



Assessment Fairness and Examination-System Effectiveness in Higher Education: Evidence from University Students in Pakistan

Sehar Mahmood¹, Maryam Bibi²

1 Independent Scholar, Bahawalpur, Pakistan. Email: chsehar748@gmail.com

2 Teacher, Government Girls Primary School Akberabad, Pakistan. Email lodhimaryam45@gmail.com

ARTICLE INFO

Article History:

Received: January 14, 2026

Revised: June 23, 2026

Accepted: June 25, 2026

Available Online: June 27, 2026

Keywords:

Assessment Fairness

Examination System

Higher Education

Invigilation

Marking Transparency

Semester System

Student Perceptions

Pakistan

Funding:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Higher education examination systems play a pivotal role in the education system, as they are critical for progressing through education, for recognizing learning, and for influencing confidence in the quality of higher education. This study aimed to investigate the effectiveness of examination system as perceived by the university students in a semester based higher education in Bahawalpur, Pakistan. The objectives were to evaluate the students' perception regarding the administration of the exam, monitoring, invigilation, the standard of the question paper, the fairness of the marking and the transparency of the results. A quantitative descriptive survey design was used. A structured five-point Likert scale questionnaire was used to gather data from 300 students from two public sector universities in Bahawalpur, Pakistan. The instrument demonstrated high reliability (Cronbach $\alpha = .913$) and data was analyzed using frequencies, percentages and mean scores. The results indicated high perception of administrative management particularly in the areas of timing, seating plans, trained staff and entry checking. Lower order perceptions were found in syllabus alignment, higher order thinking and marking satisfaction, expected grades, favoritism and objection handling. Moderation of assessments, rubrics, feedback, and a standard process for script review and grievance of results are practically suggested.

© 2026 The Authors, Published by iRASD. This is an Open Access article under the Creative Common Attribution Non-Commercial 4.0



Corresponding Author's Email: altafhussain@iub.edu.pk

1. Introduction

Assessment is one of the most powerful factors in higher education, linking the purpose of the curriculum, the way students learn, their academic progress and achievement, and the community's trust in university qualifications. Constructive alignment is based on the idea that assessment ought to be a valid representation of intended learning outcomes, not just a collection of marks (Biggs, 1996). Likewise, for long term learning, assessment is seen as a way to reinforce students' future learning and working capacity beyond merely proving their immediate recall (Boud & Falchikov, 2006; Sadler, 1989). The examination system is a core quality assurance issue at universities, especially in the semester-based educational model where several assessments, midterm exams, final exams, quizzes, assignments, grades and processes of grading and obtaining results all influence the student's education.

The research area is important because the exam systems impact students' study behavior, motivation to study, trust in educational institutions, and their perception of education justice.

Assessment is not a neutral technical process, but rather a process that students use to decide if the assessment system is fair, transparent, on-syllabus, and if it can catch what they've actually learned (Rasooli, Zandi, & DeLuca, 2018, 2019). If students feel that examination halls are well-managed, invigilation is equitable, question papers are unambiguous, marking is well communicated, and that objections are considered, they are more likely to view academic results as legitimate. On the other side, the lack of perceptions of favoritism or vague marking, insufficient design of the exam paper or delayed marks could also undermine trust despite good formal examination procedures. The problem addressed in this study is that universities' examination reform is likely to concentrate on reforms in the structure and schedule of examinations, and students' experience of examination effectiveness is a sequence of practices. The arrangements, like seating plans, examination time and facilities, are of importance in a semester system but it is not enough. A technically well-managed examination could be deemed to be weak if it contained content which was not in the syllabus, if the questions did not focus on practical or higher order learning, if students were not provided with marked scripts and if objections raised against marking were not considered. Recent research in the assessment field also indicates that for assessment to be of quality, it must influence students' willingness to use feedback and to accept grades as credible, with fairness and transparency of feedback at the heart of this (Carless & Boud, 2018; Dawson et al., 2019; Winstone & Carless, 2020).

This issue is of special importance in the context of Pakistan. Semester-based assessment is a common practice in public sector universities, but quality of implementation may differ from one to another, department to department, course to course, and examiner to examiner. The study of recent evidences in Pakistan reveals that the components influencing students' perceptions of semester system include curriculum delivery, classroom management, physical facilities, student-teacher interaction, workload, time management, and assessment practices (Chauhan, 2024; Khan, Khan, & Shah, 2024). But the evidence of students' assessment of the entire process of exam is still limited in the context of local. This study is an attempt to fill this void by focussing on the university students of Bahawalpur in Southern Punjab which is one of the important centres of higher education in this province. Therefore, this study aims to accomplish five objectives: first, to gauge students' perception in the university examination's process in terms of administration; second, to examine students' perception in terms of monitoring, invigilation and academic-integrity safeguards in the university examination process; third, to evaluate students' perception of the quality of the questions paper and its consistency with the learning objectives; fourth, to assess students' perception in the fairness of marking, access to scripts and transparency of the marking process; fifth, to identify practical areas in the credibility of the university examination system during examination. They are based on the premise that the effectiveness of the examination is multi-faceted and that the whole process of the examination fosters students' trust.

This study holds relevance for the university administration, the examination controller, teachers, quality-assurance cells, and policy makers. Administrators are able to tell from the findings if there are basic examination arrangements that are effective. The findings for teachers and departments focus on the need to enhance assessment design, rubrics, marking moderation and feedback. When students are the subject of study, the focus is on fairness, transparency, and voice in exams. The study has implications for assessment scholarship in revealing that procedural management and the fairness of assessment can be distinct, with students feeling more positive about the way in which assessment is administered, but comparatively less confident in the quality of the question paper, fairness of grades, and handling of objections. The rest of the paper is organized as follows. The literature review in Section 2 examines the theory and literature on the topic. The data and methodology are described in section 3. The results are presented and discussed in Section 5 is the conclusion of the paper and presents implications, limitations and future research directions.

2. Literature Review and Theoretical Framework

The literature on assessment in higher education starts with the assumption that assessment should be related to learning outcomes and teaching activities. Biggs (1996) posits that learning will be of higher quality if objectives, teaching methods and assessment tasks are constructively aligned. This perspective aligns with the assessment-for-learning (A4L) research which focuses on the development of evidence for learning and development, not just as a gatekeeping device (Sadler, 1989). In a higher education setting where there is an emphasis on semesters, this coherence is particularly significant as repeated assessments can either provide for ongoing learning or can enable grade-seeking superficial learning if poorly designed assessment tasks. Summative evaluations remain valuable, especially for the certification of achievement, for progression control and for the production of comparable academic records, in universities. However, summative assessment has a bigger impact than certification. Joughin (2010) points out that summative assessment influences students' learning priorities and study strategies, as well as their perceptions of what is important or valued by the institution. Boud and Falchikov (2006) also claim that the purpose of assessment should be to help prepare students for future learning, thus training them to evaluate the quality of their learning. However, an effective examination system should not only ensure that defensible grades are issued but should also motivate students to engage with knowledge, skills and standards meaningfully.

Fairness has been an important issue in modern assessment research because students' acceptance of an examination system is not only based on the result of the test but also on the integrity of the assessment system. It is more than just equal treatment, and involves clear expectations, opportunities for learning, consistent procedures, transparent criteria, impartial judgment, and respectful treatment of students (Rasooli, Zandi, & DeLuca, 2018). Based on the theory of organizational justice, Rasooli, Zandi and DeLuca (2019) define the concept of fairness in assessment in four related dimensions: distributive justice, procedural justice, interactional justice, and informational justice. Distributive fairness is about the fairness with which students are treated regarding their grades, that is, the fairness of the distribution of grades based on students' performance; procedural fairness is about the fairness with which students are treated in terms of the uniformity of assessment rules; interactional fairness is about the fairness with which students are treated in terms of the respectful and unbiased behaviour of the assessment staff and teachers; informational fairness is about the fairness with which students are treated in terms of their being given clear explanations of assessment criteria, expectations and evaluation procedures. Students' perceptions are thus a key element of assessment quality. Perceptions of fairness impact student trust of grades, acceptance of feedback, and legitimacy of academic decisions. In recent studies on assessment of Groupwork, it has been found that students' perceptions of the accuracy of contribution assessment, the transparency of the assessment process, and the quality of communication have an impact on their fairness judgements (Rasooli et al., 2025). While group work does not involve formal written testing, the larger point is valid: the students consider not just the results but also the processes by which the results are achieved as matters of fairness. This encompasses entry checking, invigilation, behaviors, monitoring of anomalies, question paper clarity, marking behaviors, and grievance systems during University examinations.

Another important aspect of the effectiveness of the examination is the quality of questions. Assessment tasks must reflect the intended curriculum, be clear in language, have well-defined criteria and be cognitively appropriate for the course. The success of rubrics relies on their criteria that reflect the purpose of the assessment and are clearly stated as levels of performance (Brookhart, 2018). Likewise, Bloxham, Boyd and Orr (2011) demonstrate that markers may have different interpretations of the criteria and in the grading process, which can present difficulties for consistency and transparency. The findings here indicate that the question papers and marking schemes need to be well moderated for students to see the assessment as valid and fair. This discussion has been heightened by the transition to authentic, higher order assessment. The goals of authentic assessment are to link academic work to authentic standards of excellence in the discipline, to authentic disciplinary practice, and to meaningful uses of knowledge (Ashford-Rowe, Herrington, & Brown, 2014; Villarroel et al., 2018). Villarroel et al. (2020) also show how methodologies of problem-solving, application and contextualized reasoning can be incorporated into even written tests to achieve authenticity. This is significant for students taking semester exams as they might think papers are less important if the questions are based on rote-learning, are out of syllabus or do not assess practical knowledge. More recent evidence also indicates a

link between the students' assessment preferences and their motivation and beliefs about relevance (Baines, Chauhan, & Otermans, 2025).

The third strand of the literature we wish to highlight is feedback and marking transparency. The fourth strand of the literature that we will explore is the transparency of feedback and marking. Boud and Molloy (2013) state that feedback should be a process that helps learners make use of the information for improvement. Carless and Boud (2018) build on this when they speak of student feedback literacy – the ability to value feedback, make decisions, regulate emotions and respond to feedback. Feedback is valued by both staff and students when it helps improve learning outcomes, and is important to be clear and timely and easy to use for students to improve their performance (Dawson et al., 2019). Winstone et al. (2017) also demonstrate that students must be given the chance and control over how to respond to the feedback, as opposed to it being handed to them. Feedback in examination systems can be provided by means of access to marked scripts, a description of grades, discussion of the answers with the teacher, a rubric and comments, formal objection procedure. Nicol (2021) highlights the role of students' internal feedback resulting from comparisons between their work and criteria, exemplars, and/or other students' work or teacher feedback. Hence, self-marking of marked scripts and questions regarding marking can be beneficial for the enhancement of the students' evaluative judgement (Tai et al., 2018). Assessment may not be transparent and may undermine confidence in the fairness of assessment if students are not able to view marked scripts or complain.

Examination credibility is closely tied up with academic integrity and invigilation. Examination staff is required to provide conditions for students to attempt papers which are as similar as possible and to minimize the possibility of the use of unfair means. But integrity protecting measures also need to be seen as respectful and not disruptive. However, research evidence for post-pandemic assessment indicates that students and teachers still view face to face assessment as fair and honest, provided assessment procedures are transparent and cheating is managed (Pozza, Costa-Pereira, & Tavares, 2025). Authentic assessment research also shows that, for tasks that call for meaningful application, rather than easily produced answers, assessment design can also dampen the integrity risks (Sotiriadou et al., 2020; Villarroel et al., 2020). The semester system in Pakistan has been encouraged to help students acquire the habit of regular study routines and better matching of teaching and assessment processes, as well as continuous assessment. However, empirical studies in recent years show that there is a dependency between student satisfaction and the semester system in terms of institutional conditions, teacher involvement, workload in assessment, facilities, and justice in academic procedures (Chauhan, 2024; Khan, Khan, & Shah, 2024). These studies will supply evidence that is useful in context but allow for further in-depth analysis of the examination process itself. Some localized evidence of how pupils view administration, invigilation, the quality of question papers, fairness of marking, students' access to scripts, and transparency of results was found, but is limited.

The lack of research is thus evident. Theories and empirical studies on constructive alignment, fairness, feedback literacy, authentic assessment and academic integrity are found in international literature, whereas, the Pakistani studies include semester-system satisfaction and challenges. But less research has focused on integrating these concepts in a student-centred examination of the entire examination chain in public universities of Southern Punjab. This study fills this void by exploring students' perception on five interrelated dimensions: Examination management, Monitoring and invigilation, quality of question-papers, fairness of marking and feedback, and preparation and declaration of the results. The theoretical perspective is based on the premise that an examination system is effective if it can be guaranteed to be reliable in administration, fair in procedure, valid in its academic assessment, transparent in marking and responsive to students' concerns.

3. Data and Methodology

A quantitative descriptive survey design was used to explore the effectiveness of examination system on the perception of students in a university in semester-based education. The design was suitable for the aim of describing and interpreting the students' reports of experiences in the various elements of the examination process, rather than to

draw causal conclusions. The selected public sector universities of Bahawalpur district in Punjab, Pakistan were the target of the study. The population frame consisted of students of Islamia University of Bahawalpur and Government Sadiq College Women University. The survey respondents consisted of 300 students of which 210 students were female and 90 students were male. Students came from various departments: science, arts, humanities, social sciences and education-related. The data were collected using a structured questionnaire with five-point Likert scale that ranged from strongly disagree to strongly agree. The questions in the questionnaire were grouped under the main phases of the examination process, namely: respondent profile, examination scheduling and management, facilities at the examination hall, monitoring and invigilation, question-paper setting, marking of answer scripts and preparation and declaration of results. The instrument covered aspects related to facilities, seating plan, examination time, training of staff, awareness of the rules, access to water and toilet facilities, checking on entry, control of unfair means, invigilator behavior, alignment of the question paper, presentation of the practical and higher order questions, fairness of marking, access to the marked papers, compilation of the results, neutrality of compilation and reevaluation, and acceptance of objections. The language of the items was kept to the point so that students would be able to assess specific practices, instead of institutional claims.

To determine the internal consistency of the instrument, Cronbach's Alpha widely used to assess the reliability of multi-item scales (Cronbach, 1951) was applied. The overall alpha value was 0.913 which showed good reliability of the questionnaire. IBM SPSS was used to enter and analyze data. The study was descriptive in nature and the purpose was to summarize the pattern of students' perceptions; therefore frequencies, percentages and mean scores were used for the analysis. Each item was classified into three categories of response (agree, neutral, and disagree). The two "strongly agree" and "agree" choices were grouped together for agreement, and the two "strongly disagree" and "disagree" choices were grouped together for disagreement. The scores were interpreted on a 5-point scale with higher scores reflecting more favorable perceptions for positively worded items. How this was interpreted for the favoritism item was handled with care as agreement was not considered a beneficial component of the examination system, but a fairness concern. Ethical issues were taken care of by informing the respondents about the academic purpose of the study, collecting responses for the use of study and reporting the findings in aggregate. No study-level student was found to be analyzed. The outcomes were analyzed in each dimension and interpreted to be useful for policy formulation and improvements of examinations at the university level. The analysis is based on the perceptions reported by 300 respondents and therefore is not a national assessment of the perceptions of all Pakistani universities but it is only student-based evidence from two public universities of Bahawalpur.

Table 1: Sample profile and instrument information

Characteristic	Description
Research design	Quantitative descriptive survey
Study context	Two public-sector universities in Bahawalpur district, Punjab, Pakistan
Sample size	300 university students
Gender composition	210 female students and 90 male students
Instrument	Structured five-point Likert-scale questionnaire
Main dimensions	Examination management, invigilation, question-paper quality, marking fairness, and result transparency
Reliability	Cronbach's alpha = .913
Analysis	Frequencies, percentages, and mean scores

Note: The table summarizes the empirical design and measurement information used in the study.

4. Results and Discussion

This section reports the results and discusses them with reference to literature that has been deemed as having assessment quality. The results are summarized dimension-wise in the questionnaire because the questionnaire includes several stages of the examination process, and then interpreted based on the evidence that has been obtained dimension-wise. It is a general trend that students gave higher ratings to the administration of the exam, monitoring of examination, as well as higher ratings related to probing grievances of students on their results than results transparency and the quality of the question-papers.

Table 2: Dimension-wise summary of students' perceptions

Dimension	Item coverage	Mean	Agree %	Neutral %	Disagree %	Interpretive emphasis
Examination-hall management	Items 1-10	3.80	73.3	10.7	16.0	Strongest area; timing, seating, trained staff, and rule awareness were perceived positively. Generally positive, though noise, distractions, and peaceful environment require attention.
Monitoring and invigilation	Items 11-18	3.76	68.6	14.2	17.2	Moderate; alignment with syllabus, optional questions, and higher-order thinking were weaker.
Question-paper quality	Items 19-25	3.45	61.5	13.6	24.9	Weakest area; concerns include favoritism, evaluation satisfaction, and grade expectations.
Marking, feedback, and grade fairness	Items 26-33	3.38	56.9	17.2	26.0	Moderate; timely compilation was stronger than objection acceptance and neutrality.
Result preparation and declaration	Items 34-39	3.55	59.4	20.1	20.5	

Note: Agree % combines agree and strongly agree; disagree % combines disagree and strongly disagree. The favoritism item was interpreted as a fairness concern.

4.1. Examination-hall management and administrative readiness

The first dimension depicts the students' positive perception of the management of the examination hall in general. Strict adherence to examination timing ($M = 4.21$), seating in accordance of seating plan ($M = 4.12$), awareness of rules among staff ($M = 3.97$), availability of basic facilities for comfortable paper attempts ($M = 3.87$) were the highest means respectively. The results indicate that examination administration was comparatively good in terms of the procedure. It is important because assessment processes must be predictable to be seen as fair; students will feel more comfortable with the results of the examinations if the processes are orderly, consistent, and professionally managed (Rasooli, Zandi, & DeLuca, 2019). Meanwhile, examination-hall environment ($M = 3.56$), water facility ($M = 3.57$), toilet facility ($M = 3.50$), and toilet access on request ($M = 3.52$) indicate that facilities and student comfort need to continue to be a focus. The nature of assessment environments impact on performance because students' ability to demonstrate learning to a large extent relies on the physical and psychological environment in which examinations are attempted. While administrative management was not the poorest performing of the three, the results indicate that universities should not simply use the examination hall to act as a space for supervision but should rather make it a space conducive to concentration, minimize

unnecessary stress and preserve the dignity during examination (Pozza, Costa-Pereira, & Tavares, 2025).

4.2. Monitoring, invigilation, and academic-integrity safeguards

Overall, perceptions of the monitoring and invigilation dimension were also generally positive. The students reported that books and materials were not kept in the examination hall (77.7%), students were properly checked while entering the examination hall (75.7%), unfair means were strictly monitored (70.7), and invigilators were fair during examination (70.0%). The results show that the universities had visible systems of safeguards against academic dishonesty. These safeguards are essential as they can compromise confidence in grades and the legitimacy of examination results (Sotiriadou et al., 2020). The results also highlight a weak area in the conduct of invigilation, albeit a softer tone. Only 55.3% of students have agreed that invigilators did not bother students with noise or distracting actions while 28.3% disagreed. Likewise, 64.3% said there was peace and quiet in the examination environment for examination staff. The results suggest that integrity control needs to be accompanied by supportive and respectful behavior. In addition to ensuring that no cheating occurs, fairness also requires that supervisory behavior does not have a detrimental effect on students trying to take the examination. Fairness means more than just not permitting cheating; it also means that supervisory behavior will not have a negative impact on students who are taking the examination honestly (Rasooli, Zandi, & DeLuca, 2018). The rule enforcement and student centred examination conduct should be therefore highlighted during examination staff training.

4.3. Question-paper quality and assessment relevance

Question-paper quality was identified as a moderate area and not a strong area. The most clear-cut positive result was that 70.0 % of students agreed that the language of the question papers was simple and clear ($M = 3.64$) and 64.7 % agreed that the printing quality was fine ($M = 3.55$). These outcomes show that overall, basic readability and presentation were satisfactory. Reliable language use is essential as language that is unclear can create construct-irrelevant difficulty for students, where they may be assessed on the ability to comprehend unclear questions instead of what they are supposed to be learning (Brookhart, 2018). The less strong findings are in terms of academic alignment and cognitive demand. Of the students, only 53.3% agreed that question papers were not out of syllabus, while 31.7% disagreed. Likewise, 60.3% noted that the questions emphasized higher order thinking and 26.7% noted that they did not. The results are significant because the tasks used for assessment are important in the development of constructive alignment, which must align with the intended curriculum and expected level of learning (Biggs, 1996). If students believe that there is content that is not in the syllabus or that they are not being assessed at a higher level, the credibility of exams might be called into question. Written assessments can be enhanced with the use of authentic, application-oriented and problem-based tasks, as proposed by (Villarroel et al., 2018; Villarroel et al., 2020). Consequently, the present results indicate that there is a need to implement question paper moderation, table of specifications, and tight coupling between the course learning outcomes and examination items. The findings that the students agreed that question papers tested their practical knowledge at 63.3% is encouraging but moderate. However, in modern higher education, the relevance and application of practice are becoming progressively considered key factors in the quality of assessment (Ashford-Rowe, Herrington, & Brown, 2014; Baines, Chauhan, & Otermans, 2025). The results suggest that the question-paper system is not the problem per se, but it needs systematic improvements. Universities need to help teachers create papers that include a balance of conceptual, application, analysis and synthesis questions, and not necessarily rely on recall-based questions.

4.4. Marking fairness, feedback, and grade transparency

The marking and feedback component was the least successful element of the system of examinations. A total of 46.3% of students found that the evaluation of papers by evaluators was satisfactory, while 49.0% found that the grades given were satisfactory as per their expectations. In addition, 50.3% agreed that the teachers preferred their favorites when giving examinations. This is a case of a negative item in a fairness sense and therefore agreement reflects a serious perception problem. The perception of favoritism is enough to

undermine institutional trust, even if there is no objective evidence of favoritism, and should be considered a quality assurance issue. This is in line with the larger literature on assessment in higher education indicating that marking is one of the most litigated areas of assessment. However, Bloxham, Boyd and Orr (2011) show that if assessment criteria exist, there can be problems of consistency as the markers' interpretation of the criteria may not be the same. According to Rasooli, Zandi and DeLuca (2019), fairness should be grounded in transparency of the procedures and sufficient information regarding the decision-making. So, lower ratings for satisfactory evaluation and for expected grades imply that students might not have fully understood the marking criteria, receive inadequate explanation or a lack of uniformity among evaluators. There are some positive findings. Most students (66.0%) felt that they could see marked answer scripts, marking was fair (62.3%) and marked scripts were made available prior to the compilation of results (62.7%). Such practices are beneficial because they provide students with feedback, opportunities for clarification, and opportunities for learning. The importance of making assessment information accessible to students and providing opportunities to engage with it is highlighted in feedback literature (Carless & Boud, 2018; Winstone et al., 2017). Likewise, Nicol (2021) and Tai et al. (2018) claim that the student's evaluative judgement is formed as they compare their work to criteria, comments, exemplars, and standards. So marked script access should not be regarded as a formality, but an established feedback mechanism.

Also, the evidence indicates that there are room for improvement in grade related practices. Students seem to think that the semester system gives them opportunities to get good grades, as evidenced by moderate agreement on the concepts of "good" and "better" with 54.7% and 63.7% respectively, but this does not completely clarify the students' concerns about fairness. Boud and Molloy (2013) warn, however, that feedback and grading systems can be educationally beneficial only if they allow learners to gain insight into their performance and contribute to their enhancement. Script viewing should therefore be used in conjunction with rubrics, marking schemes, model answers, discussion after the examination, and procedures to register objections.

4.5. Result preparation, declaration, and grievance handling

Perceptions of the result-related dimension were moderately perceived. There was a greater level of positivity when it came to timely result processes than neutrality and grievance processes. In particular, 65.0% agreed that results were compiled in the stipulated time and 62.7% agreed that there was no declaration of results delayed. These results suggest that the lack of results timelines was not seen as being a serious lack. Delayed results are of practical importance to the students because they impact their academic planning, course enrollment, scholarships and progression. Weaker responses were observed for neutrality and objection handling, however. Only 56.7% agreed that results were prepared by neutral staff, 58.7% agreed that answer sheets could be re-evaluated after declaration and 55.0% agreed that students' objections against script marking were accepted. The results of this study were consistent with the trend of the examination process that the latter examination phases are not trusted as much as the administrative phases. It is essential for informational fairness for students to be aware of the process for creating results, who verifies the results, and how objections are addressed (Rasooli, Zandi, & DeLuca, 2019). Even if there are formal grievance procedures in place, if they are not communicated, students can feel like the processes about results are "closed" or not available to them. In practice, this means that the transparency of the results should be improved by having documentation of moderation, written timelines for results, easy result script-review procedures, and documentation of objection decisions. The reforms would be in line with the literature on feedback-literacy and evaluative-judgment, which foregrounds usability and actionability of assessment information (Carless & Boud, 2018; Dawson et al., 2019; Winstone & Carless, 2020).

4.6. Integrated interpretation of examination-system effectiveness

The overall pattern of results demonstrates a difference between procedural effectiveness and assessment fairness. Pupils were fairly confident in the examination timing, seating arrangements, staff training, entry checking and monitoring of cheating. These are key strengths as they demonstrate that the administration of basic examination is working.

Weaker findings emerged in relation to question-paper alignment, higher order thinking, satisfactory evaluation, expected grades, perceived favoritism, neutrality of result preparation, and acceptance of objections, however. This indicates that there are weaknesses in the examination system in relation to its visible aspects of administration, and that it is not as strong in the more judgmental aspects of assessment design, marking, feedback, and handling of complaints. This pattern has theoretical significance. Assessment fairness requires that the conduct of the assessment be fair as well as the academic and informational quality of decisions made prior to and after the assessment event. Valid question-paper design Biggs (1996), fairness theory Rasooli, Zandi and DeLuca (2018, 2019) and feedback literacy Carless and Boud (2018) all focus on the students' access to information about performance that is useful for them. Thus, the results suggest that reform of examinations should go beyond just hall management, and should include assessment moderation, rubric clarity, feedback, and formal student appeal procedures. For the Pakistani semester-system context, the findings corroborate the existing evidence that students' satisfaction is dependent not only upon the academic but also upon the administrative conditions (Chauhan, 2024; Khan, Khan, & Shah, 2024). The contribution of this study is that it indicates the examination process is credible in those areas and risky in those areas where students have perceived risks. Administrative readiness is the strongest area and marking & transparency of results is the most important improvement area.

Table 3: Item-wise results on students' perceptions of the examination system

No.	Statement	Agree %	Neutral %	Disagree %	Mean
1	Facilities in the examination hall are appropriate for comfortable paper attempts.	73.7	9.0	17.3	3.87
2	Students are seated according to a seating plan.	83.3	6.0	10.7	4.12
3	Examination timing schedule is strictly followed.	84.7	8.0	7.3	4.21
4	Examinations are conducted by well-trained staff.	81.3	9.7	9.0	4.00
5	Examination staff are aware of rules and regulations.	80.7	5.3	14.0	3.97
6	The examination-hall environment reduces unnecessary pressure.	64.3	14.0	21.7	3.56
7	Water facility is available in the examination hall.	66.0	10.7	23.0	3.57
8	Toilet facility is available in or near the examination hall.	62.7	16.0	21.0	3.50
9	Water is provided on request during examinations.	70.3	15.0	14.7	3.69
10	Students are allowed to use the toilet on request during examinations.	66.0	13.0	21.0	3.52
11	Books and other materials are kept outside the	77.7	9.7	12.3	4.10

	examination hall.				
12	Students are properly checked while entering examination halls.	75.7	11.0	13.3	3.93
13	Use of unfair means is strictly monitored during examinations.	70.7	16.3	13.0	3.81
14	Students using unfair means are punished.	69.3	15.0	15.7	3.73
15	Invigilation is fair during the examination process.	70.0	16.3	13.7	3.78
16	Invigilators behave appropriately during examinations.	66.0	18.0	16.0	3.70
17	Invigilators do not disturb students through noise or distracting actions.	55.3	16.3	28.3	3.44
18	Examination staff provide a peaceful, noise-free environment.	64.3	10.7	25.0	3.59
19	Question papers do not contain material outside the syllabus.	53.3	14.7	31.7	3.29
20	Distribution of subjective and objective sections is appropriate.	62.0	14.7	23.3	3.47
21	Question papers test practical knowledge of the syllabus.	63.3	14.3	22.3	3.45
22	Questions focus on higher-order thinking skills.	60.3	13.0	26.7	3.41
23	Language of the question paper is simple and clear.	70.0	11.7	18.3	3.64
24	Question papers contain optional questions.	56.7	15.0	28.3	3.33
25	Printing quality of question papers is fine.	64.7	12.0	23.3	3.55
26	Teachers give favor to favorite students in exams.	50.3	13.7	36.0	3.14
27	Evaluation of papers by evaluators is satisfactory.	46.3	24.3	29.3	3.23
28	Grades received are according to students' expectations.	49.0	24.3	26.7	3.24
29	Grace marks are allowed under	54.7	20.0	25.3	3.32

	examination rules.				
30	The semester system enables students to get better marks.	63.7	14.0	22.3	3.55
31	Students have the right to observe marked answer scripts.	66.0	14.3	19.7	3.62
32	Marking of papers is fair.	62.3	13.7	24.0	3.51
33	Marked scripts are shown before result compilation.	62.7	13.0	24.3	3.47
34	Results are compiled within the stipulated time.	65.0	16.0	19.0	3.71
35	There are no malpractices during preparation of results.	59.0	22.7	18.3	3.56
36	Results are prepared by neutral staff.	56.7	19.3	24.0	3.44
37	Declaration of results is not delayed.	62.7	16.7	20.7	3.58
38	Answer sheets can be re-evaluated after result declaration.	58.7	22.3	19.0	3.58
39	Students' objections against script marking are accepted.	55.0	23.7	21.3	3.42

Note: Percentages are rounded to one decimal place. For Item 26, agreement reflects concern about perceived favoritism rather than a favorable evaluation of the examination system.

5. Conclusion, Practical Implications, Limitations, and Future Research

The present study explored perceptions of university students towards the effectiveness of examination system in semester system of higher education in Bahawalpur, Pakistan. The results show that the general attitude of the students towards the administrative and monitoring part of the examinations were positive. The responses to seating plans, examination timing, trained staff, awareness of the rules, entry checking, and monitoring of unfair means were comparatively positive. The findings suggest that the physical infrastructure and external components of the examination system are essentially pragmatic, and can serve as a good basis for the trustworthiness of the institution. The study further reveals that the examination system should be enhanced in its aspects that directly relate to the fairness and trust in assessment. Relative weaknesses were the alignment of the question papers with the syllabus, the use of higher order thinking questions, satisfaction with the evaluation process, grading to expectations, perceived favoritism, marking neutrality and objections to marking. The main finding is thus that effectiveness of examination cannot be determined just on the basis of the hall management or on the basis of declaration of results. A valid system of examinations for a semester might also incorporate the following features: valid question papers, open marking, useful feedback, and easy access to complaints.

The ramifications in practice are obvious. The moderation of question papers in universities needs to be improved by the use of course learning outcomes, tables of specification, and peer review of papers prior to the examination. Teachers need to be trained in designing a rubric, to mark with reference to the targets and to understand how to create questions which test application and higher order thinking skills. Examination staff need to be trained in rules as well as monitoring, but also on proper behavior, student dignity, and how to conduct themselves while examining students. Marked scripts should be demonstrated

using a formal process that allows the student to become familiar with the marking criteria, to compare their responses to the expected responses, and to pose queries that have been documented if the need arises. The procedures for the preparation of results must be neutral, have clear time frames and contain transparent communication on reevaluation and objection decisions. There are limitations of the study. This was conducted on students from 2 public universities in 1 district so the results need to be interpreted with caution. Descriptive statistics were used in the analysis and no gender, discipline, institution, semester and academic performance differences were tested. The conclusions reached from this study are also based on the students' perceptions, which is important to capture the students' trust and experience of the examination practice, but not to independently confirm the objective quality of each examination practice. Larger multi-university samples are needed in future research, a comparison of public and private institutions, inferential analysis, and an investigation of differences across gender, discipline, level of study, and academic achievement. Qualitative interviews with students, teachers, invigilators and examination controllers would also help in explaining why students see marking, favoritism and handling of objections as weaker aspects. The effects of rubric marking, moderated question paper setting and digital result grievance systems on students' confidence in semester-based exams may also be explored in future research.

References

- Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205-222. <https://doi.org/10.1080/02602938.2013.819566>
- Baines, S., Chauhan, S., & Otermans, P. C. J. (2025). Students' perception of authentic assessment in higher education: Exploring the relationship between assessment preference and motivation in higher education [Article 2441067]. *Cogent Education*, 12(1), 2441067. <https://doi.org/10.1080/2331186X.2024.2441067>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347-364. <https://doi.org/10.1007/BF00138871>
- Bloxham, S., Boyd, P., & Orr, S. (2011). Mark my words: The role of assessment criteria in UK higher education grading practices. *Studies in Higher Education*, 36(6), 655-670. <https://doi.org/10.1080/03075071003777716>
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399-413. <https://doi.org/10.1080/02602930600679050>
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698-712. <https://doi.org/10.1080/02602938.2012.691462>
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics [Article 22]. *Frontiers in Education*, 3, 22. <https://doi.org/10.3389/feduc.2018.00022>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Chauhan, M. S. N. (2024). Students' perspective on the semester system at higher education institutions in Pakistan. *Research Journal for Societal Issues*, 6(1), 140-155. <https://doi.org/10.56976/rjsi.v6i1.191>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. <https://doi.org/10.1007/BF02310555>
- Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education*, 44(1), 25-36. <https://doi.org/10.1080/02602938.2018.1467877>
- Joughin, G. (2010). The hidden curriculum revisited: A critical review of research into the influence of summative assessment on learning. *Assessment & Evaluation in Higher Education*, 35(3), 335-345. <https://doi.org/10.1080/02602930903221493>
- Khan, A. N., Khan, Z., & Shah, S. A. (2024). Factors affecting students' satisfaction regarding semester system: A mixed-methods study of higher education in Pakistan. *Journal of Social Research Development*, 5(3), 195-206. <https://doi.org/10.53664/JSRD/05-03-2024-16-195-206>

- Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, 46(5), 756-778.
<https://doi.org/10.1080/02602938.2020.1823314>
- Pozza, D. H., Costa-Pereira, J. T., & Tavares, I. (2025). Perceptions of students and teachers regarding remote and face-to-face assessments in the evolving higher education landscape [Article 360]. *Education Sciences*, 15(3), 360. <https://doi.org/10.3390/educsci15030360>
- Rasooli, A., Turner, J., Varga-Atkins, T., Pitt, E., Asgari, S., & Moindrot, W. (2025). Students' perceptions of fairness in groupwork assessment: Validity evidence for peer assessment fairness instrument. *Assessment & Evaluation in Higher Education*, 50(1), 111-126.
<https://doi.org/10.1080/02602938.2024.2354510>
- Rasooli, A., Zandi, H., & DeLuca, C. (2018). Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond. *Studies in Educational Evaluation*, 56, 164-181. <https://doi.org/10.1016/j.stueduc.2017.12.008>
- Rasooli, A., Zandi, H., & DeLuca, C. (2019). Conceptualising fairness in classroom assessment: Exploring the value of organisational justice theory. *Assessment in Education: Principles, Policy & Practice*, 26(5), 584-611. <https://doi.org/10.1080/0969594X.2019.1593105>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144. <https://doi.org/10.1007/BF00117714>
- Sotiriadou, P., Logan, D., Daly, A., & Guest, R. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, 45(11), 2132-2148. <https://doi.org/10.1080/03075079.2019.1582015>
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467-481. <https://doi.org/10.1007/s10734-017-0220-3>
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840-854. <https://doi.org/10.1080/02602938.2017.1412396>
- Villarroel, V., Boud, D., Bloxham, S., Bruna, D., & Bruna, C. (2020). Using principles of authentic assessment to redesign written examinations and tests. *Innovations in Education and Teaching International*, 57(1), 38-49. <https://doi.org/10.1080/14703297.2018.1564882>
- Winstone, N. E., & Carless, D. (2020). *Designing effective feedback processes in higher education: A learning-focused approach*. Routledge. <https://doi.org/10.4324/9781351115940>
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of reciprocity processes. *Educational Psychologist*, 52(1), 17-37. <https://doi.org/10.1080/00461520.2016.1207538>