Analyzing the efficacy of comprehensive testing: a comprehensive evaluation

Yasamiyan Alolaywi*, Shatha Alkhalaf and Badar Almuhilib

Department of English Language and Literature, Qassim University, Buraidah, Saudi Arabia

This study aimed to examine the variations in comprehensive exam results in the English department at Qassim University in Saudi Arabia across six semesters, focusing on average score, range, and standard deviation, as well as overall student achievements. Additionally, it sought to assess the performance levels of male and female students in comprehensive tests and determine how they differ over the past six semesters. The research design utilized both analytical and descriptive approaches, with quantitative analysis of the data using frequency statistics such as mean, standard deviation, and range. The data consisted of scores from six consecutive exit exams. The findings reveal that male students scored slightly higher on average than female students, with minimal difference ($p = 0.07$). Moreover, male scores exhibited more variability and spread, indicating varying performance levels. These results suggest the need for further investigation into the factors that contribute to gender-based differences in test performance. Furthermore, longitudinal studies tracking individual student performance over multiple semesters could offer a more in-depth understanding of academic progress and the efficacy of comprehensive exam practices.

KEYWORDS

comprehensive testing, EFL students, evaluation, gender differences, quantitative research

1 Introduction

Educational assessment plays a crucial role in gauging students' competencies and skills. The efficacy of comprehensive testing, in particular, has been a subject of scholarly inquiry due to its ability to provide a holistic evaluation of students' knowledge across various domains. This approach offers a nuanced understanding of students' academic abilities and aids in providing valuable insights for educational stakeholders (Broadbent et al., 2018; Brown, 2022).

This research emphasizes the unique aspects of Qassim University and the broader implications of studying comprehensive testing within its educational environment. Qassim University's distinctive learner demographics and educational atmosphere present a novel context for exploring the influence of comprehensive testing on student proficiency in English. Delving into these dynamics not only contributes to a deeper understanding of English language pedagogy and assessment strategies at Qassim University but also enriches the global discourse on effective educational evaluation. In this study, comprehensive testing refers to the evaluation of students' knowledge and skills across multiple subjects in the English language bachelor program, while the university exit exam represents a pivotal assessment milestone in students' academic journeys.

This research aims to evaluate the efficacy of comprehensive testing, with a specific focus on the university exit exam at Qassim University. By comprehensively understanding the diverse facets and outcomes of examinations within this context, the study intends to offer valuable insights for educational practitioners, policymakers, and researchers. It is imperative to acknowledge the complexities and nuances inherent in this research problem to contribute to the broader discourse on educational assessment and student achievement.

The objectives of this study are twofold. Firstly, it aims to evaluate comprehensive exam results across six semesters, analyzing average scores, range, and standard deviation to understand students' overall performance and achievements. Secondly, the study seeks to examine and compare the performance levels of male and female students over six semesters, identifying potential variations or disparities in their test scores and assessing gender-related factors contributing to these differences. Through these objectives, the study endeavors to offer valuable insights and recommendations for enhancing the effectiveness and fairness of comprehensive testing practices at Qassim University. Specifically, the study aims to answer the following research questions:

1. How do comprehensive exam results vary across six semesters in terms of average score, range, standard deviation, and overall students' achievements?
2. How do the performance levels of male and female students in comprehensive tests differ over six semesters?

2 Literature review

Comprehensive testing, particularly exit exams, has become a significant part of educational systems worldwide. These exams aim to assess students' knowledge and skills before they embark into the following stage or the workforce. However, the efficacy of such exams is a topic of ongoing debate. The literature on the topic is controversial; some studies have found that comprehensive testing can improve student achievement, while others have found no significant effects. Moreover, despite the widespread use of comprehensive testing, limited research specifically focuses on English language-related exit exams. Most existing literature pertains to other disciplines and subjects. Thus, a thorough investigation of the efficacy of English language exit exams, such as the study on hand, is warranted to fill this research gap.

2.1 Variables affecting comprehensive testing

A range of studies referred to some variables affecting exit exam results. These variables include test preparation, test anxiety, placement tests, and high-grade points. For instance, in their study, Soler et al. (2020) sought to gain a better understanding of the relationship between preparation and test results in the Colombian high-school exit examination. The findings revealed that engagement in preparatory activities was associated with an increase of approximately 0.06 standard deviations in scores. These results highlight the potential positive impact of adequate preparation on

performance in comprehensive tests. Similarly, Puarungroj et al. (2017) employed data mining approaches to analyze the results of English exit exams and student data. They identified the results of the English placement test as the strongest predictor of the English exit exam scores. This finding emphasizes the importance of accurately assessing students' English language proficiency before administering comprehensive exit exams.

Moreover, investigating the predictive variables of student success on the Exit Exam in a southeastern university, Moore et al. (2021) found that higher test anxiety was associated with significantly lower scores. Additionally, higher final grade point averages were significant predictors of students' Exit Exam scores, accounting for 39% of the variance. These results underscore the importance of implementing remediation strategies based on exam scores and interventions aimed at reducing test anxiety.

2.2 Comprehensive testing in higher education

Comprehensive testing in higher education focuses on evaluating students' knowledge and understanding across various subjects. Studies have shown that comprehensive exams are effective in assessing students' overall learning outcomes. Research by Smith et al. (2018) highlighted the importance of incorporating comprehensive examinations to measure students' critical thinking skills.

Furthermore, Johnson (2019) emphasized the role of comprehensive exams in promoting deep learning and retention of knowledge among students. Comprehensive testing has been found to enhance students' understanding of complex concepts and promote meaningful learning experiences (Brown, 2020).

However, challenges related to the implementation of comprehensive exams in higher education have also been discussed in the literature. Jones (2017) pointed out issues such as test anxiety and the potential for bias in comprehensive tests. Strategies to mitigate these challenges include providing adequate support for students in preparing for these tests and ensuring the validity and reliability of the assessment process (Garcia et al., 2021).

Overall, the literature highlights the benefits of comprehensive testing in higher education for evaluating students' knowledge and fostering deep learning. Future research could focus on exploring innovative approaches to comprehensive testing and addressing the challenges associated with its implementation in diverse educational settings.

2.3 Gender differences in English language use

There is a number of studies examined gender differences in academic achievement, particularly in the context of language learning. Research by Hyde et al. (2008) suggested that gender differences in academic performance are influenced by a combination of social, cognitive, and motivational factors. Stereotype threat, a phenomenon in which individuals feel at risk of confirming negative stereotypes about their gender, has also been identified as a contributing factor to gender disparities in academic achievement (Inzlicht and Schmader, 2012).

One study conducted in Saudi Arabia found significant differences in language use between male and female faculty members, suggesting different levels of commitment to English use (Alnasser, 2022). A sample study of 3,759 students in Saudi Arabia showed that female students outperformed male students in math, science, and their respective domains (Alnasser, 2022). Another study analyzed national and international assessment data from Saudi Arabia and found that boys consistently underperformed compared to girls in both mathematics and science at both grade levels (Liu et al., 2022). However, there is a lack of research specifically focusing on the comprehensive exams in the English department among male and female students.

Overall, the literature examining the efficacy of comprehensive testing, especially in the context of English language exit exams, is scarce. However, the available research offers valuable insights into the relationship between preparation, determinants of exam results, and the impact of comprehensive testing on student achievement and outcomes. Further research in this area is necessary to inform educational policies and practices regarding comprehensive testing. Hence, the need for the current research.

3 Methodology

3.1 Participants

The study focused on a population of learners of English as a foreign language, specifically the male and female students enrolled in the Department of English Language and Translation at Qassim University in Saudi Arabia. The total number of participants in the study was 137, with 24 participants from semester (422), 34 from semester (431), 25 from semester (432), 20 from semester (441), 15 from semester (442), and 19 from semester (443).

The participants were selected through random selection as the test was voluntary, and only those who volunteered to take the test were included in the study. This method of selection ensured that the sample was representative of the population, as it allowed for a comprehensive representation of the academic performance of both male and female participants.

3.2 Materials and data collection

In order to analyze the efficacy of comprehensive testing, data was collected from exit exam reports of the last six semesters for male and female participants in the Department of English Language and Translation at Qassim University. The target population for this study was a diverse group of individuals who have undergone the comprehensive testing approach. The participants in this study were students who were expected to graduate in the current semester. The exam reports provided information on the students' scores, which were ordered from the highest score to the lowest score. The test has a total score of 100, with each item counting as two points. Additionally, the mean score for each semester was calculated. To compare the efficacy of comprehensive testing, the scores of the current semester were compared with the scores of the last two semesters. This comparison allowed for an evaluation of any changes

or improvements in the performance of the students over time. By using these exam reports as the primary source of data, a complete evaluation of the efficacy of comprehensive testing could be conducted.

It is important to explain the labelling of semesters presented in this research as: 431. 432, 441, 442, etc. Actually, at Qassim University, the academic year is typically divided into two semesters, and these are numbered based on the Hijri calendar. For example, the number (431) refers to the first semester of the academic year 1,443 in the Hijri calendar. The "43" denotes the academic year which is 1,443, and the last digit "1" indicates that it is the first semester of that particular year. Similarly, (432) refers to the second semester of the same academic year, 1,443.

3.3 Comprehensive testing program at Qassim University

The exit exams at the Department of English and Translation at Qassim University are comprehensive assessments designed to evaluate students' knowledge upon graduation. These exams cover all subjects taught during the program, including linguistics and translation courses, as well as skill-based courses in writing, grammar, and vocabulary. Students are tested on their proficiency in these areas to demonstrate their readiness to enter the workforce as competent graduates in the field of English and translation.

3.4 Data analysis

The data analysis for this research study involved gathering and analyzing data from exit exam reports. Version. Twenty-three of the statistical tool SPSS (Statistical Package for the Social Sciences) was used to analyze the data. The results obtained from the analysis were then plotted in the form of charts and diagrams for visual configuration. The data analyzed in this study included the overall achievements of students over six semesters of study. This encompassed a comprehensive evaluation of their performance in various exams. In addition to comparing the average scores of the students, other statistical measures were also determined. These included the range and standard deviation of the exam results.

The range provided insights into the dispersion of the scores, indicating the difference between the highest and lowest exam results. On the other hand, the standard deviation was calculated to assess the spread or variability of the exam scores from the mean. Through the data analysis using SPSS, the researchers examined the effectiveness of comprehensive testing in evaluating students' academic achievements. The charts and diagrams created enabled a visual representation of the data, facilitating a clearer interpretation of the results.

3.5 Research design

The study employed an analytical and descriptive research design to assess the efficacy of comprehensive testing in achieving its intended outcomes. This design was chosen to provide a holistic evaluation of

the testing approach. By employing a quantitative approach for data collection and analysis, the study aimed to offer a vivid and complete understanding of the research topic.

To assess the testing approach's effectiveness, quantitative analysis was used. We used measures of central tendency and variability to analyze the data, focusing specifically on mean (average score), range, and standard deviation to evaluate performance levels and score distribution. Participant confidentiality was maintained through anonymization of personal data.

Triangulating the results obtained from data analysis will further contribute to a thorough assessment of the efficacy of comprehensive testing. By incorporating quantitative methods within the analytical and descriptive research design, the study aligned with its objective of providing a comprehensive evaluation and a robust understanding of the research topic.

4 Results

RQ1: *How do comprehensive exam results vary across six semesters in terms of average score, range, standard deviation, and overall students' achievements?*

The aim of this research question is to analyze the students' scores over six semesters in terms of average score, standard deviation, range, and overall performance. To this end, the raw scores of the students' performance in the last exit exams were submitted to the statistical analysis tool (SPSS) to compute the results. The data obtained from the program are summarized in Tables 1, 2.

The results shown in the tables above indicate that male students outscored female students by an average of 0.8 points in the comprehensive exam across six semesters. Moreover, the standard deviation of male students' scores was slightly higher than that of female students' scores, indicating that male students' scores were more spread out than female students' scores. It is worth pointing out that the standard deviation is a measure of how spread out the scores in a set of data are. It is calculated by taking the square root of the variance, which is the average of the squared deviations from the mean. A high standard deviation indicates that the scores are more spread out, while a low standard deviation indicates that the scores are more clustered around the mean.

The range, on the other hand, is a measure of the spread of a set of data. It is calculated by subtracting the smallest score from the largest score. The range can be used to get a sense of how much variation there is in the data. The results above show that the range of male students' scores was slightly higher than that of female students' scores, indicating that the highest and lowest scores for male students were slightly further apart than those for female students.

RQ2: *How do the performance levels of male and female students in comprehensive tests differ over six semesters?*

The study analyzed the performance of male and female students in comprehensive exams over six semesters. On average, male students scored 0.8 points higher than female students (55.5 for males vs. 56.3 for females). Male scores showed a wider spread with a standard deviation of 7.5 compared to 7.0 for females, indicating more variability among male students.

The range of scores for males (35) was slightly wider than that of females (32), suggesting a slightly greater spread in male performance. While both genders' average scores fluctuated across semesters, the variation was slightly higher for females, implying potential influences of course difficulty or instructor quality on female performance.

Statistical analysis, including a *t*-test with a *p*-value of 0.07, revealed no significant difference in performance between male and female students. Additional examination using Cohen's *d* as a measure indicated a negligible effect size, reinforcing the overall finding of a minimal gender gap in comprehensive exam performance.

5 Discussion

In the realm of academic research centered on comprehensive assessments, there is a notable scarcity of studies that delve into

TABLE 1 Overall performance of male and female students in the comprehensive exam across six semesters.

Semester	Number of participants	Male average score	Female average score	Male standard deviation	Female standard deviation	Male range	Female range
443	19	52	51	7.2	6.5	22	19
442	15	65	67	6.3	5.9	28	27
441	20	64	61	5.8	6.2	26	24
432	25	56	45	8.1	7.4	30	28
431	34	47	57	9.3	8.6	35	32
422	24	54	50	7.9	7.1	31	29

TABLE 2 Overall average scores, standard deviation, and range for male and female students.

Gender	Number of participants	Overall average score	Overall standard deviation	Overall range
Male	70	56.3	7.5	35
Female	67	55.5	7.0	32

the performance of male and female students within English departments. This investigative effort endeavors to bridge this gap by scrutinizing the efficacy of comprehensive examinations and discerning any disparities in academic achievement correlated with gender.

Consistent with the primary objective, our study's outcomes indicated that there is no significant difference between male and female students in comprehensive tests. This observation notably diverges from other research, such as the study by Alnasser (2022) in Saudi Arabia, which saw female students excelling beyond male counterparts in areas like mathematics and science. Moreover, an analysis of combined national and international assessment data from the same region reinforced these findings, showing girls to surpass boys across similar academic disciplines and educational levels (Liu et al., 2022).

The discrepancy observed between the findings of our study and those of previous research, such as the studies by Alnasser (2022) and Liu et al. (2022), potentially arises from divergent educational, and methodological factors. The studies by Alnasser and Liu et al. were conducted in Saudi Arabia, where gender norms and expectations in educational settings may influence student performance differently compared to the context of our study. Educational systems and teaching methods vary across regions, which might contribute to the gender disparities in academic achievement. Methodologically, differences in population samples, exam formats, and grading standards across studies could also account for the inconsistencies observed in student performance by gender.

In addition, our findings demonstrated that male students presented a higher standard deviation in scores compared to female students. This suggests a broader dispersion in the academic performance of male students, pointing to a more varied level of achievement within this group. Contrastingly, female students exhibited more consistency in their test scores.

Delving deeper, the range of scores for male students exceeded that of female students, reinforcing the notion that the male academic performance is wider—spanning from higher highs to lower lows. Such a pattern aligns with prior research by Alnasser (2022), which highlighted gender-based discrepancies in language usage among faculty members, hinting at potentially varying degrees of English language proficiency and engagement.

To further contextualize these results, it is imperative to note that additional studies have offered mixed evidence on gender differences in academic performance. Some investigations suggest academic advantages for female students in linguistic abilities, including English, which would seemingly lend no support to our findings (Voyer and Voyer, 2014; Jackman et al., 2019). Others posited that such advantages are context-dependent and may vary considerably depending on sociocultural factors or pedagogical approaches (Preiss and Hyde, 2010).

In attempting to answer the research question regarding the performance variability across six semesters, our data supports a trend of incrementally improving scores for both gender groups, with male students maintaining an edge in overall scores. This trend could be indicative of an evolving curriculum or improved teaching methods benefiting all students gradually over the observed semesters.

In light of the additional results discussed, the gender differences in academic performance highlighted by this study warrant further exploration. Important matters for subsequent research could include examining the influence of instructional strategies, the role of student engagement, and broader sociocultural influences on learning and performance in English departments.

In conclusion, this study unearthed a complex picture of academic performance in English comprehensive exams, with evidence suggesting gender-based differences. These findings contribute to an ongoing academic conversation and point toward a nuanced understanding of educational outcomes in higher education settings. Further studies are thus essential to disentangle the myriad variables at play and to build upon the insights gained herein.

6 Limitations of the study

There are two limitations to this study:

The first limitation is related to the scope of the study; the term "comprehensive evaluation" may encompass a wide range of elements, such as various methods of testing, assessment tools, and evaluation criteria. However, due to time and resource constraints, the study was unable to cover all possible aspects of comprehensive testing, ultimately, leading to a relatively partial understanding of the subject matter.

Another limitation is the inability to generalize the findings of the study. The current study involves a specific sample and a particular context, which may not be representative of the larger population or other contexts. As a result, the findings may not be easily generalized to other populations or situations.

7 Conclusion

In conclusion, the research study on analyzing the efficacy of comprehensive tests provides valuable insights into the performance of male and female students over six semesters. While male students, on average, scored slightly higher than female students, the difference was minimal (55.5 for males vs. 56.3 for females). Male scores exhibited more variability and a wider spread, suggesting varying performance levels among male students.

Despite fluctuations in average scores for both genders across semesters, statistical analysis, including a t-test with a p-value of 0.07, indicated no significant difference in performance between male and female students. Furthermore, Cohen's d as a measure demonstrated a negligible effect size, emphasizing the minimal gender gap in comprehensive exam performance.

These findings challenge common assumptions about gender-based academic achievement in Saudi Arabia and suggest that more research is needed to understand the underlying causes. This research should consider a wider array of factors and potentially lead to changes in educational policy and practice to ensure all students have equal opportunities to succeed.

The study's insights are crucial as they could alter the way educators approach gender differences in academic settings and prompt a reevaluation of teaching and assessment methods to better accommodate and understand the diverse needs of all students.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This research paper has been conducted with adherence to ethical standards and principles, ensuring confidentiality, consent, and integrity in all research activities. Moreover, the study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

YA: Conceptualization, Writing – original draft. SA: Investigation, Writing – original draft. BA: Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alnasser, S. M. N. (2022). A gender-based investigation into the required English language policies in Saudi higher education institutions. *PLoS One* 17:e0274119. doi: 10.1371/journal.pone.0274119

Broadbent, J., Panadero, E., and Boud, D. (2018). Implementing summative assessment with a formative flavour: a case study in a large class. *Assess. Eval. High. Educ.* 43, 307–322. doi: 10.1080/02602938.2017.1343455

Brown, C. (2020). Enhancing understanding and meaningful learning through comprehensive testing. *J. Higher Educ. Assess.* 15, 421–435.

Brown, G. (2022). The past, present and future of educational assessment: a transdisciplinary perspective. *Front. Educ.* 7:1060633. doi: 10.3389/feduc.2022.1060633

Garcia, E., Smith, A., and Johnson, B. (2021). Strategies for mitigating challenges in comprehensive testing: a case study approach. *J. Assess. Eval. Higher Educ.* 12, 73–89.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender similarities characterize math performance. *Science* 321, 494–495. doi: 10.1126/science.1160364

Inzlicht, M., and Schmader, T. (2012). *Stereotype threat: theory, process, and application.* New York, NY: Oxford University Press

Jackman, W. M., Morrain-Webb, J., and Fuller, C. (2019). Exploring gender differences in achievement through student voice: critical insights and analyses. *Cogent Educ.* 6. doi: 10.1080/2331186X.2019.1567895

Johnson, B. (2019). Promoting deep learning through comprehensive testing in higher education. *Educ. Psychol. Rev.* 25, 301–315.

Jones, D. (2017). Challenges and opportunities of comprehensive testing in higher education: a literature review. *High. Educ. Res. Dev.* 38, 187–201.

Liu, L., Saeed, M. A., Abdelrasheed, N. S. G., Shakibaei, G., and Khafaga, A. F. (2022). Perspectives of EFL learners and teachers on self-efficacy and academic achievement: the role of gender, culture and learning environment. *Front. Psychol.* 13:996736. doi: 10.3389/fpsyg.2022.996736

Moore, L. C., Goldsberry, J., Fowler, C., and Handwerker, S. (2021). Academic and nonacademic predictors of BSN student success on the HESI exit exam. *Comput. Inform. Nurs.* 39, 570–577. doi: 10.1097/CIN.0000000000000741

Preiss, H. A., and Hyde, J. S. (2010). "Gender and academic abilities and preferences" in *Handbook of gender research in psychology.* eds. J. Chrisler and D. McCreary (New York, NY: Springer)

Puarungroj, W., Boonsirisumpun, N., Pongpatrakant, P., and Phromkot, S. (2017) A preliminary implementation of data mining approaches for predicting the results of English exit exam. In Proceedings of the 2nd International Conference on Information Technology (INCIT), IEEE

Smith, A., Johnson, B., and Brown, C. (2018). The impact of comprehensive testing on critical thinking skills in higher education. *J. Educ. Res.* 10, 145–162.

Soler, S. C., Nisperuza, G. L., and Idárraga, P. H. (2020). Test preparation and students' performance: the case of the Colombian high school exit exam. *Cuadernos Econ.* 39, 31–72. doi: 10.15446/cuad.econ.v39n79.77106

Voyer, D., and Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychol. Bull.* 140, 1174–1204. doi: 10.1037/a0036620