

Balancing the Scales of High Stakes Public Exams in the Secondary and Higher Secondary Levels in Bangladesh: The Interplay of Reliability to Foreground Validity, Fairness, and Equity for Educational Reformation

Md Taqi Yasir*

Abstract : This paper investigates the core question: Are high-stakes tests in Bangladesh reliably assessing student abilities, especially in the context of secondary and higher secondary education? High-stakes tests, defined as examinations with significant consequences for students' futures (Marchant, 2004), can narrow the curriculum and impact education equity (Madaus et al., 2009; Au, 2023). The research contextualizes Bangladesh's educational evolution, highlighting its colonial legacy and post-liberation reforms aimed at aligning with the UN Sustainable Development Goals for quality and equitable education. A critical analysis of the 2008 reform, which integrated "creative questions" into high-stakes exams, is provided to highlight challenges such as inadequate teacher preparation and allegations of misconduct. The paper argues that for high-stakes testing to be fair, valid, and equitable, a focus on the reliability of both formative and summative assessments from early education stages is crucial. This should involve thorough research, pilot testing, and phased implementation strategies, advocating for a balanced assessment approach that prioritizes validity, fairness, and equity, deeming them essential for advancing quality education reforms in Bangladesh.

Key Words : high-stakes testing, validity, reliability, educational equity, curriculum reform, Sustainable Development Goals (SDGs), formative assessments, summative assessments

Assessment Evolution and the Quest for Quality: Navigating Changes from Historical Roots to Modern Reforms

The dramatic shift in the education assessment system in Bangladesh in the last couple of years has created a topic of contention amongst parents, schools, teachers, policymakers, and news and print media, wreaking havoc in social media. On the other hand, the students are relieved knowing that they are not required to sit for rigorous competitive examinations until they attend their pivotal school leaving examination in grade ten. The newly proposed reform consists of a significant focus on assessment, where their progress

* **Md Taqi Yasir** : Lecturer, Department of English, Jahangirnagar University.

will not be tested through any “final examination” before they are in grade four and no “public examinations” till grade ten, which is the final year of secondary-level education in Bangladesh (Rahman et al., 2023, p. 1). The quantity of examination reduction is undoubtedly a stress relief for students. Still, the alternative assessments are not tested enough in the context of our country to provide a positive outcome, which will prepare students for the high-stakes tests. Henceforth, to create an equitable, valid, and fair assessment environment for the test takers in their final years (which is at high stakes), their formative assessments (in the formative years) must be highly reliable. High-stakes test reliability must be constructed through gradual research and development, pilot projects, and spontaneity, not through coercive, rigorous, and sudden changes to achieve a positive outcome and competency.

The Government of the People’s Republic of Bangladesh have a mission and vision of attaining the sustainable development goals chartered by the United Nations, and ensuring quality education is one of the 17 goals (goal number 4). Quality education requires quality reformation. At the onset of the reformation goals, qualitative infrastructural development has a trickle-down effect on increasing nationwide outreach and dissemination.

From the historical contextualisation, it is interesting to find that from the British rule in Bengal to the Pakistani regime, from the liberation in 1971 to the present, the education curriculum in Bangladesh has undergone multiple upgrades. Current learning styles and pedagogical approaches have replaced traditional rote learning and one-way formal teaching methods. However, the assessment system has remained relatively the same until last year. Previously, in 2008, the Ministry of Education attempted to reform the assessment paradigm through a subtle change, with the introduction of “creative questions” at different high-stakes tests in the junior secondary (grade eight), secondary (grade ten) and higher secondary levels (grade twelve) such as Junior Secondary Certificate (JSC), Secondary School Certificate (SSC), Dakhil, Higher Secondary Certificate (HSC), and Alim Examinations. However, the sudden change was announced and casually implemented without training most teachers under this method and adequate piloting. According to Alamgir, M., and Nandy, D. (2022), there is a rumour that teachers used to buy questions at high prices from different teachers' associations when writing tests for the examinations, which is a testament that they were not well-trained and equipped to implement the newly introduced pattern. Furthermore, the report says that 55% of the teachers did not receive any training till 2010 for the newly introduced system, which created an extreme anomaly.

Challenges and Ambitions: Implementing Outcome-Based Assessments in Bangladesh's Evolving Educational Landscape

In the last couple of years, there has been an attempt to shift the paradigm in the educational assessment methodology. The change was from a curriculum-based paradigm towards an outcome-based paradigm, though the shift was partially done. The pedagogical structure of the syllabus largely remains curriculum-based; the assessment structure is more embedded in the quartile of the outcome-based paradigm, making it outrageously ambitious. With a significant reduction in public/national examinations, the ministry focuses more on continuous assessments, practical tests, and vocational training integrated with theory and competency-based examinations. Still, evaluating such increased demography of students in secondary and higher secondary level examinations in an outcome-based assessment (OBA) paradigm moulded by a national public examination is quite unthinkable. The number of students has increased in the last seven

years, yet a need for more resources in rural secondary and higher secondary schools and madrasas challenges implementing the changes. For example, the number of students enrolled in secondary schools (grades 9 and 10) in 2022 is 10133143, which is 390071 more than that in 2015 (Bangladesh Education Statistics, 2022). The University Grants Commission has implemented and designed Outcome Based Education (OBE) curriculum for undergraduate-level students for all the university faculties in Bangladesh. Subsequently, a shift towards outcome-based assessment in the secondary and higher secondary levels was equally pivotal to perceiving the assessment methods at the university undergraduate levels. As Baird et al. (2018) mentioned, the curriculum-based assessment paradigm faced challenges with popularising outcome-based assessment by introducing criterion referencing and competency assessments. The atomistic outcome-based assessment is a methodological approach to “assessment as learning,” deeply rooted in Taylorism and vocational assessment standards. Producing high-quality rubrics is a challenge in OBA since it lacks inter-rater reliability and depends mainly on subject experts' consistency of judgement; therefore, when such colossal teenage demography will be adjudicated under the OBA paradigm in the form of a high-stakes national public examination in Bangladesh under the assessors with little or no experience or training, the question of reliability comes under the context.

Setting the Bar: Navigating the Complexities of Standard-Setting in Bangladesh's Educational Assessments

The standard-setting process in Bangladesh is complex and the most challenging aspect of the assessment. Baird et al. (2018) define “standard-setting” as the procedure by which there are definitive benchmarks in a particular assessment paradigm. These benchmarks set the levels of performance, specific parameters of performativity, and cut scores of an exam that explicitly distinguish between different degrees of competency or accomplishment. Standard setting henceforth defines a suitable degree of knowledge or competence in a specific topic, which is a crucial step in the evaluation process. At the secondary and higher secondary schools in Bangladesh, students (coming from all economic backgrounds and studying in private and government-funded schools) must sit for a national Secondary School Certificate (SSC) or Dakhil examination (Madrasah board) and Higher Secondary Certificate (HSC) or Alim examination (Madrasah board) conducted by the nine boards of Intermediate and Secondary Education and a centralised Madrasah board. The standard is set by the authorities of nine different boards (under the same curriculum across the country for science, business, and humanities disciplines or groups). The curriculums have two English and Bengali papers in common across all the disciplines. For example, English Paper I focuses mainly on the literature part and reading, and English Paper II focuses on the grammatical aspects and writing. Since the shift is from CBA to OBA, the standard setting of the examinations must rely on what criteria are formulated (through criterion referencing), selecting and training the suitable assessors, monitoring record-keeping and processes, and ensuring inter-rater reliability. Under the criteria mentioned above, the examination questions must undergo moderation based on the combination of statistical (Bangladesh Educational Statistics) and qualitative references (the quality of students under each board through analysing their results in the previous years). Multiple boards, except the Madrasah board, help decentralise education from the capital city and create a competitive balance. However, the sudden shift to the OBA paradigm will keep challenging the boards to ensure a fair, valid, and equitable ground of competition by implementing reliable examination standards.

High Stakes English: Assessing Language Proficiency in Bangladesh's Educational System

For ease of discussion and relevance, this paper intends to exemplify English examinations in Bangladesh's public national exams of secondary and higher secondary levels. In such high-stakes English tests in Bangladesh, test assessors must have a high level of expertise in the subject as well as the usage of the language. Coming from a lower-middle-class to middle-class demography, most English teachers do not have native-like English fluency or higher order of proficiency (equivalent to the C2 or C1 levels). English is considered the second language of the commoners in Bangladesh, and most official work is conducted in both Bengali and English at professional levels. The medium of instruction and materials of the university undergraduate and postgraduate levels are in English only, with a few exceptions. Subsequently, the English examinations in the secondary and higher secondary level national board examinations have high stakes. According to Rahman et al. (2023), these English tests are crucial for assessing students' eligibility for higher education, which further influences their socio-economic prospects in the future. The high-stakes nature of these tests dramatically influences and affects Bangladesh's overall educational outcomes, social infrastructure, and educational ecology. Due to the context-responsive and conservative approaches of these tests, designed in the realm of the CBA paradigm, the assessments highly perpetuate incoherent design practices, ignoring the reliability and validity factors. As a result, competency assessment remains a conundrum for educators due to such structural barriers.

Therefore, without ensuring reliability, it is not impossible, but rather implausible, to implement the OBA paradigm in testing English at the secondary and higher secondary levels. Since communication competency is the key, it has to be well-defined through measurable learning outcomes that the test aims to assess (Rahman et al., 2023). Incorporating multiple assessment techniques to assess different sets of acquired skills and competencies holistically is vital.

Consequences of High-Stakes Testing: Curriculum Narrowing and Social Inequalities

According to Madaus et al. (2009), when a test becomes high-stakes, it reduces the focus on different non-tested items, making the curriculum irrationally dubious of its objectives. The teachers focus highly on the contents of the test, limiting their instruction skills and exterminating the other essential contents of the curriculum necessary for developing interpersonal skills and acquiring competency. As a result, the teachers' efficacy relies on "teaching to the test", which further influences the behaviour of the students and the teachers, leading to the corruption of the test. The control over the curriculum shifts from the teachers to the state (such as Education Ministries) and different federal agencies who are responsible for implementing the examination policies and practices, further alienating the role of the teachers. As a result, the teachers become delirious by changing their teaching aims, which incorporate rectifying and improving the students' test performances through positive and negative reinforcements, ignoring broader educational goals and outcomes. High-stakes testing reinstates Pavlovian classical conditioning, which is the Achilles' heel in educational pedagogy, and thus systematically disenfranchises the broader educational mission that academicians worldwide strive to achieve.

A different perspective on educational inequality through in-classroom experience and not achieving educational equality goals due to high-stakes testing was explored by

Au (2023). For the conservative nature of these tests, the accountability movements resulted in increased dropout rates. 40% of African American and Latino students in Texas dropped out from the Texas public schools in 2001 due to the white students consistently outscoring them in terms of performance gaps. This disparity is tumultuously evident in the assessments of high school exit exams (Au, 2023). In Bangladesh, the case of dropout rate is rather a gender issue than a racial one. According to reports of Bangladesh Educational Statistics (2020), there were 102, 52,126 secondary school students in Bangladesh in 2020, and the number decreased to 101, 90,022 in 2021. Of 62,104 students who dropped out of different secondary schools nationwide, 85.08% were girls, equivalent to 52,838 students of the total drop-out population (Alamgir, 2023). Due to the pandemic, most of these students were victims of child marriage and turned out to be labourers.

Testing Realities: Striving for Reliability Amidst Challenges in Bangladesh's Educational Assessments

The curve of the dropout demography is steeper in the rural than in the country's urban areas. Sarker et al. (2019) mention that 40% of dropouts result from poverty; interestingly, 37.5% of students do not like the schooling system. The socioeconomic and cultural deficiency in Bangladesh provokes the stigmatisation of liberal, outcome-oriented, vocational, and gender-neutral education and assessment in Bangladesh. As a result, examinations at high stakes determine the course of continuation of education through retention or drop-outs of many poor students in the following years. The schooling systems are highly outdated, and the pedagogy is entirely test-performance focused, resulting in the disorientation of the educational capacity, which creates doubt in the minds of parents from lower-income backgrounds whether it is economically sustainable or has any future monetary benefits to continue sending their children to school. Au (2023) mentions that high-stakes testing affects education equality goals because the students of less privileged backgrounds compete with those upper classes, with fewer opportunities and resources. As a result, educational disparity increases with the establishment of inequitable educational and social relations. Therefore, if high-stakes testing is unreliable and fails to ensure a fair and equitable ground of competition, it becomes inversely proportional to educational equality and a burden to the generation of a disintegrated population. To ensure educational equality and transferability, there must be a shift of focus towards equality of opportunities rather than the equality of outcomes. The fabricated inflation of grades and the decrease of grade boundaries in the national public exams (such as SSC and HSC, where a big pool of students score a GPA of 5 out of 5 due to the lenient nature of these exams) statistically provide temporary contentment and ego boost to portray a variegated picture of the education quality. However, such outcomes do not minutely reflect the real-life problematic scenario of the learning and competency of these students. As a result, they suffer from reality checks while receiving higher education and in professional spaces where they struggle to cope with the expectations. These preposterous GPA 5 holders can be compared to the “Bubble Kids”, referred to by Madaus et al. (2009), where they get unfair advantages over the weaklings to achieve their target.

Hughes (1996) has identified some reliability issues in English testing. The fundamental challenge in achieving reliability is inherent variability in test scores. If the same assessment is administered to a group of students, they can score differently. Reliability requires sufficient behavioural sampling. To achieve that, a good number of test item variables are necessary. Each additional item must be independent and coherent

to the test's measure of ability. However, such item multiplication can lead to fatigue, resulting in unrepresentative behaviour of the test takers. Sometimes, the unfamiliarity with test formats and the newly formulated assessment environment can lead to excessive mental pressure and non-uniform conditions.

According to Ali et al. (2018), the age-old ritualistic practices in administering and designing tests, which lack professional standards and ethical considerations, compromise test reliability and validity. Furthermore, for high-stakes English testing, one of the benchmarks of socio-educational progress, the unfair educational outcomes of impoverished and marginalised students assert an absence of reliability in the test metrics. It is only possible to attain reliability by addressing socio-political and logistical challenges and ensuring fairness and representativeness of the test contents by aligning them with the curricular goals through rigorous moderation. For English public examinations in Bangladesh, the test content must not only test reading and writing aptitudes but incorporate a broad range of skills in assessment, such as listening and speaking, which are the necessary components of communication competence.

The measure of reliability lies in several statistical analyses. Test administrators, assessors and educators must strive to make the high stakes tests reliable by analysing the test scores based on different contextual underpinnings and test components. One of the core responsibilities is to upgrade testing standards to state-of-the-art quality by applying adequate educational measures. Making English tests reliable in Bangladesh is possible if the contents of the tests align with the learning outcomes. The reformation will only be feasible if administrators penetrate the country's grassroots and marginalised rural areas. Otherwise, fragmented and urban-centric development will not accelerate the target of achieving consistent reliability.

Furr (2022) has identified the statistical method for indexing the reliability. Amongst them, the most critical factor is the internal consistency of tests. Using Cronbach's Alpha or Kuder-Richardson KR-20 (only for dichotomous items), assessors can measure if all parts of the test equally contribute to what it is supposed to measure. For example, in the English Paper I national exams, as Ali et al. (2018) mentioned, all test parts should measure and reflect communicative competence to achieve reliability, henceforth the internal consistency. Cronbach's Alpha provides an estimation of whether there is internal consistency of the test items calculating the test scores of the test takers. Cronbach's Alpha ranges from 0 to 1, where a score of 0.70 is acceptable, and any score less than that makes a test unreliable. In the later part of this paper, some random values of the internal consistency test using Cronbach's Alpha will be shown for further discussion and interpretation.

Inter-rater reliability is always ensured in Bangladeshi public exams since multiple raters check the test scripts with subject expertise. Nevertheless, training the raters is an essential part of achieving reliability. However, it is impossible to attain reliability through Test-Retest and Parallel Forms Reliability in a public national exam since the budget allocation is outrageously limited in the education sector. According to the national budget overview by the Centre for Policy Dialogue (CPD), only 11.6% of the national budget was allocated to the education sector in the 2023-2024 fiscal year.

On the contrary, since the budget disallows the test takers to assess their competency through multi-level parallel aptitude testing and retesting in public exams, it is possible to do so in the individual schools where they are required to pass a 'pre-test' and a 'test' examination in order to qualify for the national board examinations. As a result, the test-retest and parallel forms of testing can be practised by individual schools, which can also be mandated by education policymakers and implemented by practical

teacher training nationwide. The secondary and higher secondary level educators can also acclimatise themselves with necessary strategic essentialist techniques to achieve reliability in institutional tests, utilising the available resources. Nevertheless, the most sustainable way is training; there is no better alternative.

Enhancing Reliability in English Testing: Techniques and Challenges in Bangladesh

In English language testing, as rooted in Furr's (2022) psychometric principles, inter-rater reliability can ensure consistency in scoring for subjective tests such as composition writing. In public examinations in Bangladesh, if there is a significant gap in marks between the first and the second examiners, a third examiner rates the test. As a result, the test becomes highly reliable, ensuring transparency and justification of the scoring. The test takers also have the right to challenge the answer scripts, and the board assigns subsequent fourth and fifth raters to assess the test. A check and balance is ensured through such measures; therefore, the raters must be well-trained subject experts to ensure fairness and avoid the collateral damage of rating incorrectly.

According to Adie et al. (2019), assessment literacy is critical. The development of expertise in assessment is dynamic and relatively progressive and requires skills and knowledge to enhance student learning and skill acquisition through assessment. This encompasses further complex goals, such as the diagnosis of learning needs and extensive evaluation of assessment practices. Presumably, the assessors must shift from rule-based and prescriptive techniques to more intuitive, context-sensitive approaches based on their experience through a continuum of skill development. In Bangladesh, continuous evaluation in the classroom must foster student learning needs, such as identifying gaps, outlining the expected performance, and gauging their progression. Teachers must be encouraged to engage in reflective and critical pedagogical practices, foster pluralism in lesson planning and implementation, and understand the socio-historical and cultural context of their teaching, which will further impact their assessment methods. The most plausible denotation of such practices means ensuring that teachers effectively use their assessment strategies and adhere to recognised professional benchmarks.

Balancing Validity and Reliability: Enhancing Educational Assessments in Bangladesh

Students' behavioural pattern comprises a more prominent aspect of their performativity in tests. Educators can always set some standards for performance-based outcome expectations, but such outcomes might only sometimes reflect reality. Gomendio and Wert (2023) shed light on reducing student heterogeneity. Student heterogeneity perpetuates the dismal and compromised efficiency of a teacher. In countries such as Bangladesh, where educational inequality is prominent with high levels of inequity, a humongous challenge is there with varied student abilities that compromise learning gains. Therefore, alongside assessment validation, policies must be developed contextually to address ability grouping and differentiated modes of education or educational groups. Management of such educational challenges requires piloting and tracking of changes.

When reliability is ensured, the validation of high-stakes tests is required. Both concepts are highly interconnected and interplay simultaneously in setting the assessment standard. Cronbach (1990) has outlined the principles of validity, integrating all the pertinent arguments and the types such as content, criterion-oriented, and construct validity. For example, when an assessor takes context-sensitive approaches to make a test reliable, criterion-referencing will focus on how the test correlates or predicts with an

external criterion. Validity consolidates high-stakes test performance in a specific context in an affiliated domain. Furthermore, validation will help eliminate any unwarranted extenuation of test circumstances affecting the standard. Henceforth, reliability foregrounds test validation and is a stepping stone to assessment reformation.

In Bangladesh, to assess proficiency in English as a second language, multiple test items ensure internal consistency to make the test reliable. When an English teacher in Bangladesh assigns students to write an essay, criterion referencing involves selecting the item based on the likelihood of the students succeeding. The referencing would typically be done by comparing their scores on a relevant test (such as an aptitude test on paragraph writing). The validity of the test in this context would be determined by how effectively it correlates with or predicts students' success in writing.

As discussed earlier, Cronbach's Alpha is essential for measuring a test's internal consistency. Let us assume a hypothetical scenario that in an English Paper II test in Bangladesh, there are three similar items. Each item carries 20 marks. In a 60-mark examination, based on their score, if the sum of the item variances is 59.37 and the variance of the total score is 143.30, the Cronbach's alpha value would be 0.88, making the test highly reliable. The Cronbach's alpha value of 0.88 is relatively high, indicating that the test's internal consistency is strong. This further means that the test items or questions consistently measure the same underlying construct. Relevant to this case, according to Furr (2022), the Standard Error of Measurement (SEM) will represent the average size of the error scores that affect observed scores.

If the value of SEM is low, it typically lowers the average difference between observed and true scores, indicating higher test reliability. SEM is an indicator of indexing measurement error and is profusely linked to test reliability. Henceforth, in the hypothetical English test, the standard deviation is 11.9, and SEM is 4.2, which denotes the percentage of error to be 6.96%. The SEM of 4.2 suggests a moderate measurement error associated with individual test scores. The scores may fluctuate by this amount due to measurement errors. The percentage of error indicates the proportion of the total variance in the scores that can be attributed to measurement error. The percentage of 6.96% suggests negligible error in the scores, which ensures the reliability and validity of the test.

Conclusion

Incorporating comprehensive test specifications and ensuring the representativeness and validity of test items is a crucial aspect of reliability. Professional training for test developers and adherence to established testing standards are inseparable from reliability metrics. In an outcome-based assessment paradigm, it is crucial to ensure test reliability. Reliability helps to incorporate measurable learning outcomes, ensure consistent application of standards, and boost outcome-based criteria. As Baird et al. (2018) emphasise in OBA, assessments should not overshadow the learning process. To balance the paradigm between OBA and CBA, the concerned authority should identify and recognise the historical and global context of outcome-based education to imply the standard-setting project. Any rudimentary standardisation for evaluators can limit the boundary of the paradigms, holding the reliability of tests at stake. Therefore, implementing feedback mechanisms can establish a pragmatic approach to receiving and integrating opinions from multiple test stakeholders, continually improving the test's reliability and aligning with learning outcomes. An integral part of the reliability process is to statistically validate the assessment's ability to consistently and accurately measure what it is intended to. This approach will further incentivise the academicians and

policymakers to ensure data-driven decision-making to review, revise, and update assessment criteria and tools to maintain alignment with objective and subjective educational goals.

Challenges to ensure the reliability of high-stakes tests and their assessments are unfathomable. In a country with an unequally heterogeneous society, where education goals are fragmented based on ethnoracial, socioeconomic, and cultural diversity, the pledge to equitable reformation is a dream. The uncontrollable inflation of grades in public examinations to portray a false numeric representation of achieving educational goals has corrupted the testing system. The increased number of students getting GPA-5 does not represent what they are supposed to measure, invalidating the high-stakes public exams. Students who achieve these grades do not show consistency of outcome when they opt for higher education or move forward to professional vicinity. As long as these public national exams at the secondary and higher secondary levels in Bangladesh are unreliable, the invalidation will haunt the prospects of the test takers.

References

- Adie, L., Stobart, G., & Cumming, J. (2019). The construction of the teacher as expert assessor. *Asia-Pacific Journal of Teacher Education*, 48(4), 436–453. <https://doi.org/10.1080/1359866x.2019.1633623>
- Alamgir, M., & Nandy, D. (2022, November 10). *Creative question paper: Over one in 3 Bangladesh teachers out of their depth*. Asia News Network. <https://asianews.network/creative-question-paper-over-one-in-3-bangladesh-teachers-out-of-their-depth/#:~:text=DHAKA%20%E2%80%93%20The%20government%20introduced%20a,preparing%20questions%20under%20this%20method.>
- Alamgir, M. (2023, January 14). 1.7m students drop out in pandemic. *The Daily Star*. Retrieved from <https://www.thedailystar.net/news/bangladesh/news/17m-students-drop-out-pandemic-3220286>.
- Ali, Md. M., Hamid, M. O., & Hardy, I. (2018). Ritualisation of testing: Problematising High-stakes English-language testing in Bangladesh. *Compare: A Journal of Comparative and International Education*, 50(4), 533–553. <https://doi.org/10.1080/03057925.2018.1535890>
- Au, W. (2023). *Unequal by design: High-stakes testing and the standardization of inequality*. Routledge, Taylor & Francis Group.
- Baird, J.-A., Isaacs, T., Opposs, D., & Gray, L. (2018). *Examination standards: How measures and meanings differ around the world*. UCL, IOE Press.
- Bangladesh Education Statistics. (2022). Bangladesh Bureau of Educational Information and Statistics.
- Centre for Policy Dialogue (CPD). (n.d.). *National budget 2023-24: Summary education - cpd.org.bd*. cpd.org.bd. <https://cpd.org.bd/resources/2023/08/Education-budget-2023-24.pdf>
- Cronbach, L. J. (1990). *Essentials of psychological testing*. Harper & Brothers.
- Furr, R. M. (2022). *Psychometrics: An introduction*. Sage.
- Gomendio, M., & Wert, J. I. (2023). *Dire straits-education reforms: Ideology, vested interests and evidence*. Open Book Publishers.
- Hughes, A. (1996). *Testing for language teachers*. Cambridge U.P.

- Jackson, J. (2024). *Introducing language and Intercultural Communication*. Routledge.
- Madaus, G. F., Russell, M. K., & Higgins, J. (2009). *The paradoxes of high stakes testing: How they affect students, their parents, teachers, principals, schools, and Society*. Information Age Pub.
- Marchant, G. (2004). What is at Stake with High Stakes Testing? A Discussion of Issues and Research1. *Ohio Journal of Science*. 104. 2-7.
- Rahman, J., Karim, R., Alam, M. S., & Hossian, M. S. (2023, February 22). *The challenge of education reform in Bangladesh*. Dhaka Tribune.
<https://www.dhakatribune.com/opinion/longform/305384/the-challenge-of-education-reform-in-bangladesh>
- Sarker, M. N., Wu, M., & Hossin, M. A. (2019). Economic effect of school dropout in Bangladesh. *International Journal of Information and Education Technology*, 9(2), 136–142.
<https://doi.org/10.18178/ijiet.2019.9.2.1188>
- United Nations. (n.d.). *Sustainable development goals in Bangladesh*. United Nations.
<https://bangladesh.un.org/en/sdgs#:~:text=The%20UN%20and%20its%20partners,Bangladesh%20and%20around%20the%20world.>
- University Grants Commission of Bangladesh. (2020). Template of Outcome Based Education (OBE) Curriculum.
https://ugc.portal.gov.bd/sites/default/files/files/ugc.portal.gov.bd/notices/e4c1bdfd_8db9_4af8_a538_34dbc84ed2b0/OBE131019.pdf

[সার-সংক্ষেপ : এই প্রবন্ধটি যে মৌলিক প্রশ্নটি তদন্ত করে তা হচ্ছে: বাংলাদেশের উচ্চ-বুঁকির পরীক্ষাগুলি কি বিশ্বস্তভাবে শিক্ষার্থীদের দক্ষতা মূল্যায়ন করছে, বিশেষ করে মাধ্যমিক এবং উচ্চ মাধ্যমিক শিক্ষার প্রেক্ষাপটে? "উচ্চ-বুঁকির পরীক্ষা" বলতে এমন পরীক্ষাগুলি বোঝানো হয় যা শিক্ষার্থীদের ভবিষ্যতের সুযোগগুলির উপর গুরুত্বপূর্ণ প্রভাব ফেলে, যেমন বিশ্ববিদ্যালয়ে ভর্তি বা পেশাগত পথ। মার্চান্ট, জি. (২০০৪) অনুসারে, "উচ্চ-বুঁকির পরীক্ষাগুলি সেই পরীক্ষাগুলিকে বলা হয় যা শিক্ষার্থীদের বা শিক্ষকদের জন্য গুরুত্বপূর্ণ পরিণাম বহন করে।" মাদাউস এবং অন্যান্যদের (২০০৯) মতে, যখন একটি পরীক্ষা উচ্চ-বুঁকির হয়ে ওঠে, তখন এটি পরীক্ষিত না হওয়া বিষয়গুলির উপর মনোযোগ হ্রাস করে, যা পাঠ্যক্রমকে এর উদ্দেশ্যগুলির বিষয়ে অযৌক্তিকভাবে সন্দেহজনক করে তোলে। আউ (২০২৩) উল্লেখ করেছেন যে উচ্চ-বুঁকির পরীক্ষার প্রভাব শিক্ষার সমতা অর্জনের লক্ষ্যগুলিকে প্রভাবিত করে কারণ কম সুযোগপ্রাপ্ত পটভূমির ছাত্ররা আর্থিকভাবে সচ্ছল ও উচ্চ শ্রেণির ছাত্রদের সাথে প্রতিযোগিতা করে। গবেষণাটি বাংলাদেশের শিক্ষার পটভূমির একটি সংক্ষিপ্ত বিবরণ প্রদান করে, যা ব্রিটিশ ঔপনিবেশিক প্রভাব থেকে মুক্তি পরবর্তী শিক্ষা সংস্কারের দিকে প্রচেষ্টার উল্লেখ করে, শিক্ষা পদ্ধতি এবং পাঠ্যক্রম নকশার মূল পরিবর্তনগুলি চিহ্নিত করে। এটি বিশেষভাবে কম ঘন ঘন হলেও সম্ভবত অধিক অর্থবহ মূল্যায়নের দিকে একটি রূপান্তর পরীক্ষা করে, যা জাতিসংঘের টেকসই উন্নয়ন লক্ষ্যগুলির সাথে সমন্বয়ের চেষ্টা হিসাবে পরিচিত। এই গবেষণাটি পরীক্ষা পদ্ধতির দীর্ঘস্থায়ী জড়তা চিহ্নিত করার চেষ্টা করে, যতক্ষণ না সম্প্রতি পরিবর্তনগুলি চালু হয়। ২০০৮ সালের সংস্কারের একটি সমালোচনামূলক

বিশ্লেষণ প্রদান করা হয়, যা "সৃজনশীল প্রশ্ন"গুলিকে উচ্চ-ঝুঁকির পরীক্ষাগুলিতে একীভূত করে, যা শিক্ষক প্রস্তুতির অপরিহার্যতা এবং অনিয়মের অভিযোগগুলি সামনে আনে। প্রবন্ধটি যুক্তি দেয় যে উচ্চ-ঝুঁকির পরীক্ষা যাতে ন্যায়সঙ্গত, বৈধ এবং সমতায়নের হয়, তার জন্য প্রারম্ভিক শিক্ষার পর্যায় থেকে গঠনমূলক এবং সারাংশমূলক মূল্যায়নের বিশ্বস্ততার উপর মনোযোগ দেওয়া অপরিহার্য। এটি বৈধতা, ন্যায়সঙ্গততা এবং সমতা অগ্রাধিকার দেওয়ার মাধ্যমে একটি সমন্বিত মূল্যায়ন পদ্ধতির জন্য সুপারিশ করে, যা বাংলাদেশে মানসম্পন্ন শিক্ষা সংস্কারের উন্নতির জন্য অপরিহার্য।]