



SSC and HSC English Language Testing in Bangladesh: A Critical Appraisal

Md. Mahbul Alam

Senior Lecturer, Department of English, Bangladesh Cadet Colleges

Email: mahbul06@gmail.com

ABSTRACT

In Bangladesh, due to recurrent modifications in the English curriculum and strong thrust of Communicative Language Teaching (CLT), the concern of language testing has become very crucial. Administering a test requires fulfilling the major criteria for a good test, though it is convenient to think that ensuring those criteria becomes difficult or sometimes not viable due to the socio-economic and pedagogical realities, especially in Bangladesh. The present study deals with SSC and HSC English tests and critically discusses about the validity, reliability, practicality, authenticity and washback effect of the contents of the questions in particular and the testing procedure in general. The method is moderately a content analysis method. The latest pattern of SSC and HSC English question is taken as sample. The analysis shows that in most of the cases, the examinations lack the proper level of validity, reliability, practicality, authenticity and washback. The study will be significant for the curriculum designers and the ELT practitioners of Bangladesh to further probe into the language testing arena of the country.

Keywords: Language testing, validity, reliability, practicality, authenticity, washback etc.

Introduction: Reality of English as a Foreign Language in Bangladesh

Two public examinations- Secondary School Certificate (SSC) and Higher Secondary School Certificate (HSC) are landmark events for a student all through his academic life in Bangladesh. Along with other subjects, English, in the form of Paper-I and Paper-II, each carrying 100 marks, is a compulsory subject in both SSC and HSC examinations. Sultana (2018) opines that the nature of the examination system followed here is 'achievement test' (p. 02). Here, as Ali (2016) states, English language assessment principally follows summative assessment, "where the learners' abilities in memorization and comprehension are assessed" ignoring the assessment of learners' listening and speaking skills (p. 01). There is significant gap between the objectives of curriculum and realities of assessment. Except in some renowned institutions, English examination is mainly based on assessing the skills of reading and writing, paying no heed to speaking and listening skills (Ali, 2016, p. 01). Regrettably, we have "failed to put into place an assessment mechanism that would assess all the four skills of learners (Rouf & Mohammed, 2018, p. 52)". Being a skill-based subject, not knowledge-based, "all learners need to develop adequate skills in the target language (TL) so that they can use the language in real-life situations (p. 52)".

English is still used in Bangladesh as the foreign language. In the statement of the *National Curriculum-2021*, the objective of English is declared as follows: "To be able to communicate effectively using basic skills of the English language for day-to-day purposes, academic purposes and other specific purposes; to be able to exert creative as well as critical insights to express aesthetically, and to appreciate English literary text; to be able to uphold democratic practice in communication at the individual, social, national and global contexts (p. 41)". That is, the National Curriculum puts emphasis on attaining communicative, creative and ideological skill and consciousness through English. The National Curriculum upholds that learning process will ensure effective communication in English, which requires contextual understanding, cultural awareness and democratic attitude that leads to the idea of Intercultural Communicative Competence (ICC) through which "the core knowledge and skills are mediated for the applications in three major areas which are- real-life application, sense of identity, and creative expression (p. 42)".

In spite of being a compulsory subject for the students of all grades in Bangladesh, most of the learners "cannot develop the ability to use English appropriately in real-life communication even after completing twelve years of education" and in the undergraduate level "they lack communicative competencies which were the main concern of the CLT approach" (Nayeem *et al*, 2020, p.25). Our national curriculum puts emphasis on the communicative competence of the learners. But "the aims and objectives of the communicative curriculum can never be achieved without a testing system that assesses the communicative ability of learners (Haider, 2008, p. 47)".

Testing and the Criteria for Good Testing

Language testing is one of the elementary areas of applied linguistics which “tackles two of its fundamental issues: the need to define and reflect on the appropriateness of Second Language Acquisition models and constructs through data analysis from language tests, and the importance of facing the ethical challenge deriving from the social and political role language tests play nowadays (D’Este, 2012, p.61)”. According to Brown (2004) test refers to “a method of measuring a person’s ability, knowledge, or performance in a given domain (p. 03)”. Bachman (1995) opines that a test is “a measurement instrument designed, to elicit a specific sample of an individual’s behavior (p. 20)”. Some key terms related to testing are validity, reliability, practicality, authenticity and washback which will be dealt with in the present study as the prime criteria for good language testing.

Validity

Though being conceived as an ‘abstract’ concept by Giraldo (2020) validity “is the most fundamental quality of testing systems, across social, professional and educational contexts (p. 195). Messick (1989), the widely recognized validity expert, defines validity as, “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences actions based on test scores or other modes of assessment (p.11)”. To Lado (1961) validity is a “matter of relevance”; a test becomes valid if “test content and test conditions are relevant (p.321)”. Kane (2016) defines validity as the “extent to which the proposed interpretations and uses of test scores are justified (p. 1)”. According to Brown (1996), validity is “the degree to which a test measures what it claims, or purports, to be measuring (p. 231)”. A valid test never measures “irrelevant or contaminating variables (Brown & Abeyickrama, 2019, p. 36). Validity can be classified into several categories. *Content Validity*: it upholds that the contents of the test items should be appropriate and show that “they are a good example of a universe in which the investigator is interested (D’Este, 2012, p. 63)”. Sample of the test should be proper. A test bears content validity “if its content constitutes a representative sample of the language skills, structures, etc. with which it is meant to be concerned (Bachman & Plamer, 1996, p. 19)”.

Construct Validity: According to Bachman (1995), construct validity is “the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs” (p. 255). Messick (1975) shows construct validity as “a measure estimates how much of something an individual displays or possesses (p.957)”. *Score Validity*: In a test the way of scoring the responses must be valid (Bachman & Plamer, 1996, p. 19). For example: In a dictation method learners’ listening skill is usually tested. But, if, for instance, the learners’ error in spelling is taken into account, the test will lose score validity. *Face Validity*: A test is claimed to have face validity “if it looks as if it measures what it is supposed to measure (Bachman & Plamer, 1996, p. 19)”. For example, a test that attempts to examine learners’ pronunciation skill but which does not make the test taker speak, lacks face validity of that test and the candidates, teachers, education authorities or employers may not accept the test as the result does not reflect the candidates’ skill truly (ibid). *Predictive Validity*: A test possesses predictive validity “if the results accurately indicate whether or not the test taker will be able to perform a specified ‘real life’ task outside the teaching/testing context (Bachman & Plamer, 1996, p. 19)”. That is, for example, in language testing predictive validity will signify whether the result will indicate the learners’ expected performance “in terms of the level of language competence (p.19)” in real life.

Reliability

A test which is reliable is “consistent and dependable (Brown & Abeyickrama, 2019, p. 30)”. A test which is unreliable is worthless. In a test measurement should be consistent and score should be stable. Reliability refers to “the extent to which a test measures what it purports to measure consistently and accurately (Chiedu & Omenogor, 2014, p 4)”. Brown (2001, p. 386) makes it easier by stating that “If the same group of subjects takes the same test on two different occasions, results should be similar, both in individual scores, and in the rank order within the group (cited in Bachman & Palmer, 1996, p. 20)”. If a test is administered to the same candidates on different occasions, then, to the extent that it produces different results, it is not reliable (Heaton, 1976, cited in Sultana, 2015). Reliability implies that “If the same written answer in a test is scored by two different markers, the two different scores should be similar (Bachman, 1990, p. 24, cited in Bachman & Palmer, 1996, p. 20)”.

Practicality

Practicality of testing is related with administrative and other practical aspects. As Brown & Abeyickrama (2019) observe, practicality signifies “logistical, down-to-earth, administrative issues involved in making, giving and scoring an assessment instrument (p. 28)”. A test is claimed to have practicality if it is easy to design, administer, score and does not need wasting too much time. If a test-taker requires five hours to complete it, the test becomes impractical because it uses more time than available (Brown & Abeyickrama, 2019, p. 28). If a test is excessively expensive to administer, it lacks practicality. Evaluation procedure of the response should also be direct and time-efficient. If there is complicity in the evaluation procedure, that very test is also impractical.

Authenticity

In language testing authenticity has been a chief concern for long (Hoekje & Linnell, 1994, p.110). As the communicative language testing flourishes, authenticity has become “the center of many test researchers’ concerns (Zheng & Iseni, 2017, p. 13)”. In respect of evaluating and designing tests the

concept of authenticity is “a little slippery” to define (Bachman & Palmer, 1996). Authenticity as Bachman (1990) posits, signifies that “tests should be designed to reflect real life situations, or in other words, the task characteristics should be consistent with the features of the real-life situations in which the target language is used” (cited in Luo, 2019, p. 161). Brown & Abeyickrama (2019) opine that an authentic task contains natural language, consists of contextualized items and uses interesting topics (p. 39). These things must be reflected on an authentic test.

Washback

There is crucial relation between washback and validity. Alderson and Wall (1993) opine that “a test is considered valid if it has a positive washback and invalid if has a negative washback (cited in Hinai & Jardani, 2021, p. 69)”. Washback implies “the influence or reaction of testing” on teaching-learning practices in the arena of education (Huang, p. 553). Prodromou (1995, p.13) argues that backwash refers to “the direct or indirect influence of testing on teaching methods (cited in Huang, 2019 p. 553)”. Mcnamara (2003, p.73) holds the view that the backwash effect of testing signifies “the impact of testing on teaching (cited in Huang, 2019, p. 553)”. According to Messick (1996) washback is “the extent to which the text influences language teachers and learners to do the things that they would not necessarily otherwise do (p. 243)”. The effect of washback can be either beneficial (positive) or harmful (negative)” (Buck, 1988, p. 17). Roghmal (2019) shares that according to the scholars, testing which has effect based on its nature, “must drive the instruction” and a test which is high-stake “affects a national level policy and curriculum, while an achievement test affects a specific course (p. 1447).

Research Questions

Based on the analysis of the contents of SSC and HSC English questions set under the prescribed national format the research will try to answer the following questions:

- a. To what extent do the English questions of the public examinations (SSC & HSC) fulfill the criteria for being a good test?
- b. What are the reasons behind the mismatch between the criteria for good testing and the reality in the language testing field in Bangladesh?

Validity in SSC and HSC English Test

Though based on the practical reality it is true that due to some practical reasons (e.g., lack of skilled and trained teachers and resources) a good number of learners cannot complete the syllabus with proper orientation with the contents of the syllabus. But, usually, the items of SSC and HSC English questions represent the materials included in the syllabus. It means that the SSC and HSC English examinations are valid in terms of content validity. The syllabuses of the SSC and the HSC English undoubtedly follow the communicative approach, and, thereby, it (the syllabus) is supposed to put emphasis on learners’ achieving the four skills of the English language: reading, writing, speaking and listening. Communicative syllabus talks in favour of these skills uniformly. But what happens in the SSC and the HSC examination is very significant to discuss in terms language testing. Only reading and writing skills of the EFL learners (examinees) are tested, ignoring speaking and listening skills. Therefore, in both the tests the objective of the curriculum is reflected partially. Here remains considerable gap between the curriculum and the test. In this circumstance, how far the SSC and the HSC English tests are valid becomes doubtful.

Brown and Abeyickrama (2019) argue that a valid test does not measure irrelevant variable (p.36). But what happens to a typical examiner of Bangladesh is very awful. Here, due to various practical and personal issues, very often the examiner resorts to the size of a distinct answer to check and score especially the writing part, ignoring comprehensibility, organization of ideas and other necessary things. In our pedagogical culture we usually experience that in the classroom the teacher encourages the learners to make the size of the answer large. Learners are convinced by the words that if the size of the answer is large, the examiner will definitely score high in spite of having irrelevant or wrong ideas and information inside the body of the answer written in the scripts. And obviously there is practical ground behind this assurance given by the teacher. The examiner truly sees a large-sized answer with special eyes which makes the scoring procedure invalid because considering only the size of the answer ignoring other necessary things cannot confirm the fulfillment of validity of the testing.

Communicative Language Teaching (CLT) puts significant emphasis on learners’ linguistic performance. An EFL learner has to deal with the tasks which possess close proximity with the real-life performance. For example, in the SSC syllabus learners have to learn how to perform dialogue. They have to write the dialogue in their answer script. It means that this is not a part of speaking test; rather, it is included in the writing test. Consequently, the EFL teacher rarely feels it necessary to make the learners practice dialogue in the classroom. The teacher is not strongly motivated to create the context of learners’ dialogue practice in the classroom. The learners have to *write* a dialogue about a given topic, instead of performing it. They memorize the basic structure of a dialogue with the help of their teachers and guide books. Then they simply vomit it on the pages of the answer scripts. Thus, there remains discrepancy between the test and what is expected in the curriculum. It makes the test invalid because dialogue requires performing through speaking which is absent in the test.

Reliability in SSC and HSC English Test

A reliable test uses unambiguous items and tasks. This precondition is almost ensured in the English question of SSC and HSC. Question items and tasks are undeniably explicit and clear. There is also clear direction for scoring. But here lies something very weird and alarming too. In spite of having clear

direction for scoring the answer scripts, the examiners are orally instructed to be very liberal in scoring. Before distributing the answer scripts to the respective examiners, usually a meeting is held in the offices of the examination boards where the examiners are verbally instructed to give marks freely as much as possible despite, in reality, the lower standard of that very answer script. Consequently, the examiner is forced to offer excessive marks to the undeserved.

Again, the examiner has to follow the instruction that scoring should be done on the basis of the performance level of the learners living the privileged area, i.e., town/city areas, where learners perform better because of the smooth availability of academic resources. This system makes the examiner bound to offer excessive marks to the comparatively underprivileged learners who enjoy less availability of academic resources, and, thus, perform bad in the examination. As a result, the test-takers who perform deficiently in English tend to get improperly extra marks hiding their real level of competence. Accordingly, scoring becomes unreliable. When the overall national standard of English gets published, we remain blind of knowing the reality of linguistic competence of the EFL learners of Bangladesh.

Administration reliability is another reliability issue which is very significant in the language testing discussing arena of Bangladesh. The physical and psychological condition of the average learners in Bangladesh is not always supportive to cope up with the mega event like public test, i.e., SSC and HSC. Of course they are mega events in the country. During those examinations the situations on the roads like traffic jam, long distance of the examination center from the test-takers' residence, lack of comfortable journey etc. create negative influence in the mind of the test-takers. Along with these issues, the test-takers are burdened with huge expectations of the guardians, teachers and other well-wishers to do the best result. All these issues create unrest and psychological imbalance in the mind of them which consequently cause the test-takers' momentary sickness, exhaustion, nervousness and other physical, psychological issues, and thus, affect their performance in the test. Therefore, there remains gap between the test-takers' real knowledge and their actual performance in the examination hall. It makes score deviate from the true score. In almost all the examination centers in our country the reality described just now is almost same. There are some more issues like chronic noise from the near streets, faulty state of photocopying, lack of proper temperature and lighting in the examination room, uncomfortable desk and chair etc. which put very negative impacts on the mind of the test-takers in the examination hall, only to make the scoring unreliable.

Inter-rater reliability is another important concern to discuss about the test reliability of SSC and HSC English questions. If a test is claimed to have reliability, it will express stable result across two or more test-givers. Different scores by two examiners (raters) due to the lack of adherence to scoring criteria, inexperience, bias, inattention etc. create doubt in the reliability of scoring. Usually, SSC and HSC answer scripts are checked by single examiner/scorer. So, naturally, we cannot find the issues like score-difference between two scorers which signifies inter-rater reliability. But the reality of the whole testing procedure centered on the examiner suggests that if the system of two scorers is introduced in the testing of SSC and HSC English, there will be found significant gap between two scorers. This will happen because for checking and scoring the English scripts of the public examination there are no such rubrics which are practical to follow in Bangladesh. There are some objective and fixed official advices and suggestions for checking and scoring given by the experts but implementing those are rare in our context since individual examiner usually goes through ample activities and psychological restlessness created by the pressure of the examination boards to submit result hurriedly as well as continuing his/her regular activities as a teacher.

Related to the scorer's point of view there is another aspect of reliability which is called intra-rater reliability. There may be inconsistent evaluation by the same examiner caused by fatigue or bias, or other physical and psychological issues. In this case, first few tests will differ from the last few papers due to tiredness or fatigue. In Bangladesh, this is quite common a phenomenon. The multiple faces of practical reality surrounded by the examiner as language teacher always make him/her go through unstable psychological disposition which causes instability in marking/scoring. Usually, the examiner starts checking first few scripts with full enthusiasm maintaining scoring criteria properly in a positive mental orientation. But, with the course of time, due the pressure of finishing checking all the allotted scripts in the given time-frame, pressure of maintain his/her regular duties as teacher, unfavorable family affairs etc. he/she cannot proceed to check the scripts with same enthusiasm and confidence. This is quite true in our context because here teachers belong to one of the under-privileged professional groups with insufficient salary, excessive engagement in various activities, meager remuneration for script checking etc. While going through checking the scripts if it comes to the mind of the examiner that he/she will get very insufficient remuneration for this menial job, not only that, the remuneration comes to hand even after one year, he/she will definitely lose interest in doing this job and ultimately the scoring process will be unsteady and questionable in the course of time.

A last point that makes the reliability of SSC and HSC English testing questionable is related to the question-contents memorized prior to the examination by the learners and being common. Being common and repeated of the question topics especially in the writing part encourages the learners' memorization. Learners typically resort to guidebooks and short suggestions available in the market. A good number of question-contents, especially from the writing part, for example: paragraph, formal/informal letter, composition, story writing, describing chart/graph etc., which have previously been memorized by the test-takers, become common in the examination. In this case, showing learners' creativity and own skill in writing is rare because of the memorized items. But, regrettably, they get very good score by the examiner, whereas, the test-taker who writes creatively with the help of his/her own ideas and knowledge, and if he/she writes a bit shorter and less competently than the test-taker who has memorized the item, the first one will get poor marks. For example, this year (2023), in the HSC question of English "City Life vs. Rural Life" was the common topic to write a paragraph based on the technique of comparison and contrast both in the Mymensingh Board and the Cumilla Board. This topic is very common to write a comparison and contrast paragraph available both in guidebooks and online platforms. Thus, in this process, genuine linguistic skill cannot be judged and memorization triumphs over creative skill which leads to unreliable scoring.

Practicality in SSC and HSC English Test

Utilizing available human resources for administering the test in smooth and better way is a part of practicality of the test. In Bangladesh the education boards do not always aptly utilize the available human resources. Consequently, sometimes, testing procedure and tools become faulty. For instance, this year (2023) in the HSC question of English first paper of the Dhaka Board there are a good number of grammatical and spelling errors which signify that there was laxity behind utilizing time and the experts to check the quality and standard the question paper. We have a sufficient number of national ELT experts who are experienced and skilled in setting and moderating questions. But the education boards never call them to let them have at least a bird's eye view to the question to check the standard of the contents. Putting it aside, even if the education boards properly check and recheck the question contents always, it becomes possible to ensure an error-free question papers.

The evaluation procedure of SSC and HSC examination is specific but not time-efficient. The number of days (on an average, it is ten to fifteen days) fixed for completing the evaluation task given to an examiner is not sufficient to score an answer script of 100 marks. This year (2023) in the Cumilla Board, one of my colleagues got seven hundred English answer scripts of 100 marks to check and evaluate. He was instructed to finish the evaluation procedure within fourteen days, though he failed and later on he requested for additional four days. He had to do hurry to submit the answer scripts to the concerned office on time, and consequently, evaluation procedure lacked meticulousness and precision. This is quite true to other examiners who go through the same restless process of evaluating the answer scripts; whereas, the contents of English question paper require much time for scrupulous scrutiny while scoring the answer.

Authenticity in SSC and HSC English Test

Discussing authenticity in SSC and HSC English Test is quite noteworthy. Maintaining this criterion while setting English question paper remains almost missing. It is true that the term 'authenticity' itself bears ambiguity. But, at least, we can include interesting topics and contents in English questions so that the test takers are a bit psychologically relieved in the examination hall. Instead, politically biased contents, which fail to induce any interest in the mind of the test-taker, are being chosen for questions. The test items do not completely reflect real life situations. There are passages for reading test containing some communicative contents, but, the language used there is almost far away from the real-life language. Language should be more natural and contextualized to ensure authenticity of the test. There should be items for listening and speaking test. But, unfortunately, in the public examinations of Bangladesh, only two skills of the EFL learners, reading and writing, are tested. Speaking and listening skills are ignored, thus the authenticity criterion of these tests is in question. If we want to ensure authenticity in language testing we have to make sure interaction and communication in the test materials and procedures. Regrettably, for some practical as well as other related reasons, this is not possible in SSC and HSC English examinations.

Washback of SSC and HSC English Test

To ensure positive washback the contents of the testing materials should be course objective-based. If the contents of the testing materials are unrepresentative and class lesson-based there will be negative washback of the test. In the English public examinations of Bangladesh, if we consider objective-based testing materials, we will see that, a large portion of the course objective is quite absent there. Here testing speaking and listening skills is absent. They are not included in the public examinations, and thus, "a strong negative washback effect is likely to result from the absence of listening and speaking papers, weakening incentives for teachers to practise these skills in class (Rahman *et al.*, 2021, p. 28)". Truly, the EFL teachers do not feel encouraged to teach and the EFL learners to learn how to develop English speaking and listening skills if these two essential skills are ignored in the public examinations. These tests do not influence both the teacher and the learners to learn English aftermath. This negative washback makes the whole testing procedure invalid.

Conclusion

In a country like Bangladesh, that still offers less financial and logistical supports for pedagogical development, the overall standard of language testing is very difficult to improve. As implementing the best language testing requires financial, administrative, social as well cultural, along with intellectual endeavour, the government as well as other responsible wings including language policy makers, should offer extra attention and care with proper budgetary issue to ensure the best implementation of the best language testing. Otherwise, the objectives of Communicative Language Teaching (CLT), which are still far away from the reach-point, will never be achieved entirely.

Works Cited

- Ali, M. (2016). Assessment of English Language Skills at the Secondary Level in Bangladesh. *The Journal of EFL Education and Research*, 1(1), 1-7. www.edrc-jeffler.org
- Bachman, L. F. (1995). *Fundamental Considerations in Language Testing*. Oxford : Oxford Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in Practice: design and developing useful language tests*. Oxford: Oxford University Press.
- Brown, D. H. (2004). *Language Assessment: Principles and Classroom Practices. (4th ed.)*. New York: Longman.

- Brown, D.H. & Abeyickrama, P. (2019). *Language Assessment: Principles and Classroom Practices*, (3rd ed.). New Jersey: Pearson.
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.
- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10(1), 15–42.
- D'Este, C. (2012). New views of validity in language testing. *ELLE*, 1 (1), 61-76.
- Giraldo, F. (2020). Validity and Classroom Language Testing: A Practical Approach. *Colomb. Appl. Linguistic*, 22(2), 194-206. doi: <https://doi.org/10.14483/22487085.15998>
- Haider, M.Z. (2008). Assessing “Communicative” Writing Skills: An Evaluation of the SSC English Examination. *Bangladesh Education Journal*, 7 (2), 47-62.
- Hinai, M. K. A. & Jardani, K. S. A. (2021). Washback in language testing: An exploration with a focus on a specific EFL context in Oman, *International Journal of English Linguistics*, 11(1), 68-74. Doi:10.5539/ijel.V11n1p68
- Hoekje, B. & Linnell, K. (1994). “Authenticity” in language testing: Evaluating spoken language tests for international teaching assistants, *TESOL Quarterly*, 28 (1), 103-126.
- Huang, X. (2019). The backwash effect of language testing on professional English learning and teaching, *Advances in Social Science, Education and Humanities Research*, 311, 553-558.
- Kane, M. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2), 198-211. <https://doi.org/10.1080/0969594X.2015.1060192>.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman.
- Luo, L. (2019). Pursuing authenticity in ESP testing: The need for interdisciplinary collaboration, *The journal of teaching English for specific and academic purposes*, 7(2), 159-169. <https://doi.org/10.22190/JTESAP1902159L>
- Messick, S. (1975). ‘The standard problem: meaning and values in measurement and evaluation.’ *American Psychologist* 30, 955–966.
- Messick, S. (1989). Validity. In R. Linn (Ed), *Educational Measurement*(pp. 13-103). New York: Mcmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. <https://doi.org/10.1177/026553229601300302>
- Nayeen, J. et al. (2020). Testing Communicative Language Teaching (CLT) through English for Today (EFT) in Bangladesh: Challenges Faced by Tertiary Students Initially. *American International Journal of Education and Linguistics Research*, 3(2), 1-27.
- NCTB (2021). *National Curriculum Framework-2021*. Dhaka, Bangladesh: Ministry of Education
- Rahman et al. (2021). English language assessment in Bangladesh today: Principles, practices, and problems, *Language Testing in Asia* (2021) 11(12), 1-23. <https://doi.org/10.1186/s40468-021-00129-2>
- Roghmal, S. (2019). The Washback and Impact of Test on Language Teaching and Learning, *International Journal of Science and Research*, 8(9), 1447-1451. 10.21275/21091903
- Rouf, M. A. & Mohammed, A. R. (2018). Teaching English at Secondary Level: Curricula Directions and Classroom Scenario. *Jagannath University Journal of Arts*, 8(2), 39-56.
- Sultana, N. (2018). Test review of the English public examination at the secondary level in Bangladesh. *Language Testing in Asia*, 8 (16), 2-9. <https://doi.org/10.1186/s40468-018-0068-1>
- Sultana, R. (2015). Reliability of the Currently Administered Language Tests in Bangladesh: A Case Study, *Journal of Literature, Languages and Linguistics*, 17. www.iiste.org
- Zheng, Y. & Iseni, A. (2017). Authenticity in Language Testing, *Anglisticum Journal*, 6(8), 9-14.