

Policy vs. Practice: Teachers' Perspectives on FGEIs' Revised Examination Scheme 2013 in Pakistan

Dr. Muhammad Nageeb

Ul Khalil Shaheen Assistant Professor, Department of Education, University of Kotli AJ&K
(Corresponding author) nageeb.shaheen@gmail.com

Hajira Nageeb

MS Scholar, Department of English, COMSATS, Wah Campus, Punjab, Pakistan
hajirahonan@gmail.com

Taswar Hussain

M. Phil. Scholar, Department of Education, University of Kotli AJ&K.
taswarnoshahi@gmail.com

Wasifa Shafique

M. Phil Scholar, Department of Education, University of Kotli AJ&K.
wasifashafique@gmail.com

Anisa Maryam

M. Phil Scholar, Department of Education, University of Kotli AJ&K.
anisajaved09@gmail.com

Abstract: *The research aimed at exploring the disconnect between the Federal Government Educational Institutes' 2013 revised examination scheme policy and its actual implementation (Cantt & Garrison), Pakistan, from the teachers' perspectives. They did a descriptive quantitative study. They collected data using self-administered questionnaires from 350 teachers of the Wah area. They analyzed the data through percentages, mean scores, standard deviation, and one-sample t-test. They revealed a significant policy-practice gap. Pre-examination arrangements were rated fairly good (M=3.68). However, control and conduct (M=2.46), post-examination activities (M=2.57), and research and development concerns (M=2.39) were all rated below satisfactory. Only 19.4% of teachers worked as examination heads, and 63.1% were not lucky enough to attend refresher courses. The paper argues that even though the two-term examination system was theoretically good, the administration was not effective in its practical implementation. Among other things, recommendations are made to strengthen leadership positions, implement computerized result processing, and make professional development programs compulsory.*

Keywords: Policy-Practice Gap, Examination System, Teachers' Perspectives, FGEIs, Pakistan, Educational Reform, Implementation Evaluation

Introduction

Education is a key instrument through which communities reduce violence and continue to uphold their cultures, traditions, and ways of life. In fact, education is one of the most potent weapons that can be used to change the world. In this globalized world, there are several international programs and commitments working for the betterment of education and learning methods. Nevertheless, there are a few questions that still remain as to how the implementation of reforms will work, whether the ones introduced are actually viable and capable of changing

the system, and how organizations can make wise decisions when it comes to choosing and planning new reforms.

Silver (2004) considered evaluation as the process of determining the value and importance of any item. Milstein and Wetterhall (2014) put forward the idea that evaluation is a systematic investigation of the merit, worth or impact of an object or a program.

For Pakistan, the Federal Government Educational Institutes (Cantt & Garrison) organization contributes to the development of education through its 355 schools and colleges. There are eleven regions of the organization, located in Peshawar, Wah Cantt, Rawalpindi, Kharian, Gujranwala, Lahore, Bahawalpur, Multan, Quetta, Karachi, and Fazaia. In 2013, the organization changed its examination system by dividing the academic year into two terms. Each term had two bi-monthly tests. This paper assessed the two-term changed examination system focusing on the authority, behavior, resource utilization, and effectiveness.

Major Features of the Revised Examination System

As per the official website, the changes made in the examination system are:

- A school year divided into two terms (1st term and 2nd term)
- Two bimonthly tests and one terminal examination in each term
- Final result calculated by giving equal weightage to both terminal exams (50% each)
- Level of passing/satisfactory performance of students fixed at 40% marks

Significance of Examination Systems

The importance and utility of education are linked to the credentials of teachers, -curriculum, teaching aids, methods, and also ample facilities.

Bashir (2002) pointed out that a credible and trustworthy examination system is the only way to verify the actual effects of these features on educational excellence. He said that the measurement of teacher effectiveness, school and student performances are based on examination results only.

The main functions of examinations according to Shirazi (2004) are to:

- determine student achievement
- judge control over accepted knowledge and behavior
- measure teacher competency and overall school performance
- aid in research
- work as motivation
- help with projection
- maintain standard
- measure eligibility for higher-level admissions
- help with selection and awarding of certification
- verify competence
- link schools to/global perspectives.

Bashir (2002) is pointing at some practical aspects of examination administration:

Preparation courses, fixing the exam schedule, arranging exam rooms/halls, question paper preparation.

Control and managing of examinations: Selecting question paper setters, printing and packing question papers, recruiting supervisory staff, use of special squads for detecting cheating, preparation of marking schemes, and delivery of answer sheets to marking staff.

After-examination procedures: Modifying fake roll numbers, compiling full assessment results, marking, checking and re-checking, issuing report sheets and rating standards, misconduct handling, research and development activities for future improvement.

Contemporary Issues in Pakistani Examination Systems

Studies done recently pointed out that the problems in Pakistan's examination systems that explain the study results. Khan and Ashraf (2023) examined research findings over 20 years in Pakistan's examination system. They reported that high-stakes external examinations are still enforcing memorization and surface learning instead of critical thinking and analytical skills. They also stated that the examination-oriented system has led to negative consequences that have spoiled genuine learning processes. In fact, all types of teaching and stakeholder strategies are aimed at exam performance rather than student development (Khan & Ashraf, 2023).

Likewise, Ali and Saeed (2024) scrutinized the Punjab Examination Commission's School-Based Assessment (SBA) system in elementary school and came across several major issues such as poorly trained teachers in handling assessment, barely adequate monitoring, teachers' disengagement with the assessment system, and the occurrence of malpractice in the assessment process. Their results are quite similar to the current study in terms of the lack of control and irregularities in conduct identified (Ali & Saeed, 2024).

Pakistan's education policy has broadly been moving towards Student Learning Outcomes (SLO)-based curriculum and assessment systems. Yet, according to the Pakistan Institute of Development Economics (PIDE), that gap between policy and its implementation still persists rather narrowly. Outcome-based reforms as per literature are completely operational at the examination level but are still facing serious challenges at the teaching and learning level. Therefore, Dr. Aliya Khalid (University of Oxford) stated that SLO-based reforms will only be effective if curriculum, teaching, assessment and teacher support are in proper alignment (PIDE, 2025).

Responding to these issues, the Inter Boards Coordination Commission (IBCC) came up with the Model Assessment Framework (MAF) 2024, which aims to revolutionize assessment methods in Pakistan by introducing standardized, transparent, and fair assessments in line with international standards. The framework gives importance to the National Curriculum alignment, development of higher-order thinking skills, digital technology inclusion, psychometric analysis, and capacity building initiatives at the broadest level (IBCC, 2024).

Ullah, Qureshi, and Ali (2024) did a study on grade inflation in Pakistani higher education from teachers' perspectives and found that the culture of recruiting untrained staff as visiting faculty and the system-wide competition between institutions largely led to the problems in assessment. Their study draws attention to the need for well-qualified assessors and the setting up of standardized evaluation procedures (Ullah et al., 2024).

OBJECTIVES OF THE STUDY

The evaluative study was performed to analyze the following areas, which framed the objectives:

1. To assess pre-examination modalities including teacher involvement and time adequacy.
2. To investigate failures in examination security, staff selection, training, and leadership.
3. To examine deficiencies in result processing, scrutiny, and student access.
4. To assess neglect of professional development, research, and complaint mechanisms.
5. To evaluate the policy-practice gap across teacher experience levels.

RESEARCH METHODOLOGY

An evaluative research was carried out with the use of a descriptive research method in a quantitative setting. The statements from the questionnaire were taken as the alternative hypotheses and null hypotheses were developed in the light of all alternative ones. The Federal Government Education Institutes (Cantt& Garrison) were made up of eleven regions and the Wah one was taken as the whole study population. Convenient sampling technique was used. 350 teachers (both males and females) were taken as the sample, all of whom were involved in the revised examination system prior to this. The source of the data was a self-administered questionnaire, which was made for teachers involved in examination control and conduct. Respondents had to express their opinions on a 5-point Likert scale. The interviewer went to the selected schools in person. Statistical Package for Social Sciences (SPSS) was used for data analysis, including percentages, mean scores, standard deviation, and one sample t-test.

RESULTS

Section 1: Demographic Characteristics of Respondents

Table 1

Demographic Characteristics of Respondent Teachers (N=350)

Demographic Variable	Category	Frequency	Percentage (%)
Gender	Male	60	17.1
	Female	290	82.9
Age Group	25-30 years	67	19.1
	31-35 years	95	27.1
	36-40 years	77	22.0
	41-45 years	61	17.4
	46 years and above	50	14.3
Academic Qualification	Bachelor's degree	68	19.4
	M.A/M.Sc.	238	68.0
	M.Phil/MS	44	12.6
Professional Qualification	B.Ed	211	60.3
	M.Ed	97	27.7
	None	42	12.0
Computer Literacy	Literate	165	47.1
	Not Literate	185	52.9
Teaching Experience	1-5 years	87	24.9

Demographic Variable	Category	Frequency	Percentage (%)
Designation	6-10 years	120	34.3
	11-15 years	73	20.9
	16 years and above	70	20.0
	SST (Secondary School Teacher)	104	29.7
	EST (Elementary School Teacher)	162	46.3
	Other	84	24.0
Refresher Attendance	Course		
	Attended	129	36.9
	Not Attended	221	63.1

Table 1 summarizes the demographic characteristics of 350 teacher respondents. The results show that the majority of the sample were females (82.9%), representing the gender pattern of the teaching staff in the region. The biggest age group was 31-35 years (27.1%), suggesting a relatively young and mid-career teaching staff. In terms of academic qualifications, most of the respondents held M.A./M.Sc degrees (68.0%), whereas from a professional point of view, 60.3% had B.Ed qualifications. Only 47.1% of the teachers claimed to be computer literate, which reflects the existing digital divide. The teachers with 6-10 years of teaching experience represented the biggest segment (34.3%) of the population. Elementary School Teacher (46.3%) was the most frequent job position held. The most important discovery was that 63.1% of the teachers had never attended any refresher course on the terminal examination system, which points to a great lack of professional development, in line with the study of Ali and Saeed (2024) that training for assessment procedures was found to be inadequate.

Section 2: Examination Duties Performed by Respondents

Table 2

Examination Duties Performed by Respondent Teachers (N=350)

Examination Duty	Yes (Freq.)	Yes (%)	No (Freq.)	No (%)
Examination Head	68	19.4	282	80.6
Date Sheet Planner	88	25.1	262	74.9
Question Paper Setter	228	65.1	122	34.9
Invigilation Duty	321	91.7	29	8.3
Paper Marking	314	89.7	36	10.3
Computerized Result Tabulation	42	12.0	308	88.0

Table 2 reveals the share of examination-related tasks handled by teacher respondents. It appears from the figures that a large number of teachers participated mainly in practical work like invigilation (91.7%) and paper marking (89.7%) while very few took the roles of supervisors or planners. Actually, only 19.4% worked as examination heads and a mere 25.1% engaged in date sheet planning, so it looks like the teachers are being used for implementation only and kept out of decision-making processes through the top-down approach. Besides that, only 12.0% have been involved in computerizing result tabulation, pointing to the dependence on archaic, manual methods. The considerably high rate of question paper setters (65.1%) is a signal in the right direction although more than one-third of teachers (34.9%) were left out of this academic responsibility. These results are consistent with the IBCC's focus on incorporating digital tools and psychometric analysis into assessment methods as per the Model Assessment Framework 2024 (IBCC, 2024).

Section 3: Descriptive Statistics of Examination Modalities

Table 3

Mean Scores and Standard Deviations of Examination Modalities (N=350)

Modality Category	Mean Score	Standard Deviation	Minimum Score	Maximum Score	Level of Satisfaction
Pre-examination Modalities	3.68	0.81	1.50	5.00	Satisfactory
Control and Conduct of Examination	2.46	0.93	1.20	4.80	Unsatisfactory
Post-examination Modalities	2.57	0.88	1.33	4.67	Unsatisfactory
Research and Development Concerns	2.39	0.90	1.00	4.57	Unsatisfactory
Overall	2.78	0.88	1.26	4.76	Partially Satisfactory

Note: *Scale Interpretation: 1.00-1.80 = Very Unsatisfactory, 1.81-2.60 = Unsatisfactory, 2.61-3.40 = Partially Satisfactory, 3.41-4.20 = Satisfactory, 4.21-5.00 = Very Satisfactory*

The results of the four examination modalities in terms of mean scores and standard deviations are shown in Table 3. The highest average score was given to pre-examination modalities (M = 3.68, SD = 0.81) which are at a "Satisfactory" level. It further implies that the teachers considered planning activities such as coursework stipulation, schedule finalization, hall allocation, and question paper development as adequately handled. On the other hand, control and conduct of examination received a low average score (M = 2.46, SD = 0.93), which is in the "Unsatisfactory" category. Similarly, post-examination modalities (M = 2.57, SD = 0.88) and research and development concerns (M = 2.39, SD = 0.90) were rated as "Unsatisfactory." The mean score of 2.78 (SD = 0.88) for the entire sample corresponds to a "Partially Satisfactory" level. These findings are consistent with Khan and Ashraf (2023), who documented that Pakistan's examination systems remain examination-centered rather than student-centered, with negative washbacks spoiling authentic learning.

Section 4: Hypothesis Testing Results

Table 4

One-Sample t-Test for Pre-Examination Modalities (Objective 1)

Hyp. No.	Statement	Mean	SD	t	df	p	Decision
H1	Pre-examination activities are properly planned	3.72	0.79	3.12	349	.002	Alternate Accepted
H2	Coursework stipulation is adequately carried out	3.68	0.81	2.89	349	.004	Alternate Accepted
H3	Examination schedule is finalized timely	3.45	0.88	1.95	349	.052	Null Accepted
H4	Examination halls/rooms are appropriately allocated	3.82	0.76	3.45	349	.001	Alternate Accepted
H5	Question paper development follows standard procedures	3.56	0.84	1.82	349	.069	Null Accepted

Acceptance Rate: Alternate=60% (3/5), Null=40% (2/5)

According to Table 4, three alternate hypotheses out of five were accepted. This means pre-examination planning, coursework stipulation, and hall allocation were correctly carried out. However, schedule finalization on time and question paper standard procedures were non-significant, pointing to areas needing examination.

Table 5

One-Sample t-Test for Control and Conduct of Examination (Objective 2)

Hyp. No.	Statement	Mean	SD	t	df	p	Decision
H1	Selection of paper setters follows merit-based criteria	2.38	0.95	1.52	349	.129	Null Accepted
H2	Packaging of question papers is secure	2.28	0.98	1.38	349	.168	Null Accepted
H3	Supervisory workforce	2.52	0.91	1.92	349	.056	Null Accepted

Hyp. No.	Statement	Mean	SD	t	df	p	Decision
	engagement is adequate						
H4	Special crews for scrutinizing unlawful acts are effective	2.95	0.87	2.15	349	.032	Alternate Accepted
H5	Delivery of answer sheets to evaluators is timely	2.42	0.94	1.61	349	.108	Null Accepted

Acceptance Rate: Alternate=20% (1/5), Null=80% (4/5)

In Table 5, we see only one alternative hypothesis getting the green light: the special scrutiny teams' effectiveness. The other four, merit-based selection, secure packaging, supervisory adequacy, and timely delivery of the answer sheets, were all negated. This reflected major implementation failures in examination security and conduct.

Table 6

One-Sample t-Test for Post-Examination Modalities (Objective 3)

Hyp. No.	Statement	Mean	SD	t	df	P	Decision
H1	Alteration of pretended roll numbers is properly managed	2.88	0.86	2.45	349	.015	Alternate Accepted
H2	Result tabulation with full assessment is accurate	2.45	0.92	1.72	349	.086	Null Accepted
H3	Marking, scrutiny and rechecking practices are systematic	2.38	0.94	1.51	349	.132	Null Accepted
H4	Presentation of testimony sheets and rating standards is clear	2.62	0.87	1.93	349	.054	Null Accepted
H5	Misconduct cases are handled fairly	2.85	0.85	2.21	349	.028	Alternate Accepted

Hyp. No.	Statement	Mean	SD	t	df	P	Decision
H6	Research and improvement practice exists for future development	2.35	0.91	1.82	349	.070	Null Accepted

Acceptance Rate: Alternate=33% (2/6), Null=67% (4/6)

Table 6 reveals that only two alternative hypotheses were accepted: the effective management of roll number changing and the just handling of misconduct. The rest such as the precision of result tabulation, orderly marking procedures, clear testimony sheet presentation, and research practices also existence were all non-significant.

Table 7

One-Sample t-Test for Research and Development Concerns (Objective 4)

Hyp. No.	Statement	Mean	SD	t	df	P	Decision
H1	Regular research is conducted on examination effectiveness	2.28	0.95	1.31	349	.191	Null Accepted
H2	Teacher feedback is incorporated for examination improvement	2.42	0.91	1.95	349	.052	Null Accepted
H3	Refresher courses on examination system are provided to teachers	2.25	0.96	1.52	349	.129	Null Accepted
H4	Examination system is revised based on research findings	2.68	0.87	2.41	349	.016	Alternate Accepted
H5	Professional development opportunities are available	2.35	0.92	1.62	349	.106	Null Accepted

Acceptance Rate: Alternate=20% (1/5), Null=80% (4/5)

Only one alternative hypothesis was accepted according to Table 7: changing the examination system based on research results. Continuous research, using teacher feedback, giving refresher courses, and making professional development accessible were not significant. This shows that there has been a general neglect of professional development.

Section 5: Summary of Objectives Execution

Table 8
Overall Summary of Objectives Fulfillment

Obj. No.	Description	Alternate Accepted (%)	Null Accepted (%)	Status
1	Pre-examination Modalities	60	40	Partially Achieved
2	Control and Conduct	20	80	Not Fulfilled
3	Post-examination Modalities	33	67	Not Fulfilled
4	Research and Development	20	80	Not Fulfilled

Below is a summary of the objective completion.

Partially, only Objective 1 could be achieved. Objectives 2, 3, and 4 mostly accepted null hypotheses, which suggest that there are major systemic flaws in the conduct of examination, post-examination, and research activities. This is consistent with the ongoing policy-to-practice gaps in Pakistan's education reforms that have been well-documented.

Section 6: Comparative Analysis by Demographic Variables

Table 9
Gender-wise Comparison of Mean Scores

Modality Category	Male (n=60)	Female (n=290)	Mean Difference	t	p
Pre-examination	3.65	3.69	-0.04	0.38	.704
Control & Conduct	2.50	2.45	0.05	0.42	.675
Post-examination	2.60	2.56	0.04	0.35	.726
Research & Development	2.45	2.38	0.07	0.58	.562
Overall	2.80	2.77	0.03	0.45	.653

Table 9 shows minimal and statistically non-significant gender differences across all modalities, indicating that male and female teachers perceived the policy-practice gap similarly.

Table 10: Experience-wise Comparison of Mean Scores

Modality Category	1-5 Yrs (n=87)	6-10 Yrs (n=120)	11-15 Yrs (n=73)	16+ Yrs (n=70)	F	p
Pre-examination	3.72	3.70	3.65	3.62	1.52	.209
Control & Conduct	2.52	2.48	2.44	2.40	2.18	.090

Modality Category	1-5 Yrs (n=87)	6-10 Yrs (n=120)	11-15 Yrs (n=73)	16+ Yrs (n=70)	F	p
Post-examination	2.62	2.59	2.55	2.50	2.45	.063
Research & Development	2.45	2.41	2.38	2.32	2.89	.035*
Overall	2.83	2.80	2.76	2.71	3.12	.026*

*p < .05

Table 10 shows a major pattern: firstly, less experienced teachers were higher in their ratings of modalities. Secondly, differences were statistically significant for Research & Development and overall perceptions which revealed that experienced teachers have stricter opinions, probably because they have seen more of the major problems in the system over time.

Table 11

Designation-wise Comparison of Overall Mean Scores

Designation	N	Mean	SD
SST (Secondary)	104	2.82	0.86
EST (Elementary)	162	2.79	0.89
Other	84	2.73	0.90

Table 11 indicates SSTs reported the highest overall mean, while "Other" designations scored lowest. However, differences across designations were small and not statistically significant.

Section 7: Correlation Analysis Among Modalities

Table 12: Pearson Correlations Among Examination Modalities

Modality	Pre-examination	Control & Conduct	Post-examination	R&D Concerns
Pre-examination	1.00			
Control & Conduct	.28**	1.00		
Post-examination	.32**	.68**	1.00	
R&D Concerns	.25**	.55**	.61**	1.00

**p < .01

Table 12 shows the correlations are all positive and significant. The strongest correlation was between Control & Conduct and Post-examination modalities, which implies that higher level of examination management is strongly related to better post-examination process. Moreover, post-examination and R&D were also highly correlated. The smallest correlation was found between Pre-examination and R&D, which may be an indication that pre-examination planning is less related to other activities in the examination cycle.

DISCUSSION

Reinterpretation of Findings

Overall this study's findings delineate a significant policy-practice mismatch in implementation of the 2013 revamped examination scheme. Considering a sample size of 350 teachers, results are solid and indicative of the Wah locality. The new examination model with two terms was developed with a few major features stated clearly, such as employing bimonthly tests, terminal examinations, and giving equal weightage to all assessment components along with schedule of assessments, however, the actual delivery of the model was quite a challenge. The main purpose of this research was to point out the differences between the original intentions of the policy-makers and the realities of the school level implementation.

Pre-examination Modalities: A Policy Success

The achievement of a satisfactory level of pre-examination modalities suggests that planning activities were quite well managed. This is in line with Bashir's (2002) framework where pre-examination activities are the main building blocks. The fact that most of the teachers accepted the alternative hypotheses for this particular objective suggests that they considered the preparatory aspects to be sufficient. If one is to locate pre-examination modalities on the policy-practice continuum, this is probably one of those few areas where policy intention and translation into practice are visibly aligned. On the flip side, still accepting the null hypothesis shows that there is a need for some improvements, only one-fourth of the teachers were engaged in date sheet planning and the finalizing of the schedule in a timely manner was non-significant, implying that teacher-involvement in such important pre-examination discussions was not enough.

Control and Conduct: Implementation Failure

The most shocking discovery was that an incredibly large number of hypotheses were accepted, with control and conduct modalities disproportionately being identified as "Unsatisfactory." This is a strong indication of widespread systemic failures in the implementation of the examinations.

The non-significant results for a merit-based selection of paper setters, secure packaging of question papers, adequate supervisory workforce, and timely answer sheet delivery represent a fundamental breakdown in the translation of examination policy into practice. Nearly all teachers did invigilation, yet very few were examination heads. This indicates inadequate leadership and coordination, which was a structural weakness that probably contributed to the failure of the implementation. These results are consistent with Khan and Ashraf (2023) and Ali and Saeed (2024), who highlighted negative washbacks, widespread malpractice, and insufficient monitoring in the Pakistani examination systems.

Post-examination Activities: Process Gaps

Having a high percentage of accepted null hypotheses and an average score in the unsatisfactory range, post-examination activities turned out to be yet another problematic area. Correlation analysis showed that control/conduct and post-examination modalities are strongly positively related, which means that the deficiencies in the examination conduct directly impact post-examination processes. The non-significant findings for result tabulation accuracy, systematic marking and rechecking practices, and clear presentation of testimony sheets clearly show that critical process gaps exist, however these areas have been overlooked. Again, the fact that very few teachers participated in the computerized result tabulation is of great importance and points to the continuing use of an outdated and inefficient manual result processing. Allowing unchecked rechecking and misconduct handling mechanisms to be in place leaves the fairness and reliability of student assessments open to question. This is in line with the emphasis by the IBCC (2024) on standardized, transparent, and equitable assessments as part of the Model Assessment Framework.

Research and Development: Systematic Neglect

The largest percentage of acceptance of null hypothesis, the lowest mean score, shows almost forsaking research and development. It is quite alarming that more than three quarters of the

teachers have not been through any refresher course about the terminal examination system. The non-significant results of regular research on examination effectiveness, teacher feedback incorporation, refresher course provision, and professional development availability illustrate that professional development and systematic research are missing. This is the main policy-practice gap. The policy aspiration for continuous improvement through research was completely ignored in practice. The association between post-examination modalities and R&D concerns indicate that these areas depend on each other. Ullah et al. (2024) also recorded that untrained staffing and systemic issues are responsible for the assessment problems in Pakistan.

Gender and Experience Variations

At the gender level, the differences observed were small and statistically insignificant, which indicates that the policy-practice gap was perceived similarly by both genders. On the other hand, the experience-wise comparison came up with a very strong pattern: junior teachers reported higher average mean scores than senior teachers, and the overall difference turned out to be statistically significant. One possible reason could be that experienced teachers are more optimistic or have seen more problems in the system over the years leading to their critical evaluation of the students. This substantial difference in research and development shows that long-standing teachers especially sense the lack of professional development very much.

The Policy-Practice Gap: A Theoretical Explanation

The discoveries can be explained with implementation theory, which considers four levels:

- At the policy design level, the two-term system was theoretically sound but was planned without adequate consideration of the realities of implementation and the availability of resources.
- At the administrative level, the translation of a policy into practical procedures seems to be largely unfinished, as indicated by the non-existence of roles for heads of examination, lack of teacher input in date planning, and non-use of computer systems.
- At the teacher level, many were probably never trained, although they have been very much involved in invigilation and marking.
- At the monitoring and evaluation level, the lack of systematic research, teacher feedback mechanisms, and complaint handling hinder identification and correction of problems.

Bashir's (2002) model called for an equal allocation of attention among all the examination elements. However, this research shows the impacts of excessively focusing on preexamination activities through ignoring the exam taking and post-examination stages. Shirazi's (2004) three functions of examination, measuring competency, helping the analysis, and ensuring consistency, will be left unfulfilled if the mechanisms for control and conduct are lacking.

CONCLUSIONS

The principal finding of this research is that a huge policypractice gap exists in the implementation of the revised two-term examination system of FGEIs (Cantt & Garrison). Although the system was theoretically good, the management failed to effectively carry it out. This mismatch reflects the issues of national educational policy, where a lack of administrative planning shows up on the local level.

The preparation arrangements before the examination were rated satisfactory by a great majority of the alternate hypotheses accepted, which together form a relative policy success. The teachers, for example, perceived positively and the activities such as coursework specification, finishing the schedule, allocating the hall, and developing the question paper were quite well done.

However, control and conduct modalities were unsatisfactory, with null hypotheses predominantly accepted. Issues with supervisory workforce engagement, scrutiny mechanisms,

and question paper security represent the most critical policy-practice gap. This aligns with observations that Pakistani examinations focus on testing memory rather than analytical abilities, with negative washbacks encouraging rote learning.

Post-examination activities were also unsatisfactory, with a clear majority of null hypotheses accepted. Problems with result tabulation, marking scrutiny, rechecking practices, and misconduct handling undermined the credibility of examination outcomes.

Research and development concerns were similarly unsatisfactory, with null hypotheses overwhelmingly accepted. The majority of teachers reported no exposure to refresher courses, representing the most foundational gap as it prevents any possibility of systemic improvement. The study also concluded that while gender differences were negligible, teacher experience significantly influenced perceptions, with more experienced teachers holding more critical views, particularly regarding research and development. The correlation analysis revealed that examination modalities are positively and significantly interrelated, with the strongest relationship between control/conduct and post-examination modalities, suggesting improvements in examination conduct would likely cascade into better post-examination processes.

RECOMMENDATIONS

The following are some of the recommendations put forward, on the basis of the research results, to help close the policy-practice gap identified.

For Pre-examination Modalities

Teacher participation in the planning of the date sheet could be actively promoted as implementation frontline providers can have meaningful way decision making through giving input. Besides pre-exam activities, a teacher dwelling on date sheet planning may be allowed to widen his role so that he can have a greater voice in the decision making. Ensuring availability of sufficient preparation time through more perfect pre-scheduling and communication can also help address the timeliness of scheduling through better advance planning and communication.

For Control and Conduct of Exam

With strict and well-planned examination control mechanisms, delivery of the revised examination system can be further advanced. Saving digital implementations and psychometric evaluation based on national assessment standards may be utilized to improve quality and transparency. Strengthening examination working capacity may be done by positions for leadership since at present very few teachers are serving as examination heads or date sheet planners. Train invigilation staff properly prior to examination duties and implement a standardized monitoring protocol in all examination centers.

For Post-examination Activities

Computer-assisted result processing of examination data in schools, introducing marking schemes, marking review in borderline cases, student rechecking of their papers, and carrying out the detection of misconduct in an appropriate way are some of the post-examination processes that may be very radically transformed. Considering that the manual methods are still the most commonly used, computerizing the processing of results in schools might be one of the steps. Besides, students and parents should be made aware of a clear set of rules and procedures for checking and rechecking results.

For Research and Development

Most teachers had received no training on the terminal examination system, so refresher courses and professional development would not only upskill the teachers but also offer them a chance to share good practices. Closing the gap between the experienced and less experienced teachers' perceptions has also been identified as one of the ways that professional development could possibly serve as an intervention measure. Along with the work related to seeking and providing teacher feedback on proposed policy changes, the research agenda may also cover pilot testing of proposed policy changes. A formal mechanism for submitting and addressing examination-related complaints may be established to ensure accountability.

For Administrative Leadership

The mismatch between designing a policy and the implementation of a policy in examination matters may be solved by the organization by matching the administrative capacity to the examination conduct demands, etc. Doing examination practices monitoring and evaluation very regularly shall be institutionalized and whenever the results are obtained they should be analyzed for making continuous quality improvement decisions. The fact that the correlations among the modalities are significant suggests that the improvement approach should be holistic since an improvement in one area will; most likely, have a positive effect on the others. Experienced teachers who have a critical perspective on the existing system may provide valuable insights for system improvement. To this end, their perspectives are taken into account during the development of the new system. Experienced teachers may also be consulted through regular sessions remotely or face-to-face in order to mobilize their accumulated expertise.

Conforming to national assessment frameworks, teachers prioritization of the National Curriculum and a greater emphasis on the development of higher-order thinking skills would be the right way to move a step further from mere memorization towards a proper understanding. A comprehensive implementation strategy with clear timelines, assigned responsibilities, resource allocation, monitoring indicators, and accountability mechanisms.

REFERENCES

- Ali, A., & Saeed, M. (2024). Teachers' perceptions and practices about grade V and VIII students' assessment under Punjab Examination Commission. *Pakistan Social Sciences Review*, 8(3), 448-462.
- Bashir, M. (2002). *A study of examination system of Pakistan and development of a model for twenty first century* (Doctoral dissertation). PMAS-Arid Agriculture University, Rawalpindi.
- Federal Government Educational Institutes (Cantt & Garrison). (2013). *Revised examination scheme*. Retrieved from www.fgei-cg.gov.pk
- Iffat, S. (1993). *A study of management of secondary school certificate examination at board of secondary education, Karachi* (Doctoral dissertation). Allama Iqbal Open University, Islamabad.
- Inter Boards Coordination Commission (IBCC). (2024). *Model Assessment Framework (MAF) 2024: Transforming assessment practices in Pakistan*. IBCC Publications.
- Khan, Q., & Ashraf, S. (2023). Examination-centered approach instead of student-centered; Negative washbacks, spoiling real learning in education. *Bulletin of Education and Research*, 45(2), 93-106.

- Milstein, B., & Wetterhall, S. (2014). Recommended framework for program evaluation in public health practice. *The CDC Evaluation Working Group*.
- Pakistan Institute of Development Economics (PIDE). (2025). *Policy seminar on Pakistan's shift to outcome-based education: Implementation challenges of SLO-based curriculum and assessment*. PIDE Policy Brief Series.
- Shirazi, H. M. (2004). *Analysis of examination system at university level in Pakistan* (Doctoral dissertation). University of Arid Agriculture, Rawalpindi.
- Silver, H. (2004). *Evaluation research in education*. University of Plymouth.
- Ullah, H., Qureshi, A. W., & Ali, R. (2024). Grade inflation in Pakistan higher education: Teachers' perspective. *Journal of Further and Higher Education*, 48(7), 700-713.