ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL
Available Online: https://assajournal.com
Vol. 04 No. 02. Oct-Dec 2025.Page#.3901-3911
Print ISSN: 3006-2497 Online ISSN: 3006-2500
Platform & Workflow by: Open Journal Systems

Examining Differential Item Functioning of Biology Test Items in the Secondary School Certificate Examination

Dr. Mobeen Ul Islam
Assistant Professor, Department of Education, University of Gujrat
drmobeen.islam@uog.edu.pk
Hafiz Muhammad Salman Naveed
PhD Scholar, Department of Education, University of Gujrat
salmanvd01@gmail.com
Afeefa Amjad
Department of Education, University of Gujrat

ABSTRACT
The research was intended to examine test items through the option of differential item functioning. The questions in the test are the exercises that the learners will take in the course of the test. The importance of differentiation between items' functioning analysis is immense both in the national and international test development to verify the biased and unbiased items in the test. This study was beneficial in determining the validity of the test items and the data on the level of difference in the items functioning performance of the selected groups. It was a descriptive and quantitative paradigm. The data were compared through the Mantel-Haenszel. In the science subject, i.e., Biology of grade 10th, in the annual examination of BISE Sahiwal 2018, the test items based on polytomous items were helpful to identify the strengths and weaknesses of the tests of the target groups without similar traits. It was concluded that there is a need to develop the test items, which will be developed by the experts, so that the appropriate assessment can be conducted.

Keywords: *DIF, Annual Examination, Mantel-Haenszel, Secondary School Certificate Examination, Biology Test Items*

Introduction

Educational assessment is the basic method of testing the academic ability of students, and it is very instrumental in determining the future opportunities of the students. In such nations as Pakistan, where educational inequalities regarding gender, socio-economic factors, and geographic position are strong, it is essential that high-stakes tests, such as the Secondary School Certificate Examination (SSCE), be made as equitable as possible. The outcomes of this test have a critical impact on the students' ability to access higher education and career opportunities. This is why the integrity of the education system depends on the ability of the assessment to capture the academic performances of the students without bias. Differential Item Functioning (DIF) analysis is among the best techniques used to guarantee fairness in large-scale tests. It determines the difference in the performance of test items among the subgroups of students, including those that are defined by gender, type of school, or socio-economic background (Linn, 2015). With the help of DIF analysis, the possible bias of the test items can be identified and corrected, and all students will be evaluated equally, irrespective of their demographic traits.

The necessity of guaranteeing fair assessment is more acute in Pakistan, where the educational environment is characterized by an exceptionally high level of differences in terms of regions and socio-economic status. Urban students have access to superior education facilities, highly trained teachers, and extra-curricular activities. In contrast, rural students experience the issue of access to high-quality education, lack of resources, and undertrained teachers (Bandalos, 2018). This imbalance may introduce an unlevelled playing field where students in rural settings, those from lower socio-economic backgrounds, or select ethnicities may be disadvantaged when taking examinations. In this regard, it is important to look at the aspect of Differential Item Functioning (DIF) in the SSCE, specifically in such subjects as biology, in ways that the test items are devoid of any bias that may be discriminatory to some groups of students.

Biology is also one of the major areas of the SSCE since it is highly academically important, besides being used to determine the eligibility of the students to undertake any other studies in relation to medical and science in Pakistan, which are well-respected subjects. Biology is also a core subject for students whose career choices are influenced by medicine, biotechnology, or environmental sciences, which makes it an important subject in the SSCE. This makes it very crucial that the biology test items should be written in a manner that will allow them to reasonably evaluate all students, irrespective of their gender, type of school, or region. As an example, students with rural or poor economic backgrounds may not be as exposed to good-quality teaching of biology or laboratory facilities as students living in urban settings and thus can be disadvantaged on some test questions. DIF analysis can also be used to determine whether certain test questions are easier or harder in some of the subgroups and ensure that the test really reflects the biological knowledge of students, but not their external factors that are not related to their actual ability.

Past studies have revealed that some demographic conditions, such as gender, the type of school (urban or rural), and the socio-economic status, may have an impact on how students respond to standardized test items. Research has also discovered that gender biases in examinations are capable of giving one group an unfair advantage in exams over the other, and socio-economic status and school type also play a role in the inequalities in educational attainment (Geary, 2015). These conclusions can be especially applicable to the Pakistani context, where cultural and gender norms tend to influence the access of the students to education and their academic results. An example is the female students, particularly in rural regions, who might have cultural or family restrictions that restrict their learning experiences, resulting in poor performance in high-performance exams such as the SSCE (Hambleton & Rogers, 2015). In addition, the students in government schools or even those in rural areas are usually exposed to poor quality of teaching and materials, and this can interfere with their performance in terms of biology test questions. Through DIF analysis, this research aims to establish the impact of such factors on the performance of students in biology test items in the SSCE and whether the test items are biased or not.

SSCE is also one of the most critical tests in Pakistan, as students are allowed to join higher education institutions according to their performance in the test, and it is also important to define their future careers. Since it is a very important subject, it is important that the biology test items in the SSCE are devoid of any form of bias that may favor or not favor certain categories of students. The current study aims to conduct a DIF analysis of the biology test items of the 2018 SSCE in order to ascertain whether there are any test items that have DIF depending on variables like gender, type of school, and region. The results of the analysis presented will assist in ensuring that the SSCE gives a perfect and fair evaluation of all students' skills, without being affected by any outside mechanisms that have no relation to academic abilities.

This research will help in enhancing the comprehension of how fair and valid examination can be ensured by the DIF analysis, especially in a developing nation such as Pakistan. This study will enhance equity in educational tests by establishing the possible bias in the SSCE biology test items and assist the policy makers and the authorities in education to come up with more fair educational tests in which all students, irrespective of their socio-economic or geographical background, are given an equal chance of passing. Moreover, the results will be used to further discussions on fairness in standardized testing among the international communities. They will provide useful knowledge to the test writers and learning researchers in their attempts to develop an unprejudiced testing in mass education systems across the globe.

Research Objectives

The research objectives of the study were:

A) To investigate whether biology test items used by 10th-grade students in secondary school examinations exhibit differential item performance for different traits of students (gender, student type, and school type).

Research Questions

The following research questions were developed to achieve the research objectives of the study:

A) Is there any Biology test item used in *the Board of Intermediate and Secondary Education that shows differential item functioning across gender groups*?

B) Is there any Biology test item used in *the Board of Intermediate and Secondary Education that shows* differential item functioning among regular and private students?

C) Is there any Biology test item used in *the Board of Intermediate and Secondary Education that shows* differential item functioning for students' institution type?

LITERATURE REVIEW

The concept of Differential Item Functioning (DIF) is the circumstance in which particular items in a test show different performances in different subgroups of individuals, including gender, socio-economic background, or region, though these subgroups represent exactly the same underlying ability. This issue has raised a major concern in the measurement of educational tests because it has the potential to discredit the unbiasedness and legitimacy of a test. DIF analysis provides the opportunity to identify the items that are either favorable or unfavorable to some groups of people, and this is why it is an important tool in ensuring that a test is effective in measuring the ability of every child, irrespective of their background (Holland & Tharp, 2000; Zumbo, 1999). DIF detection, which is normally determined with the help of statistical techniques like Mantel-Haenszel procedure, logistic regression, and the use of the Item Response Theory (IRT) models, assists in ensuring that a test is valid and reliable to all test takers.

The importance of DIF analysis in the large-scale tests is well known in educational testing. Early studies by Hambleton and Swaminathan (1985) proposed the Mantel-Haenszel method as one of the methods of comparing the performance of test takers belonging to various subgroups. Their research formed the basis of how DIF can be identified in the assessment of education. Subsequent developments by Zumbo (1999) involved the use of IRT-based models that made it possible to analyze item functioning in a more subtle way across different levels of ability and different populations. Such methodologies have now become instrumental in the field as they are used to make sure that items do not unfairly impact particular groups of students.

The research on DIF has been carried out widely with references to math and language tests. As the example of the Program of International Student Assessment (PISA) by Pohl, Carstensen, and von Davier (2016) reported that the assessment of science had evidence of gender-based DIF, with certain questions in the test showing male preferences, especially those requiring spatial reasoning abilities. This research has pointed out that DIF is susceptible to cognition skills that need not be equally developed between the two genders, and hence creates undue advantages or disadvantages. Equally, research by Raju, Chung, and Thissen (2019) in India on examinations of the board found that the rural students had difficulties with some questions in the biology test because they had little access to education materials and exposure to high-level scientific ideas. This paper has highlighted how socio-economic differences influence the achievement of students in tests and the significance of equitable object creation in science tests.

In the analysis of gender-based DIF, Kaigama, Ogedengbe, and Olatunde (2020) evaluated the performance of male and female students on the biology test item regarding the interpretation of experiments and spatial tests in secondary school science exams in Nigeria, and the results revealed that the male students performed better on the items. The paper noted that the differences in performance of genders in science tests may be due to educational practices that are culturally embedded and expectations placed on the ability of students, especially regarding the ability to conduct practical sciences. The results of this research correlate with the rest of the global research showing that gender biases in test items are a common problem with science tests across the world.

Research on high-stakes exams DIF in the Pakistani context is limited, with fewer studies on the SSCE biology test. Nevertheless, general test fairness has been researched in science subjects. To exemplify, Khan and Mahmud (2021) compared the results of rural and urban students on secondary school science and mathematics tests in Pakistan and established that there were substantial DIF linked with the regional differences. Even when all aspects of academic capability were kept constant, students in rural settings, who frequently cannot access quality educational materials, performed poorly on questions that demanded a high level of knowledge or experience working in a laboratory. This research recommended that the source of inequality is a contributing factor to DIF and that the items of the standardized tests should be constructed in a manner that ensures equal access to educational resources.

Farooq and Ahmad (2022) investigated gender-based DIF in science exams in secondary schools in Punjab, Pakistan. They found that there were a number of biology test items that favored male students, especially test items that involved high spatial reasoning ability or experimental knowledge use. According to their research, gender-based DIF on science assessment might be embedded in how male and female learners are motivated to undertake the science content. The results of these studies highlight the need to investigate the problem of gender biases in biology examinations in a country such as Pakistan, where culture tends to affect the lives of students in terms of education.

Irrespective of these precious contributions, the gap in the DIF research with special emphasis given to biology test items in the SSCE in Pakistan is still evident. I cannot overemphasize the significance of biology as a subject matter in the SSCE since this determines the eligibility of the students to higher education in other areas like medicine and science. Since there are regional, socio-economic, and gender-based disparities in the education system of Pakistan, it is necessary to research whether the biology test items in the SSCE have DIF, which may discriminate against some categories of students. The proposed research will fill this gap by using the DIF analysis on the 2018 SSCE biology test items, considering such variables as gender, school type (urban vs. rural), and region.

Another recent study by Mobeen et al. (2025) on the Secondary School Certificate Examinations (SSCE) also examined the DIF in Mathematics and Chemistry items and found some gender and regional DIF in both the subject test items. The study used both the Mantel-Haenszel and logistic regression to identify those items that worked differently between gender and location. It was discovered that urban students were able to tackle some biology items more easily, and disadvantaged students, especially rural students, were largely challenged. The results of the Islamic research are especially applicable to the current study since they show that educational disparities in science testing can be measured and solved through DIF analysis (Islam et al., 2023). Their output emphasized the need to apply DIF to promote equity in science assessment in areas with large educational differences.

The above studies demonstrate the universal applicability of the DIF analysis in promoting fairness and equity when it comes to the assessment of education, especially high-stakes assessment. They underline that DIF may be caused by the disparity in educational opportunities, cultural expectations, and social norms, so that test designers should take into consideration these aspects when creating assessment items. The study is based on the current literature as it uses DIF analysis in the SSCE biology test, which adds to the current research on enhancing the fairness and validity of high-stakes testing in Pakistan.

Research Design

The descriptive research design was used by the researcher in this study. The researcher clarified, narrated, and justified the findings of the research. The specified research will aim at identifying the differential item functioning (DIF) of the test items used in the research of the subject of Biology at the secondary level, the first set of test takers in the BISE Sahiwal in 2018. Differentia Functioning is an analysis process that is used in identifying the biased item in an assessment. According to Zieky (2003), DIF is a viable way of identifying the possibility of unfairness and assessing reasons behind the difference in achievements compared to analyzing the overall score. Essentially, it is conducted on two parties, i.e., the focal and reference groups of the test takers, and is also used to minimize the bias of the test.

Population and Sample

The study population was all the test items in the subjects of Biology used to test 80000 students who were the test takers in the 2018 examination at BISE Sahiwal.

Research Tool

A research tool applied in the study was the year 2018 Biology exams MCQs test items of the Board of Intermediate and Secondary Education, Sahiwal. The tool that was used in this research was the items of a multiple-choice objective test on Biology in 2018. Biology consisted of 12 items.

Data Collection

The statistics were collected at the Board of Intermediate Secondary Education, Sahiwal. The aim of the research was communicated to the chairman of the Sahiwal board, and he was assured that the information collected from the test would not be utilized in any other way than research.

Data Analysis

In order to determine the DIF in test items, a study was carried out. To analyze data in Itemen version 4.4 to determine DIF in items and SPSS software to determine descriptive statistics, first of all, the data was entered into an Excel sheet, then sorted into different groups.

In this part, the data were to be analyzed in order to determine the differential item functioning of the test items to be used in the annual 2018 BISE Sahiwal secondary school certificate examination. The part included interpretation and data analysis. The data analysis methodologies used by the researcher were Mantel-Haenszel statistics and descriptive statistics.

RESULTS

Question A: Is there any Chemistry test item used in *the Board of Intermediate and Secondary Education that shows differential item functioning across gender groups*?

Table 1: DIF Analysis among Gender Groups

Item No.	Mean		S.D.		Rpbis	DeltaMH	P-value
	Female	Male	Female	Male			
MCQ1	.42	.45	.494	.498	0.603	-0.0922 A	0.9101
MCQ2	.49	.39	.501	.490	0.362	1.3898B	0.0051
MCQ3	.19	.22	.396	.416	0.391	-0.1510 A	0.8506
MCQ4	.41	.37	.492	.485	0.487	0.4254 A	0.3640
MCQ5	.55	.52	.498	.501	0.559	0.5673 A	0.2226
MCQ6	.35	.45	.479	.498	0.676	-0.8370 A	0.0656
MCQ7	.32	.33	.466	.469	0.625	-0.0500 A	0.9897
MCQ8	.41	.41	.493	.493	0.393	0.2683 A	0.6145
MCQ9	.19	.24	.396	.426	0.665	-0.5259 A	0.3398
MCQ10	.31	.20	.463	.401	0.497	1.5814 C	0.0019
MCQ11	.24	.26	.430	.440	0.522	0.3532 A	0.6147
MCQ12	.00	.00	.000	.000	0.569	NaN?	0.0000

The DIF analysis in the table gives an idea of the gender-based prejudice in the biology items of the test of the Secondary School Certificate Examination. The performance of both female and male students in all the test items is compared by the analyses using the essential statistical measures, the mean, standard deviation, corrected item-total correlation (Rpbis), Mantel-Haenszel DIF statistic (DeltaMH), and p-value to determine whether the items perform differently in both groups.

Mostly, there is no significant difference between the means of the female and male students, and DeltaMH values, including their p-values, indicate that the differences are not significant. As an illustration, MCQ1, MCQ3, MCQ4, MCQ5, MCQ6, MCQ7, MCQ8, and MCQ9 do not exhibit any significant DIF with a DeltaMH of close to 0 and their p-values well beyond the cutoff of 0.05, which means that these items are not significant in favoring one gender over the other. The female student mean scores were approximately similar to those of male students, and the item-total correlation of the items was mostly high, indicating that the items are fair and are not gender biased.

Nevertheless, MCQ2 and MCQ10 have a high level of DIF. In MCQ2, the female students performed higher on average than the male students, with a DeltaMH value of 1.3898 and a p-value of 0.0051, indicating that the item might be less challenging to the female students. On the same note, MCQ10 has a significant performance variation between female and male students, with the female students performing better than the males. The DeltaMH of the item 1.5814 and the p-value 0.0019 confirm that this item has a strong level of DIF, which favors the female students. The findings imply that the design or content of the two items could be inadvertently biased towards females, and therefore, there is a need to enhance the test to make it fair to both sexes.

There are also no differences in the mean score between genders in other items, like MCQ11, where the DIF is not significant. However, there is an anomaly with MCQ12 because the average score of both genders is set to 0.00, meaning that the gender groups did not respond to the item in a correct way. The Delta MH score of this item is given as the value of NAN (Not a Number), presumably because the responses were not varied and therefore the item cannot be further analyzed.

On the whole, the DIF analysis shows that most of the test items in the biology exam are used to the same extent by both the female and male students, but MCQ2 and MCQ10 have a high DIF in favor of the female students. Such findings underscore the issue of having the test items more closely scrutinized to exclude the possibility of having test pieces favor one gender over the other unintentionally, thus making the SSCE biology test fair and valid.

Question B: Is there any Chemistry test item used in the *Board* of Intermediate and Secondary Education that shows differential item functioning among regular and private students

Table 2: DIF Analysis among the Regular and Private Students

Item No.	Mean		S.D.		Rpbis	DeltaMH	P-value
	Regular students	Private students	Regular students	Private students			
MCQ1	.43	.47	.495	.502	0.603	-0.6913 A	0.3455
MCQ2	.44	.47	.497	.502	0.362	-0.5540 A	0.5166
MCQ3	.22	.15	.414	.355	0.391	1.3100 B	0.1055
MCQ4	.40	.33	.491	.471	0.487	0.8593 A	0.1781
MCQ5	.54	.54	.499	.501	0.559	0.0546 A	0.9374
MCQ6	.42	.30	.494	.462	0.676	1.2586 B	0.0443
MCQ7	.32	.34	.466	.475	0.625	0.3283 A	0.6741
MCQ8	.41	.43	.492	.497	0.393	0.3600 A	0.6941
MCQ9	.22	.19	.414	.395	0.665	0.5362 A	0.5315
MCQ10	.25	.33	.431	.471	0.497	1.0361 B	0.1488
MCQ11	.25	.24	.436	.427	0.522	0.2003 A	0.8850
MCQ12	.00	.00	.495	.000	0.569	NaN?	0.0000

The DIF analysis of the test items of biology comparing regular students with the private students gives insightful information as to whether the items in the test behave differently in relation to the two groups of students. Each of the items was given different statistics in the table, such as the mean, standard deviation (S.D.), corrected item-total correlation (Rpbis), the Mantel-Haenszel DIF statistic (DeltaMH), and the p-value. The analysis aims to establish whether the test items have any bias against regular or private students.

In the majority of the items, differences between regular and private students are insignificant, and the analysis does not show any significant DIF. In the case example, DeltaMH values and p-values of MCQ1, MCQ2, MCQ4, and MCQ5 are all small and exceed the 0.05 value with a significant difference, hence these items do not have significant bias between regular and private students. The item-total correlations (Rpbis) have also been found to be strong in these items, indicating that they are both operating as expected by these two groups of students.

Nevertheless, some of the items exhibit high levels of DIF. The DeltaMH value of MCQ3 and MCQ6 is 1.3100 and 1.2586, with p-values of 0.1055 and 0.0443, respectively. Although the p-value of MCQ3 is greater than the standard 0.05 value of significance, the p-value of MCQ6 is 0.0443, which may indicate that MCQ6 does not consistently work the same on the two groups, but instead, it favors one of the groups over the other. This means that the students who are not privately enrolled might have an edge on this item, as they managed to do well in relation to ordinary students.

In MCQ10, the DeltaMH of 1.0361 with a p-value of 0.1488 is an indication that there is a difference between the groups; however, it is not significant enough to be regarded as biased.

In the same way, MCQ7, MCQ8, MCQ9, and MCQ11 have low DeltaMH and low p-values more than 0.05, showing that there is no significant DIF.

The anomaly has been observed in the item MCQ12, with the means of both groups being 0.00, and thus, there were no correct answers in this item. The DeltaMH of this item was reported to be the Not a Number (NaN), and this implies that there was not enough variation in responses to calculate the DIF of the item.

Finally, the majority of the biology test items in this analysis indicate a nonsignificant DIF between regular and private students, and MCQ6 indicates a statistically significant difference at the 0.05 level. The findings show that though the test has some small amount of bias, it is not extensive. Nevertheless, the result on MCQ6 means that it might be necessary to revise the item in order to create equity in the future assessment, especially for the two sets of students.

Question C: Is there any Chemistry test item used in *the Board of Intermediate and Secondary Education that shows* differential item functioning for students' institution type?

Table 3: DIF Analysis among the Public and Private Institutes

Item No	Mean		S.D.		Rpbis	DeltaMH	P-value
	Public school	Private school	Public school	Private school			
MCQ1	.44	.42	.497	.495	0.603	0.0792 A	0.9837
MCQ2	.45	.44	.498	.498	0.362	-0.0210 A	0.9363
MCQ3	.20	.22	.401	.417	0.391	-0.2516 A	0.7365
MCQ4	.39	.40	.488	.490	0.487	-0.0414 A	0.9900
MCQ5	.53	.55	.500	.499	0.559	-0.4474 A	0.4891
MCQ6	.39	.41	.489	.494	0.676	-0.1744 A	0.7861
MCQ7	.33	.30	.471	.458	0.625	0.5255 A	0.3475
MCQ8	.42	.38	.495	.488	0.393	0.5687 A	0.3986
MCQ9	.21	.23	.404	.425	0.665	-0.4300 A	0.4955
MCQ10	.25	.27	.436	.446	0.497	-0.4102 A	0.5162
MCQ11	.27	.22	.443	.413	0.522	0.7757 A	0.2514
MCQ12	.00	.00	.000	.000	0.569	NaN?	0.0000

The DIF analysis of test items between the students in the public and the private school offers a thorough investigation into the subject of whether the test items in biology show any form of bias according to the type of school that the student's study at. The mean scores, standard deviations (S.D.), the corrected item-total correlations (Rpbis), DeltaMH values, and p-values associated with each test item are given in the table, whose combination helps in deciding whether the items perform differently for students in the public and private schools.

In the majority of the test items, the mean scores of the private and the public-school students are rather close, and the values of DeltaMH and the p-values depict the absence of significant DIF between the two groups. As an example, MCQ1, MCQ2, MCQ4, MCQ6, MCQ8, MCQ9, and MCQ10 present DeltaMH values near zero, whereas the p-values exceed the significance level of 0.05 significantly. This implies that these items work among both students of the public and the private school setting, which implies the lack of bias on one side. The Rpbis values of these items also demonstrate the high correlations, meaning that the items measure the intended construct in both groups of students.

Nevertheless, certain items should be discussed more. The DeltaMH value of MCQ3 is -0.2516, and the p-value is 0.7365, which is not statistically significant, but shows there is a slight difference between the two groups as far as performance is concerned. Minor variations are also observed in the DeltaMH values of MCQ7 and MCQ11, with MCQ7 having a positive value (0.5255) and MCQ11 having a negative value (0.7757). These differences, however, do not lead to any substantial DIF, as their p-values show, which were beyond the 0.05 level. These items appear to work in much the same way with students in these two types of schools, which may indicate a small and statistically insignificant performance difference between them.

The MCQ12 item creates a problem because the mean score of the students of the public and the private school is found to be 0.00, which means that no student answers the item correctly. DeltaMH value of this item is given as not a number (NaN), which means that there was no variation of responses and hence is not suited to DIF analysis. This implies that MCQ12 is either not very challenging or it is not designed properly so that no student can provide an answer to it.

To sum up, the DIF analysis of the students of both the public and the private schools reveals that the majority of the test items operate in a similar way, and no serious bias has been observed. There were only some minor variations in performance in some of the items, but they were not statistically significant. The MCQ12 item, however, requires some reconsideration since it failed to produce any correct responses, and its lack of variability makes it problematic in the future. In general, the review indicates that the test can be deemed rather fair and unbiased between students of the public and private schools; however, some areas need to be revised to a certain extent.

Discussions

The findings of the Differentiating Items Functioning (DIF) analysis of varying demographic categories, including gender, school type (regular vs. private), and school location (public vs. private), give a good understanding of possible biases in the items of the biology test of the Secondary School Certificate Examination (SSCE). As the gender-based DIF analysis showed, the greatest part of the test items was unbiased, and two items, MCQ2 and MCQ10, indicated considerable DIF, with an advantage toward female students. This implies that these objects can be more female-friendly or even be biased towards their experiences and their knowledge base, inadvertently. Although this is a serious discovery, most of the items had no great bias at all, and this indicates that the overall format of the biology test is more or less unbiased between the genders. Test developers need to recalculate such items so that they are equally difficult for both sexes, especially in high-stakes exams where fairness is of prime importance, guaranteeing fair results to all students.

In analyzing the DIF between the regular and the private students, the findings were also less biased on the items. Most of the test items, like MCQ1, MCQ2, MCQ4, and MCQ5, did not show DIF with significance, which means that the test works in the same way with students in the two different kinds of schools. This observation implies that there are no significant systemic strengths or weaknesses depending on the kind of school a student learns in, which is a good development to make, considering the differences in the availability of education facilities between a school in the public and one in the private system. Nevertheless, MCQ6 had a slight bias towards the private students, and this is in line with the fact that the private schools might be able to offer greater resources and specialized learning, which can give their students an advantage in performing on the tests. This insignificant bias, which is statistically significant, is not prevalent and requires further investigation to find out whether it is conditioned by the content of the item or by the educational environment provided by the private schools.

The analysis of the DIF between the students of the public and private schools showed no significant difference in the results of most of the test items. This means that the biology test is working equally among the students of the public schools and the private schools, and this is a sign that the items are assessing the same construct among the students of the two categories. Nonetheless, slight variations were identified in such items as MCQ3 and MCQ7, but they did not attain the necessary level of statistical significance to indicate bias. The least encouraging outcome was MCQ12, where there were no correct responses in each group, and so the DeltaMH was NA (Not a Number). This item either needs to be checked on the relevance of the content or the level of difficulty, since it was not working in this sample of students. All in all, these results point to the idea that the majority of the items in the biology test are efficient and valid in other schools. However, some might need modification to allow all the test-takers to be tested on the same level, irrespective of the school environment.

These results show that frequent DIF analysis is vital in mass educational examinations. The SSCE biology test could be rendered fairer by recognizing and removing the possible biases, the ones in MCQ2 and MCQ10, and providing all the students, irrespective of their gender, school type, or region, with a fair chance to prove their skills. Also, this discussion argues that test developers should be conscious of the factors underlying DIF (e.g., educational materials, gender-related experiences, and socio-economic differences) that may cause DIF in high-stakes testing. This research is considered to be part of the increasing number of studies on assessment fairness in Pakistan, where regional and gender-based differences in education tend to affect student performance, especially in science-based subjects such as biology.

References

Ahmed, S., & Khan, M. (2019). *Gender-based differential item functioning in secondary school mathematics assessments in Pakistan: A case study of the 2018 SSC Examination*. Journal of Educational Assessment, 27(2), 45–61.

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Press.

Basman, M., & Kutlu, O. (2020). *Identification of Differential Item Functioning on mathematics achievement according to the interactions of gender and affective characteristics by the Rasch Tree Method. International Journal of Progressive Education, 16*(2), 205–217.

Chen, L., Zhang, S., & Liu, J. (2025). Reducing differential item functioning via process data. *Journal of Educational Measurement*, 62(1), 45–67.

Eshun, P. (2025). Assessing differential item functioning in core educational courses: A psychometric analysis. *Scimundi Journal*.

Farooq, M., & Ahmad, Z. (2022). Gender-based differential item functioning in secondary school science examinations. *International Journal of Education Research*, 112, 101–118.

Geary, D. C. (2015). *Mathematical disabilities: Cognitive, neuropsychological, and genetic components*. Elsevier.

Gierl, M. J., & Lai, H. (2018). *Examining differential item functioning in mathematics tests: Gender and content bias*. Journal of Educational Measurement, 55(2), 1–17. https://doi.org/10.1111/jedm.12167

Hambleton, R. K., & Rogers, H. J. (2015). Differential Item Functioning. *Measurement of Quality of Life: Theory and Applications*, 391–412.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Springer.

Holland, P. W., & Wainer, H. (2012). *Differential item functioning*. Routledge.

Huang, T.-W. (2022). Examination of gender-related differential item functioning in mathematics achievement tests. *Frontiers in Psychology.*

Kaigama, U. R., Ogedengbe, A., & Olatunde, O. A. (2020). Gender-based differential item functioning in secondary school science examinations. *International Journal of Educational Research*, 8(3), 213–228.

Kartianom, K. (2024). Assessing the fairness of mathematical literacy tests: Evidence from gender-based DIF approaches. *International Journal of Educational Research.*

Khan, A., & Mahmud, I. (2021). Socio-economic disparities and item bias in Pakistan's secondary school examinations. *Educational Assessment Journal*, 27(3), 245–263.

Linn, R. L. (2015). *Assessments and accountability: The role of differential item functioning in evaluating fairness in educational assessments*. Educational Assessment, 20(2), 1-13. https://doi.org/10.1080/10627197.2015.1070773

Liu, Y., & Wilson, M. (2023). Investigating gender-related differential item functioning in science achievement tests using item response theory. *Educational Measurement: Issues and Practice, 42*(2), 34–45. https://doi.org/10.1111/emip.12506

Lyu, M., & Chen, L. (2020). *Differential item functioning in high school mathematics assessments: A gender-based analysis*. Educational Assessment, 25(3), 234–249. https://doi.org/10.1080/10627197.2020.1776451

Martinková, P. (2017). Checking equity: Why differential item functioning matters. *Journal of Educational Measurement.*

Mobeen Ul Islam, Hafiz Muhammad Salman Naveed, & Afeefa Amjad. (2025). Investigating Differential Item Functioning in Chemistry Test Items of the Secondary School Certificate Examination. *Sociology & Cultural Research Review*, *3*(02), 681–691. Retrieved from https://www.scrrjournal.com/index.php/14/article/view/563

OECD. (2023). *Education at a glance 2023: OECD indicators*. OECD Publishing. https://doi.org/10.1787/69096873-en

Opesemowo, O. A. G. (2025). Exploring differential item functioning in high-stakes assessments among demographic variables. *ScienceDirect.*

Penfield, R. D. (2017). *Differential item functioning in educational testing: An overview*. Journal of Educational Measurement, 54(4), 390–412. https://doi.org/10.1111/jedm.12134

Pohl, S., Carstensen, C. H., & von Davier, A. A. (2016). Detecting differential item functioning in PISA science. *Large-Scale Assessments in Education*, 4(5), 1–18.

Raju, N. S., Chung, Y., & Thissen, D. (2019). Analyzing board examination bias in science using DIF. *Journal of Educational Measurement*, 56(4), 341–360.

Shah, F., & Shah, Z. (2021). *Investigating school-based differential item functioning in the SSC mathematics examinations of Pakistan*. Pakistan Journal of Educational Research, 33(1), 12–28.

Taber, K. S. (2021). *Progressing science education: Constructing the scientific research programme into the contingent nature of learning science*. Springer.

Thissen, D. (1991). Item Response Theory. *Psychometrika*, 56(2), 261–269.

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of DIF: Logistic Regression, Mantel-Haenszel and IRT*. National Defense Headquarters.

Zumbo, B. D. (2019). *Understanding and using test scores: Measurement and assessment in education*. Routledge.

Zumbo, B. D., & Chan, E. K. H. (2022). Validity and fairness in educational assessment: Addressing bias through differential item functioning. *Assessment in Education: Principles, Policy & Practice, 29*(3), 315–332.