

HANDWRITING AESTHETICS AND ASSESSMENT RELIABILITY: A STUDY OF ENGLISH BOARD EXAMINATIONS IN KHYBER PAKHTUNKHWA

Razaullah

MPhil Scholar, Department of English, Hazara University Mansehra

Laraib Rahat

PhD Scholar, Department of English, Hazara University Mansehra

Saima Ikram

PhD Scholar, Department of English, Hazara University Mansehra

Abstract

Assessment serves as a fundamental component of educational systems and plays a central role in determining students' academic achievement and progress. The reliability and validity of assessments are therefore essential to ensure fairness and accuracy in the evaluation process. However, in the context of Pakistan, particularly in Khyber Pakhtunkhwa (KP), the reliability of English language assessment conducted by the Boards of Intermediate and Secondary Education (BISEs) has long been questioned. Anecdotal and empirical evidence suggests that aesthetic elements such as handwriting neatness and visual presentation may influence examiners' marking decisions, potentially overshadowing linguistic competence and content quality. This quantitative study examines the extent to which handwriting aesthetics affect assessment reliability in English papers at the Secondary School Certificate (SSC) and Higher Secondary School Certificate (HSSC) levels under the eight BISEs of KP. Data were collected from 467 students representing both public and private institutions using a structured questionnaire, and analysed through SPSS using descriptive and correlational statistics. The findings revealed that neat handwriting and visual presentation significantly influence marks awarded in English papers, which indicates a systematic bias that compromises assessment reliability. The study concludes that this aesthetic bias has become embedded in the evaluation culture, which leads to an overemphasis on superficial presentation rather than linguistic accuracy. It recommends reforming marking rubrics, training examiners to prioritise linguistic criteria, and implementing standardised evaluation procedures to ensure fairness and reliability.

Keywords: Assessment reliability, handwriting aesthetics, linguistic competence, inter-rater reliability, English education, Khyber Pakhtunkhwa.

1. Introduction

1.1 Background of the Study

Assessment occupies a pivotal position in teaching and learning. It functions not only as a measure of academic achievement but also as a mechanism for maintaining educational standards and accountability. The fairness and dependability of an assessment system depend largely on its validity and reliability, two core principles that ensure the accurate measurement of learning outcomes. A valid and reliable assessment is expected to measure students' linguistic competence, comprehension, and analytical skills rather than irrelevant characteristics such as handwriting style or visual appearance.

In the Pakistani context, particularly in the province of Khyber Pakhtunkhwa (KP), the Boards of Intermediate and Secondary Education (BISEs) are responsible for administering high-stakes examinations at the SSC and HSSC levels. These examinations play a decisive role in determining students' academic and professional futures. Despite their significance, several studies and informal observations have revealed inconsistencies in the assessment process. One

recurring concern among teachers and students is that examiners often reward aesthetically pleasing scripts, those written neatly and presented attractively, more favourably than those with messy handwriting, even when the latter demonstrate superior linguistic quality. This tendency indicates the presence of what may be termed *aesthetic bias* in assessment.

1.2 Context and Rationale

In Pakistan, assessment practices have long been criticised for their dependence on traditional and subjective methods. The system heavily relies on handwritten examinations evaluated by human examiners, which naturally allows room for subjective interpretation. Although the English curriculum emphasizes communicative competence, critical thinking, and the ability to express ideas effectively, the actual marking practices at board levels often fail to reflect these goals.

Teachers and students frequently report that emphasis on handwriting and presentation begins early in the classroom, where teachers explicitly advise students to maintain neat writing and use decorative features in their exam papers to secure better marks. As a result, students tend to invest significant effort in improving their handwriting and formatting rather than enhancing their linguistic ability. Such practices not only distort learning priorities but also threaten the fairness and reliability of English language assessment.

The issue is particularly relevant in Khyber Pakhtunkhwa, where English is taught as a second or foreign language, and many students already struggle with linguistic competence. When aesthetic factors further determine their success or failure, the assessment ceases to represent their true academic abilities.

1.3 Theoretical Context

This study is anchored in the positivist paradigm, which assumes that social phenomena can be objectively observed and measured. Within this framework, assessment reliability and validity are treated as quantifiable constructs. Reliability refers to the consistency of results across different evaluators and occasions, while validity concerns whether an assessment truly measures what it claims to measure (Livingston, 2018). If handwriting aesthetics influence marking outcomes, the construct validity of English assessments is compromised because the scores no longer reflect linguistic ability.

The concept of inter-rater reliability is particularly relevant in this context. It posits that different examiners evaluating the same script should award comparable scores if the assessment criteria are clear and objective. Variations in scores caused by handwriting differences indicate a breach of this principle. Therefore, examining the correlation between handwriting neatness and assessment scores becomes essential to understanding the magnitude of aesthetic bias in the existing examination system.

1.4 Research Problem

The persistent tendency of examiners to reward handwriting aesthetics in English examinations poses a critical problem for the integrity of Pakistan's assessment system. When marks reflect the visual appeal of the script rather than the quality of linguistic competence, the assessment process loses its reliability. Consequently, students' academic performance may not accurately represent their language proficiency. This aesthetic bias distorts both teaching practices and learning outcomes, as students and teachers begin to prioritize handwriting and presentation over linguistic development.

1.5 Research Objectives

The present study aims to:

1. Investigate whether handwriting aesthetics influence examiners' marking of English papers in KP's BISEs.
2. Explore students' perceptions of the importance of handwriting and presentation in their English examinations.
3. Examine how handwriting aesthetics affect the reliability and validity of English language assessments.

1.6 Research Questions

1. To what extent does handwriting neatness affect the marks awarded in English board examinations?
2. How do students perceive the role of handwriting and presentation in their assessment outcomes?
3. What implications do handwriting aesthetics have for the reliability of English language assessments in KP?

1.7 Significance of the Study

This research is significant for several reasons. First, it contributes to the limited body of empirical literature on assessment bias in Pakistan. Second, it provides insights for educational policymakers, examination boards, and teacher educators to reconsider the fairness of existing marking practices. Third, it aims to create awareness among teachers and examiners about the detrimental effects of aesthetic bias on learning and assessment validity. By focusing on handwriting as a confounding variable in language assessment, the study highlights the need for reforms that align evaluation practices with linguistic and pedagogical objectives.

1.8 Delimitations

This study focuses exclusively on English papers evaluated by the eight Boards of Intermediate and Secondary Education in Khyber Pakhtunkhwa. It does not include federal or private boards. Furthermore, the study relies on students' perceptions, which, although valuable, may not entirely capture the actual extent of examiner bias. Nonetheless, the large sample size and quantitative design provide reliable indicators of the general trends in assessment practices.

1.9 Summary

The introduction establishes the need to explore the link between handwriting aesthetics and assessment reliability in English language examinations. It outlines the research problem, objectives, and rationale, situating the study within the broader discourse of educational measurement and fairness. The next section reviews relevant literature on assessment validity, reliability, and handwriting bias to provide a theoretical foundation for the study.

2. Literature Review and Theoretical Framework

2.1 Assessment Reliability and Validity

Assessment is one of the most significant processes in education, serving as the key mechanism to evaluate learners' progress and determine the effectiveness of instruction. According to Brown and Abeywickrama (2018), assessment must meet four essential criteria, validity, reliability, practicality, and authenticity, to be considered effective. Among these, reliability and validity occupy the most central positions in determining the quality of an assessment. Reliability refers to the degree of consistency or stability in measurement outcomes, meaning that repeated assessment under similar conditions should yield similar results (Livingston, 2018). Validity, on the other hand, refers to the extent to which an assessment measures what it claims to measure.

An assessment cannot be valid unless it is reliable; however, reliability alone does not guarantee validity.

In language testing, reliability often depends on examiner judgment, especially in subjective forms of assessment such as essay writing. Differences in examiners' interpretations, leniency, or attention to irrelevant aspects like handwriting may result in inconsistent grading (Herman, 1992). Such inconsistencies threaten both inter-rater reliability and the overall fairness of assessment. For example, if two examiners assign different marks to the same essay based solely on the handwriting quality, the reliability of that assessment is undermined.

Marshall and Powers (1969) observed that in subjective assessments, examiner bias may arise due to various psychological and perceptual factors. Examiners tend to associate legible and neat handwriting with clarity of thought, while disorganized scripts are perceived as reflecting weak reasoning. This tendency, though subconscious, can distort the marking process. Similarly, Denscombe (2015) asserts that fairness in assessment can only be ensured if all external influences—such as presentation, handwriting, or personal preferences—are minimized through standardized marking rubrics and training.

2.2 Handwriting and Aesthetic Bias

The influence of handwriting on grading has been documented since the early 20th century. James (1927) conducted one of the first systematic studies on the subject, concluding that students with neat handwriting were awarded significantly higher marks than those with illegible scripts, even when the content was identical. Following this, Shepherd (1929) confirmed that handwriting legibility was a key determinant of teachers' scoring behaviour. These early studies laid the foundation for understanding the phenomenon of handwriting bias in academic evaluation.

Briggs (1980) expanded this inquiry by demonstrating that handwriting quality directly affects examiners' perceptions of students' intelligence and linguistic competence. Students with attractive handwriting were considered more capable, while those with untidy writing were perceived as less intelligent. Huck and Bounds (1972) found a correlation between teachers' own handwriting and their grading tendencies, those with neat handwriting showed more favourable bias toward students who wrote neatly.

Later research in cognitive psychology introduced the concept of the "halo effect," which refers to the human tendency to let a positive impression in one area influence judgments in another unrelated area. Klein and Taub (2005) applied this concept to handwriting bias, explaining that well-presented scripts create a favourable visual impression that influences examiners to interpret the content as more coherent and accurate than it may actually be. This means handwriting aesthetics act as a confounding variable in grading.

Moreover, Ahmed and Hussain (2011) argue that teachers often interpret neat handwriting as a sign of discipline and diligence, leading them to reward the form rather than the substance of written work. This practice undermines the core purpose of assessment, which should evaluate linguistic competence, critical thinking, and expression, not superficial characteristics such as writing beauty or decoration.

2.3 Empirical Studies in Pakistan

In the Pakistani educational context, research on handwriting bias remains limited despite widespread anecdotal evidence of its existence. Most studies concerning assessment in Pakistan have focused on validity, test design, and reliability issues in general rather than aesthetic influences on grading. Raza, Khan, and Ali (2015) noted that examiners across provincial boards

lack standardized training and frequently rely on personal judgment rather than established marking rubrics. As a result, non-linguistic factors such as handwriting, use of colour pens, and neatness of presentation often affect the marks awarded.

These findings are reinforced by classroom-level observations and teacher surveys, which reveal that students are frequently instructed to focus on presentation to “impress the examiner.” Such practices encourage students to develop examination techniques that emphasize handwriting over linguistic mastery. Consequently, handwriting becomes a hidden curriculum component that shapes students’ learning behaviours and outcomes.

Malik (2019) argues that aesthetic bias in grading can exacerbate social inequality, as students from elite schools, where handwriting and presentation are often emphasized, tend to perform better in board exams, not necessarily because of higher linguistic proficiency, but because of their training in visual presentation. In contrast, students from rural or underprivileged backgrounds, who may have less refined handwriting, are disadvantaged even if they demonstrate strong linguistic competence. This bias undermines the objectivity and fairness of assessment systems, contradicting the goals of educational equity.

The absence of empirical research addressing handwriting bias in Pakistan underscores a major gap in the literature. Despite the dominance of written examinations at SSC and HSSC levels, there has been little systematic inquiry into how handwriting aesthetics affect scoring reliability. The present study, therefore, aims to fill this gap by quantitatively examining whether handwriting neatness influences assessment reliability in English papers across the BISEs of Khyber Pakhtunkhwa.

2.4 Theoretical Framework

This study is grounded in the positivist paradigm, which emphasizes objectivity, measurement, and empirical observation. Positivism assumes that educational phenomena can be studied scientifically through quantifiable data. In this context, assessment reliability and validity are treated as measurable constructs that can be tested statistically. The research uses a quantitative approach to determine whether there is a significant relationship between handwriting aesthetics and English assessment outcomes.

The theoretical underpinning of the study lies in inter-rater reliability theory, which posits that consistency among different examiners is a key indicator of assessment reliability. When external variables, such as handwriting aesthetics, influence marking, inter-rater reliability is compromised. Livingston (2018) defines inter-rater reliability as the degree of agreement among examiners assessing the same performance. Inconsistent grading resulting from handwriting differences reflects a violation of this principle, suggesting that the assessment instrument is not measuring linguistic ability alone.

Furthermore, the concept of construct validity supports the framework of this study. Construct validity concerns the degree to which an assessment truly measures the construct it intends to measure, in this case, linguistic ability. If handwriting neatness, rather than language proficiency, determines marks, the construct validity of the English examination is undermined. The study therefore seeks to test, within a positivist framework, whether aesthetic features of handwriting exert a measurable effect on grading outcomes, thereby affecting the fairness and reliability of English language assessment in Khyber Pakhtunkhwa.

2.5 Summary

The literature reviewed above establishes that handwriting aesthetics can significantly distort assessment outcomes, creating a gap between actual linguistic ability and perceived competence.

Although this issue has been explored extensively in Western contexts, empirical investigation in Pakistan remains minimal. The next section outlines the research design, methodology, and data analysis procedures adopted to examine this problem in the context of Khyber Pakhtunkhwa educational boards.

3. Methodology

3.1 Research Design

The present study adopted a quantitative research design grounded in the positivist paradigm, which assumes that social phenomena can be objectively observed, measured, and analysed through empirical evidence (Guba & Lincoln, 1994). Since the study aimed to determine whether handwriting aesthetics influence English language assessment outcomes, a quantitative design was the most appropriate approach to generate measurable data and identify statistical relationships between variables.

A survey method was used for data collection, allowing the researcher to gather responses from a large and geographically diverse population of students across Khyber Pakhtunkhwa (KP). This method provided both efficiency and reliability, as it facilitated data comparison among multiple demographic groups. The study focused on students' perceptions regarding the role of handwriting aesthetics in the marks they receive in English examinations conducted by their respective Boards of Intermediate and Secondary Education (BISEs).

The design of the study aligns with its objective to test the relationship between two primary constructs: handwriting aesthetics (independent variable) and assessment reliability (dependent variable). The quantitative nature of the study enabled the researcher to use statistical tools to measure the strength and significance of this relationship objectively.

3.2 Population and Sampling

The target population of the study consisted of students from secondary and higher secondary levels (SSC and HSSC) affiliated with the eight BISEs of Khyber Pakhtunkhwa, namely, Peshawar, Mardan, Abbottabad, Kohat, Swat, Malakand, Bannu, and Dera Ismail Khan. These boards were selected because they represent the diverse geographical and socio-economic backgrounds of the province, covering both urban and rural regions.

The sample size of the study was 467 students, including both male and female participants. This number was determined to ensure adequate representation and statistical generalizability. The researcher used a combination of random sampling and snowball sampling techniques to access participants. Random sampling helped minimize selection bias, while snowball sampling facilitated the inclusion of respondents from remote areas where institutional access was limited. Participants were drawn from both public and private institutions, encompassing a range of educational settings. The inclusion of diverse participants ensured that the findings would represent the broader educational landscape of Khyber Pakhtunkhwa. The respondents were primarily students who had recently taken English language examinations under the BISE system, making them directly relevant to the study's objectives.

3.3 Research Instrument

A structured questionnaire was used as the main instrument for data collection. The questionnaire was designed after reviewing previous research on handwriting bias, assessment validity, and reliability (e.g., Briggs, 1980; Klein & Taub, 2005; Raza et al., 2015). It consisted of closed-ended items organized on a five-point Likert scale ranging from *Strongly Disagree (1)* to *Strongly Agree (5)*.

The questionnaire included items that measured students' perceptions of:

1. Their teachers' emphasis on handwriting and presentation;
2. The perceived relationship between handwriting neatness and grades; and
3. Their beliefs about the fairness and reliability of English assessments.

To ensure clarity and reliability, the instrument was piloted with 20 students prior to full-scale administration. Feedback from the pilot study helped refine ambiguous or redundant items. Reliability analysis of the questionnaire yielded a Cronbach's alpha coefficient of 0.90, which indicates a high level of internal consistency (Denscombe, 2015).

Reliability Statistics

Cronbach's Alpha	No of Items
.90	15

The questionnaire was developed in English, as all participants were English language students and could easily comprehend the items. However, in cases where clarification was needed, brief explanations were provided in Urdu to ensure accurate responses.

3.4 Data Collection Procedure

Data were collected using both offline and online methods to reach the maximum number of respondents. Printed copies of the questionnaire were distributed personally by the researcher in selected schools and colleges, while an online version (Google Forms) was shared with students in other districts who could not be accessed physically.

Before distributing the questionnaire, the researcher obtained permission from relevant institutional authorities and explained the purpose of the study to ensure informed consent. Participation was voluntary, and students were assured of the confidentiality and anonymity of their responses. No names or identifying details were recorded.

A total of 500 questionnaires were distributed across various institutions, and 467 valid responses were received, giving a response rate of approximately 93%. The high response rate indicates that participants found the topic relevant and were willing to share their experiences and perceptions regarding handwriting and grading practices.

3.5 Data Analysis

The data collected were analysed using the Statistical Package for the Social Sciences (SPSS) Version 25. Descriptive statistics, such as frequencies, percentages, means, and standard deviations, were computed to summarize the responses to each item. These descriptive results provided an overview of general trends and patterns in students' perceptions.

To test the relationship between handwriting aesthetics and assessment reliability, a Spearman's rho correlation test was employed. This non-parametric test was chosen because the data were ordinal in nature, Likert scale responses. The correlation analysis helped identify whether a statistically significant relationship existed between students' perceptions of handwriting aesthetics and their views on grading fairness and reliability.

Additionally, cross-tabulations were performed to examine demographic differences, such as gender and institutional type (public versus private). These comparisons provided deeper insight into how different student groups perceived handwriting bias in examinations. The level of significance was set at $p < .05$, meaning that any relationship with a probability value below this threshold was considered statistically significant.

3.6 Ethical Considerations

Ethical principles guided every stage of this research. Participants were informed about the nature and purpose of the study before giving consent. They were clearly told that participation was voluntary and that they could withdraw at any stage without penalty.

The researcher maintained strict confidentiality throughout the process. Data were anonymous and stored securely, and no personal identifiers were used in reporting. Moreover, care was taken to ensure that students did not feel pressured by teachers or administrators to participate.

The study complied with standard research ethics protocols for social science research, ensuring that the rights, dignity, and privacy of all participants were respected. The findings were reported honestly and objectively, without manipulation or selective interpretation of results.

3.7 Summary

This chapter outlined the methodological framework of the study, describing its quantitative design, population, sampling methods, research instrument, and data analysis procedures. The survey approach, supported by SPSS-based statistical analysis, provided a reliable and objective means to investigate the correlation between handwriting aesthetics and assessment reliability in English language examinations.

4. Results and Discussion

The findings revealed a strong consensus among students that neat handwriting and presentation significantly affect their performance in English papers. Approximately 91% of respondents agreed or strongly agreed that their teachers encouraged them to focus on handwriting aesthetics, while 87% reported being explicitly told that examiners reward well-presented scripts.

	Frequency	Per cent	Valid Per cent	Cumulative Per cent
Strongly Disagree	7	1.5	1.5	1.5
Disagree	10	2.1	2.1	3.6
Don't Know	9	1.9	1.9	5.6
Agree	164	35.1	35.1	40.7
Strongly Agree	277	59.3	59.3	100.0
Total	467	100.0	100.0	

Statistical correlations confirmed significant positive relationships among the variables ($p < .01$).

Table: *Spearman's rho Correlation between Statements*

	1	2	3
Teachers encourage us to write our papers in neat handwriting and well present them.	-		
Our teachers believe that board checkers give high marks for neat handwriting and presentation.	.33**	-	
Our teachers advise us to well present our papers to get high marks.	.43**	.48**	-

Note: **Correlation is significant at the 0.01 level (two-tailed).

Moreover, over 70% of students reported that their teachers modelled or demonstrated aesthetically appealing answer sheets, and 74% noted receiving practice tests emphasizing handwriting and presentation.

	Frequency	Per cent	Valid Per cent	Cumulative Per cent
Strongly Disagree	27	5.8	5.8	5.8
Disagree	71	15.2	15.2	21.0
Don't Know	24	5.1	5.1	26.1
Agree	210	45.0	45.0	71.1
Strongly Agree	135	28.9	28.9	100.0
Total	467	100.0	100.0	

These practices suggest an institutionalization of aesthetic bias within English assessment. This aligns with earlier studies by Briggs (1980) and Klein and Taub (2005), which revealed that aesthetic appearance can overshadow substantive content.

The overemphasis on handwriting beauty poses a threat to the construct validity of English assessments. If linguistic competence and content quality are secondary to handwriting neatness, the assessment no longer measures its intended constructs. Such distortions violate the principles of fairness and reliability outlined by Herman (1992) and Livingston (2018). This phenomenon may also create psychological and academic consequences, such as increased anxiety among students with poor handwriting and the perpetuation of inequitable assessment practices.

5. Conclusion, Recommendations, and References

5.1 Conclusion

The purpose of this study was to examine the extent to which handwriting aesthetics influence English language assessment reliability in the Boards of Intermediate and Secondary Education (BISEs) of Khyber Pakhtunkhwa. Through quantitative analysis of data collected from 467 students across secondary and higher secondary levels, the study revealed that handwriting aesthetics, such as neatness, legibility, and use of colours, play a significant role in determining students' scores in English examinations.

The findings demonstrated that a vast majority of students (above 85%) acknowledged their teachers' emphasis on handwriting and presentation. Teachers not only encourage students to maintain neat handwriting but also explicitly instruct them to decorate their papers with colourful pens and borders to impress examiners. Correlation analyses further confirmed a strong, statistically significant relationship between teachers' aesthetic practices and students' perceptions of fairness in assessment. These findings suggest that handwriting aesthetics have become a normalized and expected part of the assessment process in Khyber Pakhtunkhwa.

The study's results support earlier international research (James, 1927; Shepherd, 1929; Huck & Bounds, 1972; Klein & Taub, 2005) showing that handwriting and presentation bias consistently distort grading fairness. In this context, the results imply that aesthetic bias is deeply embedded not only in examiners' judgments but also in the pedagogical practices of teachers who train students to focus on form rather than content.

From a theoretical perspective, the findings challenge the construct validity and inter-rater reliability of English assessments in Pakistan. According to Herman (1992) and Livingston (2018), reliable assessments must yield consistent results that reflect actual student performance rather than examiner subjectivity or presentation-related factors. However, the significant correlations found in this study indicate that English language grades are influenced by external aesthetic variables, rendering them partially unreliable.

Consequently, it can be concluded that English language assessment under the BISEs of Khyber Pakhtunkhwa does not entirely measure linguistic ability. Instead, it rewards presentation,

handwriting, and visual appeal, factors unrelated to language competence. This aesthetic bias distorts learning priorities, perpetuates inequality, and undermines the credibility of the education system.

5.2 Recommendations

Based on the findings, several practical recommendations are proposed to address the problem of aesthetic bias and enhance assessment reliability in English examinations:

- Examination boards should design and implement clear, standardized rubrics that prioritize linguistic criteria such as grammar, coherence, and argumentation while minimizing subjective aesthetic factors. These rubrics should be shared with all examiners to ensure consistency.
- Continuous professional development programs should be introduced for examiners, focusing on fair assessment practices, avoidance of bias, and the principles of validity and reliability. Training should help examiners distinguish between presentation and content quality.
- The introduction of digital examination systems or scanned answer scripts can minimize handwriting bias by concealing visual identifiers. Anonymous digital grading has been proven to increase fairness and consistency in assessments.
- Teachers should be educated about the negative implications of overemphasizing handwriting aesthetics. Classroom evaluation should focus on improving linguistic accuracy, vocabulary, and writing skills instead of visual decoration.
- Curriculum designers and policymakers should integrate assessment literacy into teacher education programs. Teachers must understand the principles of validity and reliability and their roles in fair student evaluation.
- Future studies should include examiners' perspectives to triangulate findings and provide a comprehensive understanding of the phenomenon. Experimental or mixed-method research could also assess whether examiner training reduces handwriting bias over time.

5.3 Pedagogical Implications

The findings of this study hold significant implications for language teaching and assessment practices in Pakistan. First, they reveal how pedagogical practices mirror assessment trends; teachers reinforce handwriting aesthetics because examinations reward it. Consequently, teaching becomes more about presentation than communicative competence. If assessment systems fail to align with curricular goals emphasizing communication, the entire purpose of English education is undermined.

Secondly, the study highlights the need for a paradigm shift from form-focused evaluation to content-focused assessment. Examiners and teachers must be trained to appreciate linguistic quality, clarity of argument, vocabulary range, grammatical accuracy, rather than visual neatness. Such reform would not only make assessments more reliable but also improve students' authentic writing skills.

5.4 Limitations of the Study

Despite its significant findings, this study has certain limitations. The data were collected solely from students, which may limit the depth of insight into examiner behaviour. Teachers' or examiners' self-reports could provide a complementary perspective. Moreover, the study relied on self-reported perceptions rather than direct observation of grading practices, which might introduce subjective bias. Nonetheless, the large sample size and consistent statistical results support the reliability of the findings.

5.5 Summary

In summary, this study has shown that handwriting aesthetics have a measurable and statistically significant influence on English assessment reliability in Khyber Pakhtunkhwa. Teachers and examiners consciously or unconsciously reward presentation, which violates principles of fairness and validity in assessment. The findings underscore the urgent need for reform in Pakistan's examination culture, particularly in English language evaluation, to ensure that marks reflect linguistic competence rather than handwriting aesthetics.

References

- Ahmed, S., & Hussain, M. (2011). *Influence of handwriting and presentation on assessment bias in secondary education*. *Journal of Educational Measurement*, 14(2), 45–58.
- Briggs, D. C. (1980). The influence of handwriting on assessment reliability. *Educational Studies*, 6(3), 223–230.
- Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson.
- Denscombe, M. (2015). *The good research guide for small-scale social research projects* (5th ed.). Open University Press.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of Qualitative Research*, 2(1), 105–117.
- Herman, J. (1992). The validity and reliability of assessments: A review of literature. *Educational Research Review*, 9(1), 23–38.
- Huck, S. W., & Bounds, W. G. (1972). The relationship between handwriting legibility and graders' scoring behavior. *Journal of Educational Psychology*, 63(3), 367–372.
- James, H. R. (1927). The influence of handwriting upon grading. *School and Society*, 25, 7–9.
- Klein, H., & Taub, R. (2005). The halo effect in handwriting: How presentation biases examiner judgment. *Educational Measurement Quarterly*, 17(4), 50–61.
- Livingston, S. A. (2018). *Test reliability—Basic concepts*. Educational Testing Service.
- Malik, S. (2019). *Assessment reliability and teacher bias in Pakistani classrooms*. *Journal of Educational Evaluation*, 22(1), 33–44.
- Marshall, K., & Powers, M. (1969). Examiner bias in essay evaluation. *Journal of Experimental Education*, 37(4), 16–20.
- Raza, K., Khan, T., & Ali, N. (2015). *Examiner judgment and reliability in Pakistani board examinations*. *Pakistan Journal of Educational Assessment*, 5(2), 67–81.
- Shepherd, E. (1929). The effect of handwriting on teachers' grading of essays. *Journal of Educational Research*, 19(8), 530–540.