

Validity and reliability issues in virtual assessment: A mixed-method approach to investigate HSC level teachers practice in Bangladesh



Khondoker Abid Hasan Tanvir

ID: 201055008

Department of English and Humanities

June 2021

Validity and reliability issues in virtual assessment: A mixed-method approach to investigate
HSC level teachers practice in Bangladesh

A Thesis Submitted to
The Department of English and Humanities
University of Liberal Arts, Bangladesh

In Partial Fulfillment
of the Requirements for the Degree of
MA in English

Khondoker Abid Hasan Tanvir
Student ID: 201055008
Thesis Supervisor: Golam Kader Zilany
Date of Submission: 11 June 2021

Declaration

I solemnly declare that this thesis paper submitted; in partial fulfillment of the requirements for the Masters in Applied Linguistics & TESOL from the Department of English and Humanities, University of Liberal Arts Bangladesh, in the result of my own research work and written in my own language. That no part of this thesis paper consists of materials copied or plagiarized from published or unpublished work of other writers and that all materials borrowed or reproduced from other published or unpublished source have either been out under quotation or duly acknowledge with full reference in appropriate place(s). I understand that the Masters, conferred on me, may be cancelled/withdrawn if subsequently it is discovered that this dissertation is not my original work, it consists of materials copied/plagiarized or borrowed without proper acknowledgement.

Khondoker Abid Hasan Tanvir

Student ID: 201055008

M.A. in Applied Linguistics & TESOL

Department of English and Humanities

University of Liberal Arts Bangladesh

Approval

The thesis/project titled “Validity and reliability issues in virtual assessment: A mixed-method approach to investigate HSC level teachers practice in Bangladesh” submitted by Khondoker Abid Hasan Tanvir (201055008) of Spring Semester, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of M.A. in Applied Linguistics and TESOL on June, 2021.

Examining Committee:

Supervisor:

(Member)

Golam Kader Zilany
Senior Lecturer
Department of English and Humanities
University of Liberal Arts Bangladesh

Dedication

This thesis paper is dedicated to the sacrifices made by my parents who continuously support my work, reinforce my enthusiasm intrinsically and leverage me with additional time by sparing other duties.

Abstract

Over time, global, economic, social and educational diagram has transformed significantly during this pandemic. Whether this transition was smooth or not, in Bangladesh, education system has shifted online without considering the standard measurement of education such as valid and reliable assessment. Assessing students is the method of identifying, choosing, creating, gathering, evaluating, understanding, as well as utilizing data to optimize their performance and progress. In case of online assessment, certain criteria have been suggested such as employing summative assessment for formative practice, combination of different modes, methods and techniques/ instrument, authentic task, formative feedback, multidimensional approach etc. which ensure the validity and reliability of virtual assessment. But in Bangladeshi context, whether this sudden shift from offline to online is able to fulfill the criteria that ensure the coexistence of validity and reliability. To investigate the whole picture of online assessment, this study has incorporate qualitative and quantitative data collected from six purposefully selected in-service higher secondary level teachers. Semi structured online interview and document observation such as question paper and obtained marks of students have used for this study. This study explores the contemporary picture of online English language assessment in Bangladesh. This article discusses, relying on the validity and reliability criteria suggested by various author, what types of assessment they implement, how they integrate technology in assessment, what they expect from the learners and how many criteria of validity and reliability they follow. It suggests that with the blockage of authority, teacher follows the summative assessment which could not match the objectives of NCTB but their own. Being found online exam easier, MCQ exam is more valid as the teacher put more efforts in task and approach designing whereas written exam is designed for offline but conducted online in unproctored manner. I hope these initiatives can help other EFL teachers conducting exam online maintaining the standard quality by fulfilling the validity and reliability criteria.

Keywords: online assessment, summative assessment for formative practice, technology enhanced assessment, validity criteria, reliability criteria

Acknowledgement

I would like to express my all gratitude to the Almighty Allah (Subhanahu Watala) who has given me the ability to complete this thesis paper successfully. I express my deep sense of gratitude and sincere appreciation to my honorable supervisor Golam Kader Zilany, Senior Lecturer, Department of English and Humanities, University of Liberal Arts Bangladesh, for his valuable suggestion, guidelines and co-operation. Without his help and guidance, it was quite impossible to come up with this study. Especially I am really grateful to him for allowing me to work and write on the religious sector of Bangladesh.

All of my thanks go to the concerned teachers of higher secondary level, who have given me their valuable time to explore their world. For this, I thank my Institution for giving me this nice opportunity to see their life world. I am also grateful to the college management who extended their co-operation for data collection and information associate with the study and I would like to thank everyone who had been very co-operative and support me to do this thesis paper. In the writing this dissertation, I have drawn as many books which mentioned in the bibliography, I acknowledge my indebtedness to all those authors for their works which has been great use for me.

I am especially grateful to my honorable father Khondoker Abdul Motin and honorable mother Khondoker Aklima Khanam for their strong stand behind the accomplishment of the thesis paper.

Khondoker Abid Hasan Tanvir

Table of Content

Content	Page no.
Chapter 01: Background of the thesis	<u>01</u>
Chapter 02: Interest and research questions	<u>04</u>
Chapter 03: Literature Review	<u>05</u>
3.1: What Does Assessment mean?	<u>05</u>
3.2: Technology enhanced assessment	<u>07</u>
3.3: Validity and reliability issues of technology enhanced assessment	<u>10</u>
3.4: Technology enhanced assessment: opportunities or threat for a low resource country	<u>13</u>
3.5: Validity and reliability issue in Bangladeshi context	<u>16</u>
Chapter 04: Methodology	<u>18</u>
4.1: Research Design	<u>18</u>
4.2: Sample and Setting	<u>18</u>
4.3: Instruments	<u>19</u>
4.3.1: Interview	<u>19</u>
4.3.2: Document observation	<u>19</u>
4.4: Data collection procedure	<u>20</u>
4.5: Pilot study	<u>20</u>

4.6: Data analysis	<u>21</u>
4.7: Ethical Issue	<u>22</u>
4.8: Limitations	<u>22</u>
Chapter 05: Data presentation and analysis	<u>23</u>
5.1: Thematic analysis	<u>23</u>
5.1.1: Implementation of various assessment types	<u>23</u>
5.1.2: How technology was integrated in exam	<u>24</u>
5.1.3: Tentative outcome	<u>25</u>
5.1.4: Validity criteria	<u>25</u>
5.1.4.1: Task	<u>26</u>
5.1.4.2: Feedback	<u>27</u>
5.1.4.3: Approaches	<u>27</u>
5.1.5: Reliability criteria	<u>27</u>
5.2: Statistical analysis	<u>28</u>
Chapter 06: Summary of findings	<u>30</u>
Chapter 07: Conclusion and recommendation	<u>33</u>

Chapter 01

Background

The ministry of education, Bangladesh embraced communicative learning and teaching approach due to develop learners to be competent holistically in English. In order to meet the goal, the responsible board which is National curriculum and textbook board designed materials and trained teacher so that learners can be able to obtain comprehension over the four skills i.e. reading, writing, listening and speaking after completion higher secondary level (*National Curriculum-English XI-XII*, 2012). The aim of embracing this approach and merging it in education system of Bangladesh is to improve teaching and learning situation and to develop learners competence that they can synchronize with the increasing demand of English in local as well as international (Ali et al., 2020).

Islam (2015) investigated that the implementation of CLT in both urban and rural area is not extracting the meaningful outcome as it expected. In both case, though they used to teach English for 12 years, the adoption of communication language teaching approach by urban and rural schools occurred explicitly in inequality as a valid and reliable learning and assessment tools has been implemented more in urban area. CLT was better carried out the essence by the high-achiever urban school than the rural schools (S.M.Ariful Islam, 2015). In rural institutions, teachers fixated on their own learning experience and the practice of their mutually shared teaching experience among colleagues objectively and cohesively apart from the approaches mentioned in the curriculum (Alam, 2016). Al Amin (2017) has investigated rigorously influential factors of ELT in Bangladesh and exposed the mechanism of teacher's education and their professional implication, the local and foreign demand of communicative skill developments, resource gap between rural and urban, curriculum efficacy, large classroom vs short class time, influence of national assessment framework and the large-scale role of contemporary and global impacts. In rural setting, such assessment echo system has been created, in which all sorts of assessment or learning growth are regulated by the old national assessment framework. This outdated mindset toward assessment framework results in a decline in

curriculum incorporation which is not in compliance with national language policy and curriculum. This present assessment policy leads learners to pre-decided responses indicating that examining the assessment format after the establishment of CLT base curriculum, the massive disparity between what is to be learned and what is assessed in the exam also remains very important (Al Amin & Greenwood, 2018; S. Rahman, 2018). After analyzing the English public examinations of Bangladesh, sultana (2018) remarks that in spite of implementing CTL based education, the examination doesn't concern about what is supposed to be measured. Instead, it promotes memorization or knowledge of forms over communicative skills which violets the prescribed curriculum. Even despite having a high passing rate, it is found doubtful in terms of the quality control measurement such as validity and reliability as it is unable to address the objectives of curriculum following the ineffective procedure (Sultana, 2018). The findings of Mohammad Ali and Sultana (2016) reveal that achieving high score in the exams doesn't ensure the high competency of communicative skills. The basic requirements to maintain validity are not fulfilled completely. The higher secondary level English exams don't contain the content or construct validity by not representing the four skills of English and exam oriented teaching (Mohammad Ali & Sultana, 2016).

However, due to covid 19 pandemic, not only the unimaginable deterioration in health has been triggered but also every sector of human life has been infected and affected that uncertainty becomes a certainty. In response to COVID-19 pandemic, policymakers have stressed on taking several steps forbidding social gathering and concerned areas and institutions such as workplaces, stores, restaurants, libraries, film theaters, all sorts of educational institutes etc. in order to reduce the transmission of covid outbreak. Likewise, the Bangladesh government affirmed to shut down all activities of educational institutes to avoid the spread of covid. But in case of student's uninterrupted education flow and avoiding session jam, significant reforms in the education system were intended to render the system more vibrant. In this regard, the situation was marked by an important transition of many institutes to an immediate online learning situation as technology gives us the best solution for it though no reliable method has developed for this virtual learning and assessment situation (Dutta & Smita, 2020; Ramij, 2020). Ramji (2020) added that virtual classes are conducted in conference

style following the traditional method which is not designed for online learning and assessment. Analyzing the survey results, he recommends utilizing technological advantages to make the vital component such as exams more valid and reliable.

Emon, Alif and Islam (2020) analysis Motijheel Ideal School and College, Cambrian School and College, Viqarunnisa Noon School and College, Willes Little Flower School, and College, and some more Institutions who are conducting online classes and stated that online classes were identified to be both helpful and redundant by a vast percentage of respondents. They identified some problems that very little mutual engagement, disproportionate technological costs, disrupted internet connection, inconsistent participation in online classes, lacks of classroom strategies, mental health instabilities which create hindrance for teachers and render question on the effectiveness of assessment (Emon et al., 2020).

Several reports have shown that deformation between classroom activities and curriculum objectives makes the examination a predictable test. Due to this nature, teachers only focus on some specific tasks or topics from the prescribed textbooks. Though NCTB has designed effective textbooks, those are for face to face classroom and examination, not for online classes and examination which results in a washback effect. Moreover, the contextual obstacles that Ali, Hamid, and Hardi (2018) considered as being the most significant factor to the discrepancy between assessment and learning were identified. Teachers frequently have to keep in mind the conceptual frameworks, which prevent them from including statements or questions in the question paper relating to the assessment of communicative components (Ali et al., 2020; Ali & Hamid, 2020).

Chapter 02

Interest and research questions

It is to inform that in terms of precautions of covid 19 pandemic, the Ministry of Education (MOE) gave permission to colleges (govt. non-govt) to conduct online classes and examinations where teachers need to shift their pedagogy online and to conduct examinations online. Thus present study gives insight of the pros and cons of online assessment, the challenges teachers face to conduct assessment online and the factors of validity and reliability to influence assessment. This study aims to investigate the current situation of assessment through online in higher secondary level. The questions this study hopes to answer are:

1. Do the validity and reliability issues coexist in the present virtual assessment practice in a higher secondary institution during the changed circumstance of this pandemic?

Chapter 03

Literature Review

3.1: What does Assessment mean?

The judges of normality are everywhere. We are in the society of the teacher-judge, the doctor-judge, the educator judge, the 'social worker'-judge; it is on them that the universal reign of the normative is based; and each individual, wherever he may find himself, subjects to it his body, this gestures, his behavior, his aptitudes, his achievement (Foucault, 1975).

Assessment is the crucial part of life as learning takes place in every spare of life, i.e- behavior, body gestures, attitudes, achievements. Learning assessment is regulatory; the factors on which it has been evaluated are anticipated to be influenced. Learning assessment is regulatory; the factors on which it has been evaluated are anticipated to be influenced. Foucault, in his regularization capacity notion, has drawn assessment in education as a crucial attribute in terms of measuring people's holistic growth (Foucault, 1975). Education and assessment are so deep rooted and connected as to be unable to get a whole vision in case of ignoring one of them separately. Assessment is carried out in terms for drawing extrapolations about a participant's learning and progress on a predetermined framework. Assessing students is the method of identifying, choosing, creating, gathering, evaluating, understanding, as well as utilizing data to optimize their performance and progress (Erwin, 1991). Assessment entails rendering the objectives explicitly determining justifiable parameters and standards efficiency of teaching learning consistently collecting, evaluating, and examining data to assess whether the performance meets the objectives and the quality recording, describing, and enhancing performance with the information collected. If assessment is successfully incorporated into broader structural frameworks, it will allow everyone direct the consensual inclination, evaluate common expectation and generate the common educational community committed to ensuring as well as enhancing the quality of education (Alderson et al., 2017; Angelo, 1995; *Faculty Center for Ignatian Pedagogy: Loyola University Chicago*, n.d.; Baird et al., 2017; Banta et al., 2009; Entwistle, 2003; Shohamy et al., 2017). Loyola University recognizes assessment as "natural concern" which helps to identify what and how the learners are being experienced and

how successful the learning mechanism is. They have set the purposes which are to ensure the learning objectives in each academic sector, to examine whether each academic sector has particular strategies to evaluate the mentioned objectives, to produce annual progress reports and provide supplementary information on learning assessment for “assessment on regular basis”, “Reaccreditation processes” and “determination on how the department leverages the information of learners’ assessment for enhancing instructions”, to initiate a vortex of assessment practice due to enhance uninterrupted educational improvement and to reexamine tactical assessment frameworks focusing on collected information as well as existing evaluation procedures (*Faculty Center for Ignatian Pedagogy: Loyola University Chicago*, n.d.).

Evaluation enables the teacher as well as the participant due to monitor the progress in meeting educational goals which can be addressed in several aspects. Summative assessment refers to a technique of measuring the learning knowledge, proficiency, or success of students, usually against predetermined parameters carried out at the conclusion of the course. The objective of summative assessment is to evaluate learners’ understanding of course materials calculating according to the context by degree or grade at the end of the specific controlled training (Dixson & Worrell, 2016; Taras, 2005, 2009). Whereas Formative assessment is a tool which is used to detect misunderstandings, difficulties and learning gaps and evaluates how these shortfalls are resolved. Formative assessment incorporates powerful instruments which, with the proper usages, enables instructors shaping the performance as well as enhancing the capacity of learners to become involved in their own learning (to be autonomous learners) as learners will realize that rather than final examination, learning should be the main concern to be improved (Harlen & James, 1997; Nicol & Macfarlane-Dick, 2006, 2006; Taras, 2005). Systematic review of 300 journal papers on formative assessment Kingston and Nash (2011) has revealed its problematic nature that volume of result sizes makes the assessment intricate to explain the findings. Arbitrated reviews demonstrated that formative assessment is more effective and more successful for Arts faculty than that of science and mathematics faculty. Skilled and online assisted formative assessments are able to extract the most successful outcome with the largest impact sizes (Kingston & Nash, 2011). Educators were unable to make distinction between formative and summative learning assessment and in this regard, Black

(2009) came up with arguments about how to utilize those two notions cooperatively helping each other maintaining their validity (Black et al., 2009). Due to data obtained from educators such as Taras (Taras, 2001) who find it helpful to employ summative assessment for formative practices. Despite this improvement process, the existing assessment systems mandate educators to replicate and duplicate the evaluation procedures, even in case where both summative and formative assessments are required. The apparent requirement for multiplication of evaluation, without creating validity and reliability issues, has become impossible for Formative Assessment implementation: educators who are still abused and overstretched do not voluntarily consent to increase their job pressure. According to Black (2003), educators are, in some cases, convinced of formative assessment as an additional job. Professional experience of educators from several nations shows that relatively numbers of educators are not capable of or prepared for conducting simultaneous assessment schemes because of violating validity and reliability components. Least numbers successfully serve summative assessment for formative practices as per the requirement of coexisting valid and reliable instruments and methods (Black et al., 2003; Taras, 2005; Torrance, 1993; Wiliam, 2000; Wiliam & Black, 1996).

3.2: Technology enhanced assessment

New technological advancements have created a shift among stakeholders' tastes and preferences about how they participate in teaching and learning contexts which resulted in high demand for online mediums. This high demand generates a vacuum of modification in the nature of teaching pedagogy (Robles & Braathen, 2002). Moreover, the apparition of Covid-19 has made many cities phantom as this serious pandemic compelled the maximum part of the world go under quarantine and as a result, educational institutes have also been affected and forced to move from offline to online mode. This disaster has demonstrated the financial benefits of online pedagogy. All educational organizations have to struggle with diverse choices for online pedagogy and to utilize the blessings of technology more adequately (Dhawan, 2020).

Lots of educational institutions, for example universities in China, have shifted their functions online as immediately as they perceive the new requirement in this new normal situation. Teachers also move teaching methods to meet the needs of new normal situations and adjust them to this situation with different strategies due to ensure and to enhance the quality of learning and assessment. To maintain this quality of learning and assessment, how the institution utilizes the technologies to manage education in large scale is the concern, similarly valid and reliable assessment in massive scale utilizing the technology is also important in this regard (Carey, 2020). With the help of technology learning to be effective and stakeholders to be prepared for new normal, strict quality control protocols have been initiated to maintain consistency. Easy accessibility of digital tool and the adaption of them in terms of e-note, recorded lecture, e-books in learning and different evaluation methods such as quiz, game, case study, discussions, brainstorming sessions debates and higher order problem in assessment was pioneer to make the learning and assessment as valid and reliable as offline class (Carey, 2020; Dhawan, 2020). Online education, alias E-learning, can be described in terms of making such a platform where teaching and learning will be more learner centered, creative, interactive and adaptable. This E-learning can alter the teaching-learning environment utilizing various technological platforms with the internet connection where learning will happen experiencing lessons both in synchronous or asynchronous environments. Since synchronous methods open many doors to communicate and negotiate with each other, this environment can be designed in the way that learners experience online sessions, negotiate with instructors and receive immediate feedback. But an asynchronous environment is not well organized as practices are merely possible in that setting which needs different systems and forums (Dhawan, 2020; Littlefield, 2020).

Assessment (summative and formative) in virtual modes besets specific properties, correlating to face to face modes and asynchronous characteristic of functionality among the virtual participants is very distinguished which has been pointed out by Vonderwell, Liang Alderman (2007). Due to developing successful formative assessment to create and to serve asynchronous environments, which will facilitate productive and higher order learning and assessment, researchers need to re-engineer the approaches. Higher order and meaningful

assessments' approaches in a teaching and learning environment are prerequisites for cooperative engagements. This meaningful interaction is essential for socio-cognitive mechanisms, in virtual modes, that are required to promote critical thinking and problem solving which are the desired indicator in terms of higher order learning (Anderson, 1999; Eber & Parker, 2007; Vonderwell, 2007). Successively, based on rigorous review, Kigandi (2010), cited in Baleni (2015), established following ten guidelines for online formative learning and assessment which help the concurrence of validity and reliability (Baleni, 2015).

- Assessment tasks must be genuine, realistic and applicable so that participants can connect the activities with their real life.
- Assessment tasks are supposed to involve, encourage and facilitate the students so that they can develop their own understanding and allow them to be capable of meaning making utilizing their cognitive power.
- Assessment tasks must offer students the chance of developing awareness. Learners must be able to communicate knowledge to certain friends and followers.
- The assessment practices must be supplemented by means of developmental, continual and immediate input. Comprehensive and Immediate feedback is supposed to give to the learners according to their performance, not to their grades.
- Assessment tasks are supposed to be supplemented and supported by relevant, comprehensive rubrics which help learners to achieve the objectives.
- Assessment exercises must provide ways for students so that they can participate in reflective thought. Students should be permitted the opportunity to make decisions based on individual cognitive knowledge.
- The potential of constant recording and assessment of student continuous support and guidance throughout periods must be provided.
- Teachers must be very specific while reinforcing social value and importance in teaching and assessment tasks.
- Assessment tasks must include students in various ways such as rubrics, portfolio.
- Assessment processes must be scalable, including various alternatives and strategies must be appropriate. Learners are supposed to focus on their own perceptions on

particular subjects and the teachers can assess by gazing at the front window of learners' reflection on that.

However, the definition of assessment is reformed in “Evidence of innovative assessment: literature review and case studies” as the mechanism by which educators obtain an understanding of what students have experienced, how well they performed allocated assignments, and the efficacy of content, procedures, and strategies used in their learning (Kapsalis et al., 2019). It also promotes individual growth; to its extent, acknowledging one's achievements is central to establishing a person's identity that ensures long-term learning. Rather than being dependent on the conventional assessment system, the combination of different modes, methods and techniques have been suggested in the name of innovative assessment by Kapsalis, G., Ferrari, A., Punie, Y., Conrads, J., Collado, A., Hotulainen, R., Rämä, I., Nyman, L., Oinas, S. and Ilsley, P (2019).

The development of the ability to eloquently analyze facts and contexts, to develop rational hypotheses, then to function on this hypothesis in any situation is the future learning goals (Boud & Soler, 2016) and to meet the objectives European Commission (2018) has proposed new educational priorities . The assessment outlines the priorities of education since assessment stresses what is truly deserving of, the implementation that frames the learning improvement, and the organization of education. The assistance of innovative assessment (the combination of formative and summative) maximizes the validity, reliability and sustainability of education as it evaluates the complex learning growth from different lenses using different modes, modality, methods and techniques, especially digital tools such as a computer, internet (Kapsalis et al., 2019).

3.3: Validity and reliability issues of technology enhanced assessment

Any mode of learning and assessment has been affected by its validity and reliability as these terms are used to assess the quality of assessments. Validity and reliability describe the consistency of how effectively a procedure, methodology, or examination evaluates anything. The term “Validity” is to examine the measurement of what the learners are claimed to be

evaluated (Rogler, 2014). Messick (1989) stated that it is a holistic measurement of learning and assessment to which scientific and rational data supports the consistency, suitability and adequacy of objectives and assessment tasks matching the results. He extends his notion to that the objective which is basically the viewpoints or measurement indicators we have set before the assessment plays a vital role in identifying validity. With the application of it, validity implies that desired purposes and its enactment must be stated directly (Messick, 1989). Hernández, García, Padilla, García, and Leal (2012), cited in (Marciniak, 2018), described the validity as the tool that evaluates the components which are supposed to be measured. Sophisticated evidence reveals various kinds of validity and methods for defining the credibility of a measuring tool such as examination, assignments. So many protocols implement methodologies which enable consistent performance and ensure its attainment of certain objectives listed. After rigorous analysis, Gikandi proposes a statement to define validity of online assessment which is whether the assessment facilitates the learning or not and he also identified three features of the validity of- online assessment (Gikandi et al., 2011):

1. Validity of online assessment tasks rely on whether the tasks make the participant involved in applying the real life language interpretation and analysis such as problem solving, critical thinking. It also relies on the genuineness of the process and tasks. Online formative evaluation must facilitate engaging opportunities for students by means of a number of genuine activities, thus fostering involvement and practicability.
2. Successful performance feedback, that shows if feedback has the feature of dynamic, time orientation and allows students making changes accompanied by excellent rubrics, exchanged among students and facilitators. Specific traits like quickness and ability to review and rewrite the unpleasant assignments which have already been delivered are needed for online formative assessment. Digital networks have multifaceted tools for exchanging and reviewing to facilitate with openness and versatility of the rubrics.
3. Diverse techniques of assessment and Multi modal concept of tasks for assessment that will help students to illustrate ones skills and facilitate to be autonomous learner. Virtual platforms

promote multi-dimensional experiences by providing students with diverse technical tools in using different techniques for developing as well as demonstrating developmental skills.

On the contrary, reliability means to the extent that the progress of learners has been evaluated consistently using various assessment tools. Learners will secure the same grades in reliable assessment if the same test is conducted over two days or in separate but similar ways (Scheerens et al., 2007). In this regard, the exam is supposed to be sufficiently lengthy to examine all materials upon which learners have been assessed so that most of the students have enough time for it to complete. The assessment topic is not supposed to be so simple or so complex for the participants, the problem highlighted on the exam paper should not be complicated, instructions have to be straightforward and grade needs to be broad (Rogler, 2014). Driessen, van der Vleuten, Schuwirth, van Tartwijk and Vermunt (2005) sought to rethink the concept of reliability throughout the framework of formative assessment. According to them, reliability of online assessments is what is supposed to be measured needs to be consistent, stable and adequate to evaluate that level of information framework which has been formed and collected (Driessen et al., 2005). Variables that harm the reliability

1. The duration of the evaluation – a lengthy evaluation usually leads to very accurate outcomes.
2. Appropriateness of problems or activities that are being set to assess the participants.
3. The explicit instructions and terms that are used in assessing.
4. Consistency of assessment –Equal chance to show the progress and outcome, sufficient duration provided to examination, guidance offered to participants prior to the examination
5. The construction both of measuring timeline and measurement techniques for grading. The scoring should be analytical.
6. Learners' preparation for examination, for instance, immediate after disasters is not the appropriate time for examination.

Now Moss (1994) has raised a question “can there be reliability without validity” and using the hermeneutic a method, he concluded with a positive attitude if the definition of reliability is the consistency of measurement “consistency among independent measures intended as interchangeable”. But Hang Li (2003) disagree with the definition of Moss (1994) and preferred to Guilford (1954, p. 350) who has come up with psychometric definition of reliability in which it is “the proportion of true variance in obtained test scores” where the term “consistency” is missing. After analyzing Moss arguments, Hang li stated that Moss has strongly affiliated between psychometric and hermeneutic methods while constructing the argument of correlation between reliability and validity where reliability is portraying the psychometric method. In this case, no conflict has been found that denies the presence of validity. He extended his argument describing the correlation of validity and reliability to those different portions of assessment systems measure different but limited areas of behavioral domain within the assessment protocols. Assessments have to encompass a wide area of behavioral domain to maintain better validity and this can be increased by setting the units on different behavioral domains in such a way so that they don't overlap. In a group of psychometric assessing units, the interrelated units, possessing internal consistency, constructs a composite (group of selected units set for assessing the progress) which reflects that actual grades, extracted from the composite, are similarly correlated. In this case, the validity of that group (instrument for assessing) will increase, whereas the reliability will fall. In other words, the higher the validity, the lower the reliability (Guilford, 1954; Li, 2003; Moss, 1994).

3.4: Technology enhanced assessment: opportunity or threat for a low resource country.

In the technology-integrated teaching and learning situation, the measurement of students' performance through technology (e-assessment) is reliable and valid as offline assessment. In this regard, Hewson and Charlton (2019) highlight a piece of evidence where the modality (online vs offline) effect of assessment has been depicted. Two groups of students have participated in the offline exam and web-based online exam respectively and the outcome was that online evaluation would not appear to offer participants any cumulative adverse effects. He also mentioned that the online version of the assessment could be more reliable

and valid in the case of proctored computerized assessment instead of unproctored assessment. The modes in which the learning is administrated shape the validity and reliability of adopted assessment modes. Their study proved that there are no relations found between online and offline assignments grade as the online evaluation approach has not been used to distinguish students preferentially with particular computer-related attitudes such as computer anxiety and computer engagement. (Hewson & Charlton, 2019).

Similarly, In the United Kingdom, it is estimated that approximately 1000 computer-assisted evaluations are conducted per day. So, declaring Computer-based assessment as the alternative to traditional assessment, Öz, H., & Özturan, T. (2018) outline that because of the increased utilization of the internet and technology as instructional tools both within and outside the classroom, computers have become effective assessment options for teachers. Some experiments applying computerized assessment differentiating to paper based assessment have been performed to assess coexistence of its validity and reliability. According to some reports, the two experimental modes had a major impact on test results (Choi et al., 2003) and some experiments came up with the reverse conclusion. Conducting the same assessment in both computerized and paper based modes to extract the reliability and validity, Al Amri (2008) analyzes the consistency of those assessments with Cronbach's alpha and concludes with the internal reliability coefficients ranged between .57 and .70 (Al Amri, 2008) which is not impressive. But in the term of validity and reliability coexistence, Öz, H., & Özturan, T. (2018) invigilated similar types of experiment on two groups of students and using the split-half reliability method, he indicated that assessment in both modes had high levels of reliability coefficients ($r_p = .756$, $r_c = .903$). This results in that assessment have significant coexistence of validity and reliability, internal coherence and good relation on both computer and paper-based assessment. It is also noticed that the performance of participants was not influenced by computer-based or paper-based assessment. Besides, the computer-based assessment also gave a boost to the test consumers because they received better results than a hardcopy listening test (Öz, 2018).

However, to maintain the reliability and validity of assessment in online classes, one of the trickiest challenges is handling a “massive open online class” (MOOC). Here, "Massive" is the potential for learners' recruitment and the monitoring of huge amounts of participants' involvement and success results for a significant proportion of students. 'Open' applies to the close to zero participation expense that students think appropriate and resources available for anyone with sufficient internet connections to attend. Researchers see this challenge as an opportunity to develop hybrid model which incorporates MOOC's guideline of face to face training or micro-school interactive classroom that have participant input and assistance. These frameworks retain its benefits for small group participants, suggestions and engagements or teacher-learners negotiation. It also reduces the need for lecturing and facilitates teachers to invest extra time to communicate with participants, to measure their performance, as well as their desire to share insights and to receive input (Bettinger et al., 2017). The scale of the class does not affect the consistency of the course that includes grades, preservation or continuing in massive online classes. Research also indicates that significant learners' performance when retaining a decent work pressure and performance in massive open online class (MOOC) is tough to accomplish. With the aforementioned parameters such as faculty work pressure, grades, retention and performance that declare the quality of learning growth, Lowenthal et al. (2019) has examine the assessment quality on 148 teachers and more than 30 students. The extracted result is that hybrid model including multiple-choice exams, written assignments quizzes, discussion boards and group assignments is followed to document the learning objective and small group discussion and creating feedback boost the persistency. This both way measurement ensures the synchronicity of reliability and validity of the assessment. But result of the workload and technology anxiety on teachers depicted the influence on the assessment which questioned the assessment validity (Lowenthal et al., 2019).

On the contrary, the inconclusive manner of the impact of class size on assessment is documented and discussed rigorously by Gannaway (2018). Teacher extracts the maximum outcome in case of small class size where enabling teaching estimates good results for large classes. Researchers noticed that large classes are embraced more favorably by the teacher; on the other hand, some showed that small and large classes are more favorable than the

medium class size (Gannaway et al., 2018). Additionally, In terms of management and operation, online education is distinct from conventional education. Online teaching is not identical to conventional systems and so the same content measurement processes and planning cannot be used for all education modes.

3.5: validity and reliability issue in Bangladeshi contexts.

To describe the importance of Assessment literacy, Popham (2004) has characterized the teachers with the paucity of sufficient knowledge in assessment by saying “Professional suicide” (Popham, 2004). It has essential importance throughout the framework of high stake exams to select such educators who analyze and maintain the quality of exam ensuring the reliability and avoid approving them without challenging their validity. Analyzing the comprehension about assessment of the sampled teachers who have 4 to 15 years teaching experiences, Sultana (2019) has drawn the line that assessment strongly shapes the teaching and learning in highly test oriented county such as bangladesh (Sultana, 2019). When ELT is concerned in such a low resource country like Bangladesh, English language teaching and learning is stated to have been struggling dismally and Ali, Hamid & Hardy (2018) blamed this failure caused because of an inappropriate assessment system. According to these experts, various factors are working behind which are standard assessment, knowledge of assessment and lack of concern about maintaining the quality. They have also investigated the approaches of assessing and the assessing items that are basically the imitation of the same process over and over again at different stages of schooling (Ali et al., 2020).

In order to demolish the effect of the pandemic, maintaining the flow, Bangladesh needed to go under lockdown. As the education field will be the sensitive field greatly affected by this pandemic, all educational institutions have been shut down. Due to mitigate the negative effect of the pandemic on education, several Bangladeshi universities have decided to run their programmes online (M. M. Rahman et al., 2021). Majed, Jamal and Kabir (2020) identified three following obstacles with implementing the programmes online in the context of Bangladesh: educators or facilitators are being restricted within the option opened by technologies, recognizing learners’ ease of using various technological learning and assessment tools and

ensuring the quality of assessment. The Institutional Quality Assurance Cell (IQAC) in UAP has added additional policies to cope with the new normal situation. To appropriate forms of lesson transmission, IQAC offered a particularly customized guidance to avoid jeopardizing the quality of education. The strategies for assessment will be student evaluation by questionnaires initiated with particularly adapted instruments, verbal examinations, team ventures talks, open book examinations, online proctored tests accompanied with vivas, tasks, community activities, etc. these varied methods of assessment possibly increase learners' awareness of learning (Majed et al., 2020).

Hewson (2019) stated that variables which are against the online assessment can hamper the learners' performance in assessment. In that case, validity and reliability of that online assessment and the instruments are being questioned. Despite learners participating in both online and offline assessments demonstrated relevantly similar average grading, learners who took online mode for assessment did well in completing the task but slight reduction has been noticed in performance in exam rather than learners who took offline mode of assessment. As far as it is explicit in his argument that online assessment is a crucial component of this virtual education, whether it possesses validity and reliability issues as similar as offline mode (Hewson & Charlton, 2019).

The reviews of above literature give explicit overview on the coexistence of validity and reliability of assessment which reveals the measurements or factors in terms of assessment prerequisite, process, technology literacy and how validity and reliability coexist in assessment; that facilitate the quality of online assessment. Researchers encourage fulfilling the prerequisite of using summative assessment for formative practice though teachers are not ready because of workload. In the matter of validity and reliability of online assessment, religious literacy on technology delivers it opening many doors to explore the learning growth in massive scales using the combination of different modes (online/offline), methods (synchronous/asynchronous) and techniques (quiz, game, case study, group discussion, proctored exam, higher order problems). However, inconsistency in assessment activities, grade distribution, duration, instrument and instructions, caused by workload, anxiety, not being tech savvy,

massive classroom and several limitations, can harm the reliability of assessment which is contributed by the present study of Driessen. On the other hand, authentic tasks, effective feedback and multimodal approaches are the foremost components in reference to ensure the validity as Gikandi mentioned. While online assessment is embraced widely, it is questioned the sustainability of validity and reliability in online mode as similar as offline mode.

Investigating the current situation of online assessment in Bangladeshi context, this research is intended to measure the quality of assessment and sustainability of validity and reliability, using Gikandi's (Gikandi et al., 2011) requirements of validity and Driessen's factors of reliability (Driessen et al., 2005). Respectively the measurement of high level coefficient will also be investigated using the statistical framework of Hüseyin Öz and Tuba Özturan (Öz, 2018)

Chapter 04

Methodology

4.1:: Research design

Qualitative and quantitative data will be collected and analyzed following the design to understand a research problem more completely and comprehensively. This small scale mixed method research has been used in order to explore the problems in different lenses, to get valid data and result for this study. Teachers' observations, experiences and opinions on online examination are considered as Qualitative data that has decoded whether the facilitator maintains the codes of the validity of assessment as well as of reliability. On the other hand, quantitative data, extracted from the exam techniques and feedback obtained by students, has revealed the consistency level by conducting coefficient analysis.

4.2:: Sample and Setting

The sample for this study consists of six teachers selected from different parts of Bangladesh; four teachers of them are from Dhaka city, two teachers are from Madaripur.

Though the setting of the sample was randomly selected on the basis of availability, the sample teachers have been selected purposefully and recruiting of participants will occur through the database of the available teachers who have:

1. At least B.Ed qualified of relevant field;
2. at least one years experiences;
3. Completed at least English Honors
4. been working at a Higher Secondary Level.

The sample details are provided below.

Name (pseudonym)	Setting	Applied instruments	Status
Bristy	Urban	written	In service (English) – 4year experience
Raj	Urban	Written and MCQ	In service (English) – 7year experience
Rahman	Urban	MCQ	In service (English) – 5year experience
Rubel	Urban	MCQ	In service (English) – 13year experience
Asad	Semi Urban	MCQ	In service (English) – 8year experience
nayim	Semi Urban	Written and MCQ	In service (English) – 5year experience

Moreover, Question papers of online exams and obtained marks of each activity by the students are considered as samples for this study.

4.3:: Instruments

4.3.1:: interview

Interview helps to investigate problems in depth, to reconnoitre “how and why people frame their ideas in the ways that they do, how and why they make connections between ideas,

values, events, opinions, behaviours, etc.” (citation). The literature review reveals that due to investigating the identified problem, the experience, observation and opinion of the teachers who have already conducted exams online are needed to decode the necessary information. Thus, the researcher found a semi structured online interview convenient for this study. In order to conduct a semi-structured online interview, a questionnaire consisting of fourteen open- ended questions was designed which took almost 50-60 minutes to complete. The questions were designed keeping in mind the codes’ map extracted from the literature reviewed for this study. The codes’ map was generated with the help of Nvivo(qualitative data analysis computer software).

4.3.2:: Document observation

Several documents related to online exams such as question papers of written exam, instructions for assignment, MCQ question as well as the students’ obtained marks of each activity have been observed due to statistical analysis. These documents are generated by the samples following the institutional and government guideline.

4.4:: Data collection procedure

Initially the researcher needed to match the criteria to identify the best fit . As the target samples engaged with their busy schedule, the researcher needed to contact them for their convenient time slot. A synchronous digital platform “google meet” was chosen with the consent of the participants. After getting their appointment, a google meeting link has been in their private contact and Online audio interviews took place.

The interview had three parts: researcher disclosed the objectives and asked ethical consent at the introductory part, later on, after questioning and answering, researcher asked for the mentioned documents from the interviewee and they have delivered the documents in “Doc” and “Jpg” format under the condition of not disclosing it as the documents are sensitive for the institutes. The whole interview was recorded with different screen recording applications such as “Camtasia” and “Anymp4 screen recorder”.

As per the prerequisite of quantitative data, the researcher has asked for the question papers or assignment instructions and the obtained marks of seven to ten students, of course with their consent. After discussing with their authority, they provided the sample information through email which is used to generate quantitative data.

4.5:: Pilot Study

The data effectiveness relies heavily on the correct text of the statements made by the interview. Queries for the interview have been piloted to test the interview framework and to get insight about whether the information guides to achieve the research goals, whether the data are influenced by the methodology in terms of assessment's validity and reliability. It is to introspect logistical problems that can be overcome prior to the main research.

The researcher used the pre-testing on the instruments used in this study. After conducting the test on interview, a slight modification was required as the interview retrieved a lack of few information. In order to reach the objectives of this study holistically, few sub-questions were added in the questionnaire to generate the accurate codes.

On the other hand, required data has been extracted from the question paper and obtained marks. The piloted sample, consisting of 12 items which covers 100 marks, was splitted into two parts: odd-numbers and even-numbers and then the reliability coefficient was calculated utilizing the split-half reliability method to identify the internal consistency level. This pilot test has obtained effective psychometric properties ($r = .751$) (Figure 01) which is the nearest score of OZ (Öz, 2018).

All the data, displayed above, has led the researcher towards retrieving accurate codes to reach the objectives of this study. That makes the methodology accepted as reliable.

4.6:: Data Analysis

Primarily the data (qualitative), generated from the interview by purposive transcription, is converted into the codes and categorized according to their themes which are sorted following the codes' map formed by Nvivo. Later on, the data is discussed on the basis of classification of the themes and analyzed to develop the central idea of the present online exams compared to the theoretical framework of this study.

Using the SPSS software, the required quantitative data are extracted from the both samples in two ways. First, the mean, median, mode and standard deviation have been calculated with the data collected from the interview to get a central understanding about what teachers' and institutes' drive from online exams. Second, it also calculated reliability coefficient (r) from the question papers and obtained marks. In this process, researcher has calculated the reliability coefficient (r) of each sample and then all samples collectively.

4.7:: Ethical Issues

This study has been conducted following the data collections and thesis writing regulations of the "University of Liberal Arts Bangladesh". University has instructed APA 7th edition for in-text citation and reference which are properly incorporated in the process of literature writing, references and gratitude. In the data collection part, researcher has informed the objective of the study, how the data will be used and the duration of reserving the data. Then the oral consent from every potential participant has been recorded before performing the interview. The sample was chosen not by force and allowed to discard the interview in case they find something inconvenient. The identity was disclosed so that they can provide information without hesitation or manipulation.

At the end of the interview, the question papers and the obtained marks have been asked from them. They have discussed the procedure and provided me the samples under the conditions of anonymity of the provider, not disclosing or displaying the main sample as these sorts of information are quite sensitive for the institutes. And I have ensured them to use these information only generating the quantitative data, only that is displayed in the analysis section.

4.8:: Limitations

More information will generate more data for this study which will provide more lenses to explore the central situation. Due to reducing the spread of covid 19, lockdown has been embraced for pre-cautions which has created some obstacles to find and to reach more available potential samples. By this time, being completely dependent on the internet, frequent load shedding and poor internet facilities were other major hindrances for data collection.

Chapter 05

Data presentation and analysis

5.1:: Thematic Analysis

5.1.1:: Implementation of various assessment types

It is evident from the study that formative and summative assessments have distinct objectives. In terms of innovative assessment, the burden of the summative evaluation typically provides much room for implementation of formative assessment, especially if it is related to frequent quality evaluations.

When assessing the learning growth virtually, a tendency of using two techniques, i.e, Multiple choice question (MCQ) and Tradition writing exam have been found common which have, of course, several failings of the validity and reliability coexistence. In order to avoid that maximum preferred MCQ technique. For example, Rahman said “instead, questioning the subject matter, we generate questions on the mechanism of the literary terms to judge the skills in English second paper”. They conducted the MCQ exam on 30 to 50 marks compared to

100 percent. As they use google form for this, the grade is generated automatically. Though it is tough to evaluate the consistent performance of learning, meaning making and cognitive knowledge in the form of MCQ technique, Rubel claimed, “I tried to incorporate the mechanism of the literary terms with subject matter in the form of MCQ so that students need to brainstorm before answering”.

Two out six teachers follow both MCQ and written exams for assessing the english reading and writing skills despite facing technical limitations. For instance, Raj said, “We also conducted written exams keeping the webcam or camera open”. Asad said. “we provide time bound writing tasks which they need to submit when the college will be opened”. This real time exams following the old school technique are assessed summatively that indicate towards the existing assessment systems mandating teachers to replicate and duplicate the evaluation procedures. For example, Bristy, who used only written exam technique, said,” for english, we conducted exams online using the same technique we used for offline”. They provided the question paper, generated following the traditional patterns, on facebook groups and students needed to capture the written answers and to upload them in the group. However, all the exams are conducted after a specific time period such as after six months. In this case, observing their techniques, it can be concluded that they have tried to perceive learners’ grasp of course content by generating a degree or grade based on the context at the end of a specified program.

5.1.2:: How technology was integrated in Exam

Teachers have to follow the instructions provided by the institutes they are working in, which didn’t let them incorporate the vast opportunities of online mode. Bristy, Raj, Asad and Nayim said almost same, “I need to follow the institute’s instructions”. Rubel replied that they can’t try something unless the authority permits them. Bristy and Nayim said,”I suggested google classroom but senior teachers didn’t allow it. I think they find it complicated” though all the institutes have ict department that controls everything related to online. It also plays the role of Exam control management. They provided training only on how to conduct class and exam in above techniques rather than how to create both synchronous and asynchronous

environments. For example, Raj said,” the ict department has demonstrated before us how to conduct Exams through Google form and facebook group”. Rahman said.”Our ict department has provided only a video tutorial from online class and also arranged a mock test for students to show them how to participate in exams”. He has also shown the whole process. They prepare two or three sets of question papers and input them in google form with the answer. The ict department provides the link of google form to the students who only get the question and chance to submit it one time within a particular time. In case of a written exam, a facebook group is used to provide the question papers, to collect the answer scripts and to publish the grades. The whole process concludes that they are still following the traditional pattern in online mode. Nayim claimed that nothing has changed, we have just shifted from offline to online. Bristy describes that they divided the number of students into several sections and allocated each section to one teacher who is responsible to assess them. Students get the question paper through facebook group and submit the computer composed answer and also get the feedback through email.

Moreover, no teacher has received training to design tasks or to prepare questions/assignments for online exams. And surprisingly institutes didn’t intend to use the vast opportunities of technology though technology opens so many doors that facilitate teachers to explore more using different lenses. A sample has claimed that once he received professional training for designing tasks from teachers training college (TTC) which he wanted to incorporate but failed under the pressure of the institute”.

Nevertheless, teachers along with the institute find the MCQ and written techniques, in case of online exams, easier, more convenient and ,most importantly, time saving which can be utilized in combining different modes, methods and techniques.

5.1.3: tentative outcome.

While designing the tasks or preparing the questions, the teacher needs to be explicit about what is being assessed and how s/he will assess as assessment is the process of meaning making, of using cognitive knowledge and of measuring consistent performance. A common intention has been noticed that all want to make students think or analyze before answering; in

both mcq and written exams. Asad sample claims, “I design questions in such a way that none can copy from elsewhere as the internet is an open door.” Rahman said, ‘i alter the information in the question that they need to analyze it.’ Bristy also expresses that he wants to see the reflections of what he taught. But with the pressure of institutes, it is really tough to incorporate technology with course materials. Rubel has claimed that it is exam/result oriented as per the government instructed i need to follow the decorum”. Raj reveals that the institute compels on engaging the students with study rather than on the outcome.

5.1.4:: Validity criteria

The key components of exams are task, feedback and approach which ensure or enhance the validity of exams.

5.1.4.1:: Task

Task unlocks the door to measure the exact quality of learner’s cognitive and linguistic acquisition. Authentic, practical and encouraging tasks help teachers to measure what is supposed to be measured. Several tasks can be designed according to the needs of learners. In case of MCQ based exam, the MCQ questions are designed in various ways such as multiple choice, answering questions, and filling in the blanks (with clues/without clues). As Rahman said,” actually we need to design questions according to google form format as the college follows”. On the other hand, those who design questions for the writing part need to maintain the NTCB format. Bristy explained, “Need to strictly follow the NCTB instruction”.

Moreover, the MCQ sample reveals that they have set some practical problems of English grammar with the subject of current affairs. To answer the problems, learners need to incorporate the grammar rules within the time. The subject matter and implementation of grammar rules in making the questions found very effective and those who conducted MCQ exams claim that they designed the question so that students need to brainstorm or analyze the information. This makes this instrument as valid as offline assessment (Carey,2020). But in case of writing part including 1st paper and 2nd paper, they follow the traditional tasks such as, Communicative tasks in 1st paper and paragraph, email, letter etc in both 1st and 2nd paper.

For English 1st paper, they design the task based on the topic from the textbook and generate questions to evaluate that students understand the topic and can communicate with others in real life though that is not possible because of online mode, claimed by a sample. Another said that student cannot connect the task except the subject matter which questions its practicality. In writing part of both 1st and 2nd paper, tasks don't encourage students to apply their cognitive knowledge, rather it is predictable and students memorize them to answer in the exam, claimed by two of the samples. Overall, Higher order questions, problem solving, critical thinking and practical tasks were not in adequate amounts. In addition, the organisation of tasks found lower to higher order and most of them said that they used 70% standard questions and 30% below standard questions as per the requirement of NCTB.

5.1.4.2:: Feedback

Feedback explains the gap and shows the way of further improvement. All the samples used to provide numerical grade which is not comprehensible where to improve. They provide a solve-class on the exam question in lieu of comprehensive feedback. For example, Rahman said, "actually there is nothing between right and wrong in mcq technique. So we can't provide any other feedback." Nayim said, "we can't provide any other feedback except grading as we need to deal with 100+ students at a time". The students of massive open online class (MOOC) can understand their grading if the exam is supported by comprehensive rubric. But no teacher has provided such support. Bristy said, "we don't provide rubric support but few instructions such as one mark will be deducted for each grammatical mistake".

5.1.4.3:: Approach

From the study, a multidimensional approach is preferred due to ensuring its validity. The open internet gives the chance of using versatile tools to assess learners but they use only various types of questions rather than multimodal approaches.

5.1.5:: Reliability Criteria

In the MCQ exam, duration was allocated on the basis of marks. For instance, the exam of 30 marks has a time limit of 30 minutes. Besides, authority compels on completing Written exam within 2 hours though they are flexible to submit the paper within the whole day of exam if students face any technical problems which helps students to be anxiety free during the exam. However, Rahman refuses to be flexible to prevent the trend of submitting the answer late with the excuse of technical problems.

The consistency has been found in using the digital tool in case of each sample. Those who conduct exams through google form have always used this digital tool for exams.

The problems/activities used in exams are found appropriate for reading, writing skills and assess the particular topic/subject knowledge which don't match the objectives set by NCTB. As it claims to develop all four skills to make students autonomous, the exam technique, task, feedback and approaches reveal that it only concerns reading and writing skills which don't match the needs of students.

Moreover, the exam instructions are quite explicit and comprehensive.

5.2: Statistical Analysis

This section demonstrates the psychometric consistency of the performance analyzing the reliability coefficients based on both MCQ and Written exam question papers and the obtained marks in order to supplement the result of thematic analysis. Four samples (two MCQ and two written) and eight to ten students' obtained marks have been analyzed. The coefficients of reliability and validity are supposed to be near and no statistically noticeable disparities are supposed to be shown as compared to OZ's score. For each assessment technique, split-half reliability approach was used.

MCQ

Reliability Statistics

<u>Cronbach's Alpha</u>	Part 1	Value	.689
		N of Items	15 ^a
	Part 2	Value	.440
		N of Items	15 ^b
Total N of Items			30
Correlation Between Forms			.434
Spearman-Brown Coefficient	Equal Length		.605
	Unequal Length		.605
<u>Guttman Split-Half Coefficient</u>			.574

⊕

a. The items are: Question_01, Question_03, Question_05, Question_07, Question_09, Question_11, Question_13, Question_15, Question_17, Question_19, Question_21, Question_23, Question_25, Question_27, Question_29.

b. The items are: Question_02, Question_04, Question_06, Question_08, Question_10, Question_12, Question_14, Question_16, Question_18, Question_20, Question_22, Question_24, Question_26, Question_28, Question_30.

□

Figure 01

Written

Reliability Statistics

<u>Cronbach's Alpha</u>	Part 1	Value	.835
		N of Items	4 ^a
	Part 2	Value	.893
		N of Items	4 ^b
Total N of Items			8
Correlation Between Forms			.816
Spearman-Brown Coefficient	Equal Length		.899
	Unequal Length		.899
<u>Guttman Split-Half Coefficient</u>			.882

a. The items are: MCQ, With clues, Rearrange, Letter.

b. The items are: Answering question, Without clues, Paragraph, Story.

Figure 03

After analyzing with Spss, the two MCQ exams score $r_{mcq}=0.434$ which has slight differences as compared to OZ score. On the other hand, the two written exams score $r_{written}=0.816$ which is the closest scores as compared to OZ's score.

The lower score of MCQ exam possesses the lower reliability coefficient as well as higher validity. On the other hand, the higher score of the written exam possesses higher reliability coefficient as well as lower validity. It can also be concluded that the consistency of the performance has been violated as the score varied between two different techniques. But in the case of each technique, a close value revealed the internal consistencies of the tasks because of having a strong, positive coefficient. In comparison to MCQ with written exam, written exam scores are higher than that of MCQ exam which means that students perform better than that of MCQ. All the participants reveal that the written exam online has vast opportunities to answer with dishonesty as the internet is an open door and they don't get the chance to invigilate. Besides, MCQ is a time bound exam and teachers prepare questions where students need to apply and analyze the rules and cognitive understanding within that time. However, the validity of that group (instrument for assessing) will increase, whereas the reliability will fall.

Chapter 06

Findings and Discussion

After analyzing the collected data and matching the data with criteria of validity and reliability, Different observations have been gathered and examined. On the basis of analysis and information the findings which were placed below-

1. The tendency to assess the performance summatively is common to all. Though everyone finds conducting online exams, preparing questions, managing the technique and scrutinizing the answer script easier, no trace found to include formative assessment along with summative assessment. But assessment is supposed to evaluate maximum number of common objectives to ensure and to maintain the quality of education (Alderson et al., 2017; Angelo, 1995; *Faculty Center for Ignatian Pedagogy*:

Loyola University Chicago, n.d.; Baird et al., 2017; Banta et al., 2009; Entwistle, 2003; Shohamy et al., 2017). To certain that, it is required the coexistence of validity and reliability employing summative assessment for formative practice (Black et al., 2003; Taras, 2009; Wiliam & Black, 1996) which is maintained partially.

2. Though the mode of assessment doesn't affect the consistency of performance (Al Amri, 2008; Choi et al., 2003; Öz, 2018), faculty workload while conducting assessment in massive scale has significant impact on grading and feedback (Lowenthal et al., 2019). The data shows that teachers are comfortable with online modes and found it easier to conduct exam online than that of offline exam. So, having less workload, teacher has the opportunities to utilize the technology to assess learners in massive online class (Baleni, 2015) but authority didn't provide any specific instructions or training.
3. The influence of authority, which was explicitly strong on teachers, has been found blocking the path of exploring new techniques for online assessment. Teacher needs to deal with large scale of heterogeneous students. Moreover, using the full potential of technology can leverage teachers to deal this massive open online class (Carey, 2020; Dhawan, 2020) where learning will happen experiencing lessons both in synchronous or asynchronous environments (Dhawan, 2020; Littlefield, 2020). But teachers didn't get the chance to incorporate or explore new techniques to cope with new normal as the accountability towards the authority didn't permit them to do so.
4. It is evident that disregarding technology in exams as digital technique, grading, and feedback, the coexistence of validity and reliability relies on using technology in the form of a combination of modes (online or offline), methods and techniques (Boud & Soler, 2016; Kapsalis et al., 2019). Completely depending on synchronous methods has declined higher order assessment (Eber & Parker, 2007) which is ensured by asynchronous methods (Vonderwell, 2007). The amalgamation of both methods leads to generating several technology integrated techniques such as game, viva, case study, and brainstorming debate. It is found that both methods have been integrated but not completely. Asynchronous method (written) is a predictable assessment which leads the

learner towards memorising or following dishonesty while answering. The applied synchronous method (MCQ) is designed where students need to incorporate their cognitive knowledge and developed skills within a limited time (Baleni, 2015; Gikandi et al., 2011) which is effective for assessing the learning though it has no practicality. However, it is surprising that when a teacher puts effort into developing technology-integrated methods and techniques, the consistency of performance is declined. In contrast, where the teacher puts less effort, the consistency of the learner's performance is escalated.

5. The common objectives of e-assessment is to follow the process of meaning making, incorporating the cognitive power and decision making rather than grading only the subject knowledge (Gikandi et al., 2011) which is also the objectives provided by the NTCB. Rather focusing on achieving objectives, the authority imposes the teacher to make the students engaged with the study and cut a good figure. Despite this exam oriented mindset, a tendency has been noticed to prepare the question paper so that students need to perceive or apply the learning and meanwhile, cannot answer by unfair means which make this exam equally valid as compared to offline (Carey,2020). But it is explicit that to meet the objectives, a consistency among the digital tools, techniques, methods is supposed to be concurrence with the process of question making, subject matter and practicality. The reliability coefficient score reveals; that is missing in the explanation of samples.
6. Authentic, real-life engaging and encouraging activities provided with ongoing feedback and multidimensional approaches are crucial for ensuring coexistence of validity and reliability (Baleni, 2015; Driessen et al., 2005; Gikandi et al., 2011). The tasks in mcq exams are genuine and encourage critical thinking but have less practicability and diverse approaches. As students get no opportunity to answer by unfair means, the outcome is more genuine and valid because it ensures the learning happens with authentic task, comprehensive feedback and various approaches (Gikandi et al., 2011; Öz, 2018). Whereas, the written exam is completely designed with predictable

traditional technique where validity is harmed by lack of invigilation. Matching the objectives, explicit instructions, students friendly duration and appropriate tasks (Driessen et al., 2005; Öz, 2018), this technique maintains more reliability criteria than that of validity, which is also supplemented by the reliability coefficient score. However, Mcq exam is more valid than that of the written exam.

In all cases, students get only numerical grades instead of getting formative feedback along with summative grades. The consistency of grading is not maintained and lack of comprehensive feedback makes the exam process invalid.

Chapter 07

Conclusion and recommendations

This study was outlined to explore the current online assessment in terms of the coexistence of validity and reliability identifying the major aspects of validity and reliability of assessment in the context of Bangladesh. The applied instruments have been measured through various lenses such as implementation of various types of assessment, technology integration, validity and reliability criteria. This investigation reveals that rather following both assessment types(Black et al., 2003; Taras, 2009) authority compels teachers just to engage students with study, though technology offers vast opportunities to evaluate learning growth with validity and reliability by integrating technical tools (Boud & Soler, 2016; Kapsalis et al., 2019). Instead, they used traditional exam pattern and MCQ style exam in online forms which are the instruments to assess summatively. Analyzing the benchmarks of those exams, MCQ exam

follows the pre-requisites of authentic tasks and feedback (Gikandi et al., 2011) as the teacher put extra effort to design those tasks. Whereas with the lack of those pre-requisites, written exam, which is basically design for proctored offline exam, has comparatively more appropriate problem task according to needs of learners but conducting in unproctored manner violet the validity of this exam. Thus, the application of mcq instruments is more valid than that of the application of a written exam. When the validity of that group (instruments for assessing) increases, the reliability will fall (Guilford, 1954). In that case written exam is more reliable than that of MCQ exam. However, considering the overall benchmarks for coexistence of validity and reliability, the applied techniques are found with lack of maximum pre-requisites (Driessen et al., 2005; Gikandi et al., 2011; Moss, 1994; Öz, 2018; Taras, 2009) such as employing summative assessment for formative practice, combination of various modes, methods and techniques with technology enhancement, multimodal approach and consistency in meeting the needs of learners which made them invalid and unreliable. This finding reveals not only the position of validity and reliability in assessment process but also the scopes of enhancing in terms of validity and reliability which leads the researcher achieve the research objectives,

Virtual assessment can ensure meaningful learning and teaching as it provides diverse opportunities to generate pedagogical foundations facilitating various and maximum number of learning requirements. In online assessment, students get the opportunities to engage themselves in rigorous evaluation with multidimensional approaches that identify the precise learning gap and provide formative feedback that adds reflective and autonomous learning experience. After investigating the current scenario of online assessment of Bangladesh, it is evident that the authority needs to identify the demand of innovative assessment as a part of learning processes and stress on evaluating in both product and process based assessments. Those who are related to online assessment should assist teachers in using educational technology to create their own frameworks. Due to obtaining the effective and meaningful results, it is necessary to involve students to establish common knowledge regarding the purpose of assessment and their responsibilities within the process. Thus online assessment can show its true potential to achieve the needs of learners. Technology provides vast materials

and opportunities for instant feedback showing their gaps, errors and scopes of corrections and this process extracts the positive washback effect (Ali & Hamid, 2020).

Reference

- 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), 4.
<https://doi.org/10.1186/s40468-018-0060-9>
- Al Amri, S. (2008). Computer-based testing vs. Paper-based testing: A comprehensive approach to examining the comparability of testing modes. *Essex Graduate Student Papers in Language and Linguistics*, 10, 22–44.
- Alam, S. (2016). *Teachers, collaboration, praxis*: 305.
- Alderson, J. C., Brunfaut, T., & Harding, L. (2017). Bridging assessment and learning: A view from second and foreign language assessment. *Assessment in Education: Principles, Policy & Practice*, 24(3), 379–387. <https://doi.org/10.1080/0969594X.2017.1331201>
- Ali, Md. M., & Hamid, M. O. (2020). Teaching English to the Test: Why Does Negative Washback Exist within Secondary Education in Bangladesh? *Language Assessment Quarterly*, 17(2), 129–146.
<https://doi.org/10.1080/15434303.2020.1717495>
- Ali, Md. M., Hamid, M. O., & Hardy, I. (2020). Ritualisation of testing: Problematising high-stakes English-language testing in Bangladesh. *Compare: A Journal of Comparative and International Education*, 50(4), 533–553. <https://doi.org/10.1080/03057925.2018.1535890>
- Anderson, L. W. (1999). *Rethinking Bloom's Taxonomy: Implications for Testing and Assessment* (p. 25) [Reports - Descriptive]. <https://files.eric.ed.gov/fulltext/ED435630.pdf>

- Angelo, T. A. (1995). Reassessing and defining assessment. *American Association for Higher Education Bulletin*, 48, 7–9.
- Assessment at Loyola University. (n.d.). Retrieved April 9, 2021, from <https://www.luc.edu/fcip/oldfiles/assessment/assessmentatloyola/>
- Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice*, 24(3), 317–350.
- <https://doi.org/10.1080/0969594X.2017.1319337>
- Baleni, Z. G. (2015). *Online formative assessment in higher education: Its pros and cons*. 13(4), 9.
- Banta, T. W., Jones, E. A., & Black, K. E. (2009). *Designing effective assessment: Principles and profiles of good practice* (1st ed). Jossey-Bass.
- Bettinger, E., Doss, C., Loeb, S., Rogers, A., & Taylor, E. (2017). The effects of class size in online college courses: Experimental evidence. *Economics of Education Review*, 58, 68–85.
- <https://doi.org/10.1016/j.econedurev.2017.03.006>
- Black, P., Harrison, C., Lee, C., & Marshall. (2009). *Assessment for learning: Putting it into practice* (Reprinted). Open Univ. Press.
- Black, P., Lee, C., Wiliam, D., Harrison, C., & Marshall, B. (2003). *Formative and summative assessment: Can they serve learning together?* Annual meeting of the American Educational Research Association, Chicago.
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400–413. <https://doi.org/10.1080/02602938.2015.1018133>
- Carey, K. (2020). *Everybody Ready for the Big Migration to Online College? Actually, No*. The New York Times. <https://www.nytimes.com/2020/03/13/upshot/coronavirus-online-college-classes-unprepared.html>

- Choi, I.-C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20(3), 295–320.
<https://doi.org/10.1191/0265532203lt258oa>
- Dhawan, S. (2020). Online Learning: A Panacea in the Time of COVID-19 Crisis. *Journal of Educational Technology Systems*, 49(1), 5–22. <https://doi.org/10.1177/0047239520934018>
- Dixson, D. D., & Worrell, F. C. (2016). Formative and Summative Assessment in the Classroom. *Theory Into Practice*, 55(2), 153–159. <https://doi.org/10.1080/00405841.2016.1148989>
- Driessen, E., van der Vleuten, C., Schuwirth, L., van Tartwijk, J., & Vermunt, J. (2005). The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: A case study. *Medical Education*, 39(2), 214–220. <https://doi.org/10.1111/j.1365-2929.2004.02059.x>
- Dutta, S., & Smita, M. K. (2020). The Impact of COVID-19 Pandemic on Tertiary Education in Bangladesh: Students' Perspectives. *Open Journal of Social Sciences*, 08(09), 53–68.
<https://doi.org/10.4236/jss.2020.89004>
- Eber, P. A., & Parker, T. S. (2007). Assessing Student Learning: Applying Bloom's Taxonomy. *Human Service Education*, 27(1), 45–53.
- Emon, E. K. H., Alif, A. R., & Islam, M. S. (2020). Impact of COVID-19 on the Institutional Education System and its Associated Students in Bangladesh. *Asian Journal of Education and Social Studies*, 34–46. <https://doi.org/10.9734/ajess/2020/v11i230288>
- Entwistle, N. (2003, October 21). *Promoting deep learning through teaching and assessment: Conceptual frameworks and educational contexts*. Education-line.
<http://www.leeds.ac.uk/educol/documents/00003220.htm>
- Erwin, T. D. (1991). *Assessing Student Learning and Development*. Jossey-Bass.
- Foucault, M. (1975). DISCIPLINE AND PUNISH. *Random House*, 353.

- Gannaway, D., Green, T., & Mertova, P. (2018). So how big is big? Investigating the impact of class size on ratings in student evaluation. *Assessment & Evaluation in Higher Education*, 43(2), 175–184. <https://doi.org/10.1080/02602938.2017.1317327>
- Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4), 2333–2351. <https://doi.org/10.1016/j.compedu.2011.06.004>
- Guilford, J. P. (1954). *Psychometric methods*. McGraw Hill.
- Harlen, W., & James, M. (1997). Assessment and Learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365–379. <https://doi.org/10.1080/0969594970040304>
- Hewson, C., & Charlton, J. P. (2019). An investigation of the validity of course-based online assessment methods: The role of computer-related attitudes and assessment mode preferences. *Journal of Computer Assisted Learning*, 35(1), 51–60. <https://doi.org/10.1111/jcal.12310>
- Kapsalis, G., Ferrari, A., Punie, Y., Conrads, J., Collado, A., Hotulainen, R., Rämä, I., Nyman, L., Oinas, S., Ilsley, P., European Commission, & Joint Research Centre. (2019). *Evidence of innovative assessment: Literature review and case studies*. https://op.europa.eu/publication/manifestation_identifier/PUB_KJNA29882ENN
- Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, 30(4), 28–37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Li, H. (2003). The Resolution of Some Paradoxes Related to Reliability and Validity. *Journal of Educational and Behavioral Statistics*, 28(2), 89–95. <https://doi.org/10.3102/10769986028002089>

Littlefield, J. (2020). *Which Method of Distance Learning Is Best for You?* ThoughtCo.

<https://www.thoughtco.com/synchronous-distance-learning-asynchronous-distance-learning-1097959>

Lowenthal, P. R., Nyland, R., Jung, E., Dunlap, J. C., & Kepka, J. (2019). Does Class Size Matter? An Exploration into Faculty Perceptions of Teaching High-Enrollment Online Courses. *American Journal of Distance Education*, 33(3), 152–168.

<https://doi.org/10.1080/08923647.2019.1610262>

Majed, Dr. N., Jamal, A., & Kabir, Dr. M. R. (2020, July 28). *Online Education: Bangladesh Perspective, Challenges and Way Forward*. The Daily Star. <https://www.thedailystar.net/online/news/online-education-bangladesh-perspective-challenges-and-way-forward-1937625>

Marciniak, R. (2018). Quality Assurance for Online Higher Education Programmes: Design and Validation of an Integrative Assessment Model Applicable to Spanish Universities. *The International Review of Research in Open and Distributed Learning*, 19(2). <https://doi.org/10.19173/irrodl.v19i2.3443>

Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher*, 18(2), 5–11. <https://doi.org/10.3102/0013189X018002005>

Mohammad Ali, C., & 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). <https://doi.org/10.7575/aiac.ijalel.v.5n.6p.64>

Moss, P. A. (1994). *Can There Be Validity Without Reliability?* 8.

National Curriculum-English XI-XII. (2012). National Curriculum and Textbook Board (NCTB).

http://www.nctb.gov.bd/sites/default/files/files/nctb.portal.gov.bd/files/21fe6788_469a_40f4_8776_8118c144d51c/English%20XI-XII.pdf

- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218.
<https://doi.org/10.1080/03075070600572090>
- Öz, H. (2018). *Computer-based and Paper-based Testing: Does the Test Administration Mode Influence the Reliability and Validity of Achievement Tests?* 19.
- Popham, W. J. (2004). Why Assessment Illiteracy Is Professional Suicide. *Educational Leadership*, 62(1), 82.
- Rahman, M. M., Khan, S. J., Sakib, M. S., Halim, M. A., Rahman, M. M., Asikunnaby, & Jhinuk, J. M. (2021). COVID-19 responses among university students of Bangladesh: Assessment of status and individual view toward COVID-19. *Journal of Human Behavior in the Social Environment*, 31(1–4), 512–531. <https://doi.org/10.1080/10911359.2020.1822978>
- Rahman, S. (2018). English Language Policy Initiatives and Implementation in Bangladesh: Micro political issues. *The Asian EFL Journal*, 88, 59–96.
- Ramij, G. (2020). *Preparedness of Online Classes in Developing Countries amid COVID-19 Outbreak: A Perspective from Bangladesh*. 26.
- Robles, M., & Braathen, S. (2002). Online assessment techniques. *Delta Pi Epsilon Journal*, 44(1), 39–49.
- Rogler, D. (2014). Assessment Literacy: Building a Base for Better Teaching and Learning. *English Teaching Forum*, 52(3), 2–13.
- Scheerens, J., Glas, C. A. W., & Thomas, S. (2007). *Educational evaluation, assessment, and monitoring: A systemic approach*. Taylor & Francis.
- Shohamy, E., Or, I. G., & May, S. (Eds.). (2017). *Language Testing and Assessment*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-02261-1>
- S.M.Ariful Islam. (2015). LANGUAGE POLICY AND PRACTICE IN SECONDARY SCHOOL CONTEXTS IN BANGLADESH: Challenges to the Implementation of Language-in-Education Policy [PDF]. *The*

PhD Series of the Faculty of the Humanities, Aalborg University.

<https://doi.org/10.5278/VBN.PHD.HUM.00006>

Sultana, N. (2018). Test review of the English public examination at the secondary level in Bangladesh.

Language Testing in Asia, 8(1), 16. <https://doi.org/10.1186/s40468-018-0068-1>

Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in

Bangladesh. *Language Testing in Asia*, 9(1), 1. <https://doi.org/10.1186/s40468-019-0077-8>

Taras, M. (2001). The Use of Tutor Feedback and Student Self-assessment in Summative Assessment

Tasks: Towards transparency for students and for tutors. *Assessment & Evaluation in Higher*

Education, 26(6), 605–614. <https://doi.org/10.1080/02602930120093922>

Taras, M. (2005). ASSESSMENT – SUMMATIVE AND FORMATIVE – SOME THEORETICAL REFLECTIONS.

British Journal of Educational Studies, 53(4), 466–478. <https://doi.org/10.1111/j.1467->

8527.2005.00307.x

Taras, M. (2009). Summative assessment: The missing link for formative assessment. *Journal of Further*

and Higher Education, 33(1), 57–69. <https://doi.org/10.1080/03098770802638671>

Torrance, H. (1993). Formative Assessment: Some theoretical problems and empirical questions.

Cambridge Journal of Education, 23(3), 333–343. <https://doi.org/10.1080/0305764930230310>

Vonderwell, S. (2007). Asynchronous Discussions and Assessment in Online Learning. *Journal of Research*

on Technology in Education, 39(3), 20.

Wiliam, D. (2000). *INTEGRATING SUMMATIVE AND FORMATIVE FUNCTIONS OF ASSESSMENT*. 25.

Wiliam, D., & Black, P. (1996). Meanings and Consequences: A basis for distinguishing formative and

summative functions of assessment? *British Educational Research Journal*, 22(5), 537–548.

<https://doi.org/10.1080/0141192960220502>

Appendix

Literature	Codes	Themes
Black (2003), Taras (2001)	Multiplication of evaluation, summative assessment for formative practices	Implementation of various assessment types
Lowenthal (2019)	Teacher's workload, hybrid model	
Kapsalis (2019).	different modes(online or offline), methods (synchronous or asynchronous) and techniques (written exam, MCQ, game, viva, case study)	How technology was integrated in Exam
Baleni (2015),	Multimodal approaches	
European Commission (2018)	Combination of different modes, methods and techniques	
(Marciniak, 2018),	Process of consistent performance	tentative outcome.
Baleni (2015),	Process of meaning making	
(Anderson, 1999; Eber & Parker, 2007; Vonderwell, 2007)	Process of using cognitive knowledge	
Gikandi	Authentic and encouraging assessment activities and process	Validity criteria
	Timely,continuous formative useful and easy comprehensive feedback	
	Multidimensional approaches with alternative	

	activities	
Driessen, van der Vleuten, Schuwirth, van Tartwijk and Vermunt (2005)	Consistency in using digital tools, grading, matching needs, problematic activities, explicit instruction, duration	Reliability Criteria
OZ	Reliability coefficient (rc=0.903)- consistency in performance	Statistical analysis

Table 01: codes extracted from literature, by Nvivo software

Questionnaire
1. What instructions were provided by the institute regarding conducting the exam online? What roles did the institute play?
2. What types of techniques (written exam, quiz etc.) did you follow to assess the learners?
3. Why did you choose those techniques?
4. What new methods/techniques did you explore and use in case of conducting the exam online?
5. How did you incorporate your professional training/experience to cope with this new exam system?
6. In terms of question making, what are the tentative outcomes you expected?
7. What types of instruments/tasks did you choose to assess your students' learning

growth?
8. What are the criterias you followed to create and to organise the tasks in the question paper? Did you follow any particular pattern, like easy to difficult?
9. Do the tasks provide students the opportunities to connect with their real life?
10. How did you grade the answers/ how did you prepare the feedback against their answers?
11. How does feedback help learners to improve the performance?
12. What are the challenges you faced to conduct the exam online on a large scale?
13. To what extent did you facilitate your students in participating in exams without having any anxiety?
14. What are the differences, in the context of online and offline modes, you experienced during the whole process?

Table 02: Questionnaire used in semi structured interview