiency levels. By examining test scores from the English Placement Test (EPT) and the Preparatory Year Program Tests (PYPT), this research aims to provide a more nuanced understanding of learner variation in a gender-segregated, EMI context, contributing to more equitable and data-informed pedagogical strategies in Saudi higher education.

This study analyzes English Placement Test (EPT) and Preparatory Year Program Tests (PYPT) scores across three proficiency levels: beginner, intermediate, and advanced. The PYPTs are level-specific, periodic, and summative assessments designed and administered in-house by the Department of English Language (DEL). In light of the research problem, the following questions are investigated:

1.1. Research Questions

1. *What is the degree of the relationship between the EPT scores and the preparatory program tests?*
2. *Is there a statistically significant difference in the EPT scores across gender, proficiency levels, and major variables?*
3. *Is there a statistically significant difference in the final test scores of the preparatory program across gender, proficiency levels, and major?*
4. *Is there a statistically significant difference in the preparatory program final test sub-domain scores across gender, proficiency levels, and major variables?*

2. Literature Review

Several studies have emphasized the role of individual factors such as gender and academic major in shaping students' language proficiency. For instance, J. F Lee (2023) found that female students consistently outperformed male students in speaking and writing proficiency, although the gap was not highly significant. In another study, M. Perrone-Bertolotti, C. Pichat, J. F Bas, Le, A. Baciu, and M. Baciu (2011) observed that females were more efficient at translating speech sounds into visual representations and read faster than males. This cognitive efficiency was attributed to greater involvement of the inferior frontal cortex (Broca's area) in females, suggesting a potential biological advantage in language processing. While there is no conclusive evidence that such processing advantages apply uniformly across all cultural and linguistic contexts, these findings point to the possibility of a biological edge in females when processing textual data. However, it remains unclear how such advantages manifest in EFL environments, particularly when the target language is learned after the so-called "critical period" (around the age of 11–14). This study seeks to examine the performance of Preparatory Year Program (PYP) students under standardized testing conditions, focusing on the interaction effects of gender and academic major on students' language proficiency.

2.1. Gender and Language Proficiency

Recent research has consistently shown that females tend to score higher than males on language proficiency assessments. James (2010) reported that female students scored significantly higher in language usage, reading comprehension, and sentence construction. Similarly, on standardized language tests such as IELTS and TOEFL iBT, females also outperformed males (James, 2010, p. 388). These results suggest that females are generally more proficient language users, at least in the context of placement testing.

Some scholars have cautioned that gender could act as a "potential source of measurement error" (Bachman, 1990, pp. 162–163), indicating its significant influence on test outcomes. It has also been observed that female students use a wider range of language learning strategies compared to their male counterparts (Green & Oxford, 1995; Psaltou-Joycey, Sougari, Agathopoulou and Alexiou, 2014; Goh & Foong, 1997). However, Milla and Gutiérrez-Mangado (2019) found that these strategic differences tend to diminish as students' proficiency levels increase.

While female students often demonstrate stronger language proficiency, this advantage does not necessarily translate to superior performance across all academic contexts. In general, females tend to excel in arts and humanities, while males report higher self-confidence in mathematics (G. Huang, 2013; Asio, 2019). Stoeger, Almulhim and Ziegler (2022) examined the gender achievement gap in Saudi Arabia and found that female students consistently outperformed males academically. They also possessed and utilized more educational and learning capital. However, within the context of this study—where male and female students have access to the same curriculum and resources—outcomes may differ.

Although gender is considered a significant factor in language learning strategies, socio-cultural theorists argue that language proficiency is not inherently gender-determined (Hyland, 2007). Instead, various factors

influence test performance, including the quality of education at the secondary and tertiary levels (Woodfield, Earl-Novell, & Solomon, 2005), personality traits that align with certain test formats (Bishop, Bulloc, Martin and Thompson, 1997), perceptions of test fairness, and levels of test anxiety (Woodfield et al., 2005).

2.2. Academic Major and Language Performance

Academic major refers to the specialization students pursue in their undergraduate studies. According to Becher and Trowler (2001), academic disciplines can be broadly categorized into "hard" and "soft" fields, each with either theoretical/pure or applied orientations. For example, disciplines like physics and mathematics are considered hard–pure, while medicine and engineering fall under hard–applied. In contrast, sociology and anthropology are examples of soft–pure disciplines, whereas education and English language teaching (ELT) are soft–applied (Lau & Gardner, 2019).

In Saudi Arabia, this disciplinary distinction is often rephrased in terms of compliance: "highly compliant" disciplines typically refer to STEM fields, while "low compliant" disciplines encompass the humanities and social sciences (Friedrich, Prøitz, & Stensaker, 2016, p. 879). High-compliance disciplines tend to project concrete and measurable learning outcomes (Laird, Shoup, Kuh & Schwarz, 2008), whereas arts and humanities are more abstract and less aligned with standardized qualification frameworks (Friedrich et al., 2016).

This distinction between technicality and abstraction has been widely discussed in language education (Wignell, 1998). Disciplinary discourses, whether logical or ideational, and their modes of expression can significantly influence students' language use over time. Exposure to diverse academic registers and genres has linguistic consequences that may affect proficiency development. To better understand how gender and academic major influence students' language performance across specific sub-domains, this study posed the research questions outlined earlier.

The impact of language proficiency on academic performance also varies by discipline. For example, language plays a more central role in the social sciences than in technical fields like engineering (S. Y. Lee, 2024). Although limited research has examined the direct interaction between disciplinary background and language proficiency, existing studies indicate that language skills are crucial to academic success. Sánchez-Pérez (2021) suggests that language proficiency is a key factor in acquiring disciplinary knowledge. In light of these insights, this study compares student performance across multiple testing conditions. The following section outlines the methodology used.

3. Methodology

This study compared the performance of 2,913 Preparatory Year Program (PYP) students under two assessment conditions. The sample included all PYP students enrolled during one academic year. First, we examined whether there was a correlation between results on the Preparatory Year Program Tests (PYPT) and the English Placement Test (EPT). The second objective was to quantitatively investigate the effects of gender, proficiency level, and academic major on students' performance in both tests. The third research question focused on the relationship between independent variables and the results of the PYPT. Finally, we aimed to explore whether sub-domain scores varied across the different variables.

All students in the study took the EPT administered by the National Center for Assessment (NCA) and four internally developed, periodically administered PYPTs (Test 1, Test 2, Test 3, and Final Test). Based on their EPT scores, students were categorized into level-specific groups, considering both gender and major. The students were admitted into four academic majors: engineering, science, health-1 (medicine, dentistry, pharmacy and applied medical sciences), and health-2 (nursing and public health). Within each major, students were further classified into three English proficiency levels: beginner, intermediate, and advanced. Table 1 presents the frequency distribution of students across the variables—gender, proficiency level, and academic major.

Table 1: Frequency distribution of students across the variables.

Variables	Gender	Frequencies	Percentages (%)
Gender	Male	1368	47
	Female	1545	53
Proficiency Level	Beginner	678	23.3
	Intermediate	1538	52.8
	Advanced	697	23.9
Major	Engineering	513	17.6
	Science	1344	46.1
	Health-1	776	26.6
	Health-2	280	9.6

Source: Calculated by the author

The study group included 1,368 males (47%) and 1,545 females (53%). Based on EPT scores, 23.3% of students were classified as beginners (A1 level), 52.8% as intermediate (A2 to B1), and 23.9% as advanced (B1+

and above). In terms of academic majors, 46.1% of the students enrolled in the Science track, followed by Health-1 (26.6%), Engineering (17.6%), and Health-2 or Nursing (9.6%).

The sampling strategy adopted in this study involved a population-based approach, whereby all students enrolled in the PYP during the academic year 2020-2021 were included in the analysis. This comprehensive inclusion ensured that the dataset reflected the full demographic and academic diversity of the student population, including variations in gender, major, and proficiency level. Such an approach enhanced the internal validity of the study and allowed for robust statistical comparisons without the constraints of sampling bias. While the study was conducted at a single institution, the large sample size (N = 2,913) and full-population coverage offered a strong foundation for analyzing performance trends and generalizing findings within the institutional context. This sampling design was well-suited to the descriptive-comparative aims of the study and supported the use of factorial ANOVA and post-hoc tests to examine subgroup differences.

3.1. Development and Validation of PYP Tests

To ensure the validity and reliability of the in-house PYPTs, a structured process was followed during test development and validation. For content validity, the Department of English Language (DEL) collaborated with Oxford University Press to develop a bank of multiple-choice and open-ended items, using the Common European Framework of Reference for Languages (CEFR) as the guiding framework (UCLES, 2018; Alderson, 2002; Little, 2007).

Test item developers—including both native and non-native English speakers—were selected based on their experience in language teaching and assessment. Each item writer completed a 30-hour workshop led by a certified trainer from Oxford University Press and the Oxford Teachers' Academy (ELAS International, Examination and Assessment Clinic Phase 1 and 2, 2019). Participants were trained in developing listening, reading, and writing test items aligned with CEFR levels. They were given sufficient time and standard rubrics to develop an item bank.

All submitted items were reviewed by an expert panel comprising members from the DEL, Oxford University Press, and external consultants. Items were evaluated for accuracy, clarity, complexity, and relevance to course content. A blind review was then conducted by a screening team of four to five teachers representing different proficiency levels—including at least one native and one non-native English speaker—to assess inter-rater reliability. Items were rated for alignment with textbook content and assigned CEFR levels (from basic to advanced). Final selection was conducted by the Exam Committee and the DEL.

Typically, test items were randomly distributed across four test versions. Overall, 80% of the items were multiple-choice (with four options), while 20% were open-ended, requiring students to write essays based on course-related topics.

The EPT, by contrast, is a standardized, centrally designed test administered by the NCA. According to the NCA's policy documents, the EPT is CEFR-aligned and evaluated using the framework's rubrics. Table 2 outlines the classification criteria used by the evaluation team to align test items with CEFR levels from Basic to Advanced (see Appendix 1 for cut scores).

Table 2: Classification criteria used to align tests with CEFR

EPT	0 – 45 Beginner	46 – 79 Intermediate	80 - Above Advanced
CEFR	A 1	A2 – B1	B1+and Above

Source: Calculated by the author

3.2. Pedagogic Intervention and PYPT Tests

Over a 32-week academic period, students were enrolled in two English language courses: ENG 101 and ENG 102. ENG 101 focused on general English, while ENG 102 introduced academic English relevant to students' disciplinary specializations. Students who scored 80 or above on the EPT studied English for 12 hours per week, while those who scored below 45 received 20 hours of instruction per week.

The Final PYPT assessed overall proficiency at the B2 level of CEFR. In contrast, Tests 1, 2, and 3 evaluated students' level-specific proficiency at earlier stages in the course. These progressive tests deliberately excluded language components beyond the designated proficiency level to maintain assessment accuracy and validity.

4. Data Analysis

As previously mentioned, the operationalization of the PYP test constructs maintained consistent structural and content parameters across all assessments. Test validity was reinforced by ensuring consistency in item types, scoring rubrics, and the number of raters evaluating the answer sheets. This section presents a detailed analysis of the test results and explores the interaction effects across different variables.

Across all test conditions, results revealed a moderate to strong correlation, which can likely be attributed to the tests' adherence to CEFR-based rubrics. As shown in Table 3, the bivariate correlations between the EPT and the PYP tests were statistically significant at the 0.01 level, with coefficients ranging from 0.601 to 0.807—

indicating moderate to strong positive relationships. The strongest correlation was observed between the EPT and the Final PYP test, followed by Test 3 and the Final Test, and then by Test 1 and Test 2. The high correlation between the EPT and the Final Test may be due to both assessments targeting summative achievement aligned with the CEFR B2 level, whereas the earlier tests emphasized formative assessment. As a result, only the Final Test scores were included in the correlation and comparison analyses with the EPT; Test 1, Test 2, and Test 3 were excluded from further analysis.

Table 3: Correlation between the EPT and PYPTs.

	EPT	Test-1	Test-2	Test-3	Final Test
EPT	1				
Test-1	.683**	1			
Test-2	.601**	.791**	1		
Test-3	.642**	.752**	.758**	1	
Final Test	.807**	.775**	.769**	.793**	1

** *Correlation is significant at the 0.01 level (2-tailed).* Source: Calculated by the author

The bivariate correlations between sub-domains of the EPT and those of the Final Test were also significant at the 0.01 level. Correlation coefficients ranged from .482 to .622, suggesting small to moderate relationships among sub-domains. The listening component of the Final Test showed relatively weaker correlations with the EPT sub-domains, while the structure domain of the EPT demonstrated stronger correlations with the Final Test sub-domains. These findings indicate that listening assesses distinct language skills that are not as strongly captured by the EPT, whereas structural knowledge appears to contribute consistently across different types of language assessments.

4.1. What is the degree of the relationship between the EPT scores and the preparatory program tests?

4.1.1. EPT scores across genders, proficiency levels, and majors

Factorial ANOVA analyses were conducted to examine the main and interaction effects of gender, proficiency level, and academic major on EPT scores. Table 4 presents the factorial ANOVA results for variance within the EPT scores.

Table 4: ANOVA results for the EPT test scores.

Source	Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Gender & Proficiency level	556.768	2	278.384	3.833	.022	.003
Major & proficiency level	692.421	6	115.404	1.589	.146	.003
Gender & Major & Proficiency level	980.165	6	163.361	2.249	.036	.005

Source: Calculated by the author

The results indicated that gender had a statistically significant effect on EPT scores ($F(1, 2889) = 4.159$, $p = .042 < .05$, $\eta^2 = .001$). The main effect for proficiency level was also significant, with an F-ratio of $F(2, 2889) = 2271.279$, $p < .001$, indicating substantial differences among proficiency groups. In contrast, the main effect of major was not statistically significant ($F(3, 2889) = 1.022$, $p = .382 > .05$). Given these results, a Bonferroni post-hoc test was conducted to explore the mean differences between the categories of the variables. The post-hoc results, shown in Table 5, provide mean difference comparisons across gender and proficiency levels with respect to EPT scores.

4.2. Is there a statistically significant difference in the EPT scores across gender, proficiency levels, and major variables?

4.2.1. EPT scores across gender, proficiency levels, and major variables

The Bonferroni post-hoc comparisons revealed that the mean EPT score for male students was significantly different from that of female students, with female students scoring higher, regardless of their academic major.

Additionally, mean scores increased progressively from the beginner to advanced proficiency levels. The differences between each pair of proficiency levels were also statistically significant, supporting the validity of the predefined EPT proficiency level cut-scores (see Table 5).

Figure 1 displays the EPT scores as a function of proficiency level across academic majors. The figure shows that the mean scores for each proficiency level remained relatively consistent across the different majors. This suggests that EPT scores are primarily influenced by proficiency level rather than major.

Table 5: ANOVA Comparisons of the EPT scores across gender and proficiency levels.

					Bonferroni's Post-hoc test (*p* values)		
Variable	Group	*n*	Mean	*SD*	Male	Female	
Gender	Male	1368	59.032	.431			
	Female	1545	60.843	.777	.001		
Proficiency	Group				Beginner	Intermediate	Advanced
Level	Beginner	678	33.762	.333	-		
	Intermediate	1538	60.931	.294	< .001	-	
	Advanced	697	85.119	1.256	< .001	< .001	-

Source: Calculated by the author

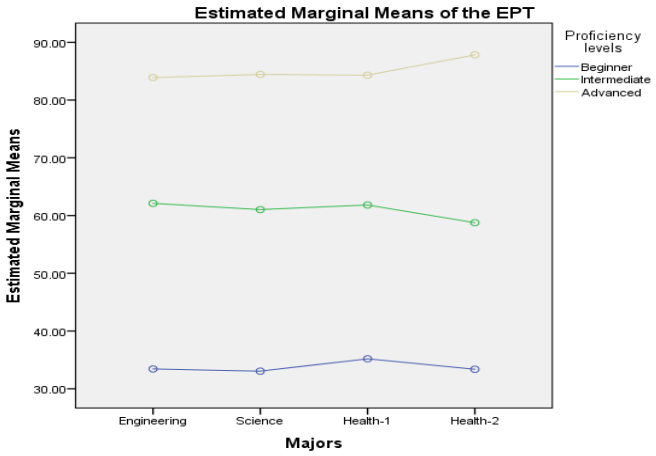

Figure 1: The EPT scores as a function of proficiency levels and major.

Figure 2 illustrates EPT scores based on gender and academic major. Visual inspection of the data suggests that the effect of gender was most pronounced in the Health-2/Nursing major. Across all majors—except for Science—female students consistently outperformed male students. In Science, male students slightly outperformed females, although this effect was relatively small compared to differences observed in other majors.

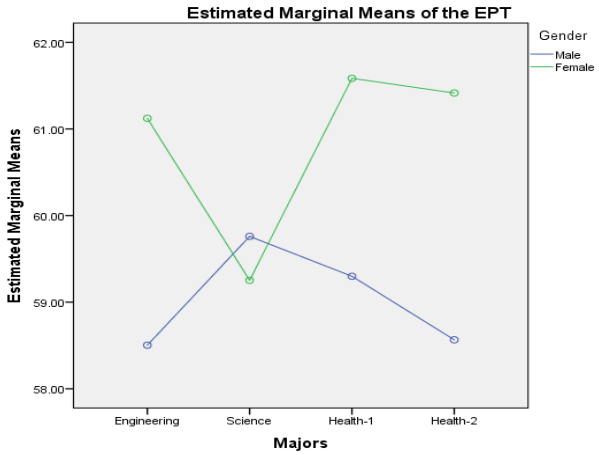

Figure 2: The EPT scores as a function of gender and majors.

As shown in Figure 3, which includes data for all three proficiency levels, the female students' mean scores were higher than the male students across engineering, health 1, and health 2 majors. Moreover, the interaction effect of gender and major on the EPT score was relatively minor, while it was somewhat more prominent for the health-1 group. Unlike the beginner level, the effect of gender on the EPT scores differed across the majors for both the intermediate and advanced proficiency levels. Moreover, the effect of gender was much larger for the health-2/Nursing group than for other majors at beginner and advanced proficiency levels. In comparison, this effect was much larger for engineering at the intermediate level.

4.3. Is there a statistically significant difference in the final test scores of the preparatory program across gender, proficiency levels, and major?

4.3.1. PYPT final test scores across genders, proficiency levels, and majors

The factorial ANOVA analyses compared the main and interaction effects of gender, proficiency levels, and major types on the final test of the preparatory program. Table 6 provides the factorial ANOVA results of the main and interaction effects of the variables on the preparatory program final test scores.

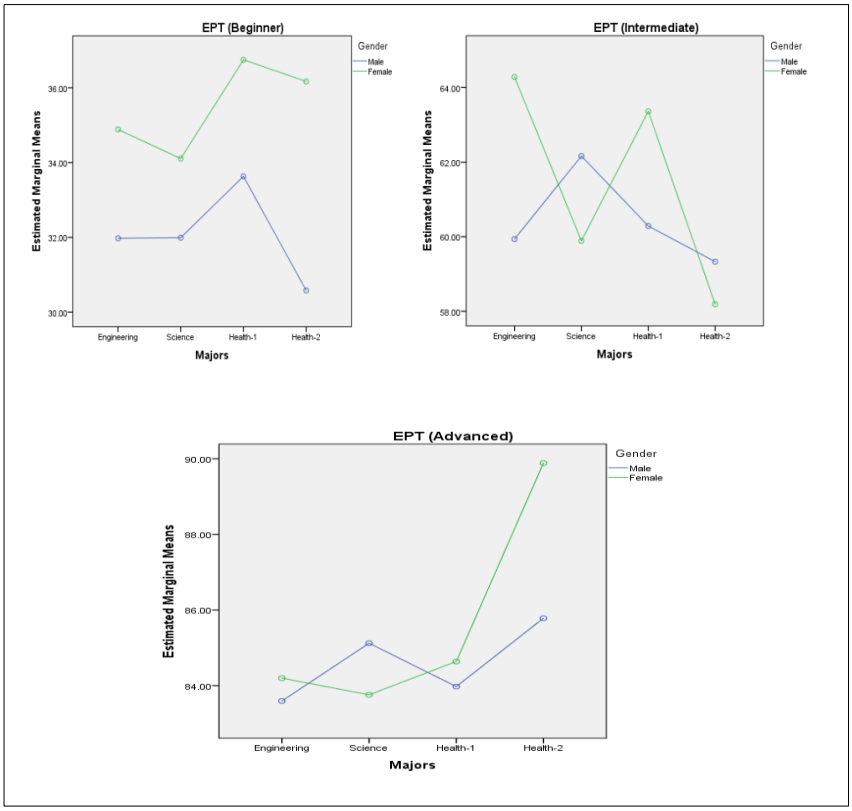

Figure 3: The EPT scores as a function of gender, majors, and proficiency levels.

The factorial ANOVA results in Table 6 indicate that the main effect of gender on the Final Test scores was significant (*F(1,* 2889*)*= 83.339*, p*=.00< *.05,, η2* = .028), indicating a significant difference between male and female students' scores in the Final Test. Likewise, the main effect for proficiency level yielded an F ratio of *F(2,* 2889*)*= 664.093*, p=0.00),* and the main effect for major variables yielded an F ratio of *F(3,* 2889*)*= 25.076*, p=0.00,* indicating a significant difference between proficiency levels and major types. Moreover, proficiency level had the largest effect size compared to other independent variables. Unlike EPT scores, the interaction effect of major types on the Final Test was statistically significant, along with the main effect of gender and proficiency levels. The Bonferroni post-hoc test was conducted to investigate the significant mean difference between the categories of the independent variables.

Table 6: ANOVA results for the preparatory program Final Test scores.

Source	Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Gender & Proficiency levels	511.188	2	255.594	8.421	.000	.006
Gender & Major	353.510	3	117.837	3.882	.009	.004
Proficiency levels & Major	632.687	6	105.448	3.474	.002	.007
Gender & Proficiency levels & Major	413.740	6	68.957	2.272	.034	.005

Source: Calculated by the author

4.4. Is there a statistically significant difference in the preparatory program final test sub-domain scores across gender, proficiency levels, and major variables?

4.4.1. Difference in the preparatory program final test sub-domain scores across gender, proficiency levels, and major variables

Factorial ANOVA analyses were conducted to examine the main and interaction effects of gender, proficiency level, and academic major on the Final Test scores of the Preparatory Year Program (PYP). Table 6 presents the factorial ANOVA results for these variables.

As shown in Table 6, the main effect of gender on the Final Test scores was statistically significant ($F(1, 2889) = 83.339$, $p < .001$, $\eta^2 = .028$), indicating a meaningful difference between male and female students. Similarly,

the main effect for proficiency level yielded an F-ratio of $F(2, 2889) = 664.093$, $p < .001$, and the main effect for academic major yielded $F(3, 2889) = 25.076$, $p < .001$, indicating statistically significant differences across both variables. Among the independent variables, proficiency level had the largest effect size.

Unlike the EPT scores, the interaction effect of major was statistically significant in the Final Test scores, along with the main effects of gender and proficiency level. A Bonferroni post-hoc test was conducted to further investigate the significant mean differences among the categories of the independent variables.

Table 7: ANOVA Comparisons of the Final Test scores across categories of gender, proficiency levels, and major.

					Bonferroni's post-hoc test (p values)		
Variable	**Group**	***N***	**Mean**	***SD***	**Male**	**Female**	
Gender	Male	1411	31.748	.200			
	Female	1625	34.837	.273	.001		
Proficiency	Group				Level-1	Level-2	Level-3
Level	Beginner	678	24.436	.370	-		
	Intermediate	1538	33.836	.183	< .001	-	
	Advanced	697	41.606	.296	< .001	< .001	-
Major	Group				Engineering	Science	Health-1
	Engineering	513	34.890	.370	-		
	Science	1344	33.927	.183	.153	-	
	Health 1	776	34.044	.321	< .001	< .001	-
	Health 2	280	30.309	.430	< .001	< .001	< .001

Source: Calculated by the author

Figure 4 shows that the effect of gender was notably larger for Engineering than for Health-1. Consistent with the EPT results, female students consistently outperformed male students across all majors, with a relatively smaller effect size for Health-1.

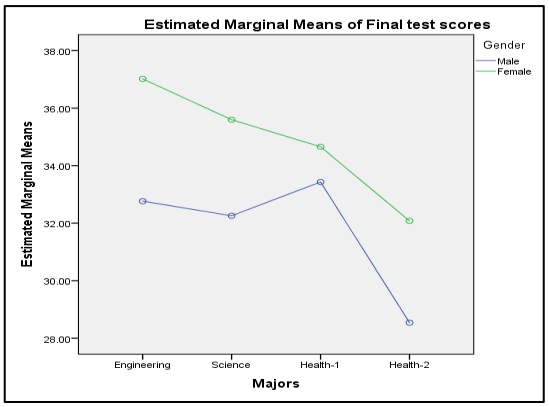

Figure 4: The final test scores as a function of gender and major.

Figure 5 presents the Final Test scores by proficiency level across majors. It shows that mean scores for each proficiency level (beginner, intermediate, and advanced) remained fairly consistent across majors. However, Engineering students consistently achieved the highest scores, while the Health-2/Nursing group had the lowest scores across all proficiency levels.

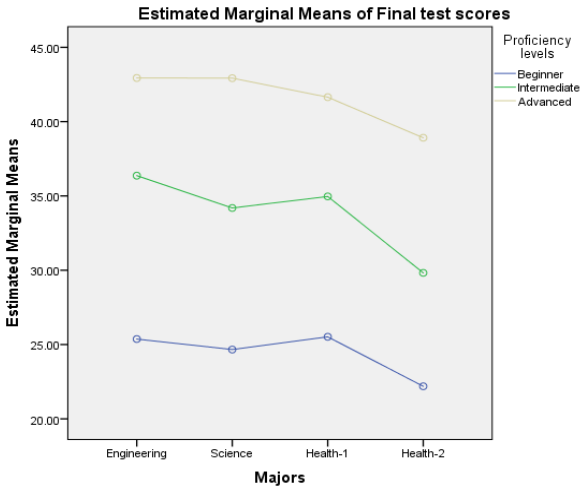

Figure 5: The final test scores as a function of proficiency levels and majors.

Figure 6 shows that, overall, female students scored higher than male students across all majors and proficiency levels in the Final PYPT. However, the mean scores of students in the Health-2 (Nursing) group were the lowest among all majors for both genders. The interaction effect of gender and major on Final Test scores was relatively smaller in this group. Notably, female students at the intermediate proficiency level scored higher than their male counterparts across all majors, with the gender effect again being most pronounced in Engineering. At the advanced proficiency level, female students also outperformed male students across all majors, with the gender effect being strongest in Engineering and least pronounced in Health-1. Overall, it is evident that female students outperformed male students across all proficiency levels and majors, with particularly strong effects observed in Engineering..

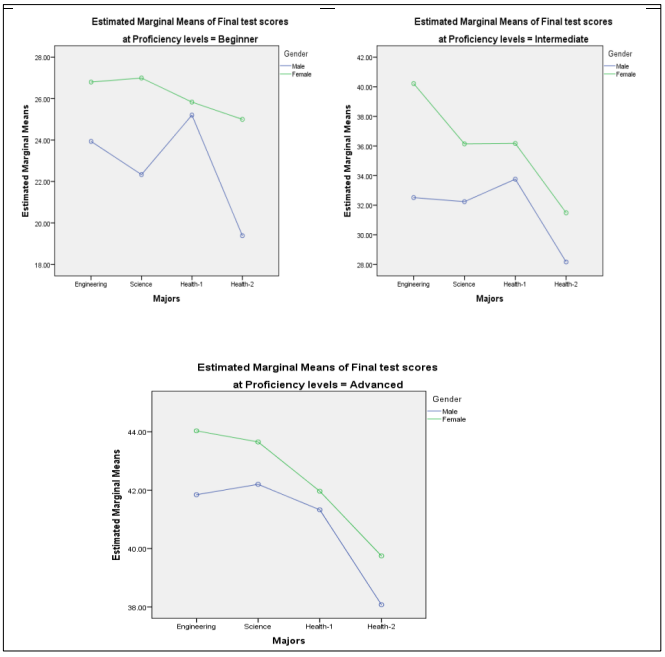

Figure 6: The final test scores as a function of gender, proficiency levels, and majors

4.4.2. Performance on sub-domains across gender, proficiency levels, and major

As previously mentioned, the final PYPT test assessed four sub-domains: reading, vocabulary, grammar, and listening. Writing was primarily evaluated through portfolio assessments. This section examines the main and interaction effects of gender, proficiency level, and academic major on each of the sub-domains of the Final Test using factorial ANOVA methods (see Table 8).

Table 8: ANOVA results for subdomains in the final test.

Language Aspect	Source	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Reading	Gender * Proficiency levels	48.641	2	24.321	7.371	.001	.005
	Gender * Major	58.895	3	19.632	5.950	.000	.006
	Proficiency levels * Major	101.912	6	16.985	5.148	.000	.011
	Gender * Proficiency levels * Major	41.346	6	6.891	2.088	**.051**	.004
Vocabulary	Gender * Proficiency levels	111.782	2	55.891	10.550	.000	.007
	Gender * Major	23.919	3	7.973	1.505	**.211**	.002
	Proficiency levels * Major	113.328	6	18.888	3.565	.002	.007
	Gender * Proficiency levels * Major	38.964	6	6.494	1.226	**.290**	.003
Grammar	Gender * Proficiency levels	30.873	2	15.436	3.394	.034	.002
	Gender * Major	56.437	3	18.812	4.136	.006	.004
	Proficiency levels * Major	34.370	6	5.728	1.259	**.273**	.003
	Gender * Proficiency levels * Major	17.264	6	2.877	.633	**.704**	.001
Listening	Gender * Proficiency levels	4.394	2	2.197	.907	**.404**	.001
	Gender * Major	5.393	3	1.798	.742	**.527**	.001
	Proficiency levels * Major	58.957	6	9.826	4.057	.000	.008
	Gender * Proficiency levels * Major	30.303	6	5.050	2.085	.052	.004

The non-significant results were written in bold. Source: Calculated by the author

The factorial ANOVA results presented in Table 8 show that the main effect of gender on reading scores was statistically significant. The Bonferroni post-hoc test revealed that female students had higher mean scores than male students, regardless of their academic major or English proficiency level. Additionally, the post-hoc results indicated that the mean differences between majors were statistically significant, except between the Engineering and Science majors.

Similarly, vocabulary scores varied significantly across gender, proficiency levels, and academic majors. All interaction effects among the independent variables were statistically significant, except for the interaction between gender and major (gender × major). Among all majors, Health-1 students achieved the highest mean vocabulary scores, followed by students in Science and Engineering.

Regarding grammar, the ANOVA results indicated that the interaction effects between gender and proficiency level, and between gender and major, were statistically significant. Bonferroni post-hoc comparisons confirmed that female students scored significantly higher than male students. With respect to the major, Health-1 students again achieved the highest mean grammar scores, followed by Science and Engineering students.

In terms of listening performance, statistically significant differences were observed between male and female students. The factorial ANOVA results also revealed significant effects for proficiency level and academic major. However, unlike the other sub-domains, only the interaction effect between major and proficiency level (major × proficiency level) was statistically significant. The Bonferroni post-hoc test showed that female students outperformed male students in listening. Among the academic majors, Science students had the highest mean listening scores, followed by those in Engineering and Health-1.

5. Discussion

This study examined the relationship between gender, academic major, and English language proficiency among Preparatory Year Program (PYP) students by analyzing performance on two assessment types: the English Placement Test (EPT) and the internally developed Preparatory Year Program Tests (PYPTs). The results indicate strong and statistically significant bivariate correlations between the EPT and PYPT scores, with coefficients ranging from .601 to .807. The strongest correlation was observed between the EPT and the Final Test ($r = .81$), suggesting that both assessments measure similar constructs of language proficiency, particularly at the summative level. These results provide evidence of test reliability and internal alignment with CEFR-based assessment standards.

The findings confirmed a consistent gender effect across all tests and sub-domains: female students outperformed male students, regardless of their academic major or proficiency level. This result aligns with prior research suggesting that female learners tend to demonstrate higher proficiency in language tests (Green & Oxford, 1995; Goh & Foong, 1997; James, 2010;). Socio-cultural perspectives on education (Hyland, 2007) help explain this trend, emphasizing that gendered social roles and expectations shape learning behaviors and outcomes. In the Saudi context, this is particularly salient. Female students often face strong social and familial pressure to succeed academically, as education is seen as a key pathway to empowerment and opportunity. National initiatives, such as Saudi Vision 2030 (Vision 2030, 2016), which aims to increase female workforce participation, may further reinforce this motivation. Recent data indicate that women's participation has already surpassed the original 30% target, reaching 35.4%, demonstrating a societal shift that likely impacts academic motivation and performance.

The study also found significant main effects of proficiency level and academic major on Final Test performance. Although major was not a significant factor in the EPT, it did influence performance on the PYPT Final Test. Notably, engineering students achieved the highest mean scores, while students in the Health-2 (Nursing) major scored lowest. These results may reflect differences in the cognitive and linguistic demands of various disciplines or variations in instructional practices across departments. While prior studies have broadly categorized disciplines as "hard" or "soft" (Becher & Trowler, 2001; Friedrich et al., 2016), our findings suggest that these distinctions may have nuanced implications for language performance. Future studies could benefit from exploring how disciplinary language practices—such as genre exposure, reading load, and writing requirements—shape learner outcomes.

Among the sub-domains, listening scores showed the greatest variability and the weakest correlation with EPT sub-domains. This suggests that listening proficiency may rely on factors not fully captured by general language placement tests, such as familiarity with accents, context-specific cues, and real-time processing skills. In contrast, grammar (structure) showed the strongest correlation between tests, indicating that structural language knowledge may be more consistently assessed and developed across test formats. The moderate correlations between other sub-domains (reading, vocabulary, compositional analysis) suggest that while they are aligned with CEFR outcomes, student performance varies across these skill areas, possibly due to differences in instructional emphasis or learning strategies.

These findings carry practical implications for language educators and curriculum designers. The consistent gender gap, particularly in productive and receptive skills, highlights the need for gender-responsive pedagogy. Male learners may benefit from targeted support, such as differentiated instruction, strategic learning interventions, or mentoring programs aimed at increasing motivation and engagement. Additionally, discipline-specific

instruction may need to be more deliberately integrated into English courses to support students in fields with distinct linguistic requirements.

While the use of large-scale, standardized assessments provides practical and reliable insights into student proficiency, educators should remain cautious of washback effects. When test performance becomes the primary goal of instruction, there is a risk of narrowing curricular focus and diminishing the richness of the language learning experience. Therefore, assessment practices must be complemented by formative feedback mechanisms, peer-review activities, and training in metacognitive strategies that encourage deeper learning.

6. Limitations and Future Research

A notable limitation of this study is the exclusion of potentially influential variables such as rater effects, classroom instruction quality, and students' individual learning histories. While the large sample size enhances the generalizability of findings, future studies should incorporate qualitative approaches, such as interviews, classroom observations, and learner diaries, to provide richer insights into the factors underlying performance differences. Additionally, Future studies could explore the motivation of both genders through personalized support. In particular, schools in Saudi Arabia might provide counseling services that address gender-specific pressures and assist students in navigating their academic goals in relation to their future societal roles. Furthermore, schools could promote diverse learning strategies tailored to the strengths of both genders, such as project-based learning and collaborative activities. It is crucial that future research incorporates qualitative data, as it may offer insights into how male students can enhance their proficiency levels and academic achievement to align with those of their female counterparts.

7. Conclusion

This study aimed to examine the relationship between gender, academic major, and student performance in two English language assessments—the English Placement Test (EPT) and the Preparatory Year Program Tests (PYPTs)—within the context of a Saudi university's Preparatory Year Program. The analysis focused on identifying significant differences in test outcomes across demographic and academic variables. The results provided consistent evidence that gender, particularly female student performance, significantly influenced language proficiency outcomes across all proficiency levels and test formats. While academic major showed some variation in specific domains, it did not exert a significant effect on general language proficiency.

The study contributes to the growing body of research highlighting gender-based differences in language achievement, particularly in EFL contexts, and reinforces the importance of addressing learner diversity in curriculum and assessment design. Although the study did not directly examine the underlying causes of performance differences, its findings point to the necessity of exploring broader learner-related factors—such as motivation, engagement, and study strategies—to better inform English language teaching practices. While the use of a large, representative cohort enhances the internal validity of the findings, the study acknowledges that variables such as classroom instruction quality, rater reliability, and testing conditions were not included in the analysis. Overall, the study underscores the need for more nuanced, learner-centered approaches to language education in preparatory programs and provides a foundation for further pedagogical reflection and research.

Acknowledgement Statement: The authors wish to extend their heartfelt gratitude to Dr. Chintalapalli Vijayakumar for his invaluable feedback, insightful critique, and meticulous review of the manuscript, which has greatly enriched its quality and clarity.

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contribution statements: HA developed the conceptual framework and methodology. ST performed the data analysis. Both authors actively contributed to the discussions, interpretation of results, and the writing of the manuscript..

Funding statements: As there was no external funding received for this research, the study was conducted without financial support from any funding agency or organization.

Data availability statement: Data is available at request. Please contact the corresponding author for any additional information on data access or usage.

Disclaimer: The views and opinions expressed in this article are those of the author(s) and contributor(s) and do not necessarily reflect JICC's or editors' official policy or position. All liability for harm done to individuals or property as a result of any ideas, methods, instructions, or products mentioned in the content is expressly disclaimed.

References

Alderson, J. C. (Ed.). (2002). *Common European Framework of Reference for Languages: Learning, teaching, assessment: Case studies*. Strasbourg, France: Council of Europe.

Asio, H. T. (2019). Characteristics of male and female spoken interaction. *Asian EFL Journal, 23*(3.4), 141–162.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Open University Press/SRHE.

Bishop, K. N., Bulloc, K., Martin, S., & Thompson, J. J. (1997). Students' perceptions of coursework in the GCSE: The effects of gender and levels of attainment. *Educational Studies, 23*(2), 295–310. https://doi.org/10.1080/0305569970230212

Friedrich, P. E., Prøitz, T. S., & Stensaker, B. (2016). Disciplining the disciplines? How qualification schemes are written up at study program level in Norwegian higher education. *Teaching in Higher Education, 21*(7), 870–886. https://doi.org/10.1080/13562517.2016.1184138

Goh, C., & Foong, K. (1997). Chinese ESL students' learning strategies: A look at frequency, proficiency, and gender. *Hong Kong Journal of Applied Linguistics, 2*(1), 39–53.

Green, J. M., & Oxford, R. (1995). A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly, 29*(2), 261–297. https://doi.org/10.2307/3587625

Huang, G. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28*(1), 1–35. https://doi.org/10.1007/s10212-011-0097-y

Huang, S. H., & Tsai, K. J. (2024). English for general purposes (EGP) as a means to improve undergraduate freshmen's language proficiency in Taiwan. *International Journal of Educational Development, 106*, 102997. https://doi.org/10.1016/j.ijedudev.2024.102997

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing, 16*(3), 148–164. https://doi.org/10.1016/j.jslw.2007.07.005

James, C. L. (2010). Do language proficiency test scores differ by gender? *TESOL Quarterly, 44*(2), 387–398. https://doi.org/10.5054/tq.2010.222215

Juan-Garau, M., & Jacob, K. (2015). Developing English learners' transcultural skills through content- and task-based lessons. *System, 54*, 55–68. https://doi.org/10.1016/j.system.2015.04.017

Laird, T. F., Shoup, R., Kuh, G. D., & Schwarz, M. J. (2008). The effects of discipline on deep approaches to student learning and college outcomes. *Research in Higher Education, 49*(6), 469–494. https://doi.org/10.1007/s11162-008-9088-5

Lau, K., & Gardner, D. (2019). Disciplinary variations in learning styles and preferences: Implications for the provision of academic English. *System, 80*, 257–268. https://doi.org/10.1016/j.system.2018.12.010

Lahuerta Martínez, A. C. (2018). Analysis of syntactic complexity in secondary education ELF writers at different proficiency levels. *Assessing Writing, 35*, 1–11. https://doi.org/10.1016/j.asw.2017.11.002

Lee, J. F. (2023). How processing instruction and structured input research has operationalized individual differences: How much do they matter? *Ampersand, 11*, 100127. https://doi.org/10.1016/j.amper.2023.100127

Lee, S. Y. (2024). Theoretical framework for the empirical analysis of English medium instruction competence outcomes. *Social Sciences & Humanities Open, 9*, 100893. https://doi.org/10.1016/j.ssaho.2024.100893

Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal, 91*, 645–685. https://doi.org/10.1111/j.1540-4781.2007.00627_2.x

National Center for Assessment. (2018, April 18). *EPT - English Placement Test*. Retrieved from http://www.qiyas.sas

Perrone-bertolotti, M., Pichat, C., Bas, J. F. Le, Baciu, A., & Baciu, M. (2011). Functional MRI evidence for modulation of cerebral activity by grapheme-to-phoneme conversion in French, and by the variable of gender. *Journal of Neurolinguistics, 24*(4), 507–520. https://doi.org/10.1016/j.jneuroling.2011.03.003

Psaltou-Joycey, A., Sougari, A.-M., Agathopoulou, E., & Alexiou, T. (2014). The role of age, gender, and L1 strategies in the L2 strategies of primary school children in Greece. *11th International Conference on Greek Linguistics: Selected Papers*, 1436–1448.

Sánchez-Pérez, M. d. M. (2021). Predicting content proficiency through disciplinary-literacy variables in English-medium writing. *System, 97*, 102463. https://doi.org/10.1016/j.system.2021.102463

Stoeger, H., Almulhim, N., & Ziegler, A. (2022). Correspondence heuristic and filter-empowerment heuristic: Investigating the reversed gender achievement gap in a sample of secondary school students in Saudi Arabia within the framework of educational and learning capital. *Education Sciences, 12*(11), 811. https://doi.org/10.3390/educsci12110811

UCLES. (2018, April 18). *Cambridge Assessment English*. Retrieved from https://www.cambridgeenglish.org/search-and-validation/fitness-for-purpose/#conceptual-perspective

Vision 2030. (2016). *Saudi Vision 2030*. https://www.vision2030.gov.sa/

Wignell, P. (1998). Technicality and abstraction in social sciences. In J. R. Martin & R. Veel (Eds.), *Reading science: Critical and functional perspectives on discourses of science* (pp. 299–328). Routledge.

Woodfield, R., Earl-Novell, S., & Solomon, L. (2005). Gender and mode of assessment at university: Should we assume female students are better suited to coursework and males to unseen examinations? *Assessment & Evaluation in Higher Education, 30*(1), 35–50. https://doi.org/10.1080/0260293042003243887

About the Author(s).

Hamdallah Abdulkarim Alhusban is a Lecturer in English Language at Imam Abdulrahman bin Faisal University. He holds an M.Phil. in English and Applied Linguistics from the University of Cambridge and an MA (Hons) in English-Arabic Translation and Interpreting from Heriot-Watt University in Edinburgh. Hamdallah has extensive experience teaching English for Academic Purposes (EAP) in the United Kingdom, Saudi Arabia, and Oman. His research interests include foreign language acquisition, corpus linguistics, e-learning, and English-Arabic simultaneous interpretation. He has published his work in peer-reviewed journals such as *Electronic Journal of e-Learning (EJEL)*, *Journal of Intercultural Communication*, and *TESOL International Journal*.

Dr Shakul Tewari is a visiting Assistant Professor in the Department of Humanities and Social Sciences at BITS, Pilani, Hyderabad Campus. Previously, Dr Tewari worked as a lecturer at Imam Abdulrahman bin Faisal University from Her research interests include technology-enhanced learning, mobile-assisted language learning, second language (L2) listening, and primarily leveraging technological tools to enhance language acquisition and teaching methodologies. She is a recipient of the Fulbright FLTSA scholarship. Besides teaching, she has also worked as a Curriculum Developer for an EdTech company and as an IELTS Examiner with IDP.

Appendix A. Cut-score and frequency distribution of proficiency levels based on the EPT scores.

	Range	Frequency	Percent (%)
Beginner	0-45	678	23.3
Intermediate	46-79	1538	52.8
Advanced	80-UP	697	23.9