

**ENGLISH LANGUAGE ASSESSMENT IN NEPAL:
POLICIES, PRACTICES AND PROBLEMS**

**By
Rabu Ranjit
(20277008)**

A thesis

**Submitted to the Department of Brac Institute of Languages in a partial fulfillment of
the requirements for the degree of Master of Arts in Teaching English to the Speakers of
Other Languages (TESOL)**

**Department of Brac Institute of Languages
Brac University
November 2022**

**© 2022. Rabu Ranjit
All rights reserved**

Declaration

It is hereby declared that

1. The thesis submitted is my original work while completing a degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

A handwritten signature in black ink that reads "Rabu". The signature is written in a cursive style with a horizontal line under the name.

Rabu Ranjit
20277008

Approval

The thesis titled “English Language Assessment in Nepal: Policies, Practices and Problems” submitted by

Rabu Ranjit (20277008)

of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirements for the degree of Master of Arts in Teaching English to the Speaker of Other Languages (TESOL) on 24 November, 2022.

Examining Committee:

Supervisor:
(Member)

Mohammad Mosiur Rahman
Former Lecturer
Brac Institute of Languages
Brac University

Program Coordinator:
(Member)

Dr. Faheem Hasan Shahed
Former Coordinator, MA in TESOL program
Brac Institute of Languages
Brac University

External Expert Examiner:
(Member)

Departmental Head:
(Chair)

Lady Syeda Sarwat Abed
Senior Director
Brac Institute of Languages
Brac University

Abstract

The present study focused on the test materials and curriculum policies, as well as the perspectives of the testers and moderators, to examine how the English language test for secondary education examinations in Nepal is developed, what principles and criteria govern test production, and what factors impact this process. Participants were selected using a purposive sampling method. The findings indicated a range of circumspection, compliance, and context-aware methods, demonstrating the continuation of questionable test design procedures and practices. These replies backed ritualized design approaches that downplay test validity and reliability worries and hide the fundamental principles of the educational system by categorizing students into achievers and failures. Importantly, the findings revealed the need to scrutinize the test design itself. Future research is also recommended based on factors and stakeholders.

Keywords: language test design, high-stakes testing, external examinations, Nepal, testing across societies

Acknowledgment

First and foremost, praises and thanks to the God, the Almighty, for his showers of blessings throughout my life particularly, during the rigorous hours spent to complete the dissertation successfully and for all the support I had received from the generous people who were involved in this process.

Similarly, I would like to express my deep and sincere gratitude to my dissertation supervisor, Professor Mohammad Moisir Rahman for enlightening me at the first glance of research and providing invaluable guidance. His vision and sincerity have deeply inspired me. He has taught me the methodology to carry out the research and present the research works as clearly as possible. It was a great privilege and honor to work and to study under his guidance. I am extremely grateful for what he has offered me during my initial research studies. I would also like to thank him for his friendship and empathy.

Moreover, my sincere thanks to Prof Dr Jaya Raj Awasthi, former chairman of National Examination Board (NEB), founder of Nepalese English Language Teachers Association (NELTA), former vice-chancellor of Tribhuvan University (TU), former English Language Curriculum Developer and Test designer, Current Dean of Kathmandu University School of Education, Former Principal in Little Angels School, for providing me insights and handful of resources on Nepalese English language education, its assessment system and practices. Besides, I would like to extend my heartfelt gratitude to NEB library staff for assisting me to locate articles and reports related to assessment in Nepal. Furthermore, I would like to remember the staff of the Ministry of Education (MoE), Nepal, for granting me access to educational resources.

In addition, I am extremely grateful to my great grandparents and parents for their eternal blessings, love and support for educating and shaping my life. Likewise, I am very much

indebted to my brother for his understanding, prayers, continuous support and sensitivity to complete this work.

Finally, my special thanks goes to my caring, loving, and supportive best friend Er. Prashish Maharjan for taking time to review my work and provide valuable feedback to complete this dissertation successfully. Your encouragement and motivation has played a significant role in completing this work. I appreciated and duly noted them. It was a great comfort and relief to know that you were willing to provide management of our household activities while I completed my work. My heartfelt thanks.

Sincerely,

Rabu Ranjit

Table of Contents

Declaration	ii
Approval	iii
Abstract	iv
Acknowledgement	v
Table of Contents	vii
List of Tables	ix
List of Acronyms	ix
Chapter 1 Introduction	1
1.1 Background	1
1.2 Problem Statement	2
1.3 Aims and Objectives	4
1.4 Significance of the study	5
Chapter 2 Literature Review	7
2.1 Basic considerations in test design	7
2.2 Understanding high-stakes testing across societies	8
2.3 Testing in Nepalese schools	10
2.4 Curriculum and Assessment policies for English	10
2.5 Teaching Artifacts	12
2.6 English Language Education in Nepal	12
2.7 The Education System in Nepal	14

2.8 Education Policy	16
2.9 Assessment principles in Nepalese curricula: the foundation of meaningful assessment	18
2.10 Reforms in English Language assessment	22
2.11 Current assessment practices: a mismatch between policy and operation	24
2.12 Overview of Nepalese practices on Students' Assessment	28
2.13 Washback: reverse navigation path	30
Chapter 3 Research Design and Methodology	32
3.1 Philosophical positioning of the Present Research	32
3.2 Research Design	33
3.3 Participants	34
3.4 Site selection and sampling	35
3.5 Instrument	36
3.6 Data Collection	36
Chapter 4 Data Analysis	38
4.1 Data analysis procedure	38
4.2 Ethical considerations	39
4.3 Delimitations	40
Chapter 5 Findings	40
Chapter 6 Discussion	47
Chapter 7 Conclusion	50

Appendix A.

List of Tables

Table 1. Profiles of the participants

Table 2. Repetition of test content across education boards and years

List of Acronyms

ADB = Asian Development Bank

AERA = American Educational Research Association

BEC = Basic Education Curriculum

BEE = Basic Education Examination

BPEP = Basic and Primary Education Programme

CAS = Continuous Assessment System

CBA = Classroom Based Assessment

CDC = Curriculum Development Centre

CELTA = Certificate in English Language Teaching to Adults

CEM = Centre for Evaluation and Monitoring

CERID = Research Centre for Educational Innovation and Development

CERSOD = Centre for Educational Research and Social Development

CLIL = Content and Language Integrated Learning

CLT = Communicative Language Teaching

CTEVT = Council for Technical Education and Vocational Training

DELTA = Diploma in Teaching English to Speakers of Other Languages

DOE = Director of Education

EDSC = Educational Support and Distribution

ELT = English Language Teaching

ERO = Education Review Office

GPA = Grade Point Average

ICT = Information and Communication Technology

MOE = Ministry of Education

MOEST = Ministry of Education, Science and Technology

NASA = National Assessment of Student Achievement

NCF = National Curriculum Framework

NEB = National Examination Board

NELTA = Nepalese English Language Teachers Association

NESP = National Education System Plan

NEPC = National Education Planning Commission

NNEPC = Nepal National Educational Planning Commission

OECD = Organization for Economic Cooperation and Development

PEC = Primary Education Curriculum.

PISA = Program of International Student Assessment

SBA = School Based Assessment

SEE = Secondary Education Examination

SLCE = School Leaving Certificate Examination

SSRP = School Sector Reform Plan

TEFL = Teaching English as a Foreign Language

TESL = Teaching English as a Second Language

TESOL = Teaching English to Speakers of Other Languages

THT = The Himalayan Times

Chapter 1

Introduction

1.1 Background

With the foundation of the School Leaving Certificate Board in 1934, the tradition of public exams officially started in Nepal. Internal assessment, a form of classroom-based evaluation, was first proposed by the National Educational System Plan (Ministry of Education, 1971). The National Education Plan 1971-1976 was created for developing knowledge-based and technical education. This strategy was comprehensive and viewed education as a dynamic force that produces human capital for nation building. In the late 1980s, the Nepali government reinstated CAS (Continuous Evaluation Method) as a classroom-based assessment system in the elementary school grades. The Curriculum Development Centre (CDC), a division of Nepal's Ministry of Education (MOE) in charge of creating the curriculum, producing and dispersing textbooks, teachers' manuals, and other relevant resources, included a brief section on assessment (CDC, 2007, p. 57).

On the other hand, public exams are often taken outside of a classroom setting and are summative. Public tests may be given to evaluate each student at several levels, such as the national, regional, district, resource center, and/or school levels. These tests are typically administered by external organizations, such as the government or independent/autonomous bodies. Public exams are used to select and advance pupils to the next grade or educational level and certify learning, but they seldom offer feedback on classroom teaching-learning or the educational system.

The exams are based on conventional testing theory, with insufficient item and testing procedure uniformity. However, the majority of these assessments revealed elements relating to the instructor, the school, and the students that affected student learning. Although assessment policies emphasize the value of all four language skills—listening, speaking, reading, and writing, it has rarely been practiced in our classroom setting and evaluation system. Therefore, the evaluation system was abandoned once the redesigned curriculum was implemented at the school level in the early 1980s because it eliminated the internal assessment system from classroom instruction.

The Constitution of Nepal (2016) amended the existing education policy highlighting the significance of access to education as a fundamental human right. Consequently, the then

government, revised the Amendment Bill to support free education policy (Bajracharya, 2016). This coverage started free education up to secondary level in school. Similar to this, all measures pertaining to education must be in line with the nation's interests and overall educational aims. Most notably, the Education Policy 2016 has given instructional assessment a special priority. In order to ensure the improvement of education quality at all levels, it is stated that the exam system would change. It also mentions the necessity of implementing appropriate and effective exam structure or straggles, as well as the requirement that successful assessment processes be ingrained within the educational system.

Aside from school-based assessments, pupils in the mainstream education system are required to take three national public examinations at the end of Years 8, 10, and 12. The results of these exams decide which schools or universities students will attend and whether they will be eligible for quotas at home or abroad . Thousands of Nepalese students apply for a relatively limited number of primary spots in higher education in metropolitan regions, district towns and international educational institutions. Only a few will secure their seats. Additionally, tests are extensively used in the labor market, both in the public and commercial sectors. They sit recruitment exams for a few hundred posts in the bureaucracy, the financial industry, and basic and secondary schools, all in the hopes of securing a permanent career and a stable future.

Despite significant improvements in the educational field over the past few years, particularly in terms of student enrollment and gender parity, educational evaluation as a mechanism for guaranteeing high standards of education and learning achievement has not received enough attention. This lack of interest in assessment directly relates to the teaching and learning of the English language.

1.2 Problem statement

Hamid (2010) sees English proficiency as fundamental for human capital development and the nation's economic growth. However, English language teaching and learning vary depending on the institution type. Although English education in Nepal began in 1854 A.D., it was not introduced as a compulsory subject from Year 1 to Year 12. Still, English as a subject is less prioritized until Grade 6 in many community schools.

On the other hand, English medium instruction has been established in most private schools. Officially, communicative language teaching (CLT) has been designated as the primary technique for enhancing students' functional English competence. However, neither

school-based nor public examinations adhere to a communicative strategy. Despite the testing prevalence, there is a concern about whether or not students gain communicative ability in this important language. More significantly, many have no idea how these English assessments are created, delivered, and verified. However, the results of these examinations determine who "achieve" and "fail" in the school and job markets.

Nepal adopted communicative language teaching (CLT) as a language teaching methodology at the basic and secondary levels in 2007 (Curriculum Development Center, 2020). It is expected that the assessment methods embedded in the CLT approach will be used by individuals and institutions. However, measurement specialists have identified several weaknesses of the single high-stakes test. One single high-stakes test is unable to help students, educators, or schools in making important educational decisions (Ritt, 2016). Students are being treated with a one-size-fits-all education approach, which is causing unreliable test scores (Popham, 2015). According to the new research by American Educational Research Association (2000), "Assessment experts oppose high-stakes testing because using a single indicator of competence to make important decisions about individuals or schools violates the professional standards of the measurement community."

Other critics are concerned that the unintended effects of high-stakes testing lead to "perverse" (Ryan, 2004) and "corrupt" educational practices (Jones, 2007; Jones & Hargrove, 2003; Nichols & Berlinger, 2007), while Pedulla et al. (2003) argue that the pressure of doing well on a test seriously compromises instructional practice. Where the high-stakes test is in practice, the teachers put considerable importance on the contents that are tested, and the students' achievement in the maximum test score because the test results are associated with the national test results (Harlen, 2009). As a result, most of the teaching time is devoted to preparing the students for the test or doing the testing. The quality of teaching decreases if the teacher spends a lot of school time preparing students for tests.

As a result, there is a deep mistrust of tests and of testers. It is observed that a great deal of language testing is of very poor quality and they have a harmful effect on teaching and learning, and often fail to measure accurately whatever it is they are intended to measure. Furthermore, international education or assessment scholarship says virtually little about tests and testing in Nepal. Research into education is not a priority for this low-income country, which must prioritize more important requirements like poverty reduction, health, and the economy.

1.3 Aims and Objectives

Given the importance of testing in Nepalese culture, it is vital to question who designs these high-stakes exams, how they are developed, what ideas and criteria lead them, and how the public perceives these exams. This study provides us with insight into the procedures of test creation and evaluation in the Nepalese education system.

I investigated the design of the English Paper for the School Education Examination (SEE) examination, which students take at the conclusion of Year 10. The outcomes of this last school test decide pupils' admittance to further education institutions. Moreover, the result distinguishes student's from their ability to use English in Nepal. Those students who are proficient in using English will opt for higher education in 'A' class schools in the country or they may opt for study abroad and those who lag behind struggles in every facet of career. For example, people get rejected from access to education in a sophisticated college or university or job interview because their English language proficiency was less than the selected candidate. This further aids to decide the potential of a student in the educational and job market. In recent years, more than a hundred thousand students have taken this public test each year. According to THT Online (2021), nearly 5,17,000 students in 2021.

The goal of the study is to add to the body of knowledge on testing across nations in a global context (Ashadi & Rice 2016; Cheng, 2008; Ramanathan, 2008; Rose, 2015; Ross, 2008; Sasaki, 2008). The study used the perspective of important assessment actors, such as education officials and test developers, to better understand the nature of the assessment process and the factors that influence the design and administration of the English exam.

Based on my findings, I contend that in countries such as Nepal, test design may have taken the shape of traditional rites done in accordance with implicit principles that go unchallenged. Professional requirements and ethical norms linked with testing (e.g., Downing & Haladayan, 2006) may seem insignificant in the context of such commonplace social behaviour. It is worth noting that neither scholarly nor popular criticism has been leveled at this rationalization of testing. Taking into account these considerations, the present study addresses the following questions:

RQ 1: How is the English test designed and developed?

RQ 2: To what extent does the test reflect the goal of English curriculum policy for this level of education?

RQ 3: What factors influence test design?

1.4 Significance of the study

Assessment is an essential component of English learning and instruction. It is defined as "any act of analyzing information regarding student performance gathered by any of a variety of sources" (Brown & Hirschfeld, 2008). The power of tests, particularly those with high stakes, encourages test takers and educational institutions to alter their educational practices and tactics in order to excel in examinations, notwithstanding their negative consequences (Shohamy, 2016). It is a test that provides a clear image of the student's ability and identifies the student's strong and weak areas in language teaching and learning. As a result, teaching and assessment are inextricably linked, and one is meaningless without the other. Nobody can deny the value of evaluation and testing in language instruction programs.

Language testing is commonly acknowledged as providing helpful inputs to instructors in order for them to be aware of the effect of their teaching, as well as some insight into whether they should continue to teach in the same manner or modify it in order to make their teaching more successful (Khaniya, 2005; Wu, 2015). Assessment is the most crucial part of the teaching and learning process since it affects all aspects of teaching and learning. An assessment allows a teacher to receive feedback on his or her instruction and the students' learning. It establishes language teaching objectives and monitors both instructors' and students' progress toward those objectives. It is obvious that learning is achievable with the assistance of instructors, classmates, parents, and their children.

It is obvious that learning is achievable with the assistance of instructors, classmates, and parents, as well as their ongoing assessment of what has already been taught, what may be acquired, and what must wait to be learnt. As a result, evaluation is a continuous process in language classes. The instructor employs a variety of assessment methods, including formative and summative evaluations. After receiving some education, the instructor constantly looks for accomplishment. The emphasis is transferred to the learning process, in which evaluation occurs continually and formatively while the student is still learning (Stiggins, 2002).

The current study is undertaken to understand how the high-stakes English test for the SEE examination is designed and developed in relation to relevant curriculum policy. As previously noted, students must pass the SEE examination with a higher grade (e.g. A) to be eligible for further studies. Based on the Curriculum Development Center, (2020), the achievement of a higher grade means test-takers have achieved several competencies in English that are useful for international communication, academic and vocational purposes (Curriculum Development Center, (2020). Nonetheless, as mentioned before, English is not necessarily informed by the curriculum policy because not all four skills are tested. This mismatch between official curriculum and assessment policy on the one hand and the stakes attached to the test on the other calls for an investigation into the processes of test design and the factors affecting the design.

Chapter 2

Literature Review

2.1 Basic considerations in test design

Language assessment is not an abstract process; it requires a goal and context to work. Language evaluations are used to elicit information about people's communicative language abilities in order to make correct and legitimate interpretations based on scores for a variety of objectives (Bachman, 1990). All assessments should be of high quality, emphasizing the importance of good design (Alderson, Clapham, & Wall, 1995; Fulcher, 2010).

According to Hughes (2003), the first stage in language test design is to define the problem (purpose) based on information about the nature of the test, its objectives, the constructs (abilities) it wants to assess, its impact, ethical considerations, and limits in conducting the test. The next stage is the creation of specifications ('specs') by establishing standards concerning test content, structure, length, media, and procedures to elicit test-takers' performance. The specifications include explanations of the performance criterion level(s) as well as scoring processes for evaluating test-takers' language use. The third stage involves writing and reviewing test items based on the test criteria so that the test accurately portrays the problem. The moderated questions are trialed and the results are analyzed in the next three steps to verify reliability, task complexity, and representativeness, as well as to minimize unforeseen difficulties. Following the completion of the study, the next stage is to calibrate rating scales, which give samples and reference points for scorers. The final version of the test is confirmed at the conclusion of this procedure.

Although the testing literature is replete with theoretical discussions of test design, reviews, and validation (as seen by the references given previously), there is a lack of attention on how high-stakes language exams are actually constructed, particularly in developing cultures. Tests used in external examinations at various stages of schooling in these civilizations are of special importance. Although public examinations in English and other courses have been utilized in Nepal for decades, there has been little study on how these tests are created, what learning or success is targeted for evaluation, and what repercussions these tests may have for students and their families, the education system, and society at large.

2.2 Understanding high-stakes testing across societies

As Cheng and Curtis (2004) point out, the design of tests can heavily influence the nature and goals of education in many parts of the world. The test is no longer 'an obedient servant' but rather 'a leader' of pedagogy, with a critical and influential role in shaping curricula and educational policies. Policymakers continue to exploit the power of exams to influence their local educational systems, regulate curricula, and enforce (or encourage) new textbooks and teaching techniques in many areas of the world. One of the aims of assessment is to make policy judgments based on test findings (Cheng & Curtis, 2004).

The use of test data (i.e., results) to influence policy choices is a key trend in education in the twenty-first century. This development is linked to a new governance model in which educational standards and their outcomes are compared according to notes of Hamilton, Maddox, & Addey, 2015; Hardy, 2015; Kamens, 2013. Test results serve as a measure of the quality of instructors' and students' work in this new institutional framework (Lingard, 2011; Wu, 2012). Thus, conceptions of audit culture, reference societies, and comparisons of one school system with another, all based on test results, now substantially influence educational discourses (Lingard, 2011; Smith, 2016).

The international 'horse race' (Kamens, 2013) among nations to demonstrate educational quality has been made possible by global education players such as the Organization for Economic Cooperation and Development (OECD) through international examinations, most notably the Programme for International Student Assessment (PISA), which evaluates students' reading, numeracy, and applied abilities (Meyer and Benavot 2013; OECD 2016). PISA statistics are used to assess educational quality and the degree of 'human capital' in participating nations. Bloem (2015) Large-scale international evaluations put national governments under pressure to alter their education systems and match them with international examinations like the PISA. In the United States, the Bush Administration enacted the No Child Left Behind (NCLB) Act in 2001, which elevated high-stakes testing to the forefront of governance and data-driven accountability (Hursh, 2007). In order to enhance accountability in its education system, Australia implemented the National Assessment Program – Literacy and Numeracy (NAPLAN) in 2008 (Klenowski and Wyatt-Smith 2012; Wu 2015). East Asian nations such as China, Japan, and South Korea fared well on PISA as a result of educational reforms and a "strong dependence on standardized assessments at the national level" (Allen, 2016; Gu, 2014; Kwon, Lee, & Shin, 2017; Poole, 2016).

The worldwide trend of data-driven accountability emerges differently in low-income and emerging economies. Many of these nations do not participate in international evaluations, but they do have national standardized tests as part of their educational systems. Davies (2016) described the high stakes nature of the Nepalese School Education Examination (SEE) test, which was held at the conclusion of Year 10. The examination results affect whether or not students may apply for higher study and what type of work they will have in the future. Jilani (2009) both reported similar high-stakes examinations in Iran and Pakistan, respectively. These studies indicate that test data are important in the lives of test takers and schools, which are held accountable for student achievement in public examinations.

While high-stakes exams are employed across countries to create data for key socio-educational choices, they are often criticized; in particular, they are viewed as constraining pedagogy that becomes concentrated on the tests, which is seen in several jurisdictions, including the United States (Au, 2009), Australia (Klenowski and Wyatt-Smith, 2012), China (Gu, 2014), Bangladesh (Das et al., 2014), Korea (Kwon, Lee, and Shin, 2017), Japan (Allen, 2016), Nepal (Davies, 2016), (Balwanz, 2016). Because high test scores are regarded as quality indicators, instructors and schools are under pressure to enhance test data (Klenowski & Wyatt-Smith, 2012), potentially raising validity difficulties in test design (Harris and Brown 2016). According to Harris and Brown (2016), when scores are utilized as a key source of accountability and as an indication of quality, society may value grades more than test ethics or testing principles such as validity and reliability. This emphasizes the need to understand test design in an audit culture where test data dominates educational choices.

Given the power of tests and their social repercussions, critical language testing researchers contend that high-stakes tests must be scrutinized in order to investigate "the uses [or, misuses] and consequences of tests in education and society" (Shohamy, 2016). The first step in understanding testing culture across nations is to understand how tests are constructed. This knowledge is critical because policy decisions may be hampered if tests "fail to assess precisely whatever it is they are designed to evaluate" (Hughes 2003, p. 2). The failure might be traced to a "misunderstanding of the nature of language testing and language test creation," which resulted in "tests that do not suit the special demands of the test users" (Bachman & Palmer, 1996, 7). As a result, policy choices based on the results of a poorly constructed test may be deceptive. However, there is a lack of knowledge on how high-stakes examinations are developed across civilizations, as demonstrated by Nepal for instance.

2.3 Testing in Nepalese schools

In Nepal, mainstream education is organized into three stages: primary, secondary, and university. Lower secondary (Grades 7–8), secondary (Grades 9–10), and upper secondary (Grades 11–12) education are the three levels of secondary education. Students must pass a nationwide public examination conducted each year at the end of each level. The last two public examinations, School Education Examination [SEE] and School Leaving Certificate Examination [SLCE] are the most essential. Students can only continue on to upper secondary school if they pass the SEE. The SEE examination is much more essential for pupils since their subject scores and GPAs decide where they will attend college. Students are examined in all disciplines, including English, at the SEE. The current research looks into the design of the English test in the SEE.

2.4. Curriculum and assessment policies for English

In line with the spirit of the new National Education Act (Ministry of Education, Science and Technology, 2019), the secondary English curriculum adopted learner-centered approaches to develop students' communicative competence (Curriculum Development Center, 2020). Thus, one of the fundamental aims of the curriculum is to help students acquire competence in four [English] language skills, i.e. listening, speaking, reading, and writing (Curriculum Development Centre, p. 36). The main objectives of the curriculum can be understood from the following extract:

“English is taught as a compulsory subject not only in school but also in almost all the undergraduate programmes at the university level. English has also been adopted as the medium of instruction in many schools and most higher education programmes. Efforts have been made to incorporate recent trends and undercurrents in the field of language learning and teaching and all four language skills are adequately addressed. Strong grammatical foundation is also given due consideration and the learners are expected to be able to communicate in the English language with confidence (Curriculum Development Center, (2020, p.1).”

There has been a shift from traditional, grammar-based language instruction to a ‘skill-based’ approach to develop learners’ ‘real-life’ functional skills so that they can access ‘higher education’ and ‘local and global employment.’ The current English language policies in Nepal, therefore, have been influenced by the nation’s desire for economic development

through English language education (Sergeant and Elizabeth, 2011). Accordingly, the authorities have adopted Communicative Language Teaching (CLT) for English language teaching and devised CLT-based textbooks for primary and secondary education.

However, it is often argued that the assessment system is partly responsible for the failure to achieve the desired English language teaching and learning outcomes (Ali & Walker, 2014; Hamid & Baldauf, 2008). This justifies the need to examine whether the assessment processes, including test design, administration, and evaluation, are consistent with English language policies, and how those engaged in developing these tests make sense of their work. At the higher secondary level, students are usually assessed by means of: (a) continuous assessment; (b) internal examination; and (c) public examination.

More informal, continuous assessment is carried out throughout the year by means of class tests and homework. An internal examination is organized by individual institutions to monitor students' progress towards the SEE examination. For the public examination, and following the guidelines for test design provided by the national curriculum authority, the National Examination Board (NEB), each education board invites selected teachers from higher secondary colleges to write test papers. Once the invited teachers have submitted their test papers to the Controller of Examinations of the relevant education board, the board invites another group of teachers to moderate the papers. The Controller of Examinations receives multiple sets of the moderated test papers. She or he can decide which one to use and which one to save as a reserve piece of paper. Kabir (2008) claims that the selection process of test writers and moderators is not fair, as teachers who maintain good relations with the education board officials are often given the opportunity to write test papers.

Finally, the public examination is administered throughout the country simultaneously on the same days by seven education boards located across the country, including Morang, Dhanusha, Kathmandu, Kaski, Rupandehi, Surkhet, and Kailali. Higher secondary college teachers work as test invigilators. After the administration of the test, the scripts are packed and sealed and sent off to the Office of the Controller of Examinations in each education board. The board then invites examiners from higher secondary level English teachers from across the country. Although both internal and external assessments are expected to contribute towards students' final GPAs, in practice their GPAs are determined exclusively by their subject grades on the SEE external examination.

2.5 Teaching artifacts

The subject of English is taught using an official textbook called *Our English for Grades VIII, IX, and X*, which is written by subject experts commissioned by the Curriculum Development Center (CDC). The textbook prioritizes developing learners' four skills in English 'with a new focus on Communicative Language Teaching incorporating tasks and activities with contextual grammar and related vocabulary, providing opportunities for language skills practice in order to foster competence in reading, writing, listening and speaking skills. Although the classroom teaching learning practice focuses more on reading and writing skills, it emphasizes the presentation of four skills by incorporating reading, writing, listening, and speaking skills as well as cultural elements required for developing learners' intercultural communicative competence, creativity, and critical interpretation. In some contexts, such as Hong Kong, it has been reported that textbooks play a major role in aligning tests with curriculum policies (Leung & Andrews, 2012). However, evidence from other contexts suggests that such a role for textbooks cannot be generalized. For example, in research from M. Rahman et al. highlighted contextual complexities including the lack of skilled teachers and technological support needed for implementing CLT tasks presented in the textbook (2018). These contextual challenges may have implications for the assessment of the English subject.

2.6 English Language Education in Nepal

Different factors played a crucial role in the establishment of English language education in Nepal since the country has no visible colonial legacy. Notwithstanding, As Mulmi (2017) notes in his article, "Even though Nepal had never been officially colonized by the English, there was some sort of colonial legacy" (p.2). Consequently, the historical backdrop of Nepal and the historical backdrop of the English language in Nepal go hand in hand.

However, the impact of English in Nepal has not been felt in terms of its usefulness (Poudel, 2016). English entered Nepal as a lingua franca in the seventeenth century (1661) when European missionaries first settled in Nepal to train Nepalese people to work for them in the country and abroad (Shrestha, 2014). Then, the rulers involved in trade used English widely while exchanging goods and services with Tibet and North India (Hodgson, 1864; Morris, 1963 as referred to in Giri, 2011).

The former state leader, Junga Bahadur Rana, commanded English language training in Nepal. English instruction was also perceived as a hot topic around here. The language entered Nepal through correspondence with the English-speaking individuals who settled in India. The elites demanded a return to English training for their youths. Also, he transferred Nepalese youth to the English fortified force. These youthful people should bear eight years of formal training in English. It played a huge part in the spread of the English language in Nepal.

One further defining moment throughout the entire actuality of English training is Junga Bahadur's visit to Britain. He was profoundly impacted by English instruction in Britain. Accordingly, he laid out Darbar School in 1853 after his visit. He believed that the western system of English training would help his youngsters keep up with the norm and support their position (Vir, 1988 as pertained to in Poudel 2016). It was the first government-run English medium academy in Nepal (Shrestha, 2014). For the Rana family, it was distinctly for the Rana family, as the Ranas considered an informed existence to be a peril to their control (Caddell, 2007).

The first tertiary educational association in Nepal was Tri Chandra School, which was established in 1918. In this academy, the language of instruction was English. Its provocation was to forestall scholars from traveling to another country. His retired design was to keep the Nepalese from getting revolutionary studies that could be useful for them. Tri-Chandra School was an annex of Patna College, India. It acquired a class and evaluation frame from that point; accordingly, there was an immediate impact of the British Indian system of education on the Nepalese system.

In 1950, a new educational policy on language was formed, considering the recommendation of the Nepal National Educational Planning Commission (NNEPC). The report of the NNEPC unequivocally upheld Nepali as the medium of instruction for lessons with an end to achieving public integration. The report championed the use of the Nepali language in all spheres of life. As a result, English is tutored as a subject. Nonetheless, NNEPC proposed an instructional approach, National Education System Plan (NESP, 1971), by following Hugh B. Wood's view and practice in his country. The policy suggested the use of Nepali as it is spoken in administration, education, and the media, whereas English is a mandatory subject in seminaries.

Currently, English is considered the way to overcome the rigors and losses of one's own life. According to Kachru, B., Kachru, Y. & Nelson (2006), "In most South Asian countries, English is viewed as a language of power and as the simplest means of profitable uplift and upward social mobility (p. 90)." It led to the establishment of numerous private seminaries and modalities and made English necessary in the Nepalese class. Although there have been numerous programs and practice positions, English has mingled with Nepalese life in such a way that it seems nearly insolvable to detach from their lives due to its long history. As Kachru (2005) opines that Nepalese learners don't learn English to communicate in their motherland but they learn to communicate in their work or study abroad. Thus, from the time of the commencement of English education, English was learned for the purpose of professional development (Shrestha, 2014).

2.7 The education system in Nepal

The political transition of 1960 ushered in a brand new urgency to the need for an appropriate education system in Nepal. It is evident that the fulfillment of the system is notably based on the ability to generate considerable prospects for this. This briefly explains the records of the prevailing plan of the National Education System that regards education as a sustainable investment in human resources for the improvement of the country.

The contemporary Nepali education system advanced after introducing an egalitarian education system under the National Education System Plan in 1971. The education sector commenced flourishing after the establishment of the National Education Plan, the education sector commenced flourishing. Despite severe political setbacks over the decades, access to education has increased significantly. National Education System Plan (NESP, 1971) may be taken into consideration as a chief reform within the records of education in Nepal. The system has delivered structural reform in terms of the institutions, regulation framework and assessment. According to the policy, internal assessments and periodic examinations will be held in close coordination. Passing terminal examinations will be held at the end of each level, and admission to the next higher level will be contingent upon passing them. After the School Leaving Certificate examination, higher education entrance tests will be held, and opportunities for higher education will be given strictly based on merit as shown in the tests.

However, the advent of the federal government system within the country in 2015 set back the educational administration. Not all local governments are purposeful and Nepal's

education system is still administered under the preceding system wherein the Ministry of Education (MoE) adjudicates five Regional Educational Directorates. These five regional educational directorates include district education offices and resource centres to enforce policies at the local level. The Federal Ministry of Education (MoE) is liable for developing standard instructional guidelines and directives for the country. But, it's unclear how precisely the position of the MoE will evolve within the federal system. The MoE is likewise responsible for curriculum and textbook improvement, instructor education and recruitment, and conceptualizing and administering the nationwide school education exam through the National Examination Board (NEB).

Nevertheless, the brand new education system of Nepal, introduced in 2016, brings three stages of schooling: compulsory basic education, secondary education, and tertiary education. Basic education consists of grades one to eight. The secondary schooling system is split into levels: lower secondary schooling, which incorporates grades nine and ten. And higher secondary schooling includes grades eleven and twelve. Under the contemporary system, those degrees had been unified right into a 4-year secondary schooling cycle. Both stages finish with a local exam and a country-wide exam. Formative and summative assessments will be used to assess students' learning.

The antique nationwide School Leaving Certificate (SLC) Examination held at the end of grade 10 will be held on the local degree and has been renamed into Secondary Education Examination (SEE). The SEE board organizes, manages, and oversees the final board test, including the announcement of the results (Curriculum Development Centre, 2016). Nationally, there will be one very last nationwide school-leaving examination at the end of grade 12. In other words, the students take the Basic Level Examination (BLE) at the end of grade eight and then, the Secondary Education Examination (SEE) at the end of grade ten. After completing the twelve standards, the school students obtain the school-leaving certificate administered by the National Examination Board (NEB). Tertiary schooling is commonly undertaken at a university, a technical or a non-public institution.

The National Examinations Board Nepal, additionally called NEB, is the board liable for engaging in examinations of grades XI and XII, in addition to grade 10, in Nepal. Previously, the examinations of grades XI and XII had been carried out through the Higher Secondary Education Board (HSEB) and the exam of grade 10 through the School Leaving Certificate (SLC) board, but as in keeping with the 8th amendment of the Education Act 2016, achieved

recently, the obligation was assigned to NEB. Due to the amendment, every other alternative that took place changed to classifying our instructional levels into two categories, i.e., basic (Grade 1 to 8) and secondary (Grade nine to 12), as opposed to four.

Previously, SLC was a success, the crowning glory of Grade 10 and regarded as a national exam. The Examination Controller's Office managed the exam. However, the Grade 10 exam has now been converted to be managed through local offices. It might be on the provincial or local level. HSEB converted into NEB in keeping with the 8th amendment to the Education Act, 1971. Biratnagar, Janakpur, Kathmandu, Pokhara, Butwal, Surkhet, Dhangadi, and five sectoral offices (Lahan, Hetauda, Dang, Kohalpur, and Dadeldhura) (NEB, 2016) administer these high-stakes examinations.

There are alternative boards: the Council for Technical Education and Vocational Education (CTEVT) and the Madrasa Board Nepal. CTEVT is a self-reliant framework under the MoE. It oversees and quality controls technical and vocational schools and colleges. It sets curricula, checking out the necessities and capabilities required in specific occupations. The most popular packages are short-term certificate programs, formal secondary level programs, and diploma and technician certificate programs. On the other hand, the Madrasa Education Board in Nepal develops a curriculum on cultural education and additionally integrates obligatory guides of general education: science, mathematics, English, and Nepali. However, more than half of the 2000 madrasa establishments in Nepal have not been registered yet (Aljazeera, 2018).

2.8 Education Policy

The country's nationwide reconstruction in 1971 introduced an egalitarian education system to counteract an elite based education system with the purpose of serving the needs and goals of Nepal. The fundamental purpose of education has become to satisfy the manpower needs for improvement through the spread of knowledge-based and technical education. And thus, the National Education Plan 1971-76 was formulated. Little development has been made till now. The youths of Nepal have some educational possibilities and process markets in contrast to the former generation.

Even though Nepal transitioned to a partyless panchayat system in 1960, releasing itself from the previous democratic system. The education sector operated through local directorates till 2016 whilst the presently purposeful National Education Policy was amended. This policy is

overarching in terms of goals and objectives and conceives education as a lively force to generate human assets for nation-building. Moreover, the Education Policy of Nepal's amended constitution in 2016 emphasizes that access to education is a primary human right. So, the government revised the education system plan within the Amendment Bill to promote a free education policy (The Himalayan Times, 2016). But it is relevant to public schools only. According to this coverage, a pupil can get free education until a secondary degree. Similarly, any instructional initiatives have to be aligned with the wider instructional goals and slender interests of the country. The policy additionally consists of the integration of all ranges of schooling (primary, secondary, vocational and technical and tertiary), curriculum update, use of Information and Communication Technology (ICT) in all instructional projects, enhancements of teacher education and teacher education programs, and concord amongst all stakeholders.

The cornerstone of the policy is to promote education at all local levels. Most importantly, the Education Policy 2016 has placed a specific emphasis on instructional assessment. It explicitly states that the system of exams will be changed to ensure the quality enhancement of education at all levels—it is essential to put into effect suitable and powerful exam structures or strategies, and it additionally mentions that successful assessment processes have to be embedded within the educational system.

Assessment influences the students' learning the most. It plays a vital role in the learning and teaching process in any realm of education. The endorsement and execution of suitable evaluation strategies and their impactful implementation can perpetuate the improvement of English language proficiency. The language curriculum goals and objectives manual for the assessment system. Hence, the assessment process should align with the curriculum goals and objectives. In addition, an education system should have the requisite mechanisms to construct the intended English language assessment system to achieve the goals and objectives endorsed in the English language curriculum of Nepal.

In several cases, assessment is used interchangeably with tests. Brown and Lee (2015) state "test" as a deliberately tailored tool comprising rating scales to systematically grade an individual's expertise in a particular learning area indicated through standardized performance. However, assessment isn't always constrained to tests. It focuses on the expertise, skills, attitudes, and ideals of individual learners. In other words, it is an eclectic procedure that contains numerous devices to differentiate the students' learning development.

Assessment amplifies students' learning through feedback, and permits powerful interactions among the academics and the students; such interactions inform students approximately their strengths, and weaknesses and, additionally, help them use appropriate learning approaches.

Despite the tremendous gravity of assessment in deciding on the English language learning of students, only a trivial extent of exploration has been conducted on English language assessment in Nepal. In any case, this observation has examined a modest bunch of given papers from the databases of Google Scholar, Springer, Academia, and ResearchGate. Articles have been investigated, coming into watchwords and afterwards, having been selected through checking and perusing the abstract, the advent, and discussion. Incorporation and avoidance measures have been carried out to select the most suitable and urgent articles. The Web crawler Google turned into a frequent browser to look for articles diagnosed in the references of the assessed articles for additional investigation.

According to this point of view and specifics amassed from the formerly cited approaches, the concentration first provides a concise definition of the education system of Nepal as endured through the communique of the most recent education plan for Nepal. Third, The unresolved circumstance of English in Nepal has been tested sequentially. Fourth, look into various factors of English language assessment in Nepal. For example, the difficulty of association among instructional application goals and assessment methodology, modifications of assessment methods, modern assessment rehearses, washback impacts, assessment proficiency, and validity and reliability features in the assessment regarding the applicable variables that affect the practice and picks of evaluation. Finally, the review winds up with numerous placements, primarily based on the ramifications of language assessment and education in Nepal.

2.9 Assessment principles in Nepalese curricula: the foundations of meaningful assessment

A curriculum is a set of guidelines that focus and steer classroom instruction and evaluation. For example, instructors must teach a travel unit if it is mandated by an English curriculum. Following the instruction, students should devote time to improving their travelog skills. If travelogue is highlighted in the curriculum, it should also be evaluated in the classroom. The amount of knowledge and competence necessary to satisfy the unit's criteria is determined via

assessments. Curriculum, curriculum-embedded assessment, and instruction are all aligned in this method.

English is the second language of Nepal, and it is taught in all schools (CDC, 2006). Similarly, the National Education Commission papers, as well as several regional and national seminars and workshops, have placed a greater focus on making English a compulsory subject in all Nepalese schools. They go on to say that it should be taught from the very beginning of a child's education. As a result, this curriculum has been created for elementary level (Grades 1-5) education in Nepal to meet the immediate requirements of Nepalese children and lay the groundwork for their future studies in and via English. Also, it tries to help students develop a broad range of communicative abilities. Listening, reading, speaking, and writing is the main focus of this curriculum.

A brief part on assessment was included in the Curriculum Development Centre (CDC), a unit of Nepal's Ministry of Education (MOE) responsible for establishing the curriculum, printing and distributing textbooks, teachers' guides, and other related resources (CDC, 2006, p. 57). Continuous evaluation of each child's achievement of curriculum items in all four skills is most suitable in the lower grades. Formal tests for Grade 5 (and possibly Grade 4) can include (a) simple, familiar conversation (b) response to aural stimulus (listening activity) and an engaging reading/writing test with new material but familiar methods, vocabulary, and language that assesses the students' ability to read and understand simple sentences and write neatly and correctly.

In Nepal, the practice of public examinations formally began with the establishment of the SLC Board in 1934. NESP (MOE, 1971) introduced classroom-based assessment, known as internal assessment. However, the assessment system was discontinued after the implementation of the revised curriculum at the school level in the early 1980s, as it dropped the system of internal assessment from school education. The government of Nepal reintroduced CAS (Continuous Assessment System) in the primary grades of school education as a classroom-based assessment system in the late 1980s. Classroom-Based Assessment (CBA) conducted by classroom teachers is an integral part of teaching-learning and assesses each student. It facilitates student learning through regular feedback, and, therefore, it is formative. On the other hand, public examinations are generally detached from the classroom, and they are summative. Public examinations are conducted at various levels, including national, regional, district, resource centre, and/or school levels to assess each

student, mostly by external agencies such as government or independent/ autonomous agencies. Public examinations serve the purpose of selection and promotion of students for the next grade or level of education and certification of learning, which generally does not provide feedback to classroom teaching-learning or the education system.

The tests are based on a traditional testing theory with inadequate standardization of items and a testing process. However, most of these assessments identified teacher, school, and student-related factors influencing student learning. Although the curriculum stressed the importance of all four language skills: listening, speaking, reading, and writing, the Primary English Curriculum did not include assessment policies. Without a defined assessment, Teachers at this level would naturally choose whichever evaluation processes suit them best, for the benefit of students' effective learning, which is aided by assessment (Shepard, 2000). The typical test items consist of matching, true/false, short answer questions, filling in the blanks with the given clues, letter writing or essay writing, and rearranging words to make sentences. These test items categorically indicate the traditional approach to assessment focusing on discrete test items.

However, the basic level English curriculum (grades 6–8) incorporates both formative and summative assessment methods (Curriculum Development Centre, 2012). The curriculum refers to formative assessment as continuous assessment system (CAS) and claims that CAS has a lot of potential to help students achieve a lot of features. CAS, for example, assists students in identifying their strengths and limitations. Furthermore, CAS enables educators to evaluate pupils in a cost-effective and time-efficient manner. Most significantly, CAS aids instructors in identifying the psychosocial effects of assessments on students, allowing them to create measures to assist pupils to overcome the negative side effects of examination. The summative assessment at this level involves a final examination where the students at the end of grade 8 sit for the public examination called the Basic Education Examination (BEE). The written test, which encompasses reading and writing abilities, serves as the basis for the final examination. Grammar is incorporated in the writing section. Although the curriculum integrates all four skills, the English test in the BEE excludes listening and speaking tests; the test includes only reading and writing proficiency along with test items on grammar and vocabulary.

The students of grades 9 and 10 are given the next level of the National English Curriculum, which culminates in the most popular public examination known as the Secondary Education

Examination (SEE). This curriculum is implemented through a set of contents such as learning outcomes, functions, and language points that are delivered to the students through a textbook known as *Our English*, which is developed based on various themes, such as universal etiquette, festivals, pastimes and hobbies, adolescence, travel, and tourism (Curriculum Development Center, 2020).

The assessment methods stated in the curriculum include class participation, formative assessment, summative assessment, and public examination. The distribution of marks is also stated in the curriculum, indicating an uneven priority. The marks allocated for listening and speaking are 16 (8 marks for each skill), whereas 75 marks (40 marks for reading, 24 marks for writing and 11 marks for grammar) have been allocated for reading and writing skills. Listening and speaking skills, however, have been scrapped from the SEE. Most importantly, listening and speaking marks can be given internally.

The higher secondary English curriculum focuses on teaching and learning English as a skill-based subject so that the learners develop competence in the language to successfully communicate in real-life situations (Curriculum Development Center, 2020). Consequently, learner-centered approaches to develop students' communicative competence were prioritized in the higher secondary English curriculum, and initiatives were taken to integrate all four English language skills: listening, speaking, reading, and writing, as well as the other components of language: grammar and vocabulary (Curriculum Development Center, 2020). This curriculum is delivered to the students through content and language integrated teaching (Curriculum Development Center, 2020). The contents of the curriculum has been developed as a textbook known as 'English'. Language development and literature are the two aspects of the book. The language development area contains a variety of current issue-based local and global topic texts aimed at improving intensive reading abilities and fostering proficiency in grammar, vocabulary, speaking, and writing in various forms.

The literature department contains genre-based literary works for both intensive and extensive reading, allowing students to distinguish distinct features of the texts and practice creative writing. The book is based on a range of themes, such as education and humanity, communication, media and society, life and love, and health and exercise. The English assessment scheme at the higher secondary level includes continuous assessment in the classroom, internal examinations, and the public examinations known as the School Leaving Certificate Examination (SLCE).

Overall, assessment is an integral aspect of English language instruction at the post-secondary level in Nepal. Assessment in tertiary public sector English foundation courses has likewise been enslaved by a rigid system. There is also a rigid system of assessing English learners' proficiency, similar to pre-university practices. Private universities, on the other hand, have a slightly different procedure. Teachers in private colleges employ several formative classroom assessment tools. However, feedback is less emphasized.

2.10 Reforms in English language assessment

Since its founding in Nepal in 1934 (1991 BS), the SEE board, formerly known as the School Leaving Certificate (SLC) board, has modified the evaluation systems of English language courses multiple times, notably the letter grading in 2014. In 2000 (2057 BS), the SLC examination board introduced English listening and speaking abilities into the test system, and the testing centers used secondary English instructors to administer the test (Santwona Memorial Academy, 2013).

To treat assessment as an integral part of learning, the continuous assessment system (CAS) was introduced as a concrete initiative of School-Based Assessment (SBA), which has been widely practiced in the basic school system of Nepal (National Curriculum Framework, 2007). The assessment methods integrated into the SBA policy include class tests, classwork, homework, class assignments, achievement tests, and summative tests. However, the NCF (2009) reported that SBA was not successfully implemented. It has been continuously recorded that very few schools implemented SBA due to extra load teachers, a shortage of trained teachers, a complex type of recording system for SBA results, and a lack of close monitoring and mentoring system (Chongbang, 2021).

When the National Curriculum Framework was revised in 2009, continuous assessment system (CAS), which was re-emphasized in the basic education curriculum, was introduced in the secondary education of Nepal. Class tests, portfolios, classwork, observation, practical work, and homework as assessment methods were recommended in the curriculum. However, CAS implementation has raised questions too because it is being implemented without effective teacher training and a lack of an adequate teacher's manual.

The report entitled "Reflecting Stakeholders' Experiences with Classroom Assessment Practice in the Complex Contexts of the School System in Nepal" presents mixed outcomes

of CAS policy (Nepalese Journal of Educational Assessment, 2021). The study collected data from students, teachers, headteachers, and parents. Of headteachers, they are not familiar with CAS let alone their effort to encourage other teachers to implement it although they are the school level leaders, supervisors, and mentors (Nepalese Journal of Educational Assessment, 2021).

The study reported that the headteachers and teachers perceived a lack of training as a barrier to carrying out CAS in practice. Similarly, the research participants, headteachers, and teachers, shared the experience of having received no monitoring from concerned authorities regarding the implementation of CAS. Likewise, the case study recorded that teachers had not heard about CAS, which signifies very rare or no talk about CAS and its practice in schools while students are engaged more for the exam than for the study and learning from the lessons.

In addition, the research findings of Wagle, Luitel, and Krogh (2019) suggest that policy guidelines are not being followed by schools. However, the interesting fact is that, according to the anecdote, the internal test is conducted every Friday and the teacher registers the students' marks. The study also demonstrates that students like CAS practices. The main reasons why students favor CAS, as reported by the teachers, include the effectiveness of CAS in improving students' learning.

Moreover, the teachers also took the socio-economic dimensions of the students into cognizance while conducting the CAS. The challenges the headteachers and the other teachers shared concerning the implementation of CAS include absenteeism; the passivity of students; inadequate training and support; fragmented collaboration among stakeholders; shortage of classroom teachers in the schools; high teaching load of the teachers; large class size; financial constraints; lack of time to check homework and tests; lack of a CAS manual; and lack of monitoring.

At present, in basic and secondary education, formative and summative assessments, as well as internal and external assessments, are used to measure students' learning achievement as required by this curriculum (National Curriculum Framework, 2019). Based on the subject, both theoretical and practical evaluations are carried out. Students' learning outcomes are tested, evaluated, and certified through regular and ongoing assessments, school-level assessments, and public examinations. From grades 1-7, school-based assessment is widely

used, while in grades 8, 10, and 12, students are required to appear in high-stakes tests or public examinations.

With the exclusion of listening and speaking skills tests from public examination, partial assessment reform from grammar-translation oriented tests to skill-focused assessment has occurred in English language assessment in Nepal ever since the communicative language teaching (CLT) approach was introduced in 2000 (Rana and Rana, 2019; National Curriculum, 2007, National Curriculum, 2019).

Despite several notable gains in Nepal's education sector, such as increased school enrollment and better gender parity, necessary attention to evaluation techniques is still lacking. In reality, in this low-income nation, allocating enough funds for education and investing in educational research is secondary, since poverty alleviation and safeguarding health and security have been prioritized in the government's development strategy. Assessment reform, however, is a complicated process that is often difficult to undertake, according to research (Cheng & Curtis, 2010). Although some Asian nations have conducted studies on various aspects of examination systems (Kwon et al., 2017; Qi, 2007; Qian and Cumming, 2017), Nepal has a dearth of research.

2.11 Current assessment practices: a mismatch between policy and operation

National assessments, public examinations, and school-based assessments are the three types of assessments used in Nepal (ADB, 2017; Poudel, 2016). The National Assessment of Student Achievement (NASA) is a national assessment program used to provide policy input in Nepal. Public examinations are student certification programs that are conducted externally after Grades 8, 10, and 12 by examination offices located at various levels. The school-based assessment (SBA), which is used at the micro-level, is considered crucial. In addition, the policy has reemphasized the importance of implementing CAS more effectively in educational settings. In the assessment policy, it was envisaged that student assessment would be changed into CAS and that feedback from CAS would be used to create and implement educational programs and that learners would be competent enough to fit into society (CDC, 2007).

During its pilot phase, the Basic and Primary Education Project II (BPEP II) used the Continuous Assessment System (CAS), which included teacher training and the development of various assessment methodologies. Despite these efforts, neither the general assessment of

students nor the evaluation of the educational system based on students' assessments was determined to be beneficial. As a result, CAS was shown to be ineffective (CDC, 2007, p. 26). The student evaluation system has not been designed as an essential component of teaching-learning activities, nor has it been linked to the intellectual level, interest, pace, or requirements of the students. For the general growth of education, systematic programs for assessing students, teachers, schools, and curricula have yet to be implemented (CDC, 2007, pp. 27-28). The procedure of evaluating teachers, schools, and the entire educational system based on students' learning success has not been proven.

The need for a comprehensive and fair test system for assessing pupils has been emphasized by educators and assessment professionals. The examination should be focused on meeting curricular objectives, enhancing teaching/learning activities, and giving feedback for national assessment policy development. It has been consistently said that Nepal's current assessment system is ineffective and undervalued as a critical component of teaching and learning. The majority of schools do not employ a formative assessment strategy. The assessment system relies solely on examinations, which has resulted in a slew of errors. Learning is not aided by classroom activities, homework, project work, or extracurricular activities.

From the primary level to the university, Nepal has demonstrated the utilization of a typical high-stakes summative evaluation system (MOE, 2016). A high-stakes testing system exists in every grade level from elementary through the end of the school year in all schools (i.e., public, technical, Sanskrit, and private educational systems), where students must pass every grade level exam to advance to the next level. This means that in Nepal, practically every grade level exam is a necessity for advancement to the next grade. Because these exams are high-stakes, pupils begin to feel exam-related pressure as early as preschool. Beginning in the 2016 academic year, however, efforts were made to soften the high-stakes evaluation system in Grade 10.

The present evaluation system is ineffective, and learning facilitation has not been realized intrinsically. Similarly, formative assessment has not been widely used and established as a learning assessment (National Curriculum Framework, 2007). An assessment approach has been established based on the examination point of view. The assessment does not take into consideration homework, class assignments, extracurricular activities, social activities, or other things. The assessment could not be tailored to the intellectual level, pace, interest, or needs of the pupils (National Curriculum Framework, 2007).

Concerning classroom-based assessment, there is a policy provision to implement a Continuous Assessment System (CAS) in the primary grades of Nepalese schools, which has now been extended to grade seven. However, in practice, some schools have been using CAS as a non-testing device while others have been using a combination of non-testing CAS, trimester and final examination, and some schools have only been using various types of tests for deciding students' grades and promotion. One of the weaknesses found in the practices of CAS in Nepalese schools is the detachment of assessment from classroom teaching-learning; as a result, the use of assessment for formative purposes is minimal. In this regard, "assessment results have hardly been used in the teaching-learning process; rather it has been a ritual of filling out the forms in many schools" (Poudel, et al., 2015). In Nepal, public examinations have been conducted at the national level for grades 10, 11, and 12—which mostly assess the lower order skills using test items that are mostly not standardized. The recent practice of grade 8 district level examinations is just a ritual in most cases.

The overall measurement approach in Nepal is test-oriented since examinations are ritualistically implemented here (National Curriculum Framework, 2007). The new formative assessment practice, with its true purpose of facilitating learning by influencing pedagogical methodology, and materials used in teaching and learning, is hard to implement in Nepal for various reasons, which are discussed and explained below. Ineffective assessment methods focusing mainly on rote learning and grammar-focused are prevailing in the testing and assessment culture in Nepal's English language teaching and learning (National Education Policy, 2016). Therefore, they risk the validity of the assessment, an essential quality criterion of assessment. There is a gap between what is "intended to be taught and what is measured." Therefore, the assessment practices are unable to examine if the learners are genuinely acquiring communicative competence, a goal of learning English set in the curricula, to function effectively in real-life contexts.

Nepal adopted communicative language teaching (CLT) as a language teaching methodology at the basic and secondary levels in 2007 (National Curriculum, 2019). It is expected that the assessment methods embedded in the CLT approach will be used by individuals and institutions. However, measurement specialists have identified several weaknesses of the single high-stakes test. One single high-stakes test is unable to help students, educators, or schools in making important educational decisions (Ritt, 2016). Students are being treated with a one-size-fits-all education approach, which is causing unreliable test scores (Popham,

2015). Assessment experts oppose high-stakes testing because using a single indicator of competence to make important decisions about individuals or schools violates the professional standards of the measurement community (AERA, 2000). Other critics are concerned that the unintended effects of high-stakes testing lead to “perverse” (Ryan, 2004) and “corrupt” educational practices (Jones, & Hargrove, 2003; Nichols & Berlinger, 2005), while Pedulla et al. (2003) argue that the pressure of doing well on a test seriously compromises instructional practice. Where the high-stakes test is in practice, the teachers put considerable importance on the contents that are tested, and the students’ achievement in the maximum test score because the test results are associated with the national test results (Harlen, 2007). As a result, most of the teaching time is devoted to preparing the students for the test or doing the testing. The quality of teaching decreases if the teacher spends a lot of school time preparing students for tests.

Alarming is the fact that these widely used English high-stakes tests exclude two essential language skills from the tests: listening and speaking. At present, continuous assessment, internal examinations, and public examinations are used to assess students in Nepal at the basic and secondary levels. Continuous assessment using tests and homework is held in the classroom during the lessons while the individual institutions conduct the internal examinations and the public examinations are administered by the educational boards mentioned above. Although the literature on assessment in the English foundation courses is not available, I, as the teacher of this level, report that assessment methods at this level include both summative and formative procedures, which are, in fact, blended with the eventual summative purpose, which is grading. Common assessment procedures include quizzes, tests, presentations, assignments, and interviews.

Because of the test-oriented culture, feedback, mostly associated with formative assessment, has hardly received adequate attention from researchers in Nepal. Feedback refers to the information regarding the gap between students’ performance in the assessment task and the intended learning outcomes (Black & Wiliam, 1998). Need-based detailed feedback helps learners minimize their lack of content understanding and also boosts their motivation in learning (Islam & Bt Stapa, 2019). Hyland (2006) maintains that teachers may enjoy the rare opportunity of exercising significant tasks in the classroom by providing the students with feedback.

2.12 Overview of Nepalese practices on Students' Assessment

Assessment is a critical tool for determining whether or not an education system is providing the desired outcomes for students, the economy, and society. The information obtained about students' achievements gives a chance to improve student learning. Furthermore, the input from the evaluation findings motivates all key stakeholders to define and modify their responsibilities in order to improve student learning. As a result, one of the primary functions of assessment is to hold the educational system accountable for the learning of children.

Three forms of evaluation are used in diverse educational systems, each with a different purpose. Classroom-Based Assessment (CBA) by classroom teachers is an important aspect of teaching and learning since it evaluates each student. It is formative in nature since it supports student learning through constant feedback. Public examinations may be held at various levels, including national, regional, district, resource centre, and/or school levels, with the goal of assessing each student, and are usually administered by an external agency, such as the government or an independent/autonomous agency. Public examinations are often held outside of the classroom and are summative in nature. On the other hand, the primary goal of large-scale national assessments is to provide policymakers, program planners, and implementers with system-level information. However, the ultimate goal of system reform is to increase student learning. Since the formation of the Education Review Office (ERO) in 2010, Nepal has launched a systematic practice of national assessment and has completed two rounds of student assessment using large samples of schools and students.

With the inception of the SLC Board in 1934, public examination procedures in Nepal effectively began. Classroom-based assessment, also known as an internal assessment, was introduced in the National Education System Plan (NESP) (MOE, 1971). However, the evaluation system that discontinued the installation of a revamped school curriculum in the early 1980s removed the system of internal assessment from school teaching. In the late 1980s, the Nepalese government reinstated CEM (Continuous Evaluation Method) as a classroom-based assessment system in the elementary grades of school instruction. In addition, Nepal initiated large-scale national examinations of student success in various grades in the mid-1990s, although not on a regular basis. Since 1995, the Nepalese government's Ministry of Education (BPEP/DOE) has commissioned a variety of student accomplishment evaluations.

From 1995 through 2011, the Nepalese government's Ministry of Education/BPEP/DOE commissioned a variety of examinations of students' learning at various grade levels. BPEP (1995, 1997), EDSC (1997), BPEP (1998), EDSC (1999), CERID (1999), EDSC (2001, 2003), CERSOD (2001), EDSC (2008), Fulbright (2008), and EDSC (2011) administered student assessments from 1995 through 2011. (see: Ministry of Education & UNESCO Office in Kathmandu, 2015; EDSC, 2011). There was a lack of institutional setup and a regular system of national evaluation because these assessments were commissioned by the MOE/BPEP/DOE and carried out mostly by external entities. The tests were based on classic testing theory, with insufficient item and testing procedure uniformity in relatively small sample numbers. However, the majority of these evaluations highlighted teacher, school, and student-related variables that influence student learning.

The Education Review Office (ERO) was established in 2010 as part of the School Sector Reform Plan (SSRP, 2009-2016) (MOE, 2009). ERO's primary duty is to undertake large-scale system-level assessments on a regular basis. ERO has been administering the National Assessment of Student Achievement (NASA) for various grades since its inception (Office, n.d.). It undertook two nationwide evaluations of student achievement for grade eight in 2011 and 2013 (Office, n.d.), as well as one national assessment of each in grades five and three in 2012 and 2013. (Office, n.d.). The second round of National assessments for students in grades three and five is also going to be completed, with data processing ongoing. Each of the previous assessments was based on a broad sample of schools and students from Nepal's diverse ecological zones and development areas. As stated in the preceding section, the primary goal of Classroom-Based Assessment (CBA) is to increase student learning or performance by enhancing the teaching-learning process. Whereas public tests are used to select and promote pupils to the next grade or level of the They are used to certify learning and not give feedback on classroom instruction, learning, or the education system.

Similarly, the primary goal of periodic evaluation of student achievements is to give policy input to the education system as well as to create evidence-based data for tracking progress over time. In terms of classroom-based assessment, there is a policy provision in Nepalese schools to use the Continuous Assessment System (CAS), which has now been extended to grade seven. In practice, however, some schools have used CAS as a non-testing method, while others have used a combination of non-testing CAS, trimester and final examinations,

and still, others have simply used various forms of tests to determine students' grades and advancement.

One of the flaws of CAS methods in Nepalese schools is the separation of assessment from classroom teaching and learning; as a result, the use of assessment for formative purposes is negligible. In this sense, "evaluation findings have seldom been utilised in the teaching-learning process; rather, in many schools, it has become a ritual of filling out forms" (Poudel, et al., 2015). In Nepal, public examinations have been held at the national level for grades 10, 11, and 12—which mostly examine lower-order abilities using non-standardized test problems. In most situations, the current practice of grade 8 district-level tests is only ceremonial. Since the establishment of ERO in 2010, some type of systematic and regular practice in large-scale national evaluation has begun.

2.13 Washback effect

Regarding the intertwined relationship between washback and teaching/learning, it is critical to investigate the effects of washback induced by English language assessment techniques in Nepal. Because of Nepal's test-dominant assessment culture, high-stakes testing causes awful washback effects that impact English language education and learners. For one reason, the washback effects impact students' exam preparation. This learner attitude jeopardizes the attainment of curricular goals and objectives because students will focus on just those aspects of curriculum information that are included in the examinations. In reality, historical-grammatical exam procedures are frequently replicated in the guise of communicative testing.

Another severe issue is that students will focus solely on scoring high on the test, forgetting the curriculum's primary purpose of learning communicative competence in the target language. And, in order to do this, students in Nepal are pushed to engage in rote learning and memorizing, which eventually leads to surface learning. Students' desire to get top scores at whatever cost encourages them to pursue alternate channels such as enrolling in private tuition for an extensive focus on test preparation, purchasing guide books with suggested examination questions and answers, and attending coaching facilities that instruct students on how to perform well in public tests.

Negative washback effects of exams are also investigated at the tertiary level in Nepal; the study suggests that negative washback affects students' learning and emphasizes the

significance of teacher training. Aside from the basic linear causal link between washback and teaching, washback has sociopolitical, economic, and psychological consequences (Al Amin & Greenwood, 2018a, 2018b).

Examinations in Nepal have a significant influence on instructors, students, parents, and society as a whole. Students and parents become keen on scoring high on the English examinations, perceiving test results as "gatekeeping devices" (McNamara, 2000, p. 74). They are also impacted by societal pressure since Nepalese society evaluates students' achievements based on their performance in public examinations (Bhattraï, 2019). Based on this information, we can infer the power of tests and examinations in Nepal.

Furthermore, the washback effects are exacerbated by governments' political agendas since exhibiting increased passing rates through standardized examinations is a typical practice in order to meet their electoral goal of improving education in the country. This practice has ramifications for English language instruction in Nepal since English teachers frequently feel forced to prepare students for the test. We contend that unless and until a conscientious pragmatic strategy of assessment toward the meaningful development of communicative competence in English is developed and implemented promptly, negative washback will continue to have a negative impact on English instruction and learning in Nepal.

Chapter 3

Research Design and Methodology

This chapter details the research approach that was used. The procedures and approaches used to conduct a study are collectively referred to as the methodology. The notion of Hitchcock and Hughes is iterated by Cohen et al. (2018, p. 3), who state that a researcher's "epistemological assumptions" are influenced by his or her "ontological positions," which in turn escort the "methodological considerations" that in turn guide "instrumentation and data collection." The examination of the researcher's philosophical views and methodological approaches is what happens throughout this chapter.

3.1 Philosophical positioning of the Present Research

This study operates on the ontological premise that there are "many realities" and that reality is relative (Cohen et al., 2018; Creswell & Creswell, 2018; Denzin & Lincoln, 2018). According to Lincoln and Guba (as stated in Mackey and Gass, 2012, p. 182), the ontology of qualitative research is predicated on the notions that "reality is relative" and "it is co-constructed via social interaction." In contrast to positivistic research, this research does not assert that its conclusions are universal or broadly generalizable.

It is appropriate to use the ontological stance for this research because it addresses the ethical complexities of language testing and evaluation and because ethical concerns are inherently contextual and relative (Fulcher, 1999; Hamp-Lyons, 2000; Rahman, 2017; Vladi, 2015). It can also be supported by the fact that ethics is primarily decided by a group or society in order to control the social and cognitive behaviors of its members. Additionally, a society's ethical foundations are strongly anchored in its sociocultural values and the traditions that its people connect with. Despite the fact that moral conundrums arise everywhere, communities establish a particular set of ethical standards and ideals to guide members' conduct at a particular period. Even if moral standards and values evolve over time (Colnerud, 2006). There are ethical conundrums in every civilization because the origin and ontogeny of ethical difficulties are always like a vibrant movement that is challenging to comprehend in its whole.

Similar to this, the researcher's epistemological viewpoint encompasses the fact that knowledge is both personal and socially related. the creation of knowledge as an individual, societal, and relative process. In order to better grasp the issue at hand, individuals' lived experiences must be understood and interpreted, which is a process that is embodied in the creation of knowledge. Qualitative research is described as being epistemologically "subjectivist" by Lincoln and Guba (as reported in Macbeth and Gass, 2012, p. 182) and that "research findings are formed via the interplay of researchers and their participants."

3.2 Research Design

Through this research, I wanted to explore the curriculum, testing procedures and classroom teaching learning practices. I followed a qualitative approach because the focus is on qualities in the curriculum, its practices in classroom teaching and learning and testing. So, the outcome is best understood contextually. Also, this method takes a variety and number of forms depending on the project at hand, and there's a number of tools available in its toolkit such as interviewing, discussing with participants, taking notes, analyzing documents, textual analysis, or drawing from related publications. And obviously all these methods provide holistic insights. To interpret specifically, first, the qualitative method focuses on lived experience in context; what are people doing, and saying, both verbally or in writing. It included participants' attitudes, behaviors, and practices. Second, it demonstrates the pattern of events along with their consequences, and unfolds the process. Third, it helps to combine disparate data and makes it easily accessible later on to a larger audience. Most importantly, I believe qualitative research helps to solve contextual problems by providing vivid explanations and lived evidence.

As I adopted a qualitative approach, I visited Curriculum Development Center, National Examination Board, Ministry of Education, Education Review Office and among others to identify required documents for analysis and find participants. According to the suggestions received from officials, I went to several schools and observed classroom teaching learning and assessment systems. I talked to the teachers and based on their interest collected their email to invite for an interview. In addition, I also collected former English test papers for analysis. Online journals like Google Scholar, Taylor and Francis etc. were frequently used to select similar research papers for references. Moreover, the researcher sent an email to potential participants to obtain permission for an interview and assumed that their honest,

accurate and unbiased responses were valued and maintained confidentiality for such significant research work.

The qualitative data was gathered in two stages. The first part entailed an analysis of curricular policy documents and test papers, which disclosed the nature of the test, including the structure and competencies measured by the test. In the second step, test developers were interviewed to determine why the exam was developed in the manner that it was.

3.3 Participants

Different stakeholders directly involved in the English language testing process were involved as the participants from whom the data for this research were collected. I sent an interview invitation email to 15 potential participants out of 30 from different schools. Finally, I received a positive response from 10 teachers. Therefore, I purposely selected them based on their paper setting and moderating experience in secondary level. Following that I interviewed five SEE English test setters and four test moderators from the National Examination Board (NEB) to learn how test questions are prepared and regulated, how test content representativeness is handled, and what variables impact test writers and moderators in their work. We used purposive sampling to discover respondents who had experience with a certain occurrence while choosing participants (Creswell and Clark 2011). In the local education system, the test design process consists of only two stages: writing the test paper and moderating it. As a result, purposive sampling assisted me in selecting experts at both rounds. Given that only a few teachers from government institutions are accepted for the tests, two rather small groups of participants from four government colleges were selected.

Table 1: Profiles of the participants.

Participants	Teaching experience (years)	Paper-setting experience (years)	Professional Qualification/Training
S1	15	6	None
S2	20	5	None

S3	17	13	None
S4	18	10	None
S5	21	15	None
M1	24	12	None
M2	25	14	None
M3	16	8	None
M4	22	9	None

Table 1 shows the participant profiles. To safeguard the participants' identities, the paper setters were designated as S1, S2, S3, S4 and S5 while the moderators were designated as M1, M2, M3 and M4. Despite having more than 10 years of teaching experience and at least five years of experience setting and moderating test papers, none of the participants had professional qualifications or training in testing or test creation. Before entering the English teaching profession, they had all studied English literature as part of their undergraduate and postgraduate studies. They only got brief in-service training related to classroom pedagogies from Teachers Service Commission, Nepal.

3.4 Site selection and sampling

Perry (2011, p. 57) asserts that “purposeful sampling is the paradigm that is most commonly used in applied linguistic research”. In line with this trend of sampling, the site and the sample in this study were selected purposely based on accessibility and convenience. Besides, to the extent possible, attempts were made to involve the information-rich informants based on my anticipation and experience. The data were collected from the different test constructors available in the Kathmandu valley only. Additionally, the Controller of NEB was also requested for a brief interview but his ideas were not included in the data analysis procedure. As to the site, Office of National Examination Board, Madhyapur Thimi, Bhaktapur, Nepal, Curriculum Development Center, Bhaktapur, Nepal, Education Review

Office, Nepal, Center for Education and Human Resource Development (CEHRD), Ministry of Education, Nepal, where required documents and participants were available were visited as often as needed.

Concerning English test paper analysis, altogether 21 question papers, three each from seven provinces from the past three years were selected as sample question papers purposely. The question papers were taken as samples mainly in terms of the latest curriculum.

3.5 Instruments

The qualitative study's data was gathered in two ways: (a) through the analysis of official documents, such as the official textbook (English), the National Curriculum (2019), former and current exam papers; and (b) through semi-structured interview. The interview guide was created using Hughes's (2003) test development framework, which was previously mentioned. The interviews yielded information on: (a) the paper setters' and moderators' impressions of the exam; (b) how they compose and moderate test items; (c) how they tie exam questions to the curriculum in terms of predictive value; and (d) what influences their work. While the documentary data provided information about the test's goals, format, and tasks, as well as assisted in the investigation of the tests relationship to the curriculum, the interviews provided in-depth information on specific factors that influenced the test design that could not be determined solely from documentary evidence.

3.6 Data collection

The task of collecting the data for the present study comprised a stepwise procedure as specified below:

Step I

1. Contacting the participants and taking their consent in advance.
2. Seeking permission for observing the examination centers, observing the test administration with the help of the observation checklist and maintaining field notes.
3. Conducting informal interviews with test constructors.
4. Collecting past three years' English question papers of seven provinces

Step II

1. Contacting the different groups of stakeholders as specified above and conducting semi-structured in-depth interviews with them separately.

The procedure is delineated in the following paragraphs.

To gather interview data (Silverman 2016), the researcher first went to numerous government schools and colleges in Kathmandu, Nepal's commercial city, explained the purpose and process of the research, took permission to observe exam centers and invited paper setters and moderators to take part in the study. She then requested them to give her permission to carry out the interview. After taking the permission from the participants, the researcher sent a questionnaire to prepare for the interview. Only those teachers who had prior experience creating and moderating SEE question papers were contacted. Before the data was collected, the respondents were given all of the information they needed regarding the study, and their written agreement was obtained corresponding via email (See Appendix A). The researcher assured them of the confidentiality in terms of research of ethics regarding the information obtained through the interview. Then, the researcher started the interview and asked them to speak without any hesitations, fear or anxiety because some information can also be kept off the record which would not be shared in the paper. Moreover, Interviews were undertaken in English, based on the choices of the participants and recorded utilizing an iPhone.

The national English curriculum was examined for information on the English Paper's curriculum objectives. The test guidelines in the prologue to English were also looked at to acquire a better understanding of the test format, techniques, target skills, and other important test design instructions.

Chapter 4

Data Analysis

4.1 Data analysis procedure

The qualitative data methodologies and analysis recommended by Richards were used in both the document and interview data analysis (2003). The systematically qualitative data analyzed and interpreted descriptively in a narrative style and a few data analyzed and interpreted are demonstrated by using simple statistical tools like tables.

I employed document analysis, an effective and efficient qualitative research method, as a plan for gathering secondary data. The study's title was determined after a thorough examination and analysis of the materials published between 1971 and 2019. These materials provided expertise and a wide range of facts, which helped to contextualize the research within its field. It has also tackled the problem of how poorly the English language is evaluated in Nepalese public exams. Materials that can be discovered in books, libraries, government offices, e-libraries, and internet search engines include public records, private documents, and physical proof. The information was acquired from sources including official Ministry of Education Curriculum Development Center releases and online news sources. The information was acquired from official reports from the Curriculum Development Center, Ministry of Education, National Examination Board, as well as various independent publications, journals, research papers, and books pertaining to assessment techniques, policies, and practices. It has looked into the factors contributing to the fall in English language testing. The study provides a thorough description of English language evaluation in Nepal, along with an analysis of its issues and procedures.

The policy documents were carefully reviewed to comprehend how the test reflects curricular objectives. Examining former exam papers, sample test questions from the SEE, test formats, and assignment descriptions were all part of the analysis of the materials. Through the analysis of the materials, the researcher identified the connection between the exam, official curriculum and classroom teaching learning practices.

According to Richards' (2003) methodology, the interviews were transcribed, and participants had access to examine them. The interview was reviewed numerous times in order to further confirm the findings, and initial coding was done to find emerging themes and important

codes (Shank 2006). Later, these themes and codes (derived inductively from the raw data as well as in light of pertinent literature on testing and test design more generally, including in Nepal), were carefully examined in order to allow the original codes to be modified in light of what is now known about English language learning policy, practice, and problem.

The information from the interview was also used to analyze the conclusions drawn from the documents. The interpretation of the interviews was also influenced by assumptions made during the document analysis. The interview data analysis, for instance, sought to examine the elements that may compromise validity as well as other technical and professional aspects of large-scale test design when the document analysis suggested that the test may have validity issues. Four dominant themes in regard to test development methods were found as a result of these analytical approaches.

One of these was related to questioning the degree to which the policy promoted communicative competence, while the other three were related to practical issues. These include a tendency for test developers to maintain current, dominant testing practices (known as "Conventionalism"), evidence of a compliant disposition on their part (known as "compliance"), and a propensity to interpret concerns about corruption and context (known as "context"). These themes helped create the "image" Richards (2003) refers to in the data.

4.2 Ethical Considerations

Research ethics are of the utmost significance. Given that Zahavi (2019), in his words, correctly alludes to van Manen's thesis that "phenomenological descriptions convey a moral weight," this research in particular regarded such questions seriously with significant relevance (p.125). I made a serious effort to complete my job in this research by completely adhering to the ethical guidelines and procedures. When and wherever necessary throughout the task, I make every attempt to eliminate prejudice and any sort of preconceived assumptions, known as "bracketing" in research (Koul, 2009; Langdrige, 2007). Participants were treated with respect, their written consent or agreement through email correspondence was obtained in advance to participate in the study, and their privacy was not at all violated. In addition, I am aware that the information collected from participants should not be used for any other purposes than those that have been made public.

Thus, during the conduct of this study, complete anonymity, uniqueness, secrecy, fairness, compliance with national and international laws, respect for the preservation of subjects' privacy and dignity, etc., were taken into consideration as the ethical criteria. The protection of the material was more important than it would have been in any other study since this one, in my opinion, is a very "sensitive" research that looks into very private topics related to public exams. I made the decision to transcribe the data in hard copies and store them in a private location where they are out of the reach of others in order to safeguard the data more effectively. Overall, I made an effort to follow ethical concepts such as autonomy, non-maleficence, beneficence, fairness, faithfulness, and propriety when collecting data and reporting it.

4.3 Delimitations

The following were the study's limitations:

1. There are many aspects of English language testing that need to be addressed, but only the three most important ones—test design, curriculum, and variables that affect test design—were included in the research.
2. The participants were all connected to institutions located in the Kathmandu valley.
3. Only Secondary level English language examinations were taken into consideration.
4. The statistical methods employed in the study were confined to the level of specific numbers because the study was created using a qualitative paradigm.
5. The level of validity of the results and conclusions depends solely on my strict commitment to the theoretical underpinnings of testing and assessment, the breadth of the information gathered through interviews, and the careful analysis of the data.
6. A wise balance has been reached between how much or how little clear the reporting should be, taking into account the "high sensitivity" of the study, in regions where the readership may demand deeper detail.

Chapter 5

Findings

The results show a significant gap between the curriculum's goals of fostering communicative competence in connection to English language learning and testing in Nepal and the degree to which this was made possible by the instructions provided to test creators and their actual practices. This section goes into further detail on each of the four major themes.

Principled policy: communicative proficiency

The SEE English test measures how much communicative competence students have attained at the conclusion of secondary studies (Years 9–10), based on the national curriculum goals and the objectives listed in the textbook's preface. In other words, the test should gauge how well students can communicate in academic and professional contexts using the four abilities of listening, reading, writing, and speaking using English.

However, only reading and writing abilities were really evaluated, and the topics that were evaluated didn't really relate to communication in real life. The NEB's requirements provide broad guidelines for paper setters on how to structure the exam, what kinds of test items to include, and how to display items to students, among other test-related topics (e.g., multiple choice questions and gap-filling). A sample test paper is given to test authors to use as a guide. However, there are no explicit guidelines about how to sample, write, ensure representativeness, or use criteria for performance levels or scoring methods.

Most importantly, the construct (ability) that the exam aims to evaluate is not clearly defined. The NEB sample test paper and the recommendations are problematic since they are not supported by the evaluation of communication skills. The sample exam is designed to evaluate only reading and writing abilities; listening and speaking abilities are not evaluated. As a result, there are issues with the curriculum's objectives and how assessments should be conducted, even at the national level. This raises doubts about the test's reliability for the nation's educational, social, and economic objectives.

Conventional Culture in practice

The interviewees reported that in writing test items, they usually sample content from the textbook, *Our English*, for this level of education. One interviewee mentioned that as test

writers, they also consulted test items or prompts from previous test papers used in different education boards. As S1 explained: We have to refer to and examine the earlier questions... to see the common passages [texts for reading comprehension]... To understand the pattern and what the exact focus points of the earlier board questions are. The paper setters' reliance on items from previous years' tests has resulted in the repetition of items over the years across education boards. Table 2 provides examples of how five topics were repeated in the SEE English test over the years in the assessment of reading skill through activities such as fill in the blanks, match the following and answer the given questions. In addition, Table 2 shows that the subject of Past and Present has been brought up four times by the various educational bodies. Other subjects have also been brought up, but not as frequently. As was already mentioned, Nepal frequently repeats exam questions. It has only lately come to some people's attention that the purpose of developing communicative competence and pedagogy may be impacted by content repetition.

Table 2. Repetition of test content across education boards and years

Topics	I wandered Lonely as a Cloud	The Chimney Sweeper	Past and Present	Habit Cultivation	Memoirs of my visit to France
Educational Boards and Years	Province 1: 2074 Province 2: 2075 Bagmati: 2075	Gandaki: 2074, 2075 Sudurpaschim : 2075	Karnali Province: 2074 Sudurpaschim:2074 Province 1: 2075 Province 5: 2075	Bagmati: 2074 Province 5: 2075	Province 1 2075 Gandaki: 2075 Bagmati: 2078

The paper setters' conventional approach of consulting things from prior years may have resulted from the fact that they lacked formal training in creating new items. S2 offers perceptions on how test items are created by unskilled paper setters:

“In our nation, we learn by practice. Certainly, skilled question writers might come up with something more valuable. The repeated use of test items was also claimed to be caused by students' anxieties about the exams.”

The following excerpt will help you understand this culture of repetition:

S3: Yes, we have to ask the same questions again. Because, you know, every time we attempted to create a fresh question paper [without repetition], all the students at the time grumbled that the test's material was not from the curriculum.

An illness, that's for sure and certain professors, setters, and moderators can be like that; thus, you must be restricted to the previous question papers.

New things are not welcomed by students, as S1 and S5 indicates. Their rote learning style necessitates predictable questions with simple solutions. The educational outcome, according to Maniruzzaman and Hoque (2010), is that certain students—especially the poorer candidates—tend to spend a lot of time memorizing such answers (to previous test questions).

It was not deemed bad because the moderators themselves enjoyed repeating things and actively encouraged it. Despite the fact that repetition reduces the sample size, calls into question its validity, and has a detrimental impact on pedagogy (Hughes 2003), M2 was of the opinion that repetition was not a concern, particularly if items were repeated across boards. However, Maniruzzaman and Hoque's (2010) study, which found that teachers "skip subjects and lessons that may not be assessed" and "narrow down the syllabus, and teach their pupils the selected topics to be tested in the examination," reveals the detrimental impacts of repetition (68). This has a wider resonance with comparable activities in connection to more generalized standard testing procedures in other cultural contexts (cf. Au [2009]; Gu [2014]; Jilani [2009]; Hursh [2007]; Klenowski and Wyatt-Smith [2012]).

Another element that encourages test item repetition is the assurance of high grades. Education stakeholders appeared to be more focused on grades than learning effectiveness or academic success. S4 offered information on this element: (-) Our attitude (-) Our inclination to get good grades (-) and our stereotypical thinking, etc. S5 nodded in agreement, pointing out that while writing exam questions, one must replicate the questions from other boards. This force comes from both students and other stakeholders. In other words, they take the students' interests into account.

Most importantly, the interview indicated that there was no standardized holistic or analytical grading system used when assessing student test assignments, indicating a cautious approach to testing methods. Furthermore, despite obvious problems with validity and reliability, it was routine practice for examiners to provide test task scores based on an overall impression. The

marking guide for the SEE test paper provides a difficulty because the supplied instructions are ambiguous and insufficient to guarantee proper marking, in short.

Compliant disposition

In addition, both test creators and moderators provided a variety of complementing instances of compliance—doing what they were seen to be instructed. The moderators' lack of a decisive role during the moderation process was exposed by the participants. They noted that realistically speaking, their responsibility was restricted to confirming that the test authors adhered to the established exam structure (i.e. sample question paper provided by the NEB). Moreover, they are told to verify the test papers for any spelling and grammar mistakes. They endorsed the test questions as they were supplied by the paper setters practically without making any significant changes and sent them to the Controller of Examinations, who would choose which version of the moderated test questions should be used. It was unheard of in the nation to trial the test paper or to produce proof for an accurate interpretation of the test and test results.

The English test does not accurately reflect what it was designed to measure. Assessment of speaking and listening abilities was not offered. The activities that are given have little relevance for communication in the actual world, even when evaluating reading and writing abilities. The majority of the participants agreed with this remark, noting that the test did not reflect curricular aims.

According to S1, the objectives are not met because we are just evaluating two talents, hence it is not operating properly. However, S1 was required to create a non-representative test paper as a paper setter for the SEE exam. He said that the test sample and the related guidelines were written in a way that they did not accurately reflect the curriculum, but there was nothing he could have done to respond otherwise, and he was following instructions. As S2 clarified that they actually have very little scope. The setters are instructed. They do everything under their authority's instructions. So they have no opinion about their role they do as per instructions through the guidelines and the model test.

The exam's lack of representativeness might be linked to its unclear objectives for the SEE English test. According to Kucuk and Walters (2009), the lack of representation "may be caused by a lack of clear, well-defined objectives, as clear objectives help test writers establish which language points to weight on achievement exams and help instructors decide

what should be taught." The moderators shared the powerlessness of the study authors in assuring test representativeness. They are unable to alter exam questions to better reflect the curriculum's objectives. Their task was restricted to fixing minor, technical faults, and they had to rely on the orders they received from the board authorities: M3 states that there aren't any rules. The first thing the board authority wants from them is any type of surface-level spelling and grammatical errors in the exam questions.

According to M4, even the authorities didn't care whether the test accurately reflected the curriculum. It is possible that covert socio-political and economic factors contributed to this compliant method to testing representativeness. According to Bachman (1990), "language testing happens in a social and educational framework, and the applications of language exams are generally dictated by political requirements" (291).

Contextual Practice

It is sometimes argued that speaking and listening skills are not assessed for "practical reasons" (Islam et al., 2001, 235), which include issues with resources, logistics, and competence for testing more than a million students throughout the nation every year. The exam is not a good representation of the curriculum, particularly in terms of content, which is a major concern. However, there were also reports of other contextual elements, such as corruption:

The Education Ministry and the NEB people assume that if they introduce listening and speaking in their assessment system, then it will become a scope for the teachers to victimize students. Even though the curriculum recommends that listening and speaking should be taught and these should be evaluated, NEB people stressed that if we introduce listening and speaking in our testing system then the teachers will make money using these tests.

M2 discusses moral and professional matters from the viewpoints of the nation's educational authority. Although placing blame on teachers might be perceived as a political ploy to avoid accountability for providing the tools and knowledge required for educational assessment, there may be some validity to the claim that children may be the victims of instructors' partial evaluations and self-serving desires. For instance, it has been stated that unethical actions driven by financial gain have occurred in connection with private tuition (Hamid et al., 2009).

Other contextual factors, such as school congestion and problems with access to high-quality education, particularly in rural places, seemed to have a significant impact on the paper setters' choices:

According to S2, the educational institutions in Nepal are overburdened with students, lack distinct examination centers, and must take into account the interests of students who live in distant places while students who live in urban regions have better educational opportunities. However, those who live in rural regions constantly fall behind, so in order to achieve a balance, they select questions that take into account and keep in mind the very level of comprehension of people who live in the rural area.

Due to pedagogical disadvantages, students in rural Nepal frequently perform poorly in school. Because of this, the question writers tried to give them an easy way out by repeating answers. Material limits connected to test materials and texts were found to be significant in further exploring why paper setters repeated elements from prior papers, leading to an emphasis on textual portions that students were previously familiar with: S2 states that occasionally questions are chosen at random but that most are chosen from prior questions. Some test questions come from many boards. They produce it because they emphasize particular sections. S3 further clarified the rationale behind the paper setters' sampling of "certain portions." There aren't many reading passages in the textbook that are instructive enough to be used as samples for writing assignments with appropriate requirements. Because there are so few sentences, the English document is quite straightforward. Limited passages (-) They have to repeat such portions in order to provide us with information. Because of this, even while it was understood that they severely restricted the potential of the tests, certain, frequently significant, contextual concerns nevertheless had an impact.

Chapter 6

Discussion

Declaring the issue (goal) of the test is the crucial first step in the design of a language test (Docherty and Corkill 2005). The English test paper I examined in this article, however, does not explicitly identify or reflect any aim, even in connection to the core objective of communicative competence; in fact, there is no clear explanation of the construct that the test measures. Furthermore, there are no particular instructions for writing the exam or guarantees of its representativeness in the official standards. Test requirements, in the words of Fulcher and Davidson (2007), "give us the logic behind the different choices that we make". The NEB's recommendations provide basic instructions, details on the exam type, test strategies, and a sample question paper. These are the sources that the test writers used to create the exam. The test's design is constrained at two crucial points: when test items are written and when they are moderated. Test objectives and writing requirements are never explicitly stated during the procedure. Fulcher (2010) asserts, however, that "creating test content is not generally the beginning point of test design" (93), and that doing so without a distinct test objective would lead to "design confusion," which would ultimately lead to "validity chaos" (96). Similarly, because the moderated things are not tested, the stages after item authoring and moderation are also neglected. Similar to the phases after item composition and moderation, construct validity, content validity, and criterion-related validity of the test have not been established and the moderated items have not been trialed or evaluated to determine test reliability.

As I stated, the test does not accurately reflect the curriculum because it excludes assessment of listening and speaking skills. The lack of specific objectives is accompanied by the dominance of cautious test specification methodologies. According to test rules and a sample NEB paper, there is no provision for evaluating speaking and listening abilities, and paper setters find it difficult to deviate from the prescribed format. This lack of test preparation was influenced by worries about student answers, and the ensuing cautious techniques made sure that items were repeated throughout educational boards and years. These worries were exacerbated by students' social pressure to get excellent marks.

Compliant methods for using relevant guidelines and example papers also helped to keep harmful practices in place. Test writers just gave in to what they saw as demands that they

continue to create tests in the same manner. The same applied to the moderators, who perceived their task as being technical in nature due to comparable demands.

These problems were made worse by worries about corruption and pedagogical distrust of instructors. Authorities seem to believe that the inclusion of listening and speaking assessments will give instructors more room to financially exploit their kids.

It's crucial to keep in mind that this might be a political ploy to withhold the necessary funding and knowledge for evaluating these talents. The decision-makers who created the exam were affected by the test's limitations, the weak economy, and the inadequate English teaching and learning facilities in Nepal. Since the exam does not test what it should in regard to the purpose of the curriculum, it may be argued that the test lacks validity. The test's resultant non-representativeness has issues for fairness and validity. The lack of a distinct test concept might be considered as indicative of Nepal's educational practices. Assessment appears to be seen as a 'ritual' that is assumed to be true and does not require professional or theoretical reason. Given how testing has been conducted in the nation for decades, what the exam assesses appears obvious and doesn't require any explanation. The continuation of present test design approaches as well as the test quality have been significantly impacted by these conservative, compliant, and context-dependent practices. Additionally, the high stakes aspect of such exams is equally acknowledged as it is in other emerging environments (Davies 2016; Jilani 2009; Safari 2016).

From the standpoint of educational or linguistic testing and evaluation, this current method of test design, delivery, and application is outstanding. The fundamental criteria for test writing and revision may not be met by an exam that decides the futures of millions of pupils (Downing and Haladyna 2006). The exam is inconsistent with neither the national curriculum's objective for English language proficiency nor the present English language education policy, which aims to improve students' communication skills for use in both school and the workplace. However, there are high expectations for English in Nepal in terms of building human capital and taking part in the global economy. If the development of high-stakes tests described in this article is representative of testing in other emerging contexts (see, for instance, Cheng [2008]; Ramanathan [2008]; Ross [2008]), it might be inferred that test development in certain countries is limited. more than a routine, ritualistic exercise that has nothing in common with outstanding practice and does little to encourage students to engage in language acquisition in a more productive, engaging, and responsive

manner, despite the fact that this is encouraged as being crucial to participating in a more extensive global context. This also echoes with the limiting impacts of the "ritual" of high-stakes testing in other contexts, indicating larger performative logics related to international competition to promote educational excellence (Kamens, 2013). Since the Nepalese educational system has been performing this "ritual" for so long, it has created implicit societal norms and regulations around the "what" and "how" of testing as well as other methods that may be used.

How this educational process of producing competition in education on a societal scale continues largely uncontested may be explained by testing as a ritualistic activity. Although the issue of education frequently appears in print, digital, and electronic media, requests for a critical analysis of the factors that determine a student's success or failure in school are rarely made; this stands in stark contrast to criticisms of similar examinations in Western cultures (Allen 2016; Au 2009; Hursh 2007; Klenowski and Wyatt-Smith 2012).

Contrarily, the ritualistic methods of identifying success and failure described here appear to be regarded as 'functioning' for a variety of social stakeholders. The nationwide celebration of the testing routine and test results reflects this. The nation-wide school-leaving examinations are conducted on the very same days with a festive atmosphere. Numerous parents and family members swarm the school grounds while the children carry out this "custom." After a few months, all national newspapers announce the test results, highlighting the academic success of various universities with vibrant images of the top students. Middle-class and affluent parents share candies with friends, family, neighbors, and coworkers to recognize their children's accomplishments. Due to the societal demand for scores and the glorification of testing itself, any inquiries regarding test tools and technical evaluation procedures may seem unimportant.

Chapter 7

Conclusion

As Downing and Haladyna (2006) notes, “the subject of language testing and assessment has advanced significantly in terms of theoretical and methodological resources, as well as in terms of professional and ethical norms for creating language tests and gauging language competency and learning outcomes in a variety of circumstances.” However, the testing practices described in this article may only be tangentially connected to these intellectual and professional advancements and more advanced testing systems in civilizations like Nepal. Instead, test creation and testing may take on a more ceremonial tone in such environments, guided by more conventional, compliant, and what are perceived as context-dependent conditions. Such processes are often not questioned, which suggests that stakeholders are on board with them. This is true, at least in terms of the lack of societal criticism of the ceremony.

The Nepalese educational system needs to adopt the theoretical, professional, and technological advancements in testing and assessment to explain the process of constructing competition. This is also relevant in other regional and global contexts more generally (which would include developed contexts in which regional and global testing practices have profound impacts). Critical examination is also required of discrepancies within regulatory correlation and causation, as well as inconsistencies between testing and curriculum.

Major repercussions for language education and language learners result from the test. The impact of tests on students' education increases with the level of risk involved. But seldom do policymakers discuss exams from a language standpoint. Instead, the language policies ingrained in high-stakes tests are frequently implicit rather than explicit, while being quite potent in... (Mencken, 2008).

The implications for policymakers are fairly clear; from the research of Shohamy (2007), “when language examinations take into account studies on language acquisition and language usage, they may be used as a tool to develop more accurate and actual language regulations that mediate and arbitrate between ideology and practice.” Such exams are used in schools to direct a wide variety of educational choices, including curriculum content, textbooks,

supplies, teaching techniques, teacher preparation, programming, and language of instruction, in addition to rewarding or disqualifying a specific student from future possibilities.

It's crucial to include instructors' perspectives while creating curricula. Following that, the CDC should invite teachers to take part in this discussion. The uncontested top-down mandate of policy makers must be questioned. For the educators, this will make things easier and improve their understanding of the curriculum. In addition, the goals of the CLT curriculum need to be evaluated in order to make them better fit the needs of the classroom.

Since their engagement will contribute to the dissemination of the curriculum, teachers must be included in the creation of new materials and the evaluation system. Encourage English medium instruction in the English classroom along with providing clear guidance to schools on how and when translations should be used.

In addition, there should be a provision of opportunities to learn for the students to ensure schools have all of the resources needed for example, the schools would require a classroom and some devices to implement listening and speaking activities.

Similarly, the assessment system has to be urgently updated to use the continuous evaluation approach. It's time to review the two fundamental skills (listening and speaking). Unavoidably, it will influence the way that students are taught in the classroom. Teachers are not encouraged to make students practice the communicative elements of the curriculum in the classroom since these skills are not employed in the assessment process. So, develop formative assessments with students for practice. These assessments should offer information that is helpful to teachers, for example, in identifying areas for future instruction.

Likewise, teachers in CAS are in charge of creating assessments that have a big influence, thus they need to be skilled at creating tests, have a basic understanding of educational measurement, and be cognizant of theory and practice.

Besides, the need for professional development for teachers should be continuously evaluated before using CLT. More frequently, in-service teachers should attend effective training sessions that take place in schools. The findings of the present study indicated positive progress in curriculum reforms nevertheless, further supportive measures, particularly in implementing CAS, should be taken to reach an acceptable level.

Research on how you can better practice instruction and testing at school which aligns with curriculum goals. For this, modifying the classroom setting is a must before putting the CLT curriculum into practice. The big class size must be decreased to a smaller size in order to apply the communicative elements of the curriculum and textbook, and the ideal class time should be extended to support activities in the classrooms.

Moreover, teachers and students should use multimedia to expedite language learning. For example, age and content appropriate digital materials like dramas, podcasts, songs, TV shows, movies, blogs, and among others. In addition, bilingual teachers should match the language needs of their students. They should demonstrate language proficiency to enable students listening and speaking abilities.

Most importantly, English language teachers should possess a postgraduate degree in English language teaching (ELT) or a related discipline, such as a Masters in TESOL/TEFL/TESL/ELT, CELTA, or DELTA, as well as prior experience instructing ELT. This language evaluation jargon will be recognizable to you if you have taken the Cambridge DELTA or a TESOL.

Limitation and Further Research

Through qualitative analysis and curricular interpretation based on specific data, I was only able to concentrate on the key aspects of the current study that are related to language testing in Nepal. Similar studies must be carried out in the future on a range of stakeholders, including policymakers, curriculum developers, school administrators, students, and parents, in order to examine and fully comprehend how these aspects affect language teaching, learning, and assessment processes.

References

- Al Amin, M., & Greenwood, J. (2018). The examination system in Bangladesh and its impact: on curriculum, students, teachers and society. *Language Testing in Asia*, 8(1). <https://doi.org/10.1186/s40468-018-0060-9>
- Alderson, J.C., Clapham, C. and Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Ali, M., & Walker, A. L. (2014). ‘Bogged down’ ELT in Bangladesh: Problems and policy. *English Today*, 30(2), 33–38. <https://doi.org/10.1017/s0266078414000108>
- Allen, D. (2016). Japanese Cram Schools and Entrance Exam Washback. *The Asian Journal of Applied Linguistics*, 3(1): 54–67. <https://www3.caes.hku.hk/ajal/index.php/ajal/article/view/338/412>.
- Ashadi, A., & Rice, S. (2016). High stakes testing and teacher access to professional opportunities: lessons from Indonesia. *Journal of Education Policy*, 31(6), 727–741. <https://doi.org/10.1080/02680939.2016.1193901>
- Asian Development Bank. (2017). Innovative Strategies for Accelerated Human Resources Development in South Asia: Asian Development Bank. <https://doi.org/10.22617/tcs179079>
- Au, W. (2009). *Unequal by Design: High-Stakes Testing and the Standardization of Inequality*. New York, NY and London: Routledge.
- Bachman, L. F., & Palmer, A.S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Balwanz, D. (2016). The Discursive Hold of the Matric: Is There Space for New Vision for Secondary Education in South Africa? In W. C. Smith (Ed.), *The Global Testing*

Culture: Shaping Education Policy, Perceptions, and Practice, 261–278. Oxford: Symposium Books.

Bajracharya, N. (2016, June 14). Big reforms ahead in Nepal's education. *Himalayantimes.Com*.
<https://thehimalayantimes.com/opinion/dynamism-education-system-new-policy-needed>

Bhattarai, G.R. (2006). English teaching situation in Nepal: Elaboration of the theme for panel discussion in the 40th TESOL conference. *Journal of NELTA*. 11. Kathmandu: NELTA.

Bloem, S. (2015). PISA for low- and middle-income countries. *Compare: A Journal of Comparative and International Education*, 45(3), 481–486.
<https://doi.org/10.1080/03057925.2015.1027513>

BPEP (1997) *The effect of new curriculum on the achievement of grade 4 student*. Kathmandu: HMG/Ministry of Education, BPEP.

BPEP (1998) *The effect of new curriculum on the achievement of grade 5 student*. Kathmandu: HMG/Ministry of Education/ BPEP.

Brown, G. T., & Hirschfeld, G. H. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice*, 15(1), 3–17. <https://doi.org/10.1080/09695940701876003>

Brown, H. D., & Lee, H. (2015). *Teaching by principles: An interactive approach to language pedagogy*, (4th ed.,). New York: Pearson.

Caddell, M. (2007). Education and change: A historical perspective on schooling, development and the Nepali nation-state. In K. Kumar, & J. Oosterheld (Eds.), *Education and social change in South Asia* (pp. 251-284). New Delhi: Orient Longman.

- Cheng, L. (2008). The Key to Success: English Language Testing in China. *Language Testing* 25(1), 15–37. doi:10.1177/2F0265532207083743
- Cheng, L., & Curtis. A. (2004). “Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning.” In L. Cheng, Y. Watanabe, and A. Curtis (Ed.), *Washback in Language Testing: Research Context and Methods* (pp. 3–18). London: Lawrence Erlbaum.
- Cohen, L., Manion, L. & Morrison, K. (2018). *Research methods in education*. New York: Routledge.
- Colnerud, G. (2006). Teacher ethics as a research problem: syntheses achieved and new issues. *Teachers and Teaching*, 12(3), 365–385. <https://doi.org/10.1080/13450600500467704>
- Continuous Assessment System. (2003). A study report.
- Creswell, J. W. & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. USA: SAGE Publications, Inc.
- Curriculum Development Centre (2068). *Nirantar bidhyar mulyankam karyanmatyan Pustika*, Bhaktapur.
- Curriculum Development Centre (2005). *National curriculum framework for school education*. Sanithimi, Bhaktapur: Nepal. JSS Printing Press.
- Curriculum Development Center (2007). *Nirantar bidhyarathi mulyakan karyanwayan pustika*. Sanothimi, Bhaktapur.
- Curriculum Development Center. (2009). Primary education grade 4-5. Sanothimi, Bhaktapur: Curriculum Development Centre, Nepal Retrieved from <http://nepaknol.org.np/cdc/elibrary/pages/view.php?ref=64&k=#>
- Curriculum Development Center. (2012). *Basic curriculum Class 6-8*. Sanothimi Bhaktapur: Curriculum Development Centre, Nepal Retrieved from

<http://nepaknol.org.np/cdc/elibrary/pages/view.php?ref=382&k=#>

Curriculum Development Centre (2007) *Secondary education curriculum (9 -10)*.

Bhaktapur, Nepal: Author.

Curriculum Development Center. (2020). *Secondary education curriculum class 9-10 part one*. Sanothimi, Bhaktapur: Curriculum Development centre, Nepal Retrieved from

<http://nepaknol.org.np/cdc/elibrary/pages/view.php?ref=567&k=#>

Curriculum Development Center. (2019). *Basic level (Grade 1-3)curriculum*. Sanothimi, Bhaktapur Curriculum Development Centre, Nepal Retrieved from

<http://nepaknol.org.np/cdc/elibrary/pages/view.php?ref=2457&k=#>

Curriculum Development Center, (2020). *National curriculum framework of school education-2076*. Sanothimi Bhaktapur Curriculum Development centre, Nepal

Retrieved from <http://nepaknol.org.np/cdc/elibrary/pages/view.php?ref=2451&k=#>

Creswell, J. W., & Clark. V. L. P. (2011). *Designing and Conducting Mixed Method Research*. London: Sage.

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

<https://doi.org/10.1080/09585176.2014.909323>

Davies, D. (2016). The ‘Iron-Gate’: High-Stakes Assessment at the Age 16 in Nepal and England. *Compare: A Journal of Comparative*

Denzin, N. K. & Lincoln, Y. S. (2018). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 29-71). Thousand Oaks: SAGE Publications, Inc.

DOE, (2012). *Education planning in Nepal*. Bhaktapur : Department of Education.

- Downing, S. M., & Haladyna. T. M. (2006). *Handbook of Test Development*. Mahwah, NJ: Erlbaum.
- Education in Nepal: An Anti-Colonial Critique on Problems and Prospects. *Dhaulagiri Journal of Sociology and Anthropology*, 13, 31-39.
- EDSC, (2011). National Assessment of Grade 10 Students. Kathmandu: Educational and Developmental Service Centre.
- Fulcher, G. (1999). Ethics in language testing. *TAE SIG Newsletter*, 1(1), 1-4.
- Fulcher, G. (2010). *Practical Language Testing*. London: Hodder Education.
- Ghimire, B. (2016, May 26). House panel endorses bill on education act amendment. *Kathmandu Post.Com*.
<https://kathmandupost.com/national/2016/05/26/house-panel-endorses-bill-on-education-act-amendment>
- Giraldo, F. (2018). Language Assessment Literacy: Implications for Language Teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195.
<https://doi.org/10.15446/profile.v20n1.62089>
- Giri, R. A. (2011). Languages and language politics: How invisible language politics produces visible results in Nepal. *Language Problems & Language Planning*, 35(3), 197-221.
- Gu, P. Y. (2013). The unbearable lightness of the curriculum: what drives the assessment practices of a teacher of English as a Foreign Language in a Chinese secondary school? *Assessment in Education: Principles, Policy & Practice*, 21(3), 286–305. <https://doi.org/10.1080/0969594x.2013.836076>
- Hamid, M. O. (2010). Globalisation, English for everyone and English teacher capacity: language policy discourses and realities in Bangladesh. *Current Issues in Language Planning*, 11(4), 289–310. <https://doi.org/10.1080/14664208.2011.532621>

- Hamid, M. O., & Baldauf, R. B. (2008). Will CLT bail out the bogged down ELT in Bangladesh? *English Today*, 24(3), 16–24.
<https://doi.org/10.1017/s0266078408000254>
- Hamid, M. O., Sussex, R., & Khan, A. (2009). Private Tutoring in English for Secondary School Students in Bangladesh. *TESOL Quarterly*, 43(2), 281–308.
<https://doi.org/10.1002/j.1545-7249.2009.tb00168.x>
- Hamilton, M., Maddox, B. & Addy, C. (2015). Introduction. In M. Hamilton, B. Maddox, and C. Addey (Ed.), *Literacy as Numbers: Researching the Politics and Practices of International Literacy Assessment*, xiii–xxxiv. Cambridge: Cambridge University Press.
- Hamp-Lyons, L. (1997). Ethics in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education*, vol. 7: *Language testing and assessment*, 323-333. The Netherlands: Kluwer Academic Publishers.
- Hardy, I. (2013). Education as a ‘risky business’: Theorising student and teacher learning in complex times. *British Journal of Sociology of Education*, 36(3), 375–394.
<https://doi.org/10.1080/01425692.2013.829746>
- Harlen, W. (2009). Improving assessment of learning and for learning. *Education 3-13*, 37(3), 247–257. <https://doi.org/10.1080/03004270802442334>
- Harris, L. R., & Brown. T. L. (2016). The Human and Social Experience of Assessment: Valuing the Person and Context. In G. T. L. Brown and L. R. Harris (Ed.), *Handbook of Human and Social Conditions in Assessment* (pp. 1–17). London and New York, NY: Routledge.
- Himalayan News Service. (2018, November 3). National curriculum framework revised. *The Himalayan Times.Com*.
<https://thehimalayantimes.com/kathmandu/national-curriculum-framework-revised>

- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (2010). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hursh, D. (2007). Assessing no child left behind and the rise of neoliberal education policies. *American Educational Research Journal* 44 (3): 493–518. doi:10.3102/2F0002831207306764.
- Hyland, K. (2006). *English for academic purposes. An advanced resource book*. London: Routledge.
- Islam, J. Majid, I. A. N. Shahidullah, M. & Shams, N. (2001). *Teacher's Guide for English for Today: For Classes XI-XII*. Dhaka: NCTB and British Council.
- Jilani, R. (2009). Problematizing High School Certificate Exam in Pakistan: A Washback Perspective. *The Reading Matrix* 9 (2): 175–183.
- Jones, B. D. (2007). The Unintended Outcomes of High-Stakes Testing. *Journal of Applied School Psychology*, 23(2), 65–86. https://doi.org/10.1300/j370v23n02_05
- Jones, M. G., Jones, B., & Hargrove, T. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman and Littlefield.
- Kabir, M. H. (2008). How Validity Is Ensured in Our Language Test: A Case Study. *IIUC Studies* 5: 37–52.
- Kachru, B., Kachru, Y., & Nelson, C.L. (Eds.). (2006). *The handbook of World Englishes*. Malden, MA, and Oxford, England: Blackwell.
- Khaniya, T.R. (2005). *Examination for enhance learning*. Lalitpur : Millennium Publication.

- Khan, R. (2010). English language assessment in Bangladesh: Developments and challenges. In Y. Moon, & B. Spolsky (Eds.), *Language assessment in Asia: Local, regional or global?* (pp. 121–157). South Korea: Asia TEFL.
- Kamens, D. H. (2013). Globalization and the Emergence of an Audit Culture: PISA and the Search for ‘Best Practices’ and Magic Bullets. In H. Mayer & A. Benavot (Ed.), *PISA, Power and Policy: The Emergence of Global Educational Governance*, 117–140. Oxford: Symposium Books.
- Klenowski, V., & Wyatt-Smith, C. (2012). The impact of high stakes testing: the Australian story. *Assessment in Education: Principles, Policy & Practice*, 19(1), 65–79. <https://doi.org/10.1080/0969594x.2011.592972>
- Koul, L. (2009). *Methodology of educational research*. Noida: VIKAS Publishing House PVT LTD.
- Kucuk, F., & Walters, J. (2009). How good is your test? *ELT Journal*, 63(4), 332–341. <https://doi.org/10.1093/elt/ccp001>
- Kwon, S. K., Lee, M., & Shin, D. (2015). Educational assessment in the Republic of Korea: lights and shadows of high-stake exam-based education system. *Assessment in Education: Principles, Policy & Practice*, 24(1), 60–77. <https://doi.org/10.1080/0969594x.2015.1074540>
- Langdrige, A. (2007). *Phenomenological psychology: Theory, research and method*. London: Pearson Education Limited.
- Leung, C. Y., & Andrews, S. (2012). The mediating role of textbooks in high-stakes assessment reform. *ELT Journal*, 66(3), 356–365. <https://doi.org/10.1093/elt/ccs018>
- Lincoln, Y. S., Lynham, S. A. & Guba, E. G. (2018). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. K. Denzin & Y. S. Lincoln

(Eds.), *The Sage handbook of qualitative research* (pp. 213-263). Thousand Oaks: SAGE Publications, Inc.

Lingard, B. (2011). Policy as numbers: ac/counting for educational research. *The Australian Educational Researcher*, 38(4), 355–382.
<https://doi.org/10.1007/s13384-011-0041-9>

Mackey, A. & Gass, S. M. (2012). *Research methods in second language acquisition: A practical guide*. West Sussex: Blackwell Publishing Ltd.

Maniruzzaman, M., & Hoque, M. E. (2010). How Does Washback Work on the EFL Syllabus and Curriculum? A Case Study at the HSC Level in Bangladesh. *Language in India* 10 (12): 49–88.
<http://www.languageinindia.com/dec2010/hoquewashback.pdf>.

McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.

Menken, K. (2008). *High-Stakes Tests as de facto Language Education Policies*. SpringerLink.
https://link.springer.com/referenceworkentry/10.1007/978-0-387-30424-3_189?error=cookies_not_supported&code=a5cf7aa6-3dca-4858-9f65-b88329c2a03e

Meyer, H., & Benavot, A. (2013). PISA and the Globalisation of Education Governance: Some Puzzles and Problems. In H. Mayer and A. Benavot (Ed.), *PISA, Power and Policy: The Emergence of Global Educational Governance* (pp. 9–26). Oxford: Symposium Books.

Ministry of Education. (2019). *National Education Policy*. Kathmandu: Ministry of Education, Government of Nepal.

Ministry of Education & UNESCO Office in Kathmandu. (2015). Education for all national review report (Nepal). In *Education for All 2015 national review: Nepal* (Vol. 114p). UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000232769>

- Ministry of Education. (2016). *Flash report*. Ministry of Education: Kathmandu
- Ministry of Education (2059). *CAS manual*. Kesharmaha : Kathmandu. National Education Planning Commission (2011).
- Ministry Of Education. (2016). *School sector development plan 2016-2023*. Kathmandu Government of Nepal, Ministry of Education Retrieved from <http://doe.gov.np/assets/uploads/files/3bee63bb9c50761bb8c97e2cc75b85b2.pdf>
- Ministry Of Education Science and Technology. (2019). *National education policy-2076*. Kathmandu Ministry of Education, Science and Teachnology Retrieved from http://moe.gov.np/assets/uploads/files/Education_Policy.pdf
- Ministry of Education, (2007). *School sector reform core document: Policy and strategies*. Kathmandu: Author.
- Ministry of Education, (2009). *School sector reform plan 2009 – 2015*. Kathmandu: Author.
- Ministry of Education, (1971). *National education system plan 1971-76*. Ministry of Education
- Mulmi, A. R. (2017, October 1). Why did the British not colonize Nepal? - The Record. *Recordnepal.Com*.
<https://www.recordnepal.com/why-did-the-british-not-colonize-nepal>
- National Curriculum. (2019). *English: Classes 9 & 10*. Kathmandu: Curriculum Development Center (CDC).
- NEPC, (1992). *National Education Commission Report- 1992 (In Nepali)*. National Education Commission, Kesharmahal, Kathmandu, Nepal.
- NNEPC, (1956). *Education of Nepal*. Bureau of Publications, College of Education, Kathmandu. Nepal.

- Neupane, D. (2019, May 2). Continuous assessment system: From paper to pedagogy. *Thehimalayantimes.Com*.
<https://thehimalayantimes.com/opinion/continuous-assessment-system-from-paper-to-pedagogy>
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.
- Office, E. N. R. (n.d.). *Education Review Office, Nepal*.
https://www.ero.gov.np/post/6_5f16e3298b914
- Organization of Economic Cooperation Development, (2016). *PISA for Development Brief 1*.
<http://www.oecd.org/pisa/aboutpisa/PISA-FOR-DEV-EN-1.pdf>.
- Opinion | *The myth and methods of teaching English in Nepal*. (2020b, May 24).
<https://kathmandupost.com/columns/2020/05/24/the-myth-and-methods-of-teaching-english-in-nepal>
- Shore, A., Pedulla, J., & Clarke, M. (2001). *The building blocks of state testing programs*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy.
- Perry, F. L. (2011). *Research in applied linguistics*. New York: Routledge.
- Popham, J. (2015). *Why Standardized Tests Don't Measure Educational Quality*. *Education Leadership*, 8-16.
- Poole, A. (2016). 'Complex teaching realities' and 'deep rooted cultural traditions': Barriers to the implementation and internalisation of formative assessment in China. *Cogent Education*, 3(1), 1156242. <https://doi.org/10.1080/2331186x.2016.1156242>
- Poudel, P. P., & Choi, T. H. (2021). Discourses shaping the language-in-education policy and foreign language education in Nepal: an intersectional perspective. *Current Issues in Language Planning*, 23(5), 488–506.
<https://doi.org/10.1080/14664208.2021.2013063>

- Poudel, Lekh Nath. (2016). Reviewing the practice of national assessment of student achievement in Nepal. *Nepalese Journal of Educational Assessment*, 1(1), 1-16.
- Position Statement on High-Stakes Testing. (2000).
<https://www.aera.net/About-AERA/AERA-Rules-Policies/Association-Policies/Position-Statement-on-High-Stakes-Testing>
- Ramanathan, H. (2008). Testing of English in India: A developing concept. *Language Testing*, 25(1), 111–126. <https://doi.org/10.1177/0265532207083747>
- Richards, K. (2003). *Qualitative Inquiry in TESOL*. Basingstoke: Palgrave Macmillan.
- Ritt, M. (2016). *The impact of high-stakes testing on the learning environment*. Retrieved from Sophia, the St. Catherine University repository website:
<https://tinyurl.com/38dzuype>
- Rose, P. (2015). Is a global system of international large-scale assessments necessary for tracking progress of a post-2015 learning target? *Compare: A Journal of Comparative and International Education*, 45(3), 486–490.
<https://doi.org/10.1080/03057925.2015.1027514>
- Ross, S. J. (2008). Language testing in Asia: Evolution, innovation, and policy challenges. *Language Testing*, 25(1), 5–13. <https://doi.org/10.1177/0265532207083741>
- Rahman, M. S. (2017). The advantages and disadvantages of using qualitative and quantitative approaches and methods in language “testing and assessment” research: A literature review. *Journal of Education and Learning*, 6(1), 102- 112.
- Rahman, M., Pandian, A., & Kaur, M. (2018). Factors Affecting Teachers’ Implementation of Communicative Language Teaching Curriculum in Secondary Schools in Bangladesh. *The Qualitative Report*. <https://doi.org/10.46743/2160-3715/2018.3220>

- Rahman, M. M., & Pandian, A. (2018). A Critical Investigation of English Language Teaching in Bangladesh. *English Today*, 34(3), 43–49.
<https://doi.org/10.1017/s026607841700061x>
- Roach, E. (2022, March 30). *Education in Nepal*. WENR.
<https://wenr.wes.org/2018/04/education-in-nepal>
- Sasaki, M. (2008). The 150-year history of English language assessment in Japanese education. *Language Testing*, 25(1), 63–83.
<https://doi.org/10.1177/0265532207083745>
- Shin, S. Y., & Lidster, R. (2016). Evaluating different standard-setting methods in an ESL placement testing context. *Language Testing*, 34(3), 357–381.
<https://doi.org/10.1177/0265532216646605>
- Seargeant, P., and Erling, E. J. (2011). *The Discourse of “English as a Language for International Development: Policy Assumptions and Practical Challenges.”* In H. Coleman (Ed.), *Dreams and Realities: Developing Countries and the English Language* (pp. 248–267). London: British Council.
- Shank, G. (2006). *Qualitative Research Methods: A Personal Skills Approach*. Upper Saddle River, NJ: Pearson.
- Shepard, L. A. & Dougherty, K. C. (1991, April). *Effect of high-stakes testing on instruction*. Unpublished paper presented at the American Education Research Association and the National Council on Measurement in Education, Chicago.
- Shohamy, E. (2007). *Language tests as language policy tools. Assessment in Education: Principles, Policy & Practice*, 14(1), 117–130.
<https://doi.org/10.1080/09695940701272948>

- Shohamy, E. (2016). Critical Language Testing. In L. Shohamy & S. May (Ed.), *Language Testing and Assessment: Encyclopedia of Language and Education*, (pp. 1–15). doi:10.1007/978-3-319-02326-7_26-1.
- Shrestha (2014). *Perception of Teachers towards Continuous Assessments System*. M.Ed thesis Tribhuvan University, Kirtipur Ktm Siksha Magazine (2007)
- Silverman, D. (2016). *Qualitative Research*. Los Angeles, CA: Sage.
- Smith, W. C. (2016). An Introduction to the Global Testing Culture. In W. C. Smith (Ed.), *The Global Testing Culture: Shaping Education Policy, Perceptions, and Practice* (pp. 7–23). Oxford: Symposium Books.
- Sultana, R. (2015). Reliability of the currently administered language tests in Bangladesh: a case study. *Journal of Literature, Language and Linguistics*, 15(2422–8435), 76–85. <https://www.iiste.org/Journals/index.php/JLLL/article/view/27665>
- Stiggins, R. J. (2002). Assessment Crisis: The Absence of Assessment for Learning. *Phi Delta Kappan*, 83(10), 758–765. <https://doi.org/10.1177/003172170208301010>
- THT Online. (2021, June 16). Editorial: Grading SEE students. The Himalayan Times.Com. <https://thehimalayantimes.com/opinion/editorial-grading-see-students>
- Vladi, A. C. (2015) Developments in language testing with the focus on ethics. *Journal of Education and Practice*, 6(32) 2015, 9-13.
- Wu, M. (2015). What National Testing Data Can Tell Us. In B. Lingard, G. Thompson, & S. Sellar (Ed.), *National Testing in Schools: An Australian Assessment* (pp. 39–51). New York, NY: Routledge.
- Zahavi, D. (2019). *Phenomenology: The basics*. Oxon: Routledge.

Appendix A

Participant selection email

Subject: Connecting for ELA in Nepal!!!

Dear educators,

It was my pleasure to meet you all. Per our consensus concerning the language testing system in Nepal, I am requesting that you share a little bit about your professional background following the questions listed below and provide consent to talk about your experiences of test setting and moderating for my research.

This questionnaire is part of my research study entitled "English Language Assessment in Nepal for the partial fulfillment of my Master of Arts in Teaching English to the speakers of other languages (TESOL), Dhaka, Bangladesh.

I am carrying out this research under the guidance of Mr. Mohammad Mosiur Rahman, former lecturer at the department of Brac Institute of Languages, Brac University.

Your cooperation in completing the questionnaire will be of great value to me. I assure that your response will merely be used as information for the research and would be kept confidential.

Please feel free to put your response required to the questionnaire and correspond to show your consent for using the information you provided here and during the interview.

Once I review your response, I will further reach out to you with the interview date and time.

I would also attach an interview question package in my next email for your reference and to ease our conversation a little bit.

I look forward to your positive response and active participation.

Best regards,

Rabu Ranjit

MAT at Bard College, USA
MA TESOL-BRAC University, Bangladesh
BA English, Tribhuvan University, Nepal
B.Ed English Tribhuvan University, Nepal
E-mail: rabu.ranjit@bracu.ac.bd.
Mobile No.: 9860879087

The following questions are related to your personal information. You can provide a short response to each question.

1. Are you a paper setter or moderator?
2. How many years of teaching experience do you possess?
3. How many years of paper setting experience do you possess?
4. Have you taken any professional development training or qualifications for setting papers for moderation? If yes, mention the training.

Interview Questionnaire Package

1. How do you set a question paper in general?
2. What might be the reasons for having repetitive questions?
3. How do you respond to repetition?
4. Won't there be any problems because of repetition?
5. Why does it happen?
6. Is it a compulsion to repeat questions?
7. What kind of force encourages you?
8. Do you think our test aligns without curriculum goals? To what extent our objectives are met?
9. Have you been provided with any guidelines to set paper?
10. Why do you think listening and speaking skills are excluded?

11. When our curriculum recommends teaching and testing these two skills, why is it not followed?
12. How do you select items for the test?
13. What happens when you try to include something new this time?