



Stakeholder perspectives of test administration in a Bangladeshi university: A qualitative study

ABSTRACT - Learner achievement benchmarks are typically operationalized through tests and assessments; however, students at newly established public universities in Bangladesh encounter a range of obstacles that hinder their academic progress. Despite receiving equivalent instruction under comparable teaching and learning conditions, learner performance varies markedly. Students, teachers, and administrators constitute the three primary stakeholder groups in any testing procedure, and their respective social and personal backgrounds, institutional contexts, and situational realities fundamentally shape assessment practices. This study investigates factors affecting test administration through semi-structured interviews with 35 participants: 26 students, 7 faculty members from the Department of English at a public university pseudonymously referred to as Padma Science and Technology University (PSTU), and two administrators—one from the department and one from the institution at large. Employing a qualitative design, the study examines both undergraduate and postgraduate programs through thematic analysis of interview data gathered from the three stakeholder groups. The results reveal substantial complexities permeating the testing process from the perspective of test takers, while also indicating that teachers and administrators encounter obstacles and appear to be operating under considerable strain. The findings provide analytic insights for anticipating testing and administrative challenges in analogous contexts, as well as for identifying areas requiring targeted intervention. The findings suggest that the prevailing situation demands heightened attention to the testing environment, reform of assessment policies, reinforcement of stakeholder roles and responsibilities, and systematic attention to other contributing factors shaping test administration as a holistic process. The research underscores the importance of context-sensitive inquiry, recognizing that each educational setting presents unique pedagogical demands. In so doing, the study emphasizes the critical role of context-specific formulations in fostering more effective and equitable test administration within Bangladeshi higher education.

Snigdha Sarmili Ria^{1*}
Sheikh Md. Muniruzzaman²

¹Department of English, University of Global Village, Barishal-8200, Bangladesh

²Department of English, Gopalganj Science and Technology University, Gopalganj-8105, Bangladesh

*Corresponding email;

snigdha.18eng067@gstu.edu.bd

Article History

Submitted: 25 December 2025

Accepted: 25 March 2026

Published: 31 May 2026

Keywords

Test administration; Assessment; Backwash; Stakeholder; Bangladesh

Recommended APA Citation

Ria, S. S., & Muniruzzaman, S. M. (2026). Stakeholder perspectives of test administration in a Bangladeshi university: A qualitative study. *Englisia: Journal of Language, Education, and Humanities*, 13(2), 157-173.

<https://doi.org/10.22373/englisia.134>

1. Introduction

In academia, tests are the most commonly used form of assessment. According to Hughes (2020), a test is a tool used to measure learners' development and language proficiency. Although students may study under the same learning conditions, their performance and achievements vary in real-world settings. Tests help to reveal these differences, and Harding and Fulcher (2021) defined them as a strategy of extracting language performance that is aimed at enabling inferences about expertise in a target domain. The American Psychological Association (APA, 2014) stated that tests should be designed to be valid, reliable, and unbiased. A well-managed test administration helps to ensure both the validity and reliability of test results (Walkin, 1990). Test administration refers to the process of organizing, conducting, and supervising tests to ensure validity, fairness, and reliability. Hughes (2020) emphasized the importance of a well-managed testing environment. Hughes warned that even the best-designed test could yield unreliable results if administered improperly. They also suggested that individual differences and test-taking contexts must be considered.

Stakeholders in testing include all individuals who are directly or indirectly involved in the whole procedure. The key stakeholders are primarily teachers, students, and institutional administrators. According to Ajzen's (1991) theory of planned behavior, an individual's perception of a situation can shape their attitude and behavior toward intended affairs. In assessment, stakeholders' experiences and educational environments shape their perceptions. To build sustainable and effective academic strategies, understanding stakeholders' perceptions is crucial for institutional goals (Ott & Todea, 2020).

A perspective is an interpretive lens through which an individual makes sense of educational motifs and practices, influenced by their personal experiences, beliefs, and social contexts. Constructivist theory proclaims that learners actively construct knowledge rather than passively absorb it. In education, this emphasizes that individuals conclude concepts and practices from their own perspectives. It shows socially and historically established ways of understanding experiences, which are consistent with constructivist theory (Barkhuizen et al., 2013).

Properly managed tests help invigilators and test takers to conduct. Alderson and Wall (1993) argued that the way a test is administered can shape how teachers teach and students learn. Effective test administration can help ensure better test performance. Different tests bring different types of challenges. Some tests may help learning, while others may create pressure. Magambo and Rukundo (2016) stated that when a test is properly administered, it reduces errors. When everything is done well, the test results become more reliable and have a positive effect on learning, which is considered "positive backwash." All the problems affect how we feel during tests. Krashen (1982) talked about "affective filters." He notes that if students feel uncomfortable, they cannot perform well. Thus, a well-maintained test administration can reinforce candidates' performance.

Although washback and test administration have been widely studied in Western and high-stakes contexts, limited research has examined stakeholder perspectives in newly established public universities in Bangladesh, particularly from a multi-stakeholder qualitative

lens. Alderson and Wall (1993) claim that tests do not cause transformation directly; basically, effects depend on factors such as teachers' beliefs, stakes, and resources. Messick (1996) shows washback within consequential validity, emphasizing that the effect of tests must be evidence-based and connected to test use. Briefly, tests may influence education in diverse ways, and those consequences are central to validity. Kumaravadivelu (2001) introduced the concept of "particularity," which holds that every context presents its own challenges and possibilities. What works in one setting may not work in another. In our department, students face a variety of specific problems that affect their test performance. This qualitative study examines the effects of various assessment types in undergraduate and postgraduate programs within our department and aims to identify practical solutions to improve test administration. The goal of the study is to explore how learners, teachers, and administrators view the process and effectiveness of test administration in the English Department at PSTU. Bronfenbrenner's (1979) ecological theory emphasizes that an individual is not merely a recipient of environmental influences but actively shapes self-development through the system. To do this, the research will examine key test-administration factors that may help students perform better.

2. Literature review

2.1. Importance of tests

To assess a candidate's competence in a specific area, testing plays a pivotal role. Fremer and Wall (2003) highlighted assessments, as they serve multiple purposes, such as identifying student progress, evaluating teaching strategies, and tracing learning gaps. By clarifying how the teaching procedure is effective and where improvements are necessary, a well-designed assessment can assist teachers in understanding these. Marzano (2006) argued that assessments are tools for measurement and powerful instruments that drive educational attainment. Tests are vital in education, as they help assess student progress and inform better instructional decisions. Brown (2010) supports that tests are important for understanding both the theory and practice in education. It helps in designing the curriculum and making changes in teaching methods, which offers feedback to both teachers and students

Structured tests are needed to place students at an appropriate academic level (McNamara, 2020). Regarding cultural diversity, the importance of tests varies. Here, Stobart (2008) points out that high-stakes tests are used frequently for system goals, such as accountability and rankings, rather than for exact educational benefits. In this condition, teachers may focus solely on test scores and ignore creativity and critical thinking. Effective tests should encourage learners to think deeply and creatively (Brookhart, 2010).

2.2. Test administration

Cheng (2005) agrees and argues that the cultural, linguistic, and educational contexts must reflect the testing methods in which they are implemented. Educational contexts are unique in their needs and outcomes. According to Cohen et al. (2011), understanding the particular context helps disclose the real situation. Linn and Miller (2005) explain that offering each student an equal opportunity to show what they have acquired is the most important rule in test

administration. Some scopes of test administration can be considered paramount according to several researchers, and those are as follows:

a. Spatial Environment

The testing environment affects test takers' comfort and attention, thereby shaping the validity and fairness of performance. A supportive environment must be designed to improve outcomes; however, it hinders candidates' performance (Hembree, 2018).

b. Procedural Fairness

Equal treatment for all candidates must be ensured through transparency and consistency. It promotes fair perception, trust, and acceptability of testing (Hamid et al., 2019).

c. Ethical Invigilation

Adequate proctoring prevents misconduct, bias, or binary treatment. Hosseinnia and Kafi (2024) claim that a foundation structure for conducting transparent, fair, and accountable tests reduces misuse in educational assessments, as ensured by a code of ethics.

d. Institutional Culture

Organizational protocols, norms, and values frame assessment practices. Treviño et al. (1998) examined organizational ethical culture, in which shared norms, values, and systems influence ethical behavior and consistent administration.

A calm and well-managed space helps students focus. Clariana and Wallace (2002) noted that stress and emotional barriers are often caused by poor test conditions. Shepard (2000) noted that tests must match the learning environment. A supportive setting helps test takers do better and feel confident. For instance, a student might underperform in a noisy environment not because of a lack of learning but because of stress from the surrounding environment. Therefore, every aspect of the testing environment can affect the test taker's performance. A calm and well-managed space helps students focus. Clariana and Wallace (2002) noted that stress and emotional barriers are often caused by poor test conditions. Shepard (2000) noted that tests must match the learning environment. A supportive setting helps test takers do better and feel confident. For instance, a student might underperform in a noisy environment not because of a lack of learning but because of stress from the surrounding environment. Therefore, every aspect of the testing environment can affect the test taker's performance.

2.3. Backwash

Backwash refers to the effect that testing exerts on teaching and learning, which can be positive or negative. Cheng et al. (2004) argue that washback is shaped by situational factors such as test type, educational level, and institutional culture, suggesting that better test performance (RQ1) also depends on administrative provisions and implementation practices. Alderson and Wall (1993) show that tests can influence both the topic and the teaching approach; however, this influence is mediated by the particular teachers' beliefs and procedural constraints. This idea informs RQ2, as various kinds of assessments in the bachelor's and master's programs may yield varied effects on teaching strategies, motivation, and learning focus.

Messick (1996) emphasizes that washback must be measured ethically, ensuring that appraisals are valid, reliable, and free of negative consequences. Similarly, Green (2007) emphasizes how test format shapes learner behavior. These interpretations underpin RQ3, which seeks empirical measures to elevate test administration and stimulate constructive outcomes. Moreover, Andrews (2004) elaborates on the discussion through the lens of “test impact,” and Fan, Frost, and Liu (2020) emphasize the importance of incorporating teachers’ and candidates’ perspectives and promote the need to evaluate administrative protocols holistically.

Tests should be administered properly so that the results accurately reflect learners’ abilities. This study aims to examine how test administration affects learners’ test performance in the English Department at PSTU and to identify ways to improve test administration and create positive washback, which means ensuring that tests support effective teaching and learning. The following research questions guided this study:

1. What factors in test administration contribute to improved test performance in the English Department at PSTU?
2. What are the effects of the various assessments administered in the bachelor's and master's programs in the English Department at PSTU?
3. What measures can be taken to enhance test administration practices and positive backwash in the English Department at PSTU?

3. Method

3.1. Research design

This study followed a qualitative research approach using thematic analysis. According to Creswell (2009), qualitative research is used to understand a social or human problem by studying how people experience or feel about it. For this study, the researchers consider that qualitative research is ideal when the aim is to explore and understand the meaning individuals or groups ascribe to a social or human problem (Creswell and Poth, 2018). The following table presents the research design:

Table 1
Study design.

Research tools	Participant	Nature of data	Sample	Sampling procedure	Mode of analysis	Approach of analysis
Questionnaire with open-ended questions	Students (10 female, 16 male)	Qualitative	26	Purposive	Organising, Identifying	Thematic analysis
Semi-structured interview schedule	Teachers (3 female, 4 male)	Qualitative	7	Purposive	Organising, Identifying	Thematic analysis
Semi-structured interview schedule	Administrators (2 male)	Qualitative	2	Purposive	Organising, Identifying	Thematic analysis

3.2. *Participants*

The participants included 26 students and 7 teachers from the Department of English at PSTU and 2 administrators from the department and the university. Among the 26 students, 16 were undergraduates, and 10 were master's students. Almost all students came from a Bangla medium background. They had studied English for an average of 15 years. All of them began learning English in school; however, some began at home. Their ages ranged from 18 to 25 years. Among them, 16 were male students, and 10 were female students. Of the seven teachers, four were male, and three were female. Among the teachers, three were associate professors, one was an assistant professor, and three were lecturers. Some teachers specialized in Applied Linguistics and English Language Teaching (ALELT), while others focused on Literature and Cultural Studies. In addition to the students and teachers, two administrative officers, one from the English Department and the other the examination controller from the university, were also interviewed. The participants were selected using purposive sampling because it helps select individuals who can provide useful answers for the study (Patton, 2002).

3.3. *Research instruments*

To collect information, three main research tools were used: semi-structured questionnaires for students, teachers, and administrators. According to Roopa and Rani (2012), a questionnaire is a set of questions used to collect data from participants. The questionnaire contained 34 open-ended questions to elicit information from students. The questions were divided into three sections. The first section collected basic information about the students, such as their age, gender, academic level, and educational background. The first section helped to understand who the participants were. The second section focused on the difficulties students faced during test-taking. They were asked about the problems they experienced in the testing environment, such as noise, lack of space, poor instructions, or examiners' behavior. The third section was about the effects of test results and experiences over time. Then, semi-structured interviews were conducted with seven teachers. There were 22 questions in the interview guide. A total of 14 questions were asked of both the department administrators and the university exam controller's office. The questions focused on their roles in test administration. The answers from these administrators helped the researcher understand the broader picture of test administration at PSTU. These tools helped explore how test administration is conducted, its effects on learning, and how it can be improved in the context of a Bangladeshi university.

3.4. *Data collection procedures*

Pilot tests were conducted with three types of stakeholders—three students, two teachers, and one administrative staff member—to assess whether the questionnaires were suitable for addressing the three research questions. There were 27 in-person and 8 online meetings, according to participants' convenience. Some candidates were unfamiliar with the roles of test administration and backwash; therefore, the researcher explained each term in brief sessions and reassured participants of confidentiality to encourage honest responses.

3.5. Data analysis procedures

Each dataset was analyzed using thematic analysis (Braun & Clarke, 2006) with an inductive approach. For identification, responses were reread multiple times, and significant segments were coded. Then, similar codes were categorized into themes, which were also reviewed and refined. For example, codes connected to “stress from exams” and “time pressure” framed the theme of test-related stress. The researcher followed a reflexive journal to acknowledge personal assumptions and reduce bias. All participants and the institution were anonymous, including students (S), teachers (T), administrators (A), and the university (Clark-Kazak, 2017).

3.6. Ethical considerations

Standard research ethics practices were followed while conducting this study. All participants were informed about the study’s purpose, the use of data, and confidentiality. Participants’ written consent was obtained before data collection, and no personally identifiable information was collected. Participation in this study was voluntary, and participants were allowed to skip questions if they felt uncomfortable, and they had the right to withdraw at any stage without any negative consequences. To maintain anonymity, responses were anonymized, and the data were stored securely and accessed only by the researchers for data analysis purposes. The study posed no risk to the psychological or academic well-being of the participants.

4. Findings and discussion

The existing barriers in terms of test administration and the issues of assessment backwash, as unveiled from the perspectives of test takers, invigilators, and administrators, are as follows:

4.1. Testing environment

Participants’ responses demonstrate that the examination hall is an important consideration while administering a test. Noise nuisance, uncomfortable seating arrangements, poor ventilation, unhygienic conditions, and unstable lighting and electricity are among the issues that are ever-present in the testing environment at PSTU. Noise level in and around the testing hall is the initial concern of the candidates. Various noises, ranging from side talk among peers and supervisors’ “off-topic conversations” to external disturbances from construction work, political rallies, and campus events, disrupted candidates’ concentration while taking tests, as they described. Such interferences diverted their attention and also led to mental fatigue and anxiety; here, it becomes difficult to maintain focus throughout the examination. McDonald (2001) argues that a well-organized, distraction-free examination hall is essential for reducing cognitive overload. S-12 described additional logistical discomfort, including unstable desks and chairs, load shedding, poor seating arrangements, and poor ventilation. The heat wave demotivates participants, and migraine symptoms are prevalent among test takers. An uncomfortable physical setting disrupts candidates’ engagement and reduces the overall quality

of the assessment (Maxwell, 2000). The basic amenities, such as drinking water and clean washrooms, imply deeper administrative oversight in logistical standards. Moreover, the tradition of showing respect to teachers requires discomfort when candidates lack the courage to change their seats in case of necessity. The test setting at PSTU not only undermines test fairness but also reflects all the resource-related issues and authority-oriented pedagogical practices in Bangladesh.

4.2. Procedures and protocols

Participants conveyed the need for fairness and integrity to secure better testing conventions. The lack of active supervision is a concerning issue, as some invigilators were observed using mobile phones, while others were seen checking previous scripts in the examination hall. Participants also reported unprofessional behavior by invigilators, such as publicly scolding test takers or allowing snacks during exams. S-5 commented that correcting candidates loudly “creates discomfort for others during taking tests,” while S-20 asserted that the food aroma during tests was “highly distracting when one is hungry and under pressure.”

Invigilators admitted many of the claimed issues as practical. Authoritative nature among teachers appears to be a traditional attribute in academic environments, as noted by T-4. T-5 stated that “change must come from within us.” Others emphasized professional growth and self-improvement. Another invigilator, T-2, highlights the benefits of training programs and several academic workshops for better test management.

The findings reveal inconsistencies in terms of procedural and instructional support. This aligns with Hughes' view (2003), which highlights the types of tests, familiarity, and instructional washback. However, administrators emphasized teacher autonomy by maintaining adherence to the University Grants Commission of Bangladesh (UGC)-approved curriculum while acknowledging limitations in monitoring practices. In contrast with Isbell et al. (2023)'s perspective, administrative vigilance is necessary for institutional integrity. Thus, the administrative body primarily bears certain responsibilities that should not be overlooked. According to Max Weber, authority in modern institutions is predominantly based on a legal-rational administrative apparatus, in which power is exercised through formal rules, regulations, and executive structures. In test administration, the authority to supervise tests, enforce regulations, and evaluate test takers is governed by the bureaucratic discipline of the educational policy (Weber, 1978).

4.3. Materials and equipment

The test administration can be efficient through the proper distribution of materials and functional equipment. However, there are various management lapses related to the timing, availability, and organization of the discussion among participants. Such treatments negatively impacted the testing experience. Candidates receive answer sheets regularly on time, and there is an active wall clock in the front. Naturally, not all students receive the question paper at the same time, but they have to stop writing on time. Participant S-24 argued that candidates seated at the back of the examination hall received the question papers late and sometimes failed to

finish the test on time. That participant viewed this as "...a kind of systematic inequality." Wall clocks in different rooms do not display the same time, which is confusing, as noted by S-14 and S-21. Moreover, delayed arrivals, insufficient question papers, and other logistical problems were reported by test takers. Frequently, shortages of additional answer sheets cause anxiety among candidates and supervisors and distract from mental preparation.

Invigilators are also concerned about the reported issues, and here they emphasized resources and personnel. T-5 argued, "The lack of skilled and insufficient manpower creates obstacles to conducting all the responsibilities accurately." Although the department's administrative body ensured the accuracy of all materials, the university exam controller group considered these issues beyond its purview.

4.4. Examiner, invigilator, test taker: roles and relationships

As invigilators are directly involved in testing, they play a vital role in ensuring assessment integrity; otherwise, this may adversely affect test takers' performance. Invigilators' supervision is a fundamental component of valid and equitable assessments. Candidates reported some invigilators engaging in personal deeds, as it was claimed before, for instance, using phones or grading papers during supervision, which stimulates misconduct and creates unfair opportunities among particular candidates. Observing peers engage in unfair means "disrupted concentration and affected the flow of writing," claimed S-11. It mirrors McMillan's (2014) notion of "psychological noise," in which interference hinders students' concentration and cognitive performance. Encouragingly, teachers are willing to improve through collaboration and self-regulation. Stiggins (2005) similarly discussed that professional development in assessment ethics and emotional intelligence can improve the integrity of test administration. Michel Foucault's (1977) theory of power explains how examiners and teachers exercise authority, shaping test-takers' behavior within institutional assessment culture (Ball, 2013). This emphasized the integrated power relations in pedagogy, as surveillance and evaluation drive both teaching and learning consequences.

4.5. Integrity and security policies

Maintaining integrity is pivotal to securing test administration policy. Candidates experienced certain issues that raise questions about the validity of tests. There are certain days when teachers arrive late to the examination hall, as reported by S-15 and S-19. They found this incident confusing, which aroused anxiety and hampered their preparation. Then, the wall clocks in different examination halls display the time differently, which S-4 opined that "Most of us never take it easy; rather, we create disbelief in the invigilator." Hosseinnia and Kafi (2024) support that a transparent, accountable, and vivid structure of conducting tests hinders misuse in academic assessments, which a code of ethics can ensure. S-4 argued about the punishment, where the test takers who are caught with wrong deeds are allowed to continue the tests, somehow. S-4 concludes that it grows a sense of supervision, which is fragile, and motivates others to do such things. The scenario of confusion and incredulity among students portrays adverse experiences of attending tests. Bias-free treatment for candidates promotes

administrative fairness, trust, and adequacy, leading to procedural fairness (Hamid et al., 2019). Experienced invigilators and administrative officials admitted such issues, and they are trying their best to overcome them. Assessment practices are shaped by organizational values in academia. Basically, the ethical governance culture, where shared values, systems, and norms reinforce consistent administration (Treviño et al., 1998).

4.6. Support staff

In administrative practice, effective testing conditions are fundamental. Skilled personnel are necessary to perform required responsibilities efficiently, as teachers and administrators have also claimed. The candidates argued about inadequate access to necessities, such as drinking water and restroom facilities, during prolonged tests. Such a scenario emphasized a vital gap in test-taker comfort. Though there is concern about cheating, the administrative body can address those through appropriate measures rather than minimizing candidates' basic needs. Here, Stobart (2008) points out, ethical standards and mutual trust are necessary in test administration, and neglecting students' physiological needs undermines both performance and broader educational goals.

Poor hygiene in washrooms and examination halls is further reflected in S-4 and S-15. The learning environment strikingly affects learners' affective filters and can increase harmful backwash (Krashen, 1982). However, some teachers and students had undertaken voluntary initiatives and established regular cleaning programs, but these efforts did not last long. The inconsistency of these efforts underscores systemic drawbacks. In essence, the existing contextual difficulties are critical to addressing and upgrading administrative quality to ensure a supportive educational environment through teacher-student mutual respect and trust, which aligns with the philosophy of humanizing pedagogy (Freire, 1970).

4.7. Washback effect

The department employs four types of assessment: continuous, formative, summative, and viva voce. Test takers experience both positive educational impacts and notable challenges.

4.7.1. Continuous assessment

Continuous assessments include presentations, assignments, creative tasks, and class tests. Generally, it is viewed as an effective tool for promoting self-directed learning. To grow communicative competence and confidence, students valued presentations. Yet several candidates expressed dissatisfaction with the grading process, as participants sometimes received similar scores regardless of their performance. As S-19 reported, "This is a sort of injustice towards the presenters who tried their best," suggesting that the similarity discourages motivation. Brookhart (2010) also valued that fair and motivating assessments can boost learners' enthusiasm to perform better.

The practice to foster creative writing, independent learning, or critical thinking can be an assignment. Nonetheless, candidates are admitted for plagiarism or copying peers' writing, which usually undermines the scope for growth. Park (2003) noted that teachers today can use

available digital tools to ensure academic integrity. Learners further emphasized project reports, fieldwork, and literary reviews, which help connect academic learning to real-world implementation. They also acknowledged that only a few teachers attempt to include class tests and other creative assessment tasks.

4.7.2. *Formative assessment*

Academic strengths and weaknesses are broadly acknowledged by formative assessments, which also advance self-regulated learning. S-12 reported feeling "confidence after taking formative assessments," while S-9 emphasized the importance of providing progress reports through both self-reflection and teacher feedback. Black and Wiliam (2009) discussed formative assessment as a way of improving achievement and fostering inclusive learning environments.

Despite these benefits, various obstacles are also raised from participants' opinions. Candidates feel stressed because there are too many questions, the syllabus is extensive, and there is insufficient time. Basically, candidates get tired or even sick because of such stress. Moreover, the lack of timely and transparent commentary reduces opportunities for meaningful improvement. One of the participants, S-11, emphasized, "If I get my progress report in due time, it helps me to take the next initiatives consciously." Therefore, the effectiveness of formative assessments depends on consistent utilization and manageable workloads, as it is progressive in principle.

4.7.3. *Summative assessment*

Summative assessment holds the utmost weight in determining academic achievement. Though this assessment is the most criticized one due to certain limitations in practice. Most of the participants argued that tests with strict time limits do not truly measure how proficient or creative candidates are. S-1 defines the process as "a time game where students have to prove that they understand and not what they understand." Guskey (2003) is also against relying on a single assessment method, stressing that test takers' diverse aptitudes require multiple approaches to unveil.

The candidates also question the potential of summative assessments to measure learners' practical knowledge. They noted that low-scoring students can do better in practical competencies, which remain unrecognized in "time-bound" written exams. As a result, many of them emphasized adopting hybrid evaluation systems that blend both practical and theoretical dimensions. S-7 points out traditional summative assessment as "closed-door knowledge practice." This consideration is supported by Garrison and Vaughan's (2008) idea that blended assessment models offer more reliable and holistic measures of candidates' proficiency.

4.7.4. *Viva voce*

In assessing oral proficiency and conceptual clarity, viva voce is perceived as crucial but is also controversial. The main challenges are with high stress and irregular or unpredictable rules. Test takers' concerns were expressed over unequal questioning, time allocations, and

instances of subjective remarks. S-25 coded viva voce as “a battle of patience,” citing long waiting periods and poor scheduling. Moreover, S-19 reported feeling humiliated as they faced personal attacks from examiners.

Invigilators also acknowledge the difficulties discussed and recognize their limited autonomy in designing assessments, owing to institutional barriers. T-2 explained, “We are free to design our tests and testing norms, but only within a certain area.” Teachers hope for updating the testing policy by integrating practical and theoretical elements. They also want to utilize technologically advanced classrooms and organize mock viva sessions to help reduce anxiety and prepare candidates more effectively for their career enrichment. Watts (2012) and Wellington (2010) also advocate for peer preparation, mock sessions, and mentoring to contribute to effective viva experiences. To conduct a viva voce fairly, examiners further prioritized the importance of ethical commitment and institutional support. So, T-3 noted for teachers that “an individual’s honest intention and effort can help in overcoming most of the barriers among test takers.”

4.7.5. *Questionable testing practices*

In the responses to three key stakeholders, there are various weaknesses in test administration. Specifically, in resource-constrained contexts like Bangladesh, where limited facilities and administrative drawbacks compromise testing integrity (Shohamy, 2001). Participants shared that frequent mid-test instructions and clarifications disrupted their concentration. However, intended to assist them, such interventions undermine accurate measurement, aligning with Fulcher and Davidson (2007), who opined that test takers must be allowed to think and respond autonomously. Some candidates observed that some received question papers late or continued writing after the given time was over, resulting in inequality. These inconsistencies arise from a lack of fairness and reliability (Cheng & Curtis, 2004; Bachman & Palmer, 1996).

Lack of identity verification implies weak invigilation, which is also evident. Some supervisors overlooked the need to check IDs because they knew the students personally. It compromises the security and fairness of the test (Kunnan, 2004). Here, ensuring appropriate identity is pivotal to maintaining the integrity of assessment outcomes. Some instances of preferential services, such as invigilators explaining word meanings to particular students, were also explained by candidates. Perceptions of bias and inequity are fostered by such treatment. Shohamy (2001) argues that favoritism undermines trustworthiness in testing. Thus, the students suggested anonymous, ID-based coding systems that would reset each semester to prevent discrimination in assessment.

Administrative formalities just before the examination period, for example, delays to get the admit card, fee clearance, and logistical disorganization, are also common. S-1 suggests that a well-designed website could resolve problems with fees, yet candidates continue to face difficulties. These points show that the institution lacks proper systems and digital management, which could make procedures easier and reduce stress for candidates. The University of

Edinburgh (2025) emphasized professionalism and consistency as pivotal to ethical supervision, while Green (2014) highlighted that effective assessment requires early logistical planning.

4.8. Pedagogical implications

The findings reflect the actual state of many Bangladeshi educational contexts and also highlight key issues in test administration and backwash. By reflecting on its own practices in academic settings, this study enables stakeholders at other institutions to examine their own contexts. Their improvement can also be identified by comparing these perspectives. The study primarily promotes awareness and brings positive changes to the assessment system. The insights gathered here contribute to a broader discussion on educational assessment in Bangladesh, especially at the tertiary level. They show that testing practices must consider local challenges and find realistic solutions. Policies must be developed to ensure testing is fair and consistent. Training for invigilators and administrators should be strengthened, and involved individuals should be able to communicate freely. By responding to the issues raised by stakeholders, particular departments can reduce negative backwash effects and implement tests more efficiently to measure and support candidates' growth. Moreover, other universities or departments can compare and contrast the findings with their contexts and consider self-improvement in assessment and administration. Overall, a shared effort to improve test administration will likely lead to better academic results and a more supportive educational environment.

5. Conclusion

This study explored the perspectives and experiences of three key stakeholder groups—test takers, invigilators, and administrators—regarding test administration and its implications within the Department of English at Padma Science and Technology University (PSTU). The findings indicate that successful test administration plays a critical role in supporting candidate performance. When assessments are well organized, test takers exhibit heightened confidence and reduced anxiety, thereby enhancing academic achievement and motivation. Stakeholders reported that a quiet and comfortable physical environment, clearly articulated instructions, and timely communication serve to mitigate examination-related stress and foster a more positive testing atmosphere. Concurrently, this study identified several problems associated with test administration, including disruptions during examinations, poorly scheduled viva-voce sessions, and inadequate facilities and materials. Such deficiencies adversely affect both students and the overall quality of assessment. Conversely, well-managed testing supports effective teaching and learning practices and promotes greater alignment between classroom objectives and evaluation methods. In this manner, adequate test administration facilitates positive backwash effects and drives improvements in both teaching and learning.

Furthermore, this study highlights how different assessment types—namely formative, summative, and oral—influence student behavior and academic achievement. Stakeholders indicated that the appropriate implementation of formative assessments helps identify learning gaps and better prepares students for final evaluations, thereby enabling candidates to make more

progressive advances in their academic careers. Summative assessments are regarded as significant and thus require careful design to accurately reflect actual learning outcomes. Oral assessments, while valuable, necessitate more systematic scheduling and administration to achieve their intended effectiveness. In essence, administrative completeness ensures both institutional and individual growth in candidate performance.

Despite these contributions, the findings should be interpreted within the context of several limitations. The study is qualitative in nature and focused exclusively on a single department at one university (PSTU), with a relatively small participant sample comprising seven invigilators, two administrative staff members, and twenty-six undergraduate and graduate students. Consequently, future research would benefit from employing mixed-methods approaches, incorporating larger and more diverse samples across one or more departments or institutions, and extending inquiry to primary and secondary education levels in Bangladesh. Additional promising avenues include examining the role of digital tools and online platforms in supporting testing in second-language (L2) contexts, investigating the causes and consequences of low motivation among test administrators, and undertaking cross-cultural and intracultural comparisons to understand how cultural factors shape testing and learning. Ultimately, further research should aim to assist academic institutions in aligning their administrative and assessment policies with their broader teaching and learning objectives.

Acknowledgement

We are grateful to the participants of this research who contributed their time and data. We are also grateful to the academic and administrative staff of PSTU for their spontaneous support and encouragement. We are thankful to the editor and the reviewers for their constructive feedback.

Declaration on the use of AI

The authors used AI tools for grammatical and linguistic refinement and to enhance the manuscript's clarity. AI tools were not used for content generation, data analysis, interpretation, or formulation of conclusions. The authors retain full responsibility for the originality of the publication.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Alderson, J. C., & Wall, D. (1993). *Does washback exist?* *Applied Linguistics*, 14(2), 115–129. <https://doi.org/10.1093/applin/14.2.115>
- American Psychological Association. (2014). *Standards for educational and psychological testing*. American Educational Research Association. <https://www.testingstandards.net/open-access-files.html>

- Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 37–50). Lawrence Erlbaum Associates.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Barkhuizen, G., Benson, P., & Chik, A. (2013). *Narrative inquiry in language teaching and learning research*. Routledge.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.
- Brown, H. D. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Education.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge University Press.
- Cheng, L., & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. Lawrence Erlbaum Associates.
- Clariana, R. B., & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with test mode effect. *British Journal of Educational Technology*, 33(5), 593–602. <https://doi.org/10.1111/1467-8535.00294>
- Clark-Kazak, C. (2017). Ethical considerations: Research with people in situations of forced migration. *Refuge*, 33(2), 11–17. <https://doi.org/10.7202/1043059ar>
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. Routledge.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. SAGE Publications.
- Fan, J., Frost, D., & Liu, Y. (2020). Washback of a high-stakes test on teachers' perceptions and practices: The case of the TOEFL iBT. *Language Assessment Quarterly*, 17(2), 172–192.
- Foucault, M. (1977). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). Pantheon Books.
- Freire, P. (1970). *Pedagogy of the oppressed*. Herder and Herder.
- Fremer, J., & Wall, D. (2003). Use of tests and assessments. In W. J. Popham (Ed.), *Educational assessment: Principles, policy, and practice* (pp. 159–174). Allyn & Bacon.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge.
- Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. Jossey-Bass.

- Green, A. (2007). *Washback to the learner: Learner and teacher perspectives on IELTS preparation courses in China, India and Nepal*. Cambridge University Press.
- Green, A. (2014). *Exploring language assessment and testing: Language in action*. Routledge.
- Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership*, 60(5), 6–11.
- Harding, L., & Fulcher, G. (2021). Epilogue: Language testing: Where are we heading? In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed., pp. 609–614). Routledge. <https://doi.org/10.4324/9781003220756>
- Hamid, M. O., Hardy, I., & Reyes, V. (2019). Test-takers' perspectives on a global test of English: Questions of fairness, justice and validity. *Language Testing in Asia*, 9, 16. <https://doi.org/10.1186/s40468-019-0092-9>
- Hembree, R. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, 227, 483–493. <https://doi.org/10.1016/j.jad.2017.11.048>
- Hosseinnia, M., & Kafi, Z. (2024). Constructing and validating a code of ethics in testing inventory: Investigating EFL instructors' perspectives. *Language Testing in Asia*, 14, Article 56. <https://doi.org/10.1186/s40468-024-00331-y>
- Hughes, A. (2020). *Testing for language teachers*. Cambridge University Press.
- Isbell, D. R., Kremmel, B., & Kim, J. (2023). Remote proctoring in language testing: Implications for fairness and justice. *Language Assessment Quarterly*, 20(4–5), 469–487. <https://doi.org/10.1080/15434303.2023.2288251>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Kumaravadivelu, B. (2001). Toward a postmethod pedagogy. *TESOL Quarterly*, 35(4), 537–560. <https://doi.org/10.2307/3588427>
- Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context: Proceedings of the ALTE Barcelona Conference* (pp. 27–48). Cambridge University Press.
- Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching*. Pearson Education.
- Magambo, J., & Rukundo, Z. (2016). The effect of the examination environment on students' academic performance. *African Journal of Educational Studies in Mathematics and Sciences*, 12(1), 1–11.
- Marzano, R. J. (2006). *Classroom assessment & grading that work*. ASCD.
- Maxwell, J. A. (2000). *Understanding and validity in qualitative research*. *Harvard Educational Review*, 62(3), 279–301.
- McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology*, 21(1), 89–101. <https://doi.org/10.1080/01443410020019867>
- McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction* (6th ed.). Pearson.
- McNamara, T. (2000). *Language testing*. Oxford University Press.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. <https://doi.org/10.1177/026553229601300302>

- Ott, M., & Todeva, E. (2020). Technology-mediated language learning and teaching. In E. Todeva & M. Ott (Eds.), *Language learning in the digital age* (pp. 1–18). Springer. https://doi.org/10.1007/978-3-030-40634-0_1
- Park, C. (2003). Engaging students in the learning process: The learning journal. *Journal of Geography in Higher Education*, 27(2), 183–199. <https://doi.org/10.1080/03098260305675>
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. SAGE Publications.
- Roopa, S., & Rani, M. S. (2012). Questionnaire designing for a survey. *Journal of Indian Orthodontic Society*, 46(4_suppl1), 273–277. <https://doi.org/10.5005/jp-journals-10021-1104>
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. <https://doi.org/10.3102/0013189X029007004>
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Longman.
- Stiggins, R. J. (2005). *Student-involved assessment for learning*. Pearson Merrill Prentice Hall.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.
- Treviño, L. K., Butterfield, K. D., & McCabe, D. L. (1998). The ethical context in organizations: Influences on employee attitudes and behaviors. *Business Ethics Quarterly*, 8(3), 447–476. <https://doi.org/10.2307/3857431>
- University of Edinburgh. (2025, April). *Mains invigilator handbook*. Registry Services. https://registryservices.ed.ac.uk/sites/default/files/2025-04/Mains%20Invigilator%20Handbook_0.pdf
- Walkin, L. (1990). *Planning for continuing education and training*. Stanley Thornes Publishers.
- Watts, J. H. (2012). Preparing doctoral candidates for the viva: Issues for students and supervisors. *Journal of Further and Higher Education*, 36(3), 371–381. <https://doi.org/10.1080/0309877X.2011.632819>
- Weber, M. (1978). *Economy and society: An outline of interpretive sociology*. University of California Press.
- Wellington, J. (2010). More than a matter of cognition: An exploration of affective writing problems of postgraduate students and their possible solutions. *Teaching in Higher Education*, 15(2), 135–150. <https://doi.org/10.1080/13562511003619961>