

Principles of Language Assessment: A Critical Evaluation in the Context of a Selected University in Bangladesh

Tamimur Rahman Anik¹

Mohammad Golam Mohiuddin²

Esmot Ara Choity³

Abstract

A conducive environment for language teaching and learning, and assessment are interconnected.

Assessment is essential to evaluate students' language skills and teachers' knowledge and pedagogical skills. To maximize reliability and validity of language assessment, this study evaluates the language assessment procedure of a selected university in Bangladesh. In terms of assessing language courses, there are five established principles. The principles are practicality, reliability, validity, authenticity and washback. The language assessment instruments were critically evaluated based on two objectives so that the language courses offered by the university affect students' learning motivation in a positive way. To evaluate the principles of language assessment critically, a mixed method approach was adopted. The researchers collected quantitative data from undergraduate and graduate students who have knowledge regarding the principles of language assessment. The researchers collected qualitative data from the teachers through semi-structured interviews. The teachers were selected by their experience instead of the instructors of the different language courses. Because teachers' ability to implement assessment principles is another factor. The study reveals that the university incorporates all the five principles of language assessment but with a few deficits. Further attention needs to be provided to minimize these deficits associated with the principles of current language assessment practices by teachers and authority. The study provides valuable insights and numerous suggestions which will help the authority and teachers who conduct language courses and assessment to better align with the principles of language assessment.

Keywords: assessment, authenticity current assessment practices, evaluation of principles of language assessment, reliability, validity

Assessment is a crucial process which happens regularly through a variety of methods and techniques. It is the process of gathering and discussing information from various sources to increase a deep understanding of students' knowledge, comprehension, and application of that knowledge. The primary goal of assessment

¹Graduate, Department of English, Green University of Bangladesh

²Assistant Professor, Department of English, Green University of Bangladesh (*Corresponding Author)

³Officer, Center for International Affairs, Green University of Bangladesh

is to supply reliable and relevant information that can enlighten both teachers and students. Assessment, the ongoing process, is used by the teachers to make well-informed decisions regarding their students' progress based on some predetermined standards (Chapelle et al., 2015; Bajuti, 2018). It is also used to provide feedback so that students can develop their knowledge, skill and attitude. Not only that but also it enables teachers to understand what methods and techniques they can use in the classroom to improve teaching and learning process. Assessment influences the learning process more than that of teaching. As we know the development of proficiency happens gradually, appropriate assessment methods need to be adopted and implemented to facilitate the gradual process in a correct order. Assessment plans should be guided by the language curriculum goals and objectives. As a result, it is necessary to establish a suitable alignment between the design of assessment methods and the language curriculum goals and objectives. With the help of appropriate assessment, teachers can categorize and evaluate their students, provide feedback and plan their lesson properly (Tosuncuoglu, 2018). Feedback is necessary for the teachers so that they can shape their practices according to the learning styles of their learners. These days, testing (the subset of assessment) and assessment are more prominent and necessary for a number of reasons, including curriculum and teaching support, collecting information about students, teachers, and schools, selection and certification. To conduct testing and assessment teachers' knowledge regarding language tests are essential. According to Islam et al. (2021), Bangladeshi teachers have very little knowledge of designing and implementing the assessment instruments. They analyzed that "English language learning in Bangladesh" has miserably failed due to lack of proper knowledge regarding assessment and ineffective assessment methods. They also analyzed that "it was assessment, not teaching, that influenced the students most in their learning." The achievement of curriculum goals and objectives depends solely on how they are designed, implemented, and what kind of knowledge the teachers have. There are five principles of language assessment-practicality, reliability, validity, authenticity and washback. In short, assessment plays a crucial role in terms of achieving and developing the quality of teaching and learning process with adequate feedback which actually helps to meet the goals and objectives of the curriculum.

As assessment measures students' knowledge and ability, the effectiveness of assessment depends on the quality of the assessment instrument which actually collects information regarding students' knowledge and ability. For that it is important to use high quality assessment instruments. To ensure the high quality of the instrument, pre-evaluation of the quality of the instrument is much required. An effective and high-quality instrument maintains the principles of language

assessment. There are five principles in total. They include practicality, reliability, validity, authenticity, and washback (Brown & Abeywickrama, 2018) Every institution should maintain these five principles in terms of designing a practical, reliable, valid and authentic assessment instrument. But there are some institutions that do not maintain the quality of the assessment instrument. To ensure high quality assessment instruments for this institution, this paper aims to investigate the scope of developing the quality of the assessment instrument in terms of the five principles so that the institution designs and provides the high-quality assessment instrument that will be more practical, reliable, valid and authentic.

Since assessment plays a crucial role in teaching and learning, it is important to design assessment plans that should meet the curriculum goals and objectives. As the assessment plans are made and guided by the curriculum goals and objectives, it is important for the teachers and administration to maintain all the principles of language assessment. The quality of assessment depends on the knowledge of the test givers who design the test. It is seen that in most cases test givers use a test from a publisher or textbook to measure their test takers' performance (Fulcher, 2012). However, those tests or task items do not go with the learners' level. Therefore, it is very important to design the test items themselves to fit the learners' level perfectly by maintaining the five principles of language assessment. The objective of the study is to determine the outcomes of the assessment by maintaining the five principles of language assessment. The following research questions served as the basis for the current study which will be answered in the subsequent section in light of the research findings:

1. To what extent have the principles of language assessments been incorporated in this institution?
2. Does this institution lack in maintaining any principles of language assessment? If yes, to what extend does this institution lack in maintaining the principles of language assessment?
3. What are the scopes of development of assessment instruments in this institution?

Literature Review

Five principles of language assessment should be considered in terms of designing an assessment instrument so that the assessment instrument can meet its goals, purposes and objectives of the curriculum. Test developers including test givers should keep them in consideration in terms of designing an assessment instrument. These five principles include practicality, reliability, validity, authenticity, and washback (Brown & Abeywickrama, 2018).

Their operational definitions are included in the following.

Practicality

Practicality refers to the “logistical, down-to-earth, administrative issues involved in making, giving, and scoring an assessment instrument” (Brown & Abeywickrama, 2018, p. 26). This means the instrument will be easy to design, administer and score. There are four factors related to practicality and they are: 1) Costs, 2) The amount of duration it needs to construct, 3) How easily it can be scored and 4) How easily it can be interpreted (Mousavi, 2009, p. 516). So, a practical test is a test which considers time, budget, resources and administrative issues. Bachman & Palmer (2009) also include the elements related to a practical test which stays with the limits of time and proper use of existing human and material resources. Haris (1969) introduces the financial aspects of test by taking into consideration of cost per copy, remuneration of test administrators and other budgeting factors.

Reliability

Reliability refers to the assessment results which are “consistent and dependable” (Brown & Abeywickrama, 2018). In terms of giving the same test to the same or matched test takers several times, the tests will yield a similar result. If they cannot yield the similar (more or less) result the assessment will lose its reliability and the assessment will be considered as unreliable (Hughes, 2003, p.3). Issues related to reliability can be broken down depending on a number of factors including student, rater, test administration and the test itself (Brown & Abeywickrama, 2018).

Student Related Reliability: Every student has a specific level that they can gain particular scores in an assessment. So, results may vary depending on a student's knowledge. Sometimes it depends on the preparation period of a student or sometimes the physical or psychological conditions. It is impossible to get the desired results if the student is dealing with physical difficulties. So, it is important to ensure student's physical and psychological wellbeing to get a consistent result on different occasions. Test developers including teachers should inform their test takers about the question patterns and rubrics for scoring so that they can develop strategies for completing the assessment without any fear and with more confidence.

Rater Related Reliability: Normally rater is the person who scores the assessment. For scoring subjective tests it is important to employ the raters who are capable of doing these jobs perfectly. Otherwise, rater unreliability can occur. As the rater is a human being, there is a chance of human error and which sometimes creates rater related unreliability. The reliability of raters is divided into two parts. They are inter-rater

reliability and intra rater reliability. The first one refers to the similarity between the ratings provided for the same assessment by two or more distinct raters (Brown & Abeywickrama, 2018). Inter-rater unreliability happens if the raters are unaware of the scoring criteria, if they do not have any previous experience in scoring, if they are not fully attentive while scoring and if they have predetermined biases. The last one refers to a single rater's consistency throughout several assessments. Intra rater reliability can also fail. It happens if the rater is exhausted in scoring for a long time, if the rater is careless and if the rater is biased towards some test takers.

Test Administration Reliability: Test administration reliability describes the situation all about its environment where the assessment goes on. To make a test reliable regarding administration it is crucial to make sure that no noise is distracting test takers at the time of assessment. Test takers should get the required test papers and question papers and there is no ambiguity in the printed copy. The room should be enlightened perfectly so that all corners have the same amount of light. The temperature of the room should be manageable by all the test takers. The desks and chairs should be adequate for all the test takers.

Assessment Reliability: Assessment reliability talks about the design and question patterns of the assessment. Some factors can contribute to making the assessment itself unreliable. Assessment unreliability may be caused by rater biases. In terms of subjective tests, the test may have multiple correct answers. Test takers may be confused regarding the answers they should include. Raters also struggle to determine if the answers are right or wrong. This can make rater biases and the assessment itself unreliable. While designing objective tests, predetermined answers and distractors should be added carefully. As almost every test has its own time constraints, it may create assessment unreliability when the test takers are not capable of answering all items within the allotted time constraints.

Validity

The most complex criterion of an effective test and arguably the most important principle is validity. Gronlund (1998, p. 226) said that "the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment". Most importantly, validity focuses on assessment's purpose. In the context of assessment, a valid assessment measures what it aims to measure and exactly "does what it is intended to do" (Davies et al. 2002, p. 221). Brown also said exactly the same that a valid assessment measures exactly what it proposes to measure. To design a valid test or assessment, test designers need to focus on the following evidence. These content validity, criterion validity, construct validity, consequential validity and face validity.

Content Validity: The course topic is the primary consideration in content validity. If test administrators are able to precisely specify the achievement they are assessing, they can assert that their evaluation has content validity. Content validity measures exactly the same contents covered in the classroom. If the assessment covers all the objectives, it can claim 100% content validity. If one or more objectives are not covered by the assessment, content validity is compromised. Content validity is also compromised when the test gives assess a test taker's ability in a form it should not be. It is considered that content validity is more attainable through direct testing.

Criterion validity: Criterion validity is dependent on criterion referenced tests. Because these tests are designed to give test takers feedback, usually in the form of grades, on specific course or lesson objectives. As criterion validity focuses on objectives, this type of test has emphasis on classroom-based assessment. Criterion validity is best demonstrated through a comparison of results of an assessment with results of some other measure of the same criterion.

Construct Validity: A construct is any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perceptions. Construct validity is about how well a test measures the concept it was designed to evaluate. It is related to theoretical constructs. Theoretical construct refers to an explanatory concept that is not itself directly observable but that can be inferred from observed or measured data. Construct validity focuses on all the elements that should be assessed.

Consequential Validity: Consequential validity refers to the validity that encompasses all the consequences of a test. It includes the capacity of an assessment to measure intended criteria accurately, its impact on test taker preparation, and the societal consequences of how a test is taken into account and applied

Face Validity: Face validity refers to the extent to which a test appears exactly the same test giver purposes to measure. This type of validity mainly focuses on how the test looks and appears. Mousavi's (2009, p. 247) definition says, "Face validity refers to the degree to which a test looks right, and appears to measure the knowledge or abilities it claims to measure, based on subjective judgement of the examinees who take it, the administrative personnel who decide on its use, and other psychometrically unsophisticated observers". "Many assessment experts view face validity as a superficial factor that is too dependent on the whim of the perceiver", stated Brown & Abeywichrama (2018). The level of difficulty should meet the objective of the test or assessment.

Authenticity

Authenticity is a concept that includes real life situations, tasks or language items. “The degree of correspondence of the characteristics of a given language test task to the features of a target language task”, defined Backman and Palmer (1996, p. 23) If an assessment fails to attach the real-life situations, it will be considered as an unauthentic assessment.

Washback

Washback refers to the effect of assessment on teaching and learning. (Hughes, 2003)The effects of assessment on teaching and learning can be categorized in two terms. They are positive washback and negative washback. Washback is one of the important principles of language assessment which actually helps teachers to provide feedback if their students lack in any topics and allow test takers to perform better by incorporating the feedback from their teachers. It also offers test takers to get prepared for the upcoming assessment so that they can get the highest scores.

Islam et al. (2021) analyzed different published papers from the database using the search engine Google by keywords. They reviewed the previous education system in Bangladesh at the same time current education system in Bangladesh, unsettled status of English in Bangladesh, English language assessment in Bangladesh and several context-based implications of language assessment and education of Bangladesh. They showed the gap between the curriculum objectives and the principles of language assessment. Summative assessment was being implemented all over the country without a few formative assessment practices.

Ali (2016) worked on the assessment of English Language skills at the Secondary level in Bangladesh. The researchers conducted semi-structured interviews with six secondary school teachers and analyzed documents. The findings are divided into five sections. The first section is about the “English teachers’ perception of the roles of assessment”. The second section is about the “teachers’ perception on curriculum guidelines”. The third section is about the “teachers’ perception in assessment”. The fourth section is about the “major challenges that English teachers faced”. The last section is about “test driven teaching and learning”. While analyzing the first section, it was found that all the six teachers had an adequate understanding of the components of language assessment. From this finding, it is comprehensible that, normally teachers have elementary knowledge and understanding about the student

assessment/evaluation process as it is often discussed in every teacher training session.

Although the assessment literacy among teachers do vary based on their work experience, training experience, age, motivation, location of school etc. Rahman & Khan (2021N) drew a number of insights from their study on English Language testing practices at the secondary level. The insights are categorized into five sections. The first insight is on the objective of the curriculum and the course contents. The second insight is on the informal formative assessment. The third insight is about the lack of assessment literacy. The fourth insight is about the concept of washback which plays a crucial role for learners' development. The last insight is about the test impact which should be traced beyond the classroom. The study has a priority to the principles of language assessment to be maintained perfectly so that all four skills are acquired the level-best.

Sumarsono et al. (2023) examined the test against the principles of language assessment. The research findings are categorized into five categories according to the five principles of language assessment. It was found that the practicality principle of this assessment was high (rated 5 out of 5). When dealing with reliability issues, the research got 3.6 out of 5. Content and construct validity were scored 4 out of 5 separately. Authenticity was scored high (5 out of 5). Because the assessment was as natural as possible and integrated real-life situations. Washback principle got 4.5 out of 5 for its constructive feedback. The researchers concluded the research by considering itself as a good assessment tool with some minor issues as it satisfied all the principles of language assessment. Hudaya (2017) found that the respondents who are the teachers in different stages are qualified with a bachelor degree having teaching experience approximately 1-5 years. The researchers divided their findings into three categories. The first category is about the teachers' level of preparation for evaluating learners' performance. The second category is about the teachers' practices in implementing the principles of language assessment. The last category is about the usefulness of the questionnaire that includes the principles of language assessment for teachers to prepare a test.

In traditional approach of paper-pencil test for language assessment, normally the 3 skills of reading, vocabulary and listening are evaluated together. Soatov and Togaymurodov (2024) discovered that these tests had more weaknesses than strengths. Among the weaknesses, they found the second task lacks authenticity. Because the second task on reading and the task did not require any reading skills to complete the task. Finally, they ensured i + level. The teachers should not make the test practical; rather they should design the test by considering the objectives of the course. Ranjit (2022) identified a huge gap between the

curriculum goals and the testing system. Test objectives and writing requirements were never applied during the procedure. All kinds of validity were compromised. Similarly, there had been issues with reliability. At the same time the test did not reflect the curriculum properly as there was “no provision for evaluating speaking and listening abilities”. Test items were repeated “throughout educational boards and years”. Rahimi et al. (2021) found that the majority of the teachers were not adequately introduced to assessment and learning principles. Moreover, the findings showed that the majority of the teachers were ignorant to the learning principles.

Several papers have been published covering the five principles of language assessment in terms of teachers’ made test or teachers’ assessment literacy in applying the principles of language assessment. But there are very few studies in the context of Bangladesh regarding the principles, practices and problems of language assessment. (Islam, 2021) Perhaps, this study is the first which is conducted directly by evaluating the assessment instruments of the selected university against the five principles of language assessment. As the researchers’ target is to evaluate the assessment instrument of the university, the researchers are going to conduct the study on students, teachers, administrators and test developers. The above researches investigate the benefits of maintaining the principles of language assessment. Some researchers pay attention to evaluate the principles of language assessment. As the research is going to evaluate the assessment instruments, the researchers will be showing that how much they are incorporating the principles of language assessment, if there are any lacks, it will be found out. . If all the assessments maintain high quality, language skills of the students will be evaluated in a perfect way. Then the lacks of the particular students will be found out. If everything is maintained according to the principles of language assessment, the lacks will be minimized and students will be benefited and there will be a smooth teaching and learning process.

Methodology

This study employed a mixed-methods approach, combining quantitative data from online surveys and qualitative insights from semi-structured interviews.

Participants

The study involved 54 participants- 50 undergraduate and graduate students (ages 20–30; 25 females, 25 male) from a private university and 4 teachers from the Department of English from the selected university. Students were selected through purposive sampling, focusing on the completion of course "Assessment and Evaluation". The teachers, with different levels of experience and

specialized in different language modules, participated in face-to-face interviews.

Data Collection

Quantitative data was gathered via a Google Form questionnaire shared through email and messaging apps, including both open- and close-ended questions with a Five-Point Likert Scale. Qualitative data was collected through recorded one-on-one interviews with teachers, lasting 15–20 minutes. Ethical considerations ensured participants' consent and confidentiality. This study used both quantitative and qualitative methods. A questionnaire was designed for students to gather standardized numerical data, while semi-structured interviews with teachers provided in-depth qualitative insights. These methods ensure triangulation, enhancing data validity and reliability.

Survey Questionnaire for Students

The questionnaire included open- and closed-ended questions, divided into three sections: student demographics, principles of language assessment, and study-relevant questions. Closed-ended questions used Likert scales and close-choice formats to streamline data collection.

Semi-Structured Interviews for Teachers

Interviews focused on teachers' knowledge and practices regarding language assessment principles. The questions, developed from research and supervisor guidance, explored teachers' background, challenges in language assessment, and potential improvements.

Data Analysis

Quantitative data was analyzed using Microsoft Excel for frequency, percentages, and mean values, visualized with bar and pie charts. Qualitative data was transcribed and thematically analyzed to identify recurring themes, providing detailed insights while ensuring flexibility.

Findings

The findings firstly show the quantitative data received from the questionnaire surveys. Qualitative results have been shown later. The results presented in this study were limited to those that directly addressed the study topics.

4.1 Findings of Survey Questionnaire

The researchers shared the Google form link via messenger, email or WhatsApp of the questionnaire with the students of the Department of English and

asked them to fill out the form according to their choices. The questionnaire included 16 structured questions aimed at understanding the perspectives on the principles of language assessment.

Question 01: I have taken a language test like IELTS or TOEFL.

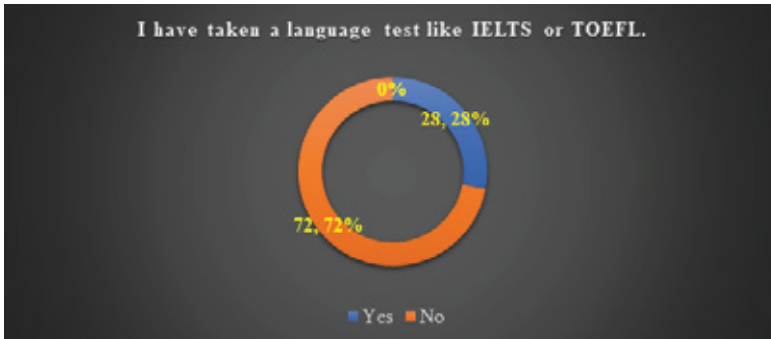


Figure 1: The students have taken a language test like IELTS or TOEFL.

The pie chart shows the number of students from the Department of English has taken a language test like IELTS or TOEFL. Notably, 72% of the student respondents have not taken any language without attending the language tests. Only 28% student respondents have taken language tests outside the university.

Question 02: The segments of the language test I have taken include:

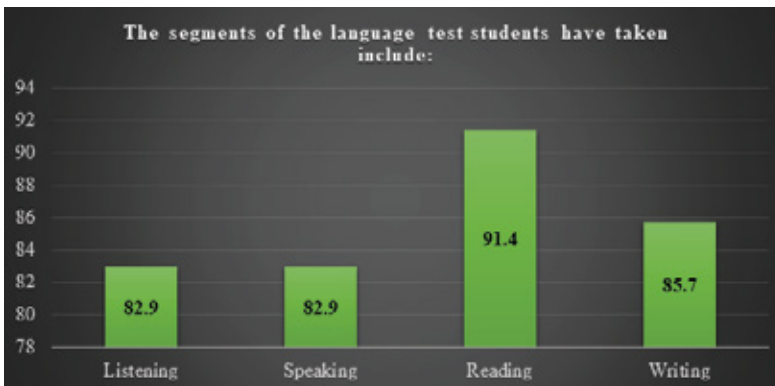


Figure 2: The segments of language tests

The column chart shows the modules of the language test taken by the students. Those who took the language tests like IELTS or TOEFL agreed that they had attended all the four segments in parallel. 28% respondents who have taken language tests acknowledged that 82.9% attended the listening module. The same

percentage attended the speaking modules as well. The highest number which is 91.4% respondents attended the reading module. The second lowest participants, 85.7% respondents attended the writing module.

Question 03: I have completed the course “Testing and Evaluation” from the Department of English.



As this purposive research requires specific knowledge regarding Testing and Evaluation or Assessment or Evaluation for investigating the principles of language assessment, the researchers needed to ask the question whether they have completed the course from the Department of English or not. All the respondents agreed that they have completed the course “Testing and Evaluation” or “Assessment and Evaluation” from the Department of English. Not a single respondent is included in the survey who has completed the required course.

Question 04: I am aware of the language assessment practices at my university.



Figure 4: Students’ awareness regarding the language assessment practices at their university

The pie chart represents the number of respondents who are aware of the language assessment at this selected university. All of the 50 respondents agreed that they are aware of the language assessment practices at their university. No students put a tick mark in the no option.

Question 05: Language assessments at my institution are practical (manageable in terms of making, giving, and scoring).

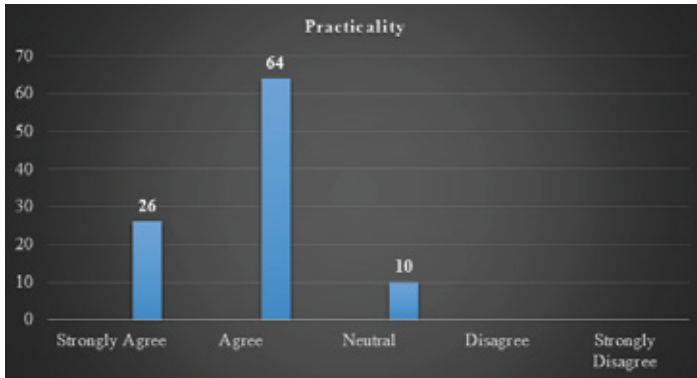


Figure 5: Practicality

The column chart depicts the percentage of the practicality principle. 26% students strongly agreed that the practicality principle is incorporated. The highest number 64% respondents agreed that their language assessments are practical. Only 10% of respondents remain neutral in the matter of practicality. No respondent strongly disagreed or disagreed about the assessment practicality.

Question 06: Language assessments at my institution are reliable (consistent and dependable on several occasions).

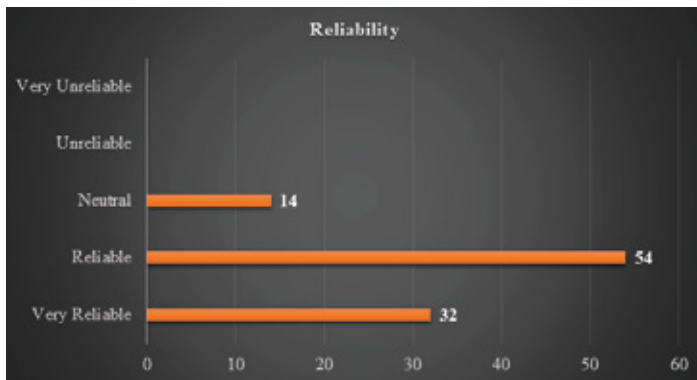


Figure 6: Reliability

The bar chart outlines the percentages of the reliability principle in terms of language assessment. 32% of respondents voted that the university offers a very reliable language assessment. The majority 54% votes are given that the university offers reliable language assessment. 14% respondents remain neutral in terms of the reliable language assessment practices. Not a single respondent gives their consent to unreliability and very unreliability.

Question 07: Language assessments at the university align with the principle of validity (the degree to which the test measures what it proposes to measure).

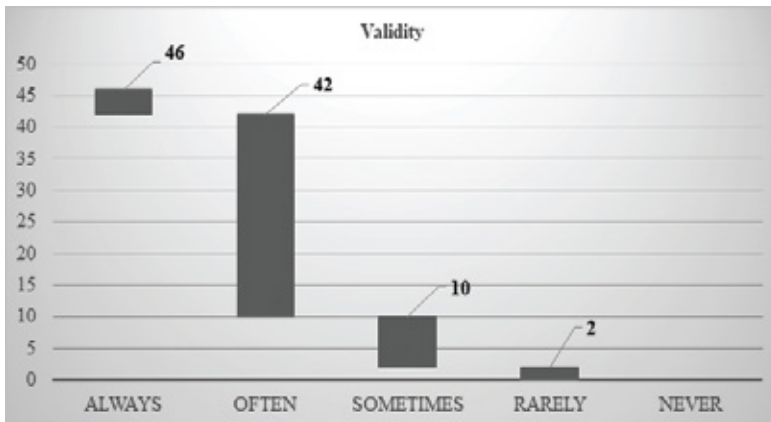


Figure 7: Validity

The stock chart illustrates the validity principle for language assessments. The chart illustrates that 46% of respondents agreed that the university always offers valid language tests. 42% of respondents agreed that the university often offers valid language assessments. Only 10% agreed that the university sometimes offers valid language assessments. Only 2% respondents gave their opinion that this university rarely provides valid language tests. No respondents agreed that the university never offers valid tests.

Question 08: Language assessments at this institution are authentic (applicable language in real life situations).

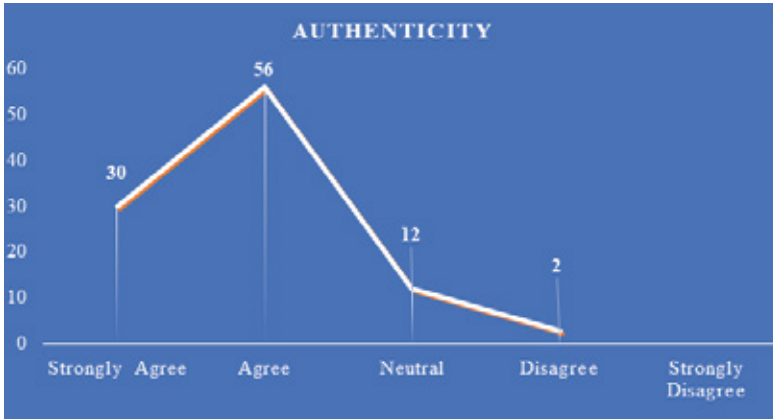


Figure 8: Authenticity

When the question comes to the authentic language tasks during the assessment, the line chart represents that 30% respondents strongly agreed that the university provides authentic tasks during the tests. The large percentage, 52, agreed that the university provides authentic tasks. A smaller number, only 12%, respondents remained neutral in the matter of authenticity. Only one respondent disagreed that the university does not provide any authentic tasks during the language tests. No respondents strongly disagreed that the university does not provide any authentic tasks.

Question 09: The formative assessment systems at the university are beneficial.

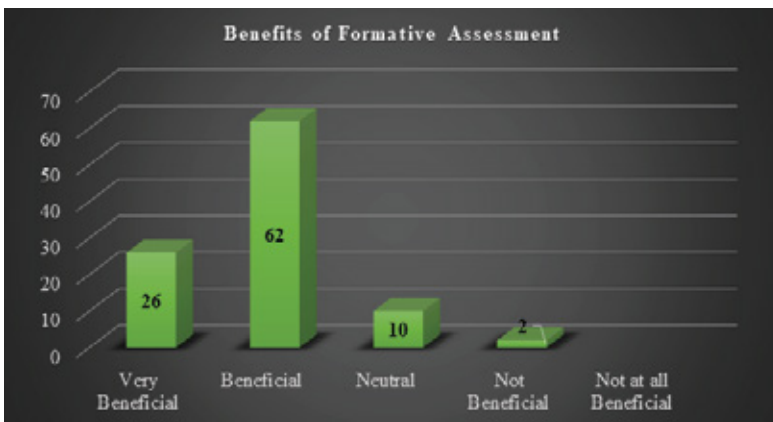


Figure 9: Benefits of formative assessment

The column chart reflects the benefits from the formative assessment procedure. The majority 62% voted the formative assessment as very beneficial. 26% agreed that the formative assessments are beneficial. Only 10% respondents remained neutral regarding the formative assessment practices of the university. Only 2% respondents reported that the formative assessment procedure is not beneficial though no one reported the formative assessment as not at all beneficial.

Question 10: The feedback provided after language assessments is beneficial.

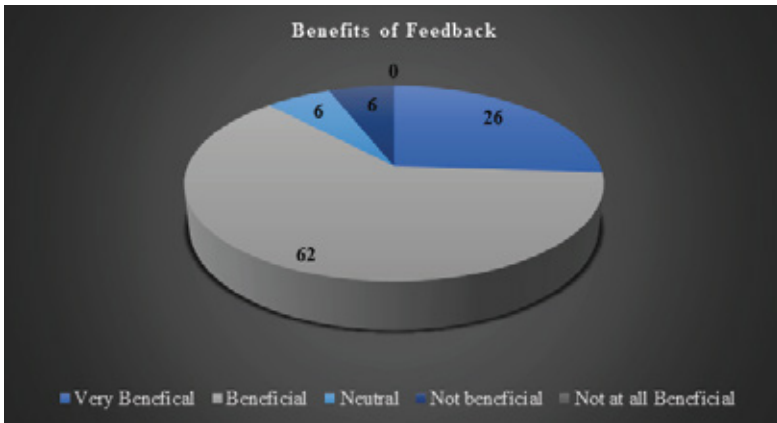


Figure 10: Benefits of feedback provided by the teachers

The pie chart shows the distribution of responses regarding the feedback provided by teachers. The chart reveals the percentages of the beneficial feedback provided by the teachers of the Department of English. 26% respondents agreed that the feedback provided by the teachers are very beneficial. 62% respondents agreed that the feedback given by the teachers is beneficial. 6% respondents remained neutral regarding feedback after each test. The same amount reported that feedback given by the teachers is not beneficial. No respondents reported that feedback given by the teachers is not at all beneficial.

Question 11: The university incorporates all the five principles of language assessment.

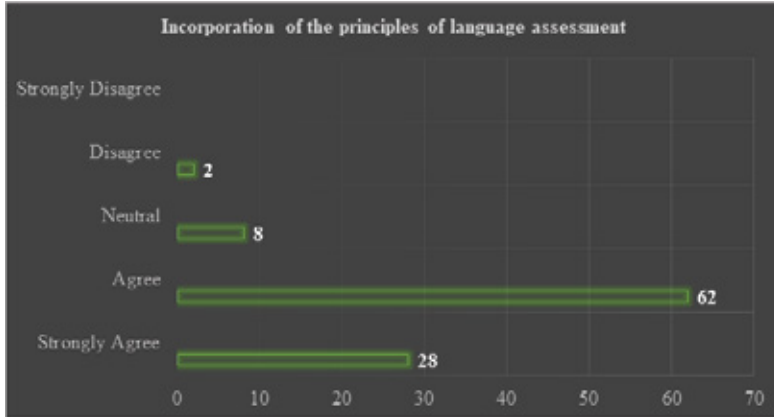


Figure 11: Incorporation of the principles of language assessment

The bar chart highlights the extent to which the university incorporates the principles of language assessment. 28% respondents strongly agreed that the university incorporates all the five principles of language assessment. The majority 62% agreed that the university incorporates the five principles of language assessment. Only 8% respondents remained neutral about the incorporation of the principles of language assessment. Only one respondent disagreed. No respondent strongly disagreed about the incorporation of the principles of language assessment.

Question 12: This institution lacks in maintaining any principles of language assessment.

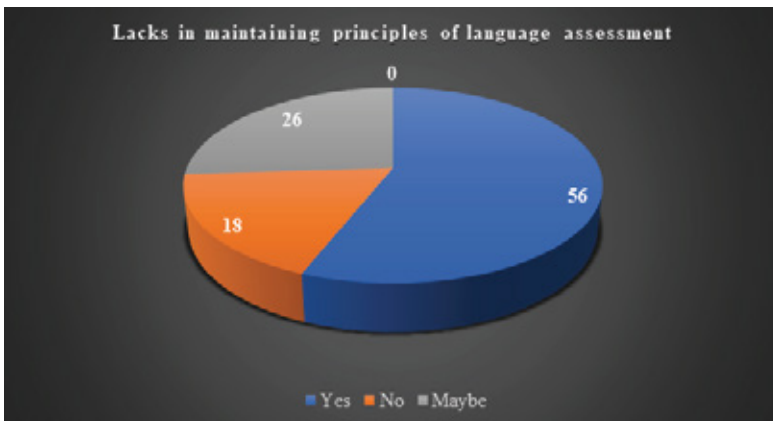


Figure 12: Lacks in maintaining principles of language assessment

The bar chart determines the number of responses about the lack of maintaining any of the principles of language assessment. The chart shows that 56% agreed that the university lacks in maintaining the principles of language assessment while 18% respondents were not sure that the university lacks or does not. 26% respondents disagreed that the university does not lack in maintaining the principles of language assessment.

Question 13: If yes, the principles that this institution lacks in maintaining are: (You can choose more than one option)

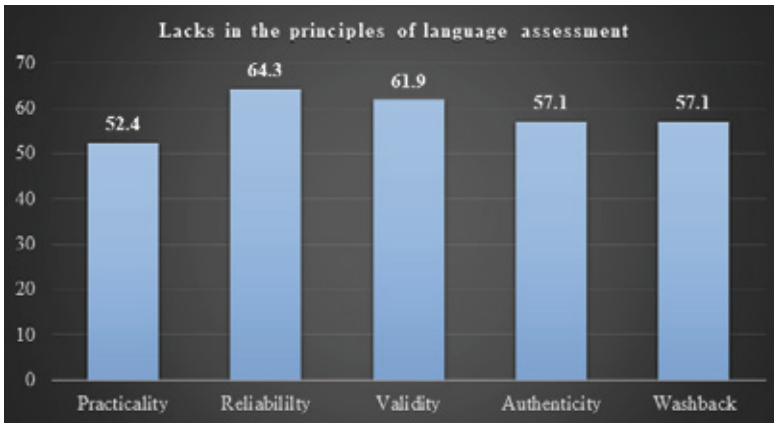


Figure 13: Lacks in the principles of language assessment

The column chart comes with the opinion of the respondents that the university lacks in maintaining all principles. 52.4% respondents reported that the university lacks in maintaining practicality principles. 64.3% respondents agreed that the university lacks in maintaining reliability. 61.9% respondents agreed that the university lacks in maintaining the validity principle. Lack on authenticity and washback is the same and which is 57.1%.

Question 14: The current language assessment instruments used in this institution are rated as:

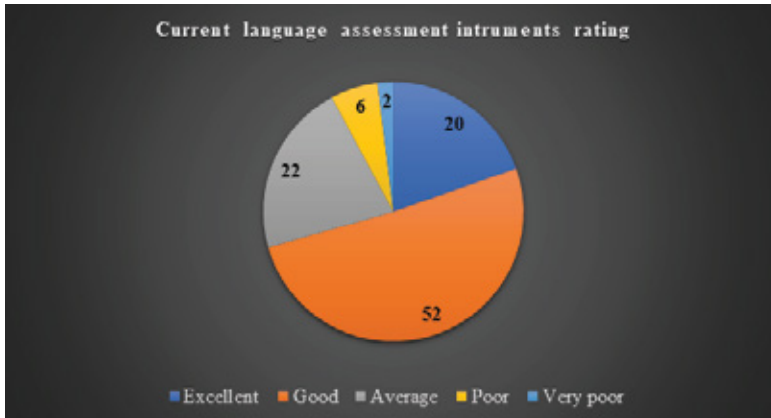


Figure 14: Ratings of the current language assessment instruments

The pie chart highlights the current language assessment instruments used at the university. 20% respondents gave their opinion as excellent. 52% respondents reported the current assessment instruments as good. Average was the choice of 22% of respondents. Only 6% reported that the current language assessment instruments are poor. No respondent reported the current language assessment instruments as very poor.

Question 15: I am confident in the institution's ability to improve its language assessment practices.



Figure 15: Students' confidence regarding assessment improvement ability of the university

The line chart shows confidence regarding the university's ability to improve its language assessment practices. 36% respondents were very confident that the university has the ability to improve its language assessment practices. More than fifty percent of respondents were confident. Only 12% of respondents remained neutral. No respondent was not confident or not at all confident.

Question 16: The scopes of development for language assessment instruments in this institution include: (You can choose more than one option)

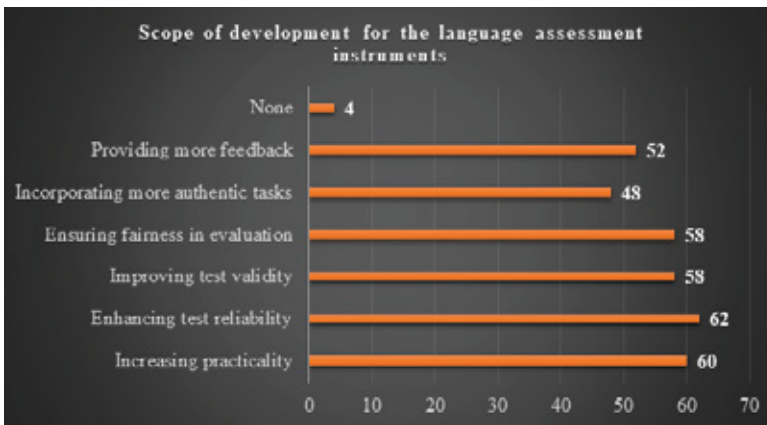


Figure 16: Scope of development for the language assessment instruments

The bar chart represents the scopes of development for language assessment instruments. 60% of the respondents reported that the university has the scope of development for language assessment instruments by increasing practicality. Enhancing test reliability is reported by 62% of respondents. 58% respondents were the side to improve test validity which will help to develop the language assessment instruments. 58% student respondents wanted the authority to ensure fairness in evaluation. 48% respondents wanted more authentic tasks in learning and assessment procedures. Furthermore, 52% of student respondents want their teachers to give more feedback after every test.

4.2 Semi Structured Interview with Parents

Q. Would you please describe the current assessment practices of your institution for language courses?

Almost all institutions incorporate the assessment process of reading and writing module of language through traditional paper-pencil test whereas, the listening and speaking module of language are evaluated through continuous assessment and practical test formats. Although a group of teachers have mentioned

that, their institutions follow traditional assessment practices that include the paper-pencil test only.

Q. Do the current assessment practices of this institution align with the principles of language assessment? If so, how?

Most of the interview participants have mentioned that the reading course is significantly more aligned with the principles of language assessment and the assessment. The institutions are trying to emphasize on the formative assessment system to align with the principles of language assessment.

Q. Can you describe any recent changes that have been incorporated to better align your institution's assessments with these principles?

A number of interview participants have informed about the introduction of formatting moderation board, designing evaluation rubric, arranging practical test for listening and speaking module. Lack of proper literacy or understanding about the recent changes of the evaluation process of language courses have been observed among some newly appointed teachers.

Q. Which principles of language assessment are most strongly incorporated in your institution's assessments, and why?

All teachers agree that validity is prioritized in assessment design and implementation, often guiding test construction and question selection, especially for curriculum alignment and content competence. Although practicality, reliability and authenticity of assessment are marked as fair, washback assessment is not up to the mark as per the findings of this research question.

Q. Do you think this institution lacks in maintaining any specific principles of language assessment? If so, which ones and why?

Almost all interview participants have expressed their concern regarding the assessment process of the speaking test. The less availability of logistics support for conducting listening test have also come up during the interview session of this particular research question. There are also concerns regarding seat arrangement which allows students to engage in cheating during exam. Also, due to lack of experience / training, junior teachers often fail to design test modules that align with the course objective. The class size also becomes a barrier in terms of providing feedback to all students individually that results in washback.

Discussion and Recommendations

This study highlights several important aspects of language assessment practices at the university. A notable finding is that 72% of students had not taken

standardized language tests like IELTS or TOEFL, presenting an opportunity for the university to expand its language courses and testing options. Students reported engaging more in reading and writing tests compared to listening and speaking, raising concerns about the pen-and-paper assessments and the decrease of practical testing components.

Teachers expressed that their assessment procedures align with language assessment principles theoretically but acknowledged lacks in practical application. While students rated 90% of assessments as practical, teachers estimated the figure at 70%, averaging at 80%. Reliability was another area of concern; students rated reliability at 86%, whereas teachers reported it at 60%, resulting in an average of 73%. Large class sizes and limited teacher qualifications were identified as contributing issues. Validity was the most effectively implemented principle, with 98% of students and 80% of teachers recognizing its incorporation, averaging at 89%. Authenticity, however, scored lower, with students rating it at 86% and teachers at 70%, leading to an average of 78%. Formative assessments and feedback were highly valued by students, who rated them 88% effective. However, teachers noted weaker washback effects, particularly among lower-level students, rating them at 60%. Students expressed the need for more detailed and constructive feedback, suggesting room for improvement in this area.

Both students and teachers agreed that the university's assessment practices have gaps, with students estimating 90% adherence to principles whereas teachers report 68%. Teachers also identified challenges such as inconsistencies in rubrics, unreliable scoring, and inadequate test environments. Furthermore, while the university offers several language courses, no clear data was found on the selection process, and limited information exists on the evaluation of speaking tests. Despite these challenges, 88% of students expressed confidence in the university's ability to improve its language assessment practices, indicating potential for development. Teachers indicated that while their assessment procedures align with language assessment principles theoretically, practical implementation is inconsistent. Students rated 90% of assessments as practical, while teachers estimated 70%, leading to an average of 80%. Reliability was another concern, with students rating it at 86% and teachers at 60%, averaging 73%. Contributing factors include large class sizes and insufficient teacher training in evaluation. Validity emerged as the most effectively applied principle, with 98% of students and 80% of teachers acknowledging its incorporation, resulting in an average of 89%. Authenticity was rated at 86% by students and 70% by teachers, averaging 78%. These findings suggest that while the university's language assessments include authentic tasks, improvements are necessary to align with best practices.

Formative assessments and feedback were rated highly, with 88% of students finding them beneficial. However, teachers observed weaker washback effects, particularly among lower-level students, which they rated at 60%. Students called for more detailed feedback, presenting an area for growth in assessment strategies. Both students and teachers agreed that the university's assessments have gaps. Additionally, the selection process for language courses lacked clarity, and little information was found on how speaking tests are conducted and evaluated. Despite these issues, 88% of students expressed confidence in the university's ability to improve its assessment practices, suggesting significant potential for growth and development in language assessment strategies. A new and fresh strategy can be developed to identify the areas of improvement for the current assessment system of the university and move forward by making impactful changes to align with the principles of language assessment.

Conclusion

To sum up, the study evaluates the principles of language assessment in the context of a selected university in Bangladesh. The findings suggest that the university incorporates all the five principles of language assessment with a few drawbacks. Some areas require improvements. The practicality principle is well-oriented and maintained according to needs. But administrators need to re-evaluate the assessment process of writing courses. The subjective tests require more time for proper evaluate. The number of students enrolled in the language courses is significantly high whereas the standard teacher-student ratio should not cross 1:25 for smooth teaching, learning and assessment. The teachers who design and conduct language tests should consider transparency regarding evaluation. For this, well-oriented rubrics are required. Student and rater-related reliability should be incorporated for increasing the reliability principles. Among the five principles of language assessment, the validity principle is incorporated perfectly for the theoretical parts. For practical parts like speaking and writing, the university has a few lackings. The findings show the concern of teachers regarding the accurate evaluation process of speaking module. The researchers have provided numerous suggestions to improve practical tests and increase validity for the practical parts. The university has been incorporating new strategies to maintain authenticity in the evaluation of language courses. The research also suggests increasing more authentic tasks so that the students can communicate for real purposes. The findings from the teachers' interview suggest that students with low performance record are not ready to accept their mistakes. They just want to complete the course without acquiring the competence of language. The main purpose of studying language courses is communicate effectively for real purposes. For this reason, high performing students want more feedback from their teachers. But the university

does not have any scope to provide feedback after the Final Examination. The researchers suggest the authority to maintain transparency of grades after the results are published. If not possible, the course teachers can come up with the screencast feedback. The authority should look into the researchers' suggestions so that maximum benefits come from the language assessment of the university and teachers and students get the opportunity to smooth teaching, learning and assessment environment at the university for the language courses.

References

- 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.
- Bachman, L., & Palmer, A. S. (1996). *Language testing in practice*. New York, NY: Oxford University Press.
- Bachman, L. F., & Palmer, A.S. 2009. *Language Testing in Practice: Designing and developing useful Language tests*. New York. Oxford University Press.
- Brown, H. D. & Abeywickrama, P. (2018) *Language assessment: Principles and classroom practices* (3rd ed.). New York: Pearson Education Inc.
- Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, 32(3), 385-405. <https://doi.org/10.1177/0265532214565386>
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (2002) *Dictionary of language testing*. Beijing: Foreign Language Teaching and Research Press.
- Fulcher, G. (2012) *Assessment Literacy for the Language Classroom*, *Language Assessment Quarterly*, 9, 2, 113-132
- Gronlund, N.E. *Measurement and Evaluation in Teaching* (3rd Ed). New York: Mcmillan Publishing Co
- Gronlund, N. E. (1998). *Assessment of student achievement* (6th ed.). Boston, MA: Allyn & Bacon.
- Harris, David, p. *Testing English as a Second Language*. New York: mc. Graw Hill Book Company.
- Hudaya, D. W. (2017). Teachers Assessment Literacy in applying principles of language In *Proceedings Education and Language International Conference* (Vol. 1, No. 1).
- Hughes, A. (2003) *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Islam, M.S., Hasan, M.K., Sultana, S. et al. English language assessment in Bangladesh today: principles, practices, and problems. *Lang Test Asia* 11, 1 (2021). <https://doi.org/10.1186/s40468-020-00116-z>
- Mousavi, S. A. (2009). *An encyclopedic dictionary of language testing* (4th ed.). Tehran,

Iran: Rahnama Publications.

- Puspawati, I. (2019, October). Understanding teachers' knowledge, skills, and principles on language assessment: a survey on teachers' language assessment literacy. In *Third International Conference on Sustainable Innovation 2019–Humanity, Education and Social Sciences (IcoSIHESS 2019)* (pp. 70-75). Atlantis Press.
- Rahimi, S. A., Razmjoo, S. A., Sahragard, R., & Ahmadi, A. (2021). Development and validation of language assessment literacy scale for high school EFL teachers: A mixed methods approach. *Teaching English Language*, 15(2), 229-264.
- Rahman, A., & Khan, R. (2021). English language testing practices at the secondary level: A case study from Bangladesh. *Challenges in Language Testing around the World: Insights for language test users*, 447-464.
- Ranjit, R. (2022). English language assessment in Nepal: policies, practices and problems (Doctoral dissertation, Brac University)
- Richards, J. C. (2001). *Curriculum development in language teaching*. New York, NY: Cambridge University Press.
- Soatov, I. A., & Togaymurodov, D. D. (2024). Developing Language Assessment: Test Modification project based on five major principles of testing and assessment. *Academic research in educational sciences*, 5(5), 456-467.
- Sumarsono, D., Arrafii, M.A., & Imansyah. (2023). Evaluating the Quality of a Teacher's Made Test against Five Principles of Language Assessment, *JOLLT Journal of Languages and Language Teaching*, 11(2), pp. 225-237. DOI: <https://doi.org/10.33394/jollt.v%vi%i.7481>
- Tosuncuoglu, I. (2018). Importance of Assessment in ELT. *Journal of education and training studies*, 6(9), 163-167.
- H. (2012). *Second Language Assessment for Classroom Teachers*. Online Submission.