Inconsistencies and Biases in English Language Testing in Bangladesh Civil Service Preliminary Examination

Zohur Ahmed[1,2]*, Shahin Sultana[3] and Kaur Manjet[1]

[1]*School of Languages, Literacies and Translation, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia*
[2]*Department of English, East West University, 1212 Dhaka, Bangladesh*
[3]*Department of English, Goa University, 403206 Goa, India*

ABSTRACT

Bias and unfairness of test items in language proficiency tests have been an issue of scholarly investigation. Unfair test items and cultural biases in international English language proficiency tests such as TOEFL and IELTS have received considerable attention. However, such unfairness and biases in proficiency tests conducted at the national level such as Bangladesh Civil Service examination (henceforth referred to as BCS) have not been investigated. This paper attempts to examine item biases and inconsistencies in English language testing of BCS preliminary examination. Qualitative data has been collected from stakeholders using a semi-structured interview technique. A detailed content analysis of the English section of the BCS examination is conducted. Most test items have been found to be biased towards testing knowledge of English literature which is inconsistent with validity constructs of the language proficiency test. Stakeholders have been found to perceive the test as ineffective in measuring their English language proficiency.

Keywords: Bangladesh civil service, bias, English proficiency test, validity

INTRODUCTION

Bangladesh Civil Service examination is one of the largest public examinations conducted by Bangladesh Public Service Commission for recruiting government civil servants. The fact that jobs in Bangladesh Civil Services are regarded as highly prestigious makes this examination fiercely competitive. Only 10 in every 10, 0000 make it to the final selection (Debnath, 2017). Within this context, this paper

attempts to examine the test validity of the existing civil service examination in Bangladesh in the light of English language proficiency testing. Firstly, this paper introduces civil services in Bangladesh and its recruitment and selection processes. Secondly, it interrogates the validity of the English language testing section in the BCS preliminary examination. Thirdly, it provides a description of data instruments. This is followed by an analysis and findings of the collected data. Finally, it moves to the conclusion of the study.

Language testing has been regarded as one of the core areas of applied linguistics partly because data analysis from language tests help in defining and reflecting on the appropriateness of Second Language Acquisition models. Also, because it is important to address the ethical issues coming from the social and political roles played by language tests (D'Este, 2012).

Fulcher (1999) argued that language testing thus had a twofold impact in a variety of contexts. In the first instance, it constitutes a scientific impulse for which research is needed to provide an accurate measure of precise abilities. Secondly, language testing has also become a subject of debate because the use and interpretation of test results introduces ethical issues concerning the concept of 'fairness' in the construction, administration, evaluation, and interpretation of language tests: the powerful effect of the '(mis)use' of the test that might have harmful-unintended or intended-consequences for test-takers or society.

Language tests have been categorized in terms of purposes for test uses such as diagnostic, placement, achievement, and proficiency tests (Hughes, 2003). According to the definitions by Hughes, the English section of the BCS preliminary test can be termed as a language proficiency test since the test does not follow any specific content from academic teaching/learning contexts. Proficiency tests are constructed and administered throughout the world for several purposes (Mustafa & Anwar, 2018) including granting immigration to foreign nationals (for instance IELTS general module), awarding citizenships to immigrants (for instance USCIS), issuing licenses for professional practices (for instance TOEIC, PLAB) and for granting students access to higher education (for instance IELTS academic module, TOEFL). However, the BCS preliminary test does not assign any fixed pass mark. The main purpose of the BCS English preliminary test seems to be eliminating candidates through test scores so that the elimination process appears to be justified.

BCS Recruitment and Selection Process

The Bangladesh Civil Services examination is a nationwide competitive examination for recruiting civil service cadres including admin, taxation, foreign affairs, and police among others. According to Ahmed and Khan (as cited in Jahan, 2012), BCS is structured into four classes, namely class-1 class-2, class- 3, and class-4. All class-1 and part of class-2 officers are treated as 'gazetted' officers, the rest are considered

as 'non-gazetted' officers. The gazette officers are provided with power, prestige, responsibilities and consequently enjoy greater opportunities.

Elaborated recruitment rules are set up for selecting the cadres of BCS. To be eligible for appearing at the BCS examination, an applicant must be a Bangladeshi citizen and a graduate from a university. The age limit is 21 to 30 years for the general cadre service and 21 to 32 years for the health care service (Khan & Ara, 2005).

Eligible candidates sit for an MCQ preliminary test of 200 marks. Upon qualifying the preliminary test, candidates need to sit for a written test. The written test consists of 900 marks. To go to the next step, the *viva voce*, a candidate has to secure an aggregate of at least 45% marks in the written examination and 40% in the viva voce.

The BCS preliminary test carries 200 marks comprising 200 MCQ questions each carrying 1 mark for the correct response and negative marking of 0.5 for the wrong response. Each question provides a 'correct answer' or in testing terminology a 'key' and three 'distracters' commonly known as 'options'. Marks distribution of the test is given in Table 1.

Literature Review

The literature concerning 'validity' in language testing, test biases and Bloom's taxonomy will be discussed in the following section to situate the study in a broader conceptual framework.

Validity in Language Testing. In 1961, Lado provided the first significant contribution by applying the term 'validity' to language testing. He conducted his research on the basis of a question-statement which summarized his concept of validity: Does a test measure what it is supposed to measure? If it does, the test is considered valid. Validity is thus regarded as one of the most important qualities of a language test. Validity is a matter of relevance, a test is considered valid when test content and

Table 1
Marks distribution of the BCS preliminary test

Subjects	Marks
Bangla Language and Literature	35
English Language and Literature	35
Bangladesh Affairs	30
International Affairs	30
Geography, Environment and Disaster Management	10
General Science	15
Computer and Information Technology	15
Mathematical Reasoning	15
Mental Ability	15
Ethics, Values and Good Governance	10

test conditions are relevant and there are no irrelevant problems which are more difficult than the problem being tested (Lado, 1961).

Validation is agreeably an important procedure to certify a test. The validity of a test refers to whether or not it measures what it claims to measure (Hughes, 2003). For the certification and recruitment tests, the test components must be highly correlated to the particular requirements. A test with poor validity does not measure the profession-related content and necessary competencies. In such situations, there is no approval for using the test for the purpose. Since theoretical constructs such as reading ability, fluency in speaking, control of grammar, and writing ability are essential components of valid language proficiency tests, the term 'construct validity' is used to mean 'validity' in general. Hughes (2003) and Bachman (1990) discussed several categories of the validity construct in the following paragraphs.

Content Validity. When a test includes a representative sample of the language skills, structures, functions related to a particular purpose for which the test is conducted, the test is said to have content validity. It includes a principled selection of test items. Judgments regarding the content validity of a test are ideally made by people with knowledge of language teaching and testing who are not directly concerned with test production. A test with higher content validity is more accurate than the one with lower content validity.

Criterion-related Validity. Criterion-related validity refers to the degree to which scores received by a candidate agree with scores from other independent assessment of the candidate's ability. Such an independent assessment is used as the criteria measure to validate the test. Criterion-related validity is understood through two kinds which are predictive validity and face validity.

Concurrent Validity. Concurrent validity is established by administering the test and the criteria are at the same time. Concurrent validity is a method that uses correlation. For the purpose of practicality, when an alternative and independent shorter version of testis administered to check if scores in the shorter test can result in 'a high level of agreement', the test is said to have concurrent validity. Once a test has been scored, the relationship is assessed between the examinees status as either pass or fail based on the test scores. This validity provides evidence that the test is classifying examinees correctly.

Predictive Validity. This method is similar to concurrent validity. it measures the relationship between examinee's future performance and the test scores obtained. An example could be how accurately a proficiency test can predict a candidate's future performance in terms of language use in the target domain for which the test is administered. This type of validity is especially useful for recruitment and admission tests.

Face Validity. Face validity of a test is achieved when a test looks as if it measures what it claims to measure. A pronunciation test that does not involve test takers to speak is an example of a test without face validity. However, a test with higher face validity does not necessarily ensure better measurement of the target skills. Unlike content validity, this validity is not measured through formal settings and also not determined by subject experts. Instead, anyone who supervises the test, including candidates, teachers, education authorities or employers, may develop an informed opinion as to whether or not the test is measuring what it is supposed to measure.

Bloom's Taxonomy. The taxonomy of cognitive domain which is widely considered to be an effective framework for categorizing conceptualization of thinking was suggested in the 1950s by an American educational psychologist Benjamin Bloom. The taxonomy has been used by several researchers from diverse disciplines to evaluate test items (see Ebadi (2015) for a comprehensive review) Bloom classified the cognitive domains into six hierarchical levels namely *knowledge, comprehension, application, analysis, synthesis, and evaluation*.

Level one, *knowledge*, deals with test-takers' ability to recall and remember information such as specific facts, dates, events, names of persons, principles, theories, etc. Level two, *comprehension*, involves demonstration of understanding including interpretation, inferences,

explanation, estimation, prediction, and translation of one symbolic form to another. Level three, *application*, assess the ability to use abstract ideas, rules, or methods in actual situations. Level four, *analysis,* focuses on one's capacity to break down an entity into its components and to understand the relationship among different constituents. Level five, *synthesis*, deals with competence in organizing and assembling constituent components into new patterns or structures. Level six, *evaluation*, assess the ability to judge the quality, merit, or value of something according to given criteria.

Evaluation of English Language Proficiency Tests in Bangladesh. In order to understand the existing depth of knowledge base in the area of English language proficiency testing in Bangladesh, several online databases of specific journals dedicated to language testing and assessment were searched. More general and frequently consulted databases such as JSTOR, Google Scholar, ERIC and Scopus were also searched with search terms such as "language testing in Bangladesh", "testing language proficiency in Bangladesh", "biases in language testing in Bangladesh". A number of search attempts returned several studies most of which deal with achievement tests conducted in the context of secondary, higher secondary and tertiary level English education in Bangladesh. Despite the fact that English language proficiency is tested in almost all executive-level recruitments, in our online search, we have not found any empirical study addressing assessment issues

in any of these English language proficiency tests administered by the government and non-government recruitment agencies in Bangladesh. Evidently, English language proficiency tests conducted in recruiting professionals in Bangladesh have mostly remained unexplored. Although high-stake tests are common in Bangladesh (for instance Secondary School Certificate Examination, Higher Secondary School Certificate Examination, Teacher Recruitment Examination, University Entrance Examination and others), there is a dearth of sufficient information about the guiding principles of these tests, their design and construction (Ali et al., 2018).

We have found only one study by Khan (2006) which investigated the issue of test bias in administering IELTS speaking test in Bangladesh. The need to involve stakeholders, particularly in language assessment, has been emphasized in discussions on cultural biases in testing. Khan (2006) urged to consider ways of minimizing cultural biases of language tests and suggested that it was essential to seek more awareness of such issues among the test designers. Findings show that there are a number of culturally inappropriate topics, vocabulary items and phrases that are alien to candidates. In addition to creating a stressful test environment, these items add to undesirable test anxiety among test takers. Thus, examiners in this study are reported to have avoided certain topics and refrained from asking problematic questions.

Language Assessment Practices in Bangladesh. In the existing relevant literature, public examinations conducted by Secondary and Higher Secondary education boards across the country have been reported to be unsuccessful in matching curriculum goals and test formats (Ali & Sultana, 2016; Das et al., 2014; Haque, 2016; Sultana, 2018). Studies investigating nature of assessment have reported a tendency among test setters for assessing students' lower-order skills ignoring higher-order skills (Sultana, 2018), teachers' preferences for close-ended questions (Rahman et al., 2011) and content-driven assessment practices (Banu, 2009).

Highly 'ritualistic' practices have been found among the test setters for test design, construction, and development (Ali et al., 2018). It has been argued that such practices compromise the necessary qualities of test validity and reliability. Studies have also reported a high degree of the negative washback effect exerted by secondary school examination in Bangladesh. Teachers have been reported to have been pressured by the headmaster and other stakeholders to teach for the test (Al Amin & Greenwood, 2018).

Information regarding necessary qualifications and expertise required to become language test setters for the BCS and other public examinations is unavailable. Although teachers often design, construct and administer language tests, they receive no in-service training in language testing and have no academic and professional knowledge about language assessment (Sultana, 2019).

Fairness in English Language Proficiency Tests. A large and growing number of studies in the field of language testing have pointed out to considerable controversies around the fairness of tests used for selection decisions in job recruitment (Auer, 2018; Diamond et al., 2012; Knoch et al., 2015; Kim & Elder, 2015). Traditionally, concerns have been raised primarily around the cultural, ethical and validity biases at the test format and item levels (Cronbach, 1971). The concern of a more significant nature is what to do with bias once it has been detected. Should 'biased' items be eliminated, neutralized, completely replaced by fair items, or ignored as representative of the same necessary target behaviour? Certainly, to give an answer to these questions, a number of studies have been carried out (Chen & Henning, 1885; Djiwandonos, 2006; Vijver & Tanzer, 2004).

If such crucial issues in a test are not treated carefully, the tendency of bias found can be high (Vijver & Tanzer, 2004). To address the problems of these biases, Djiwandonos (2006) proposes methods such as Item Response Theory (IRT) and Confirmatory Factor Analysis (CFA) which generally employ a statistical approach to detect biased items in a test. After biased items are detected the items are revised or removed from a test. If biases in a test have been taken care of, the aim of a test at measuring the proficiency of test takers regardless of their cultural and educational background can be achieved.

We assume that much like what happens with achievement tests investigated in the above studies, inconsistencies may exist between principles of language testing and the test items used in the English section of BCS preliminary examination. We also assume that test makers for these examinations have not been able to incorporate the fundamental constructs of language proficiency testing.

Research Questions

The study addresses the following research questions:

i. What is the nature of bias in the test items used in English language proficiency testing in BCS preliminary test?

ii. To what extent does the English section of the BCS preliminary test conform to the validity constructs of language proficiency tests?

iii. What are the perceptions of the test takers regarding the effectiveness of the English test in BCS preliminary examination?

METHODS

Both qualitative and quantitative approaches were used for data collection and analysis. The semi-structured interview method was used to obtain stakeholders' perceptions regarding the test. A content analysis of the test items from previous test papers was performed. Hughes' (2003) and Bachman's (1990) category of the validity constructs was used to examine the validity of the test. Bloom's Taxonomy (1956) was consulted to identify the cognitive domains that the test was capable of assessing.

Data Collection Procedures

Four examination papers of the previously conducted preliminary examination were collected from candidates who had appeared in the examination. A semi-structured interview was conducted to collect former test-takers' and aspiring candidates' opinions regarding the effectiveness of the test. The interview method was used in this paper to provide supporting and supplementary information on BCS recruited class-1 officers' (the former test-takers', O1-O5) as well as the aspiring candidates' (C1-C5) opinions and perceptions concerning the English section of BCS preliminary test. In-depth, supplementary information is obtained through the interview. Semi-structured interviews are used to enable the interviewees to freely express their opinions about the existing preliminary English language proficiency test. According to Nunan (1992), a semi-structured interview gives the interviewees full control and power to handle questions in free and flexible environments.

The interview was individually conducted with each respondent in English. The interviewers (the second author of the article) used English to conduct individual interviews with the respondents. However, the interviewees could choose either Bangla or English to respond to. The interviews were tape-recorded and later transcribed and analyzed. Strict confidentiality regarding the identity of the interviewees was maintained to ensure authentic and genuine responses.

Data Analysis

Analysis of the Examination Papers. For evaluating the test items, a content analysis (Krippendorff, 2018) was performed. The test papers for the 35th, 26th, 37th and 38th BCS preliminary examinations were closely read several times to identify emergent themes and patterns in them. Several categories emerged which are defined and exemplified below:

English Literature (EL). Items asking for factual information regarding English literature such as identifying the author's name, genre, quotations, and literary periods. For instance, "To be, or not to be, that is the question- is a famous dialogue from-
    (A) Othello
    (B) Romeo and Juliet
    (C) Hamlet
    (D) Macbeth"

Meta-linguistic Knowledge (MK). Items requiring declarative knowledge of grammar such as recognizing parts of speech, determiner, passive voice etc. For instance, "Depression is often <u>hereditary</u>. The underlined word is a/an-
    (A) Adverb
    (B) Adjective
    (C) Noun
    (D) Verb"

Vocabulary in Isolation (VI). Items dealing with candidates' knowledge of the conceptual meaning of English words such as antonyms, synonyms, the analogy between pairs of words and others without

providing any meaningful context. For instance, "What would be the right synonym for "initiative"?

    (A) Apathy
    (B) Indolence
    (C) Enterprise
    (D) Activity"

Vocabulary in Context (VC). Items attempting to assess candidates' knowledge of the contextual meaning of English words used within a meaningful context provided in a sentence. For instance, "It is time to review the protocol on testing nuclear weapons." Here the underlined word means-

    (A) Record of rules
    (B) Summary of rules
    (C) Procedures
    (D) Problems"

Language Usage (LU). Items requiring candidates to identify instances of accurate language use according to rules of grammar and vocabulary usage. For instance, "------ amazing song haunted me for a long time.

    (A) These
    (B) Those
    (C) Thus
    (D) That"

Idiomatic Expressions (IE). Items asking for candidate's knowledge of English idioms, proverbs, and phrasal verbs. For instance, "The idiom "A stitch in time saves nine"- refers the importance of --

    (A) Saving lives
    (B) Timely action
    (C) Saving time
    (D) Time tailoring "

Analysis of the Interviews

With regards to the third research question, BCS class -1 officers'(former test-takers) and the aspiring candidates' interviews were conducted to know whether the English section of BCS preliminary test can truly measure the candidates' English language proficiency level.

    Interviewees were required to respond to the following questions:

1. How well does the English section of the BCS preliminary test determine candidates' status as eligible or non-eligible?
2. Do you think the BCS preliminary English test measures the candidate's language proficiency?
    (a) If YES, to what extent can it measure the candidate's language proficiency?
    (b) If NO, what alternative(s) would you propose?

    Responses were recorded and transcribed later for interpretation and analysis.

RESULTS

Applying Hughes' (2003) and Bachman's (1990) definition the validity constructs presented in the literature review above, it can be seen that the BCS preliminary English test does not conform to any of the validity constructs required for a valid language proficiency test.

    The test does not qualify as valid in terms of its content as they do not constitute a representative sample of language structures and communicative functions the successful

candidates may require in their future profession for which the test is conducted.

No items in the four examination papers analyzed in this study assess English language skills such as listening, speaking, reading and writing. Absence of data regarding test specifications, proficiency construct and the tests' ability for satisfactory prediction of successful test-takers' future language use makes it impossible to accept the test as valid language proficiency tests. By all means, it may seem to qualify only as a general test for vocabulary, grammar and general knowledge of English literature. The total marks allocated for English section is 35 which is not based on any justifying correlation.

In light of Bloom's taxonomy, apparently, the test is entirely based on the lower order concerns of cognitive abilities. Most of the MCQ items of the English section in the BCS preliminary examinations from 35th to 38th are based on recalling factual knowledge which does not require higher-order thinking abilities of the candidates. Instead of testing language proficiency, the test assesses memorization of factual information. Table 2 shows the content distribution of the English section from 35th to 38th preliminary tests.

The average of the content distribution is presented in Figure 1.

Table 2

Percentage of content distribution of the English section from 35th to 38th preliminary tests

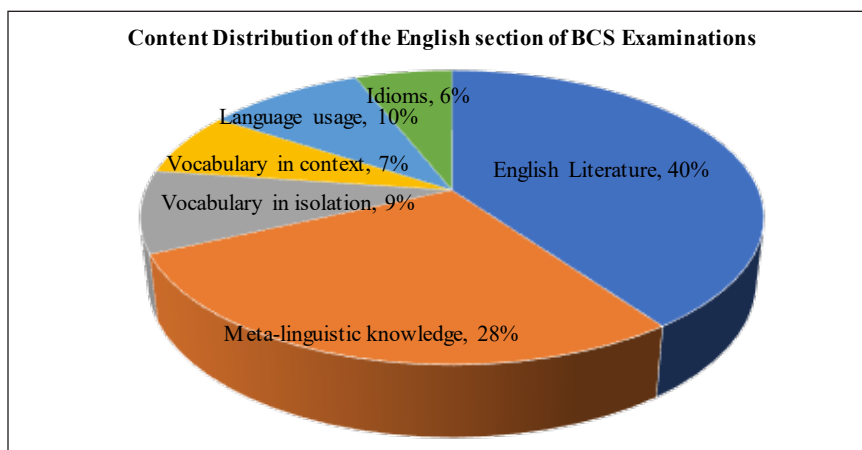| BCS preliminary examinations	Targeted domains of knowledge					
	EL	ML	VI	VC	LU	IE
35th	34%	17%	12%	23%	11%	3%
36th	50%	17%	5%	0%	20%	8%
37th	43%	32%	9%	5%	3%	8%
38th	32%	46%	12%	0%	5%	5%

Figure 1. Content distribution of the English section of BCS examinations

The test is apparently biased towards candidates with a background in English literature. Almost 40 % of the test items require candidates' factual knowledge of English literature such as the name of authors, source of quotations, identification of literary genre and periods etc. Items requiring candidates' metalinguistic knowledge constitute the second-largest category with an average of 28%. Formal aspects of English grammar are tested through items that represent more than one-fourth of the test items.

Approximately one third (32%) of the total items involve assessment of vocabulary knowledge, language usage and idiomatic expressions. Items of this category again mostly target candidates' factual knowledge through testing vocabulary and idiomatic expression in isolation.

All the BCS officers (O1-O5) and the aspiring candidates (C1-C5) explicitly expressed that the existing English language proficiency test is not an effective test for measuring a candidate's proficiency level in English. They also expressed their preference to modify the test and suggested to add new items such as reading comprehension, uses of day-to-day English, and practical uses of grammar which are all related to higher-order abilities. As O1 expressed,

In this condition, [sic] English proficiency test is not enough to measure a candidate's proficiency level. Most of the questions are from the literature. For the grammar

section, it also gives emphasis on direct grammar than the practical uses of grammar.

Thus, it is evident that the test gives more importance to testing meta-linguistic knowledge than the uses of language. C1's opinion supports the BCS officer's opinion that this English proficiency test cannot appeal to be a fair test to measure a candidate's merit and potentials. She says,

A person's English language proficiency cannot be measured through the current format of BCS preliminary examination, as it is mainly a memory-based examination which gives significance to a person's knowledge of English literature rather than his/ her ability to use the language properly.

It is found from the interview that candidates in preparation for the test memorize the MCQ items from the question banks commercially produced and marketed in the country. This indicates that the present examination type is based on memorization and reproduction. It does not assessment-takes true ability of language use. As O4 said,

I think the existing English language proficiency test is not enough to measure a candidate's proficiency level because most of the questions of the English section are based on memorization.

Besides, both BCS officers and the aspiring candidates pointed out serious faults in the test. As C3 said,

> *I don't think the BCS preliminary test can determine the best candidates. The reason behind is that major parts of the preliminary test are[sic] depend on memorizing capabilities of the candidates. So, it may not identify an ideal candidate.*

The level of English competence that BCS test attempts to measure is not sufficient. The BCS class-1 officers and the aspiring candidates expressed their views on the required level of English for speaking and listening. As O3 suggested,

> *To measure the candidate's proficiency level effectively, there should be some questions that include some basic grammar and pronunciation.*

It is also echoed by (C3) too, who believed that the current BCS preliminary test items cannot truly measure a candidate's proficiency level and he proposed to revise the test items. He expressed his views:

> *Instead of testing memorizing ability, the test should give emphasis on analytical ability, critical thinking and creative questions.*

In the same vein, C1 also believes that these test items mainly encourage memorization of factual information instead of testing language proficiency. She observes:

> *I feel the examination should change its current format and base it on international English examinations such as IELTS, TOFEL.*

It seems that the existing English language proficiency test is perceived as ineffective in measuring candidates' proficiency level in English. Therefore, based on these non-expert opinions, the test does not qualify for its face validity too.

The interviewees also expressed their preference for modifying the test and suggested to add items to include reading comprehension, uses of day-to-day English, practical uses of grammar, linking ideas, application of language use and so on. They also advised reducing the number of questions from literature, proverbs, and idioms. Only one BCS class-1 officer claimed that the present test format is acceptable as it is a test of screening out candidates from a large number of applicants. However, he also clarified later that the existing English proficiency test needs to be revised and items for testing reading comprehension and control of grammar could be included.

DISCUSSION AND CONCLUSION

The key findings of the study reveal that the English section of the BCS preliminary examination does not conform to the validity constructs and therefore does not qualify as a valid test of language proficiency. The test mostly involves assessment of memorization and is considerably biased with respect to its inclination for testing

knowledge of English literature and formal aspects of English grammar and vocabulary. The test takers and the aspiring candidates are evidently aware of the ineffectiveness of the test in measuring their language proficiency. Apparently, the test takers are in favour of modifying the test to make it more like other widely used proficiency tests.

As the findings reveal, test makers have selected test items for testing lower-order skill of remembering. This tendency clearly reflects Hughes (2003) observation that test writers tend to test what is easy to test rather than what is really important to test. Bachman (1990) argued the evidence was required to claim that scores in MCQ tests typically used for selecting clerk and secretaries were indeed relevant. As test design becomes a matter of convenience rather than accurate measurement, we need concrete evidence to claim that the test as valid.

Surprisingly, one participant still found the test acceptable. This is perhaps due to the fact that tests are often accepted uncritically. Tests are regarded as justified and unquestionable their own rights (Shohamy, 1998). Although most test-takers found the test ineffective, it still found its acceptance on the ground that the sole purpose of the test is to screen out excessive candidates and thus any test format that can successfully eliminate candidates may even appear fair to some stakeholders.

In Bangladesh, the system of assessment has always been guided by enthusiasm for assessing memorization and comprehension skills. Our findings are similar to findings from other studies conducted in Bangladeshi school settings which maintain that higher-order skills such as application, analysis, synthesis, and evaluation are hardly included in the assessment (Begum & Farooqui, 2008).

The nature of test and assessment used in the BCS preliminary English test not only fails to successfully assess candidates' creativity, imagination, and critical thinking skills but also risks selecting candidates who are incompetent in English for positions that are crucial to the nation's socio-political and economic development. It may be possible that many candidates who are not good at memorizing facts and figures or who find the whole process somewhat meaningless seek illicit ways of passing the exam. Therefore, a change in the nature of test items of English and assessment procedures may prevent any unwanted and illicit means such as leakage of question papers.

This change should also ensure that candidates' critical thinking skills and problem-solving abilities are emphasized in the tests. We need to keep in mind that knowing factual information is no longer considered important in the age of social media, Google and Wikipedia. It becomes important to consider test takers' abilities to apply knowledge and information in practical life. Therefore, it is important that the English part of the BCS examination shifts its focus from lower-order to higher-order thinking skills and places greater emphasis on test-takers' ability of critical thinking, and in-depth understanding of how the English language is used in real life.

The dominance of MCQ items from English literature and the metalinguistic aspects of the English language have serious repercussions in terms of learning practices, strategies and outcomes. Aspiring candidates are very likely to memorize factual information from English literature which will consequently lead to candidates' cultivation of memorizing strategies and avoidance of language learning strategies. Excessive use of items that test metalinguistic knowledge can eventually prompt test takers to pay more attention to the formal aspects of language and to ignore the functional aspects of language necessary to develop proficiency in the target language. Indeed, the impact of high stakes assessment has a severe washback effect on English language teaching in Bangladesh (Sultana, 2018) and consequently increases a form of shadow education (Hamid et al., 2009) as both 'micro (classroom)' and 'macro (education system and societies)' level aspects affect washback (Khan et al., 2019). It is important to note that some of the essential qualities such as oral presentation, leadership, tolerance, and social values are not included in the assessment system in Bangladesh (Begum & Farooqui, 2008).

As Khan (2006) suggested, there was a pressing need to show sensitivity to educational and cultural contexts in which language proficiency tests were administered. Language should not only be taught but also be tested in a manner consistent with the local context. Test designers, therefore, need to consider the social, cultural and educational environment of the users and adopt tests accordingly.

It may be argued that since the main objective of the BCS preliminary examination is to sort out potential candidates, attempting to measure the lower order skills at this initial stage is acceptable. Thus keeping the higher-order skills to assess in the later stages seems somewhat reasonable. However, even for assessing the lower order skills of language proficiency, it is necessary to provide test setters with a guideline for the selection and construction of test items to ensure a valid language proficiency assessment.

Looking at the poor qualities of test items in the BCS preliminary English examination, it can be assumed that test designers and question setters have little understanding about the fundamental principles of language proficiency testing. It is also evident from the content analysis that the BCS English test setters seem to have a conspicuous inclination for candidates' knowledge of English literature. One explanation why civil servants' factual knowledge of English literature and idiomatic expression is given so much undue importance is perhaps the age-old colonial entanglement of the test setters which eventually may have prompted them to assume that the future civil servants of the country have to know the literature of the masters of the past.

are also deeply thankful to the editors and the anonymous reviewers for their constructive comments and suggestions.

REFERENCES

Al Amin, M., & Greenwood, J. (2018). The examination system in Bangladesh and its impact: On curriculum, students, teachers and society. *Language Testing in Asia*, *8*(1), 1-18.

Ali, C. M., & Sultana, R. (2016). A Study of the validity of English language testing at the higher secondary level in Bangladesh. *International Journal of Applied Linguistics and English Literature*, *5*(6), 64-75.

Ali, M. 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*, *1*(1),1-21. doi:10.1080/03057925.2018.1535890

Auer, D. (2018). Language roulette - The effect of random placement on refugees' labour market integration. *Journal of Ethnic and Migration Studies*, *44*(3), 341-362.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Banu, L. A. (2009). Problems and misconceptions facing the primary language education in Bangladesh: An analysis of curricular and pedagogic practices. *BRAC University Journal*, *1*(1), 1-10.

Begum, M., & Farooqui, S. (2008). School-based assessment: Will it really change the educational scenario in Bangladesh. *International Education Studies*, *2*(1), 45-53.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* New York: David McKay.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (pp. 443-507). Washington, D. C.: American Council on Education.

Chen, Z., & Henning, G. (1985). Linguistic and cultural bias in language proficiency tests. *Language Testing, 2*(2), 155-163.

Das, S., Shaheen, R., Shrestha, P., Rahman, A., & Khan, R. (2014). Policy versus ground reality: Secondary English language assessment system in Bangladesh. *Curriculum Journal*, *25*(3), 326-343.

Debnath, B. K. (2017). *BCS exams.* Retrieved January 2, 2018, from http://www.theindependentbd.com/home

D'Este, C. (2012). New views of validity in language testing. *EL.LE. Educazione Linguistica, 1*(1), 61-76.

Djiwandonos, P. (2006). Cultural bias in language testing. *TEFLIN Journal, 17*(1), 81-89.

Diamond, L. C., Luft, H. S., Chung, S., & Jacobs, E. A. (2012). " Does this doctor speak my language?" Improving the characterization of physician non-English language skills. *Health Services Research*, *47*(1), 556-570.

Ebadi, S. (2015). Exploring the cognitive level of final exams in Iranian high schools: Focusing on Bloom's taxonomy. *Journal of Applied Linguistics and Language Research*, *2*(4), 1-11.

Fulcher, G. (1999). Assessment in English for academic purposes: Putting content validity in its place. *Applied Linguistics 20*(2), 221-236.

Hamid, M. O., Sussex, R., & Khan, A. (2009). Private tutoring in English for secondary school students in Bangladesh. *TESOL Quarterly*, *43*(2), 281-308.

Hoque, M. E. (2016). Teaching to the EFL curriculum or teaching to the test: An investigation. *The EDRC Journal of Learning and Teaching*, *1*(1), 1-25.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). New York: Cambridge University Press.

Jahan, M. (2012). Recruitment and selection process in Bangladesh civil service: A critical overview. *Public Policy and Administration Research, 2*(5), 29-36.

Khan, A. B. M. A., Aziz, M. S. A., & Stapa, S. H. (2019). Examining the factors mediating the intended washback of the English language school-based assessment: Pre-service ESL teachers' accounts. *Pertanika Journal of Social Sciences & Humanities*, *27*(1), 51-68.

Khan, R. (2006). The IELTS speaking test: analysis of cultural bias. *Malaysian Journal of ELT Research, 2*(1)*,* 61-79.

Khan, R. M., & Ara, F. (2005). Direct recruitment in the care services in Bangladesh. *Pakistan Journal of Social Science, 3*(2), 915-821.

Kim, H., & Elder, C. (2015). Interrogating the construct of aviation English: Feedback from test takers in Korea. *Language Testing*, *32*(2), 129-149.

Knoch, U., McNamara, T., Woodward-Kron, R., Elder, C., Manias, E., Flynn, E., & Zhang, Y. (2015). Towards improved language assessment of written health professional communication: The case of the occupational English test. *Papers in Language Testing and Assessment*, *4*(2), 60-66.

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. London: Sage publications.

Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman.

Mustafa, F., & Anwar, S. (2018). Distinguishing TOEFL score: What is the lowest score considered a TOEFL score? *Pertanika Journal of Social Sciences & Humanities*, *26*(3), 1995-2008.

Nunan, D. (1992). *Research methods in language learning*. New York: Cambridge University Press.

Rahman, M. F., Babu, R., & Ashrafuzzaman, M. (2011). Assessment and feedback practices in the English language classroom. *Journal of NELTA*, *16* (1-2), 97-106.

Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, *24*(4), 331-45.

Sultana, N. (2019). Language assessment literacy: an uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, *9*(1), 1-14.

Sultana, N. (2018). Test review of the English public examination at the secondary level in Bangladesh. *Language Testing in Asia*, *8*(16), 1-9.

Vijver, F., & Tanzer, N. (2004). Bias and equivalence in cross-cultural assessment: an overview. *Revue Europeenne de Phychologieappliquee*, *54*(1), 119-135.