



Beyond the Bell Curve: A Mixed-Method Evaluation of Examination & Assessment Reforms and Digital Transformation in the Board of Secondary Education Karachi (BSEK)

Noman Shahzad¹, Sadaf Raheem², Dr. Aftab Ahmad Shaikh³

¹ Assistant Director, Directorate Sindh Education Policy Commission(s) & MPhil Scholar, College of Education, Ziauddin University

Email: nshehzad38@gmail.com

² Assistant Director, Directorate Sindh Education Policy Commission (SEPC-S)

Email: sadafrahim94@gmail.com

³ Director, Directorate Sindh Education Policy Commission (SEPC-S)

Email: aftabirfan@hotmail.com

ARTICLE INFO	ABSTRACT
Submission: July 09, 2026 Revised: July 25, 2026 Accepted: August 26, 2026 Available Online: September 09, 2026 Keywords: <i>Grade Inflation, Examination Reform, Bell-Curve Distribution, Assessment Validity, Digital Transformation, Sindh Education Policy Commission, Board of Secondary Education Karachi</i> Corresponding author: Noman Shahzad Email: nshehzad38@gmail.com	Secondary Education system in Sindh has recently experienced significant assessment reforms including Uniform Examination Syllabus (UES), an internationally standardized letter-grade scale and digital transformation through introducing e-marking, OMR and QR-coded scanning. Despite these reforms, the examination system still needed a systematic overhaul. Recent, SSC-II (2026) results of Board of Secondary Education Karachi (BSEK) showed unusual high numbers of top-grades (A1 & A) comprising approximately ~49.72% of the total. The result reignite the concerns over grade inflation and unusual behavior of students' performance along bell curve trajectory. The study employed a mixed method design to examine whether the BSEK's result distribution reflect a genuine performance and achievement of students or just a grade inflation due to systemic anomalies. The study examine the key factors of this unusual behavior of top achievers through the six years results trends, key informative interviews and open-ended questionnaire from key stakeholders comprising paper setter, invigilators, core BSEK team, teachers & Head Teachers. The quantitative data were analyzed to see the longitudinal trends and a stratified weighted subset was constructed from 2026 result-band data to test the normality of the data. The results reject the normal bell curve with <i>P value</i> <.001 whereas the data skewed to -0.98 (negatively skewed i.e., high concentration on the left side i.e., top achievers). The further disaggregation of results (sector-level, gender-wise & district-wise) showed a widening performance gap between government and private institutions. Whereas qualitatively data set (n~127) identified examination-centre malpractice rather than genuine learning improvement, as the most cited explanation for high achievement, alongside a pronounced trust deficit in digital examination reforms. The findings suggest that statistical conformity to a bell curve should be treated as a diagnostic signal rather than a grading target, and that credible assessment reform requires strengthening examination governance, centre security, examiner capacity, and standards-based, evidence-anchored grading rather than relying on digital tools or distributional targets alone. The study concluded with the policy recommendations that includes independent examination oversight, professionalized invigilation, risk-based centre monitoring, and a longitudinal assessment analytics system for Sindh's examination boards.

Introduction

School education system in Sindh operates through a dual governance structure in which the School Education & Literacy Department (SELD) is responsible for sector policy and public-school administration, while statutory examination boards conduct secondary and higher-secondary examinations and award certification (Nawaz, et al., 2024). The learning outcomes in Sindh remain shaped by significant socio-economic and geographic disparities that include rural-urban divide in access and quality, OOSC, a large and loosely regulated private-school sector concentrated in urban centres like Karachi, and capacity constraints within the public examination boards themselves from item-writing expertise to secure logistics for printing, transport, and marking. These structural factors form the backdrop against which any assessment reform, however well designed, must be implemented (Imran, Akhtar, & Khan, 2026; Zaidi, et al., 2024). They also help explain reforms that demonstrate positive results under controlled pilot conditions may produce different outcomes when implemented at the provincial scale. Sindh's secondary and higher-secondary examinations are conducted by six Boards of Intermediate and Secondary Education (BISE), each with jurisdiction over a defined set of districts:

Table 1: Examination boards operating in Sindh

Board	Primary Jurisdiction	Notes
BSEK (Board of Secondary Education Karachi)	Karachi division	Established 1950; largest board by candidate volume; SSC only (Intermediate handled by BIEK)
BISE Sukkur	Sukkur division	Serves northern Sindh districts
BISE Larkana	Larkana division	Serves upper Sindh
BISE Mirpurkhas	Mirpurkhas division	Serves lower Sindh
BISE Shaheed Benazirabad (Nawabshah)	Shaheed Benazirabad division	Central Sindh
BISE Hyderabad	Hyderabad division	Second-largest urban board

Historically, each board has set and marked its own papers within a common curriculum framework, which produced persistent perceptions of inter-board variation in difficulty and grading leniency one of the principal drivers behind the province's move toward a Uniform Examination Standard. But recently the Sindh School Education Department approved a province-wide Uniform Examination Syllabus (UES). Under the UES, papers in English, Mathematics, Physics, Chemistry, and Biology are to be set to a common Table of Specifications (ToS) across all seven boards rather than each board following its own textbook-based tradition, with the stated aim of eliminating perceived disparities between "easier" and "harder" boards and shifting emphasis from rote recall toward understanding, application, and critical thinking (Akhtar & Kayani, 2024; Akhtar, et al., 2020; Anwar, et al., 2019).

Similarly the Sindh government approved the replacement of the decades-old raw-marks reporting system with an internationally aligned letter-grade scale, in line with directives from the federal Inter-Board Coordination Commission (IBCC) aimed at national grading uniformity. The approved scale runs A++ (96–100%), A+ (91–95%), A (86–90%), B++ (81–85%), B+ (76–80%), B (71–75%), C+ (61–70%), C (51–60%), D "Emerging" (40–50%), and U – "Unsatisfactory" (below 40%), with 40 percent set as the minimum passing threshold per subject and a re-appear opportunity built in for candidates graded (Imran, Akhtar, & Khan, 2026; Haider, et al., 2025).

Moreover, the BSE boards are being equipped with e-marking software, Optical Mark Recognition (OMR) systems, scanners, and QR-coded answer-sheet workflows intended to replace fully manual checking with a scanned, digitally tracked, and more auditable marking pipeline through the Sindh Secondary Education Improvement Project (SSEIP) implemented by the School Education & Literacy Department with Asian Development Bank (ADB) support. SSEIP's examination-system component explicitly targets the human-capital constraint behind weak item quality and inconsistent marking. It includes training cohorts of subject-specialist assessors in item writing, paper moderation, and exam development, and building board staff capacity in the new scanning and data-analysis tools.

Problem Statement

The recent reforms and a well-functioning high-stakes examination is expected to discriminate meaningfully among candidates of differing ability, so that grades carry information for universities, employers, and policymakers. The concern arises when a very large share of candidates cluster at the top of the distribution and negatively skewed the bell curve.

In the Science Group, 175,200 of 176,152 registered candidates appeared, of whom 141,688 (80.87 percent) passed. Within that passing cohort, 35,529 candidates (about 25 percent of those who passed, and one-fifth of all appearing candidates) **secured the** top A-1 grade, and a further 51,022 secured A together accounting for roughly 61 percent of all successful candidates in the group. The General Group result, announced separately with an overall pass ratio of 65.76 percent, is understood to show a broadly similar skew toward the upper bands, though pending a full grade-wise disaggregation, the two groups together produced roughly 88,000 of 177,000 candidates securing A-1 or A grades that has reopened questions about grade inflation that recur over a multiple years.

Two major concerns motivate the study. First, a statistical-validity problem i.e., a distribution this heavily skewed toward the top departs from the normal (bell-shaped) distribution that underpins classical assumptions about how ability is distributed in a large, heterogeneous cohort, raising the question of whether item difficulty, marking leniency, or a genuine improvement in preparation explains the pattern (Ansari, Akhtar & Hafeez, 2024; Akhtar, et al., 2021).

Second, an irregularity problem i.e., the 2026 examination cycle was also marked by public allegations of irregularities, including reports of bribery and paper leaks and a probe that found the relocation of more than 170 examination centers, which if substantiated would independently inflate results and confound any purely statistical explanation of the skew.

The reforms were envisioned to improve the quality of examination system as whole and gain the public confidence in the examination results. Despite significant reforms in Board of Secondary Education Karachi, the empirical evidences are limited to examine the impact of examination reforms.

Research Question

The study is constructed around the overarching concern to identify either the examination results are actual achievement of students or grade inflation. The research question elaborated as

- To what extent does the current grade distribution reflect the true gains in student achievement or does it indicate systemic grade inflation?

Literature Review

Grade Inflation and the Validity of Student Achievement Measures

Grade inflation define as awarding of higher grades without the actual performance or knowledge gain in students, highlighting its potential to weaken the reliability and validity of grades as indicator of academic achievement (Kaliszewski, 2025). The validity of assessment outcomes can be examined through the lens of distribution of examination grades. The normal or bell curve distribution provide a statistical benchmarks to identify the unusual patterns in the data set and to evaluate performance of students. Stankov (2024) argued that an assessment should reflect the underlying distribution of student abilities while comparing the effects of assessment reform in his study (Khoso, et al., 2024; Sultana & Imran, 2024; Ahmad, Bibi & Imran, 2023).

Kiziltas, 2024 believes that at secondary education level, the inflated grades are deeply rooted in a complex system of internal administrative demands and stakeholder pressure. Educator usually competing priorities, ranging from intent to motivate struggling students and prevent dropouts. When examination data deviates from expected distribution, it blurs the boundary between high and low achievers or somehow it creates a chaos to validate the actual capability of the system (Yeritsyan et al., 2022; Kiziltas, 2024). This study examine these deep rooted causes and evaluate either the higher grades shows the actual improvement in students' performance or it may indicate grade inflation.

Bell-Curve Validity & Educational Assessment

The "bell curve" term refer to the statistical approach used in social research to identify whether the data distribution spread symmetrical around the central value. Technically refer to norm-referenced grading, a practice in which grade boundaries are set so that a fixed proportion of candidates receive each grade regardless of absolute performance (Danish, Akhtar & Imran, 2025; Mankash, et al., 2025; Hafeez, Yaseen & Imran, 2019). At the same time, empirical research on how students themselves perceive bell-curve grading a correlational survey of 211 undergraduates across three universities finds that students hold a reasonably well-formed conception of bell-curve grading and generally perceive it as procedurally fair, even while associating it with more negative attitudes toward learning and increased inter-student competition (Tan Yuen Ling et al., 2020).

Corroborating this fairness concern with first-person student testimony, a qualitative account of Bell Curve grading collected at an Asian university found that students experienced forced-distribution grading as demotivating and inequitable regardless of individual effort, describing a shift from collaborative to competitive peer relationships and a perceived loss of transparency in how final grades were determined (Murad et al., 2021). Although this account concerns university-level grading rather than a public secondary-school examination board, it reinforces the present study's caution against treating "bell-curve" conformity as an implicit quality marker for BSEK, where a curve is imposed for its own sake rather than because it reflects a genuinely heterogeneous cohort, the wider assessment literature suggests it can undermine, rather than protect, the perceived fairness of results (Murad et al., 2021; Tan Yuen Ling et al., 2020).

Methodological Approach & Statistical Tests

Data Sources

The primary data source of the study is Six Year (2021-2026 annual results of Board of Secondary Education Karachi has been used for trend analysis.

Statistical Tests for the 2026 results of BSEK

The 2026 results of Board of Secondary Education, Karachi used as a base line to check the normality of the aggregated grade band distribution on the bell curve. The statistical results shows that the results are negatively skewed and high concentration lie on the high achievers rather than following the normality. Hence, the data distribution rejects the normality of the results.

Normality distribution for the 2026 results-BSEK

Test	Statistic	Result	Interpretation
Skewness	-0.98	Moderately-to-strongly negative	Left-skewed: mass concentrated above the mean, with a long lower tail (fails/low grades) – not symmetric
Excess Kurtosis	-0.39	Slightly platykurtic	Flatter-than-normal peak; grades are more spread across bands than a bell curve would predict
Shapiro-Wilk	W = 0.844, p < 0.001	Reject normality	Distribution is statistically distinguishable from a normal (bell curve) distribution
Kolmogorov-Smirnov	D = 0.173, p < 0.001	Reject normality	Confirms a significant departure from the normal CDF

Longitudinal Trend Analysis (2021–2026)

Table 2: Yearly aggregate BSEK results, 2021–2026

Year	Schools	Appeared	Pass %	High Achievers % (A1+A)	Fail %
2021	7,212	176,413	97.2%	42.9%	2.8%
2022	6,954	171,736	93.7%	57.6%	6.3%
2023	6,962	185,499	83.9%	42.9%	16.1%
2024	7,120	195,102	82.3%	37.6%	17.7%
2025	7,171	204,334	83.6%	41.4%	16.4%
2026	7,058	191,926	79.3%	46.2%	20.7%

The data revealed that total candidates appeared in the Examination has been grown recently with 15.6%, showing the improvement in enrollment in schools. However, the high concentration of high achievers remain above the normal distribution i.e., >40% of total population appeared for the exams.

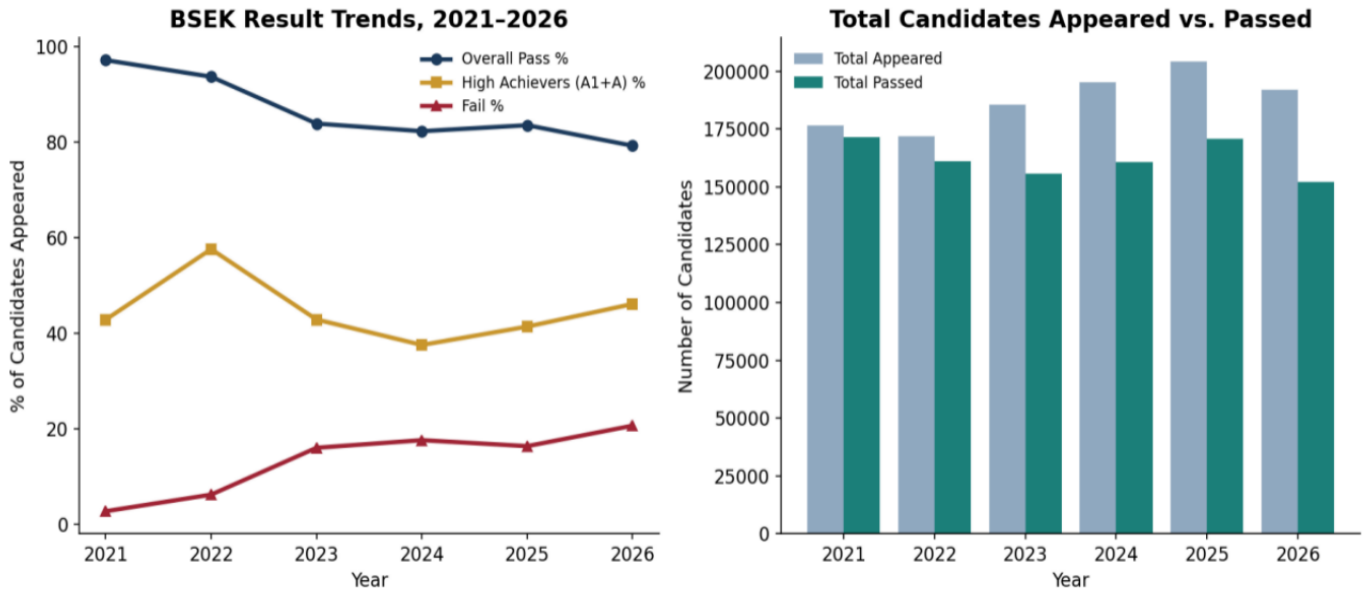


Figure 1: Yearly aggregate of BSEK Results (2021–2026 trends)

Bell-Curve Hypothesis Testing

The data distribution reject the null hypothesis of a normal distribution (Bell Curve) with $p\text{-value} < 0.001$. Whereas the distribution is skewed to the left side with value -0.98 showing the high concentration of the students lie on the high grades (A1 & A) rather than a systematic bell curve.

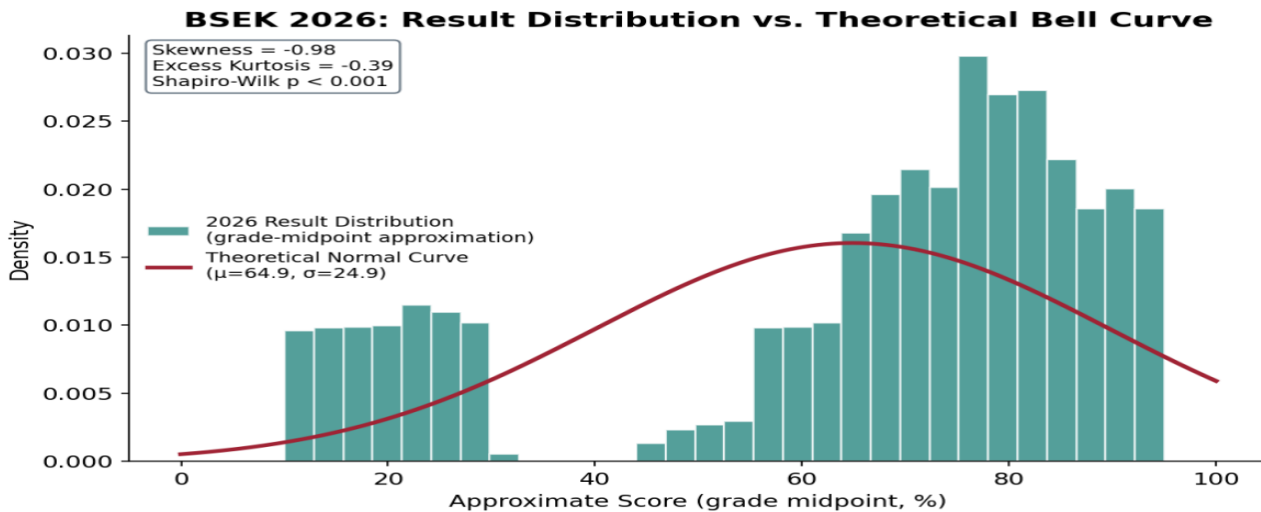


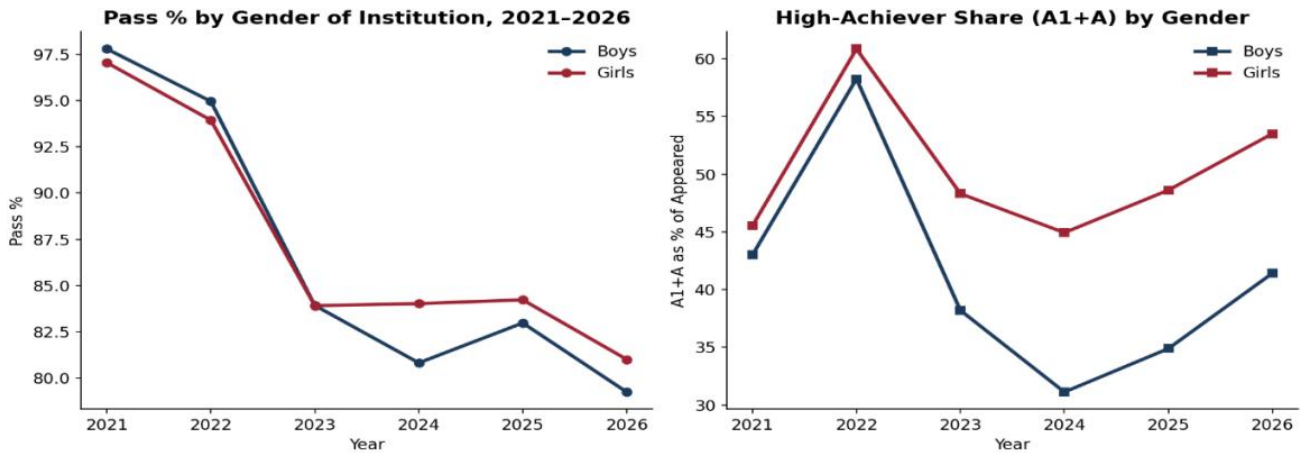
Figure 2: 2026 Result Distribution against the Theoretical Normal Curve

Gender-Disaggregated Analysis

The gender analysis used as one of the key factor to evaluate the performance gap between the results of girls and boys appeared in BSEK examination system. The results shows that girls' grades are comparatively higher than boys in every years.

Table 3: Pass rate and high-achiever share by institutional gender

Year	Boys Pass %	Girls Pass %	Gap (Girls-Boys)	Boys High-Ach %	Girls High-Ach %
2021	97.8%	97.0%	-0.7 pts	43.0%	45.5%
2022	95.0%	93.9%	-1.0 pts	58.3%	60.9%
2023	83.9%	83.9%	0.0 pts	38.2%	48.3%
2024	80.8%	84.0%	+3.2 pts	31.1%	44.9%
2025	83.0%	84.2%	+1.2 pts	34.9%	48.6%
2026	79.3%	81.0%	+1.8 pts	41.4%	53.5%

**Figure 3: Gender wise Pass % (left) and high-achiever (right), 2021–2026.**

Government vs. Private Sector Comparative Analysis

The government sector's pass rate has collapsed by 32.0% since 2021 (94.3% → 62.3%), more than double the private sector's 13.5% (97.9% → 84.4%) over the same period. The government/private pass-rate gap has widened from 3.6 points in 2021 to 22.1 points in 2026 – the sharpest structural divergence identified in this analysis.

Table 4: Government vs. private-sector examination results, 2021–2026

Year	Govt Schools	Govt Pass %	Private Schools	Private Pass %
2021	1,001	94.3%	6,823	97.9%
2022	960	87.3%	6,555	95.4%
2023	722	71.0%	6,240	87.6%
2024	747	69.0%	6,377	86.2%
2025	774	70.1%	6,397	87.9%
2026	771	62.3%	6,301	84.4%

Notably, the number of registered government schools fell from 1,001 (2021) to 771 (2026) – a 23% reduction – while average cohort size per government school rose from 34 to 57 candidates, implying consolidation of candidates into fewer, larger government examination centres. Private-sector school counts fell more modestly (6,823 → 6,301, -7.6%) with comparatively stable average cohort sizes (21 → 23).

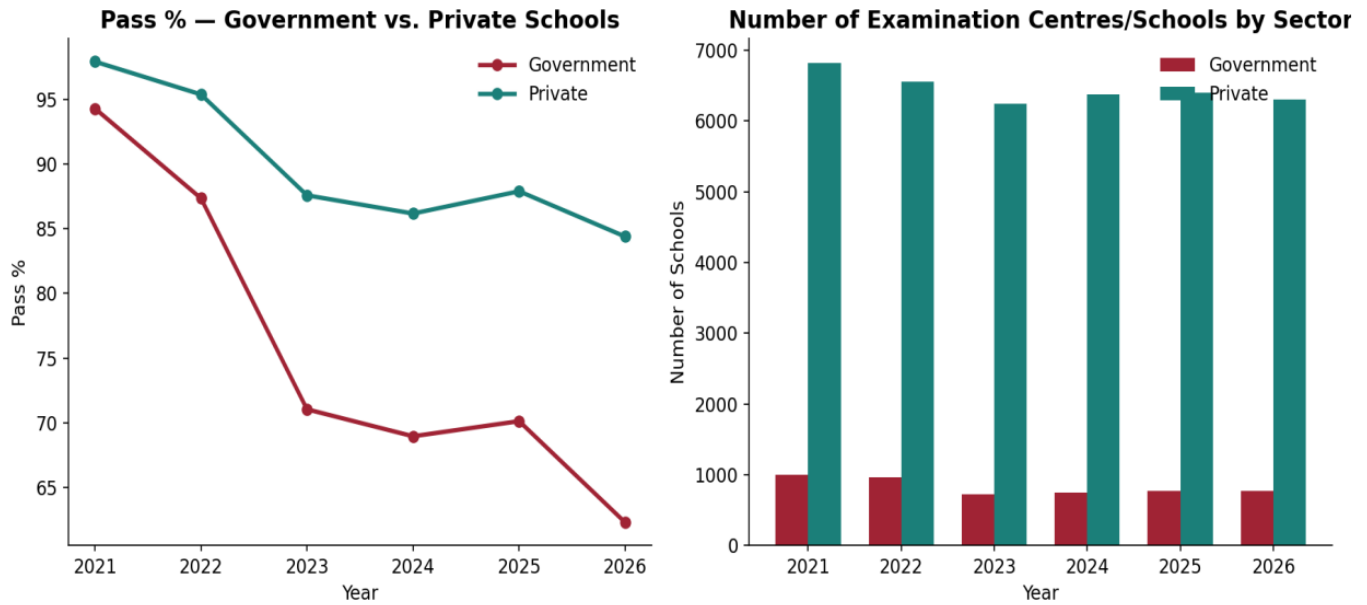


Figure 4: Pass % (left) and number of examination-registered schools (right), Government vs. Private, 2021–2026.

Correlation between a school's candidate volume and its pass rate is negligible in both sectors (government: $r = -0.09$; private: $r = +0.01$), indicating that school size is not a meaningful driver of results in either sector, the government decline is a systemic capacity and quality issue, not simply an artefact of larger, harder-to-manage cohorts.

The government schools facing continues challenges to the post-pandemic tightening of assessment standards reflecting several gaps including teacher shortages, limited capacity, larger class sizes per teacher. These challenges left them less able to absorb the stricter grading system than better-resourced private schools.

District-Wise Participation & Performance Trends

The seven districts of Karachi have been mapped to analyze the performance trend among the districts. The 2021-2026 results data shows that District West Karachi has experiences a decline of 7.6 points in these years while District South Karachi's performance decline by 31.2points. Therefore, based on the performance, these districts are grouped into two; resilient districts included District West, Korangi and Malir Karachi that show consistently rank at or near the top of the annual pass percentages whereas the second group that reflect high decline in pass percentages in data set years include District Central, East and South.

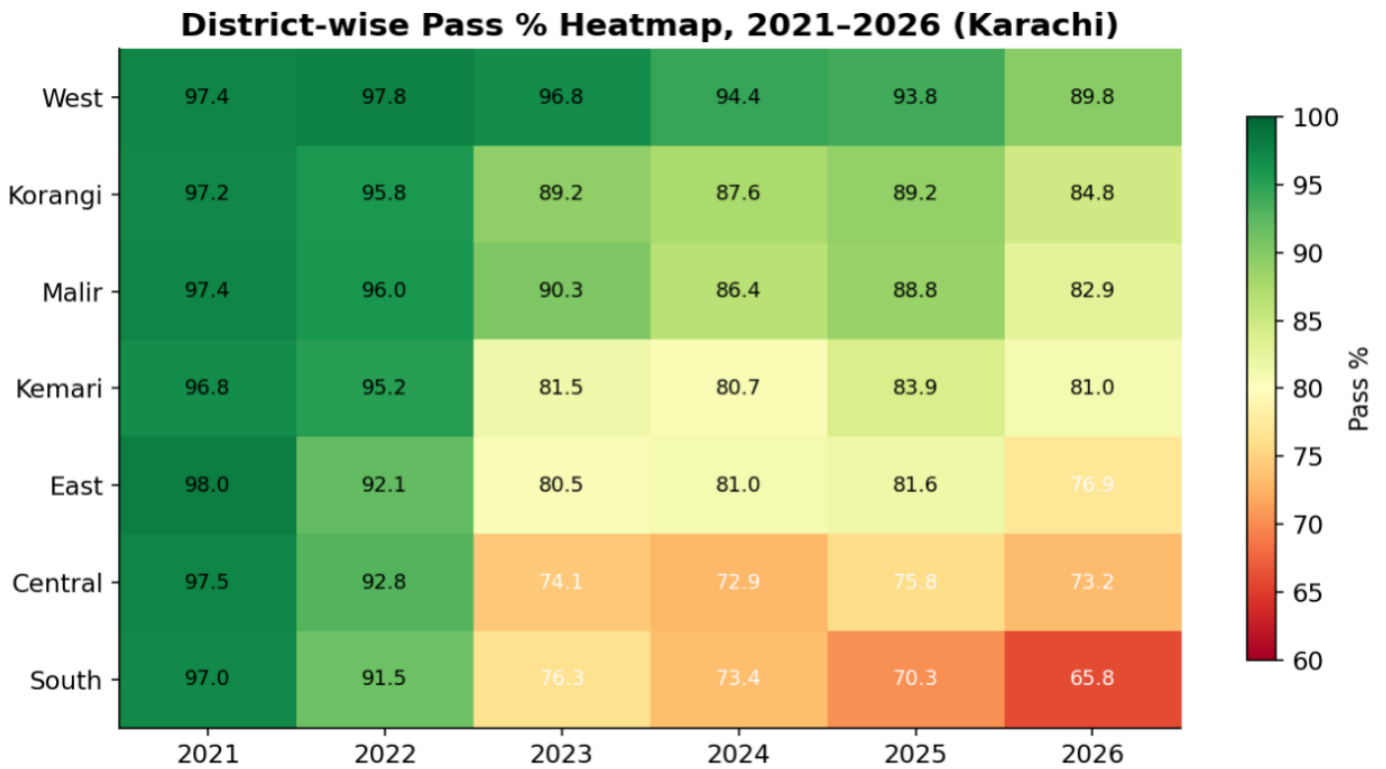


Figure 5: Mapping of District-wise pass percentage, 2021–2026

Table 5: District-wise pass-rate trajectories

District	2021 Pass %	2023 Pass %	2026 Pass %	Change 2021→2026
West	97.4%	96.8%	89.8%	-7.6 pts
Korangi	97.2%	89.2%	84.8%	-12.4 pts
Malir	97.4%	90.3%	82.9%	-14.5 pts
Kemari	96.8%	81.5%	81.0%	-15.8 pts
East	98.0%	80.5%	76.9%	-21.1 pts
Central	97.5%	74.1%	73.2%	-24.3 pts
South	97.0%	76.3%	65.8%	-31.2 pts

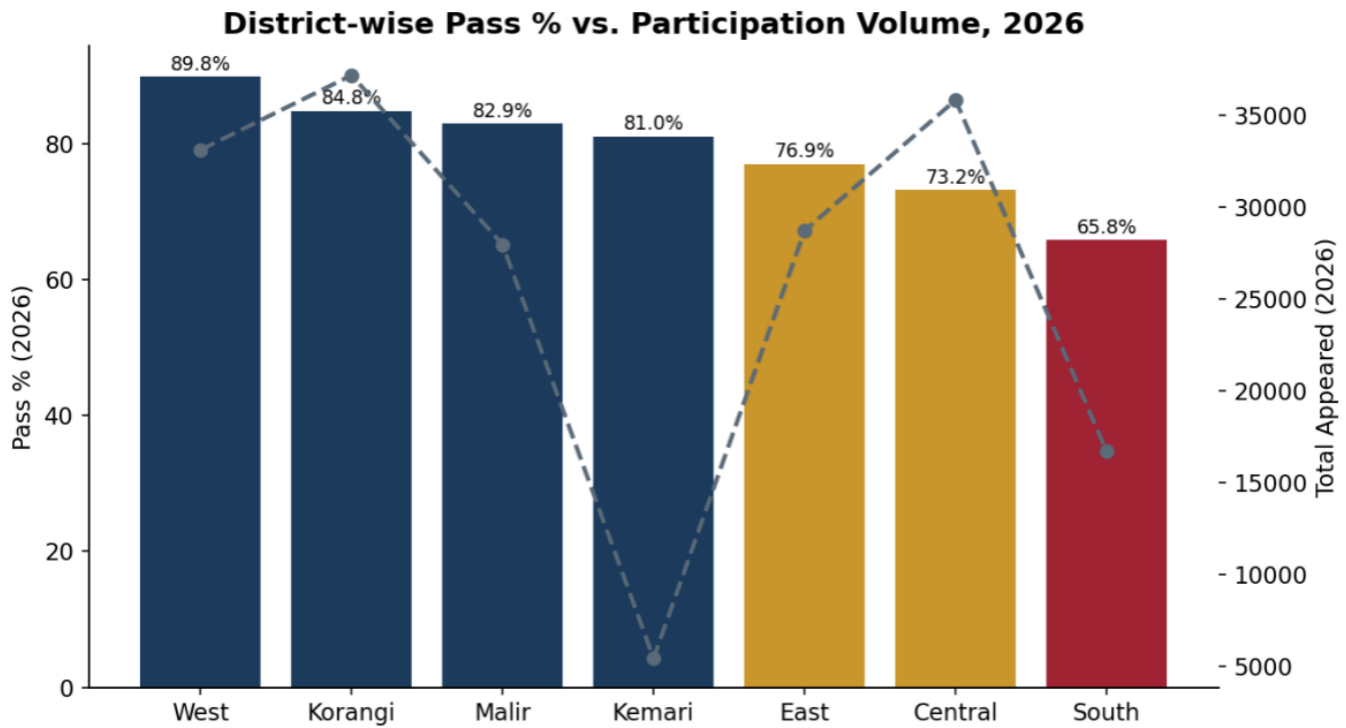


Figure 6: District-wise pass % and participation volume, 2026

Qualitative Analysis: Stakeholder Perspectives on Examination Reform

The qualitative approach is consistent with the existing study design, in which survey is designed to collect the response from the key stakeholders including secondary school teachers, head teachers/principals, examiners/markers, examination and assessment experts, JESTs, and Paper setters, providing perspectives from both school-level and examination-system contexts. The 127 responses are reviewed and interpret across questions focusing examination reforms including digital system, reasons of high concentration of high achievers (A1 & A), and areas requiring improvement to strengthen the validity and reliability of the results. The responses consolidated into broader themes so that data could be interpreted and helps to design the recommendations of the study.

Theme 1: Perceived Reasons of High Achievers including Malpractice & Non-meritorious activities

The data shows 52% of the respondents perceived cheating in examination centres are the reason of higher grades whereas only 41 % of the respond to effective teaching and learning practices. In addition to this, some of the perceived reasons that respondents highlight include paper leakages, mas shifting of centres and grace marks (Table 6).

Table 6: Perceived Reasons for High Achievers (A1 & A Grades)

Sub-theme	Respondents (n ~ 127)
Cheating in examination centres	52%
Teaching and learning	41%
Paper leakages	30%
Marking / moderation practices	11%
Shifting of Centres	7%
Grace marks	4%

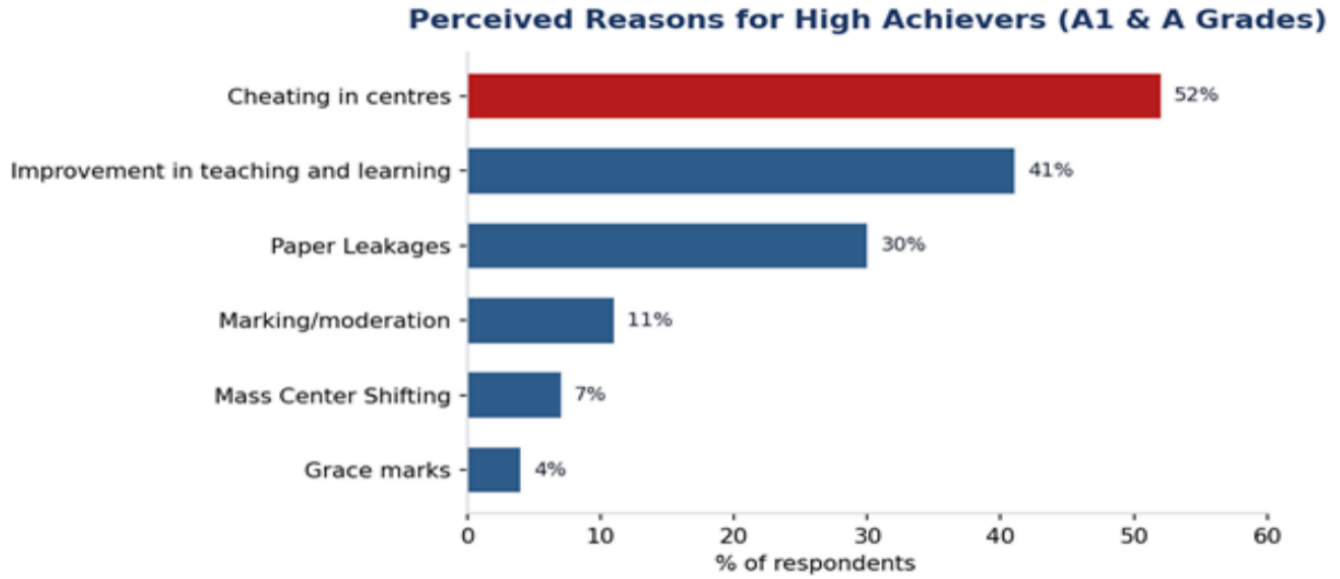


Figure 7: Perceived Reasons for High Achievers (A1 & A Grades)

Theme 2: Digital Transformation – Real Efficiency Gains, but a Deep and Internally Inconsistent Trust Deficit

Respondents rated the effect of digital transformation (e-marking, OMR, scanning, QR-coded answer sheets) on seven dimensions of the examination system *table 8*.

Table 7: Efficacy of recent Digital Transformation in BSEK

Dimension	Increased / Sig. Increased	No Change	Decreased / Sig. Decreased
Reliability of marking	37%	7%	56%
Transparency	37%	11%	48%
Examination security	26%	15%	59%
Accuracy	26%	22%	52%
Efficiency	30%	22%	48%
Accountability	30%	22%	48%
Public confidence	33%	15%	52%

After digital transformation in board by introducing OMR sheets, QR Coded answer sheet and scanners, the survey data still highlighted significant challenges. Stakeholders' responses to digital tools are positive and look at them as a means that cannot stand alone without the necessary human, pedagogical and institutional support. Two of the responses are given below;

"Digital systems should be supported by strict security measures, proper supervision, and strong action against malpractice to ensure a truly fair and transparent examination system." – Secondary School Teacher

"Unless you control the cheating culture, none of these techniques will affect examination integrity." – Head Teacher/Principal

Theme 3: Examination Security as a base fault line

The data indicates that despite numerous measures taken by the Board of Secondary Education Karachi, significant gaps remain in the examination system that limit the examination security and integrity. The key factor highlighted in the survey data included alleged external influence over the allocation of examination centres, the use of mobile phone within examination centres, the circulation of paid answer keys, paper leakage and untrained invigilators and examination staff integrity. Respondents viewed these

factors as potential enablers to organized malpractices and serious threats to the credibility and fairness of the examination process in BSEK.

"Neither students nor the invigilators should be allowed to have mobiles in the centre premises, because the answers of MCQs are provided to the students through mobile." — Head Teacher/Principal

"Examination centres must be out of reach of school authorities." — Examination/Assessment Expert

"Co-examiners should be replaced, as marking becomes a side business for them rather than a duty focused on students' future." — Examiner/Marker

Theme 4: Capacity Gaps in Assessment Design, Marking Standardization and Examiner Support

A second, technically-oriented cluster of concerns focused on the professional infrastructure behind the exam rather than on misconduct. Respondents particularly examiners/markers and assessment experts called for:

- Stronger item-writing capacity and a properly piloted Table of Specifications (TOS), so that papers are built to a defined blueprint of content/cognitive-level coverage rather than being predictable or repetitive year to year.
- Create marking criteria for constructed-response items using rubrics.
- To improve examiner training and rotate center superintendents.
- Greater pay and reduced workload for marking, as cited by one respondent as a catalyst for hurried, incomplete marking.
- Papers design that test application/creativity rather than rote recall, on the logic that conceptually demanding items are both more valid and harder to "leak" usefully.

"Papers must be designed on creativity/application levels so students cannot use short-cuts — answers should reflect cognition rather than rote memorization." — Examiner/Marker

"Centre Superintendents have a key role in transparency; after every two papers each Centre Superintendent should be exchanged with another at least at the town level — it will reduce the grip of certain groups at examination centres." — Examiner/Marker

Theme 6: Stakeholder-Prioritized Reform Agenda

In this theme, when respondents were ask "what should be improved" the following priority areas emerged, ranked by frequency:

Table 8: Stakeholder-Prioritized Areas for Improvement

Improvement area-Sub theme	Respondents (n~127)
Strict examination monitoring	70%
Transparent centre allocation	59%
Secure paper management	56%
Examiner training	48%
Question-paper quality	33%
OMR / scanning process improvement	26%
Table of Specifications	22%
Accountability / punitive action	22%
Item-writing capacity	22%
E-marking process improvement	22%
Moderation	11%

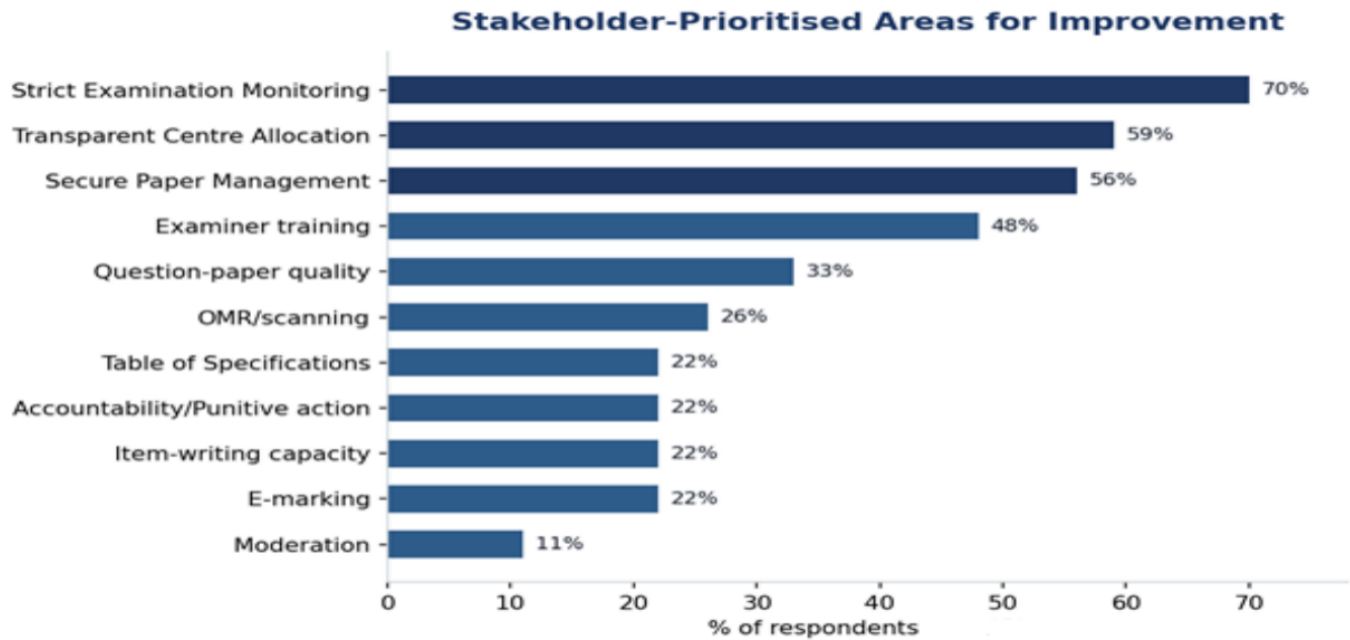


Figure 8: Stakeholder-Prioritized Areas for Improvement

The ranking provides a useful basis for sequencing policy recommendation for the Board of Secondary Education Karachi. Strict Examination monitoring, transparent examination-centre allocation, and strict paper security are three highest-ranked priority areas. These areas provide support related to process, governance, and examination integrity rather than technology or curriculum reform.

Discussion

The findings provide a strong basis to support the hypothesis that the examination results are not following the bell-curve graph. The 2026 distribution is substantially different from a conventional bell-shaped distribution, with negative skewness (-0.98), a concentration of candidates in the upper grade bands, and a distinct lower tail of low grades and failures; both the Shapiro-Wilk and Kolmogorov-Smirnov tests rejected normality ($p < .001$). Whereas, the qualitative findings emphasized that BSEK grades reflect students' actual knowledge and skills at only 2.70 out of 5, while agreement that high grades represent genuine improvement rather than grade inflation was only 2.89 out of 5. More importantly, when respondents were asked directly about the reasons for the high proportion of A1/A grades, cheating in examination-centre was the most frequently identified factor (52%), followed by improvement in teaching and learning (41%) and paper leakage (30%). This ordering is significant because it places examination integrity ahead of genuine learning improvement in stakeholders' explanations of unusually high achievement. The open-ended responses further identified invigilator collusion, unauthorized mobile-phone use, circulation of answers, favourable examination-centre arrangements, payments to examination personnel, and paper leakage. These perceptions do not themselves show the extent of malpractice, but when shared with other teachers, heads, examiners and assessment specialists, they can be a significant qualitative indicator that the grades observed are not solely due to student learning.

This study is in line with the distinction between genuine achievement and unearned grades made by Kaliszewski (2025) and Kızıldaş (2024) who stated assessment practices and institutional pressures can set a stage for grade inflation. The evidence does not indicate that all high grades are inflated; it indicates that there is a sufficiently strong integrity concern regarding the grades that they are not necessarily indicative of genuine achievement without further validation.

Not only is it a marker of "Beyond the Bell Curve" thinking, but this study offers empirical support for a grading goal of statistical conformity. At the same time, the qualitative findings reveal a deep examination-system trust deficit. Despite the introduction of e-

marking, OMR, scanning and QR-coded answer sheets, 56% of respondents perceived a decrease or significant decrease in marking reliability, 59% stated a decline in examination security, 52% for reduced accuracy, and 52% stated reduced public confidence.

The findings also identifies technical weaknesses in assessment design, including inadequate item-writing capacity, insufficient use of Tables of Specifications, limited assessment of higher-order thinking, inconsistent marking criteria, inadequate examiner training, and concerns regarding examiner workload and accountability.

Nevertheless, the convergence between the quantitative distributional anomaly and the qualitative evidence of weak confidence in examination validity raise serious policy attention. The study therefore supports a shift from asking whether BSEK results should "look normal" toward asking whether the grades provide credible evidence of what students know and can do. The qualitative evidence points particularly strongly toward examination monitoring, transparent centre allocation, secure paper management, examiner training, stronger assessment design, standardized marking, and accountability as the areas requiring priority intervention. The digital transformation should not stand in for governance but rather be seen as an enabler to governance as part of this reform agenda. The results thus indicate that, if the validity of BSEK is to be increased, attention needs to be paid to the standards of assessment, security of exams, the design of assessment, the reliability of marking, institutional accountability and the professional capacity of the examination staff members.

Conclusion

The results of this study indicated that there are credible evidence of an assessment-validity problem and plausible evidence of grade-inflation effects. The most critical implication is that BSEK should implement a standards-based assessment system, and assign grades to specific and clearly defined learning outcomes, while continuing to monitor them statistically and qualitatively. The bell curve is best used as a diagnostic tool. The following areas should be prioritized: examination security, transparent and independently managed centre allocation, examination monitoring, secure paper handling, examiner training and rotation, standardization of marking and moderation, greater focus on application and higher-order thinking skills, and higher quality examiners' items and Tables of Specifications. Further examination reforms should take place, but the effectiveness of these reforms should be measured within a wider governance framework as technology cannot address poor examination integrity. The significant variations between institutional sectors and districts also suggest that requiring stronger grades is an inadequate response to the need for assessment reform, and that more specific attention to the needs of disadvantaged and underperforming school contexts is also required. Finally, changes in grades would need to be related to changes in learning, which would require an increased amount of evidence.

Policy Implications & Recommendations

Establish an Independent Examination Governance and Oversight Mechanism

Enhance the institutional governance of Board by introducing a properly defined and independent examination oversight mechanism where administration of examination, allocation of centres, invigilation, marking, moderation and processing of results are kept separate. The results show that issues of students' conduct are not the only cause for concern; issues of the governance environment in which examinations are administered are also a concern such as perceived influence over examination centres, weak accountability and possible conflicts of interest. A more robust governance structure should, therefore, implement with clear selection and allocation criteria, keep a record of how decisions are made, record conflicts of interest, make inspections risk-based and include post-examination audits. Complaints which have arisen in connection with examinations should be directed to a separate and expedient grievance process and, where malpractice is substantiated, appropriate disciplinary action should be taken. This will move the examination administration role from a more administrative role to a risk-based system of quality assurance and accountability.

Introduce Independent and Professionally Trained Invigilators for High-Stakes Examinations

The Board should propose to establish an independent invigilation pool consisting of trained staff who are not directly employed by the school(s) or invigilation centre to which they are assigned. The qualitative results reveal that the main threats to the credibility of examinations identified are invigilator collusion, the use of mobile phones, unauthorised assistance and poor supervision at the centre level. An independent invigilation model, customized to the context of Sindh and based on the model adopted by international examination systems would minimize conflicts of interests and help to bolster the credibility of examination supervision. Invigilators are to be certified in an accredited manner and trained on a regular basis in examination procedures, Protection of confidential information, emergency procedures and professional ethics. They would also be assigned at

random via a secure digital allocation process and not as a result of a school or centre's choice. BSEK might consider using independent chief invigilators or external inspection teams in higher-risk centres to offer an extra level of security.

Strengthen Assessment Design, Marking Reliability, and Examiner Professional Capacity

The Board should design a programme of assessment quality that aims to incorporate validated Tables of Specifications, enhance item-writing skills, include more application, reasoning and higher-order cognitive skills items, have consistent marking rubrics, offer examiner training, and conduct moderation and adequate examiner workload and remuneration. The qualitative results show that some respondents were not only concerned about negligence in the examination centres, but also indicated that the design of the question paper, marking consistency and examiner capacity are poor. Assessment papers need to assess evidence of learning rather than mainly rote recall and standardization and moderation should minimize examiner variation. This would enhance the trustworthiness of the grades irrespective of the form of the grade distribution.

Establish a Risk-Based Examination Centre Monitoring and Rotation System

A systematic Examination Centre Risk Index should be instigated to identify centres that need greater supervision due to past malpractice issues, atypical grades, complaints, examiner/candidate ratios, sudden deviations in performance or other integrity indicators. High-risk centres should be subject to unannounced inspections, more independent invigilators, more stringent entry restrictions and better surveillance; centre superintendents and supervisory staff should have periodic changes of role to avoid the formation of local networks or long-term relationships. The allocation of centres should be automated, and should be determined by clear and auditable processes, and schools should have no influence on the choice of centres. This recommendation is directly related to the qualitative result that 70% of the respondents identified as strict examination monitoring, 59% as transparent centre allocation, and 56% as secure paper management as their three top priorities.

Develop a BSEK Assessment Analytics and Transparency System

The Board shall implement a longitudinal evaluation-measurement system that combines candidate level marking data, grade level distribution, test item performance, school and district level characteristics, gender-stream data, differences by stream and indicators of irregularities in the examination procedures. The system should provide an annual assessment validity and integrity system based more heavily on the distributional patterns, unusual concentration of grades, achievement gaps and integrity risks rather than just percentages passing and failing. This architecture should take into account the digital examination reforms such as e-marking, OMR, scanning or QR based systems, using auditable digital trails and independent Quality Assurance. This would help BSEK and policymakers to know if there are real improvements in the process of learning and in assessment or if there are some distortions in the assessment process.

References

1. Burton, K. J. (2006). Designing criterion-referenced assessment. *Journal of Learning Design*, 1(2), 73–82. <https://eprints.qut.edu.au/17721/>
2. Danish, F., Akhtar, N., & Imran, M. (2025). AI-Driven Personalization in Educational Marketing: A Framework for Enhancing Student Recruitment and Retention. *Journal of Political Stability Archive*, 3(2), 559-590.
3. Mankash, M. A., Ahmed, S. T., Shabbir, N., & Imran, M. (2025). Second Language Learning in the Digital Age: How Technology Shapes Language Acquisition at Universities in Karachi, Pakistan. *Liberal Journal of Language & Literature Review*, 3(1), 182-199.
4. Hafeez, A., Yaseen, G., & Imran, M. (2019). Management Paradigm Change in Pak--Turk (International Schools & Colleges) After a Failed Military Coup in Turkey: A Case Study.
5. Khoso, F. J., Shaikh, N., Dahri, K. H., & Imran, M. (2024). Educational Nurturing in Underdeveloped Contexts Unraveling the Dynamics of Student Teachers' Holistic Development. *Spry Contemporary Educational Practices*, 3(1).
6. Sultana, Z., & Imran, M. (2024). Challenges Faced by English Teachers in Pakistan. *Spry Journal of Literature and Linguistics*, 2(1).
7. Ahmad, D. N., Bibi, N., & Imran, M. (2023). Effects of Teacher's Motivation On Students' academic Performance At Public Secondary Schools In Karachi Pakistan. Available at SSRN 5165741.

8. Imran, M., Akhtar, N., & Khan, N. (2026). Ethical Leadership in Early Years Education: Implications for Teacher Professional Ethics and the Learning Environment–A Qualitative Study. *Journal of Political Stability Archive*, 4(1), 260-281. <https://doi.org/10.63468/jpsa.4.1.15>
9. Haider, S., Mirwani, R., Imran, M., & Haider, H. (2025). The Emotional Landscape of Organizational Change: A Qualitative Study of Teachers in Lasbela, Balochistan. *Journal of Political Stability Archive*, 3(3), 1482-1502. <https://doi.org/10.63468/jpsa.3.3.98>
10. Zaidi, S. S., Imran, M., Khoso, F. J., & Sultana, Z. (2024). Impact of Technologies and Co-Curricular Activities on Students' Academic Achievement at the Undergraduate Level. *Journal of Social & Organizational Matters*, 3(3), 75-90. <https://doi.org/10.56976/jsom.v3i3.99>
11. Anwar, U., Raja, F. U., Sadia, H., & Akhtar, R. N. (2019). AN INVESTIGATION OF REQUEST STRATEGY EMPLOYED BY ESL UNDERGRADUATE STUDENTS. *Journal of Social Sciences and Media Studies*, 3(2), 1-8.
12. Akhtar, N., & Kayani, M. M. (2024). Navigating The Pedagogical Landscape: A Qualitative Inquiry Into Teachers' Perceptions And Preparedness For Teaching License In Sindh, Pakistan. *Voyage Journal of Educational Studies*, 4(2), 197-223.
13. Akhtar, N., Hussain, S. Q., Raja, F., Naeem, T., (2020). Using Oral Communication Strategies to Enhance Students English Writing Skills of Grade VII Students: Experimenting ESL Strategies. *International Journal of Disaster Recovery and Business Continuity* 11(3), pp. 1127-1137.
14. Akhtar, N., Aslam, N., Hussain, S. Q., & Memon, A. (2021). Qualitative Analysis of Diffusing ERE cycle in Teaching at Higher Education: Experiential Learning Practices in Pakistan. *Ilkogretim Online*, 20(5), 7525-7539.
15. Ansari, R., Akhtar, N., & Hafeez, A. (2024). Exploring the Perceptions of Secondary School Teachers on Teaching Licenses in Sindh: A Qualitative Study. *International Journal of Social Science & Entrepreneurship*, 4(3), 126-147.
16. Corcoran, S. (2024). Beyond the bell curve. *Journal of Psychologists and Counsellors in Schools*, 34(2), 212-222. <https://doi.org/10.1177/20556365241251681>
17. EBSCO Research Starters. (n.d.-a). Norm-referenced testing. EBSCO Information Services. <https://www.ebsco.com/research-starters/social-sciences-and-humanities/norm-referenced-testing>
18. EBSCO Research Starters. (n.d.-b). Criterion-referenced testing. EBSCO Information Services. <https://www.ebsco.com/research-starters/social-sciences-and-humanities/criterion-referenced-testing>
19. Frigerio, A. (2026). Inequity by design: How Bell Curve grading undermines student success [Preprint]. ResearchGate. <https://www.researchgate.net/publication/393318241>
20. Kaliszewski, N. (2025). Ready or not, here they come: Investigating grade inflation in rural high schools and the GPA-ACT gap [Doctoral dissertation, Youngstown State University]. The dissertation identifies grade inflation as a discrepancy between awarded grades and demonstrated academic mastery.
21. Kızıltaş, Y. (2024). An important problem we need to face in schools: Inflated grades and grade inflation. *European Journal of Education*, 59(4), 1-17. <https://doi.org/10.1111/ejed.12744>
22. Stankov, P. (2024). The death of exams? Grade inflation and student satisfaction when coursework replaces exams. *International Review of Economics Education*, 46, 100289. <https://doi.org/10.1016/j.iree.2024.100289>
23. Nawaz, H., Ali, M. A., Rai, S. I., & Maqsood, M. (2024). Comparative analysis of cloud based SDN and NFV in 5g networks. *The Asian Bulletin of Big Data Management*, 4(1), 206-216.
24. Tan Yuen Ling, L., Yuen, B., Loo, W. L., Prinsloo, C., & Gan, M. (2020). Students' conceptions of Bell Curve grading fairness in relation to goal orientation and motivation. *International Journal for the Scholarship of Teaching & Learning*, 14(1), Article 7. <https://doi.org/10.20429/IJSOTL.2020.140107>

25. Yeritsyan, A., Mjelde, J. W., & Litzenberg, K. K. (2022). Grade inflation or grade increase. *Journal of Agricultural and Applied Economics*, 54(2), 375–393. <https://doi.org/10.1017/aae.2022.15>

Online Media Sources

26. 24NewsHD. (2026, January 20). Sindh govt replaces 'numbers system' with grading in all education boards. <https://24newshd.tv/20-Jan-2026/sindh-govt-replaces-numbers-system-grading-education-boards>
27. Daily Times. (2026, July 10). Karachi Board declares Class 10 Science results. <https://dailytimes.com.pk/1520900/karachi-board-declares-class-10-science-results/>
28. Eduvision. (2026, July 8). Big change for Sindh Board students: E-marking and digital exams coming soon. https://www.eduvision.edu.pk/edu_news/big-change-for-sindh-board-students-e-marking-and-digital-exams-coming-soon-news-4455
29. Express Tribune. (2025, November 13). Sindh boards to get e-marking, scanning, and QR code software for 2026 exams. <https://tribune.com.pk/story/2577301/sindh-boards-to-get-e-marking-scanning-and-qr-code-software-for-2026-exams>
30. Express Tribune. (2026a, January 2). Sindh to have uniform exam standards. <https://tribune.com.pk/story/2585188/sindh-to-have-uniform-exam-standards>
31. Express Tribune. (2026b, July 8). Sindh moves to digitise exams across all boards. <https://tribune.com.pk/story/2617218/sindh-moves-to-digitise-exams-across-all-boards>
32. Geo News. (2026, January 2). Sindh introduces uniform examination standard across all boards. <https://www.geo.tv/latest/642867-sindh-introduces-uniform-examination-standard-across-all-boards>
33. Minute Mirror. (2026, January 2). Sindh launches uniform exam system across all boards. <https://minutemirror.com.pk/sindh-launches-uniform-exam-system-across-all-boards-483562/>
34. Murad, A., Jefri, H., & Le Ha, P. (2021, October 23). Why the Bell Curve system for giving grades needs reform. *University World News*. <https://www.universityworldnews.com/post.php?story=20211019082700738>
35. Pakistan Truth. (2026, July 1). Karachi Board announces Matric General Group results 2026. <https://www.pakistantruth.com/matric-general-group-results-2026/>
36. ProPakistani. (2026a, July 11). Over 86,000 students get A-1 and A grades in Karachi matric exams. <https://propakistani.pk/2026/07/11/over-86000-students-get-a-1-and-a-grades-in-karachi-matric-exams/>
37. ProPakistani. (2026b, January 3). Sindh introduces uniform examination system across all education boards. <https://propakistani.pk/2026/01/03/sindh-introduces-uniform-examination-system-across-all-education-boards/>
38. ProPakistani. (2026c, February 12). Sindh Matric and Inter 2026 exam dates announced. <https://propakistani.pk/2026/02/12/sindh-matric-and-inter-2026-exam-dates-announced/>
39. ProPakistani. (2026d, July 10). Karachi Board announces Matric Science Group Result 2026. <https://propakistani.pk/2026/07/10/karachi-board-announces-matric-science-group-result-2026/>
40. The Nation. (2026a, July 11). Karachi matric board announces SSC Part-II Science Group results. <https://www.nation.com.pk/11-Jul-2026/karachi-matric-board-announces-ssc-part-ii-science-group-results>
41. The Nation. (2026b, January 21). Sindh govt approves 'grading system' in place of 'marks' in educational boards. <https://www.nation.com.pk/21-Jan-2026/sindh-govt-approves-grading-system-place-marks-educational-boards>
42. The Nation. (2026c, July 9). Sindh govt decides to provide special IT training for all educational board staff. <https://www.nation.com.pk/09-Jul-2026/sindh-govt-decides-provide-special-training-educational-board-staff>
43. The Pakistan Connect. (2026, January 20). New grading system replaces traditional marks in Sindh schools. <https://thepakistanconnect.com/sindh-grading-marks-schools/>



2026 by the authors; Journal of Global Social Transformation (JGST). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).