

**Review Article**

English Language Assessment in Nepal: Policies, Practices and Problems

Rabu Ranjit

Brac Institute of Languages, Brac University, Dhaka, Bangladesh

Email address:

raburanjit@gmail.com

To cite this article:

Rabu Ranjit. (2023). English Language Assessment in Nepal: Policies, Practices and Problems. *International Journal of English Teaching and Learning*, 1(1), 22-33. <https://doi.org/10.11648/j.ijetl.20230101.14>

Received: November 1, 2023; **Accepted:** November 17, 2023; **Published:** December 22, 2023

Abstract: The study investigates the high-stakes English test for the SEE examination, emphasizing a mismatch with the curriculum's language skills and prompting exploration into test design processes and influencing factors. This research employs a qualitative case study approach. The findings reveal a gap between regulations valuing language skills and practical implementation in assessments. The tests are based on traditional testing philosophy, with inadequate standardization in the test items and test administration. However, the bulk of these evaluations indicated factors including the teacher, the institution, and the students that had an impact on student learning. The four language skills—listening, speaking, reading, and writing—are valued by assessment regulations, but our classroom environment and evaluation system seldom put this into reality. Summative public exams, administered externally, lack feedback on teaching and primarily serve student progression. Despite recognizing the efficacy of communicative language instruction, the prevalence of non-communicative approaches in exams raises questions about students' communicative skills development. 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. The researcher concludes by advocating for the adoption of theoretical advancements in testing within the Nepalese educational system and globally, emphasizing the importance of critically examining discrepancies within regulatory correlation, causation, and inconsistencies between testing and curriculum.

Keywords: Curriculum, Language Teaching, Language Test Design, High-Stakes Testing, School Leaving Examinations, Testing Across Societies, Nepal

1. Introduction

Throughout the years, mounting concerns have surfaced regarding the design and development of English language tests in Nepalese education, particularly in relation to their synchronization with secondary school curricula. This study compellingly illuminates a spectrum of factors that wield considerable influence over the intricate landscape of test design in this context.

The principal method for improving students' functional English competency has been formally identified as communicative language instruction [21]. However, neither

public nor school-based exams use a communicative approach. The question of whether or not children develop communicative skills in this significant language exists despite the presence of testing. More importantly, many are unaware of the processes involved in developing, distributing, and verifying these English tests.

The tests are based on traditional testing philosophy, with inadequate standardization in the test items and test administration. However, the bulk of these evaluations indicated factors including the teacher, the institution, and the students that had an impact on student learning. The four language skills—listening, speaking, reading, and

writing—are valued by assessment regulations, but our classroom environment and evaluation system seldom put this into reality.

Summative public exams are frequently taken outside of a classroom environment. Each student may be evaluated publicly at many levels, including those at the national, regional, district, resource center, and/or school levels. The Himalayan Times [95] reports that 5,17,000 students participated in SEE in 2021. Typically, independent/autonomous groups, the government, or other external entities administer these exams. However, they seldom provide feedback on classroom teaching-learning or the educational system. Instead, they are used to select and progress students to the next grade or educational level and verify learning. Additionally, the outcomes of these tests decide who "achieves" and "fails" in the educational system and labor market.

The employment market, in both the public and private sectors, makes considerable use of exams. In the hopes of landing a secure job and a long career, they take recruiting examinations for a few hundred positions in the government, the financial sector, and primary and secondary schools. Many language tests are found to be of extremely poor quality, impede teaching and learning, and frequently fall short of correctly measuring the things they are meant to.

To better understand the nature of the assessment process and the variables that affect the design and administration of the English exam, the study incorporates the perspectives of significant assessment players, such as education authorities and test makers. The study's objective is to advance our understanding of testing across nations in a worldwide setting [4] Based on my research, I argue that test design may have taken the form of ancient ceremonies performed in line with unquestioned implicit principles in nations like Nepal. Professional standards and ethical guidelines related to testing [24] may appear unimportant in the context of such banal practices.

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.

1.1. 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 [8]. All assessments should be of high quality, emphasizing the importance of good design [26].

The first stage in language test design is to define the problem 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 [38]. The next stage is the creation of specifications 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.

1.2. Understanding High-Stakes Testing Across Societies

The design of tests, as highlighted by Cheng and Curtis [16], significantly shapes education globally, transforming tests from mere servants to influential leaders in pedagogy, playing a critical role in molding curricula, educational policies, and local systems, as policymakers leverage their power to regulate curricula, introduce new textbooks, and endorse innovative teaching methods, ultimately aiming to make policy judgments based on test findings.

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 [33, 49]. Test results serve as a measure of the quality of instructors' and students' work in this new institutional framework [54, 96]. 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 [92].

The international 'horse race' [49] 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 [58]. PISA statistics are used to assess educational quality and the degree of 'human capital' in participating nations. Bloem [12] 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 [39]. In order to enhance accountability in its education system, Australia implemented the National Assessment Program – Literacy and Numeracy (NAPLAN) in 2008 [50, 96]. 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" [3, 30, 52, 72].

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 [23] 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 [42] reported similar high-stakes examinations in Pakistan. This study indicates 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 [6], Australia [50], China [30], Korea [52], Japan [3], Nepal [23]; [9]. Because high test scores are regarded as quality indicators, instructors and schools are under pressure to enhance test data [50] potentially raising validity difficulties in test design [36]. Furthermore, Harris and Brown [36] noted, "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" [88]. 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" [38]. 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" [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 the Nepal instance.

1.3. Testing in Nepalese Education

In Nepal, mainstream education is organized into three stages: basic, secondary, and tertiary. Basic education (Grades 1–8), secondary education (Grades 9–10), and upper secondary education (Grades 11–12) are the three levels of compulsory education. Students must pass a nationwide public examination conducted each year at the end of each level. The last two public education 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 or compete in the job market either home or overseas. 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.

1.4. Curriculum and Assessment Policies for English

In line with the spirit of the new National Education Act [21], the secondary English curriculum adopted learner-centered approaches to develop students' communicative competence. 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. The main objectives of the curriculum can be understood from the following extract:

Therefore, the curriculum has focused not only on language and language functions, but also on a variety of fiction and non-fiction texts that provide a meaningful context for language learning. For some students, secondary education serves as a basis for preparation for a university education, whereas for some other students, it may be a preparation for entry into the world of work. This curriculum tries to address the linguistic requirements of both types of students.

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 [85]. 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 [2, 31]. 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 SEE examination, each education board invites selected teachers from higher secondary colleges to write test papers. Kabir claims that the selection process of test writers and moderators is not fair. Teachers who maintain good relations with the education board officials are often given the chance to write tests [45].

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. 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.

1.5. Teaching Artefacts

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 [53]. However, evidence from other contexts suggests that such a role for textbooks cannot be generalized. For example, Rahman and Pandian [76] and Rahman, Pandian, and Kaur [75] highlighted contextual complexities including the lack of skilled teachers and technological support needed for implementing CLT tasks presented in the textbook. These contextual challenges may have implications for the assessment of the English subject.

2. The Study

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. 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 [65]. 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. 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?

3. Research Methodology

Yin asserts that the study adopts a qualitative case study approach [97]. The qualitative data collection occurred in two phases: the initial stage involved analyzing curricular policy documents and test papers, revealing insights into the test's nature, structure, and assessed competencies. In the subsequent step, interviews with test developers were conducted to elucidate the rationale behind the exam's development.

3.1. Participants

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 [22]. 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. Test setters and moderator profile.

Participants	Teaching experience (years)	Paper-setting experience (years)	Professional Qualification/Training
S1	15	4	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.

3.2. 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 [21], former and current exam papers; and (b) through semi-structured interview. The interview guide was created using Hughes's test development framework, which was previously mentioned [38]. 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.

3.3. Data Collection

To gather interview data the researcher first went to numerous government schools and colleges in Kathmandu, Nepal's commercial city, and invited paper setters and moderators to take part in the study [91]. 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 via email. 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.

3.4. Data Analysis

The qualitative data methodologies and analysis recommended by Richards were used in both the document and interview data analysis [78]. 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.

In the second phase, 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 [86]. 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 "corruption and context"). These themes helped create the "image" refers to in the data [78].

4. Findings: Proficiency, Conformity, Conservatism, Context

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.

4.1. 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.

4.2. 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.

Table 2 also indicates 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
Education 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:

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 is that certain students—especially the poorer candidates—tend to spend a lot of time memorizing such answers (to previous test questions) [55].

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 [38]. M2 was of the opinion that repetition was not a concern, particularly if items were repeated across boards. However, 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 [55]. This has a wider resonance with comparable activities in connection to more generalized standard testing procedures in other cultural contexts [6, 30, 42, 39, 50].

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

4.3. *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. Activities that are given have little relevance for communication in the actual world. The majority of the participants agreed that the test did not reflect curricular aims.

According to M2, 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 M2 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. 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" [51]. 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. Bachman states, "language testing happens in a social and educational framework, and the applications of language exams are generally dictated by political requirements" [8].

4.4. *Contextual Practice*

It is sometimes argued that speaking and listening skills are not assessed for "practical reasons", which include issues with resources, logistics, and competence for testing more than a million students throughout the nation every year [41]. 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 [32].

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.

5. Discussion

The English test I looked at 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.

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.

Contrary to common belief, Fulcher argues that "generally, creating test content is not the starting point of test design" [26]. He contends that embarking on this process without a clear test objective could result in "design confusion," [93] ultimately leading to "validity chaos [96]." Furthermore, the subsequent stages following item authoring and moderation, much like construct validity, content validity, and criterion-related validity, are often overlooked. The moderated items remain untried and unevaluated for 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 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 conventional, compliant, and context-dependent practices. Additionally, the high stakes aspect of such exams is equally acknowledged as it is in other emerging environments [23, 42].

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 [24]. 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, [16]; Ramanathan, [77]; Ross, [81]), it might be inferred that test development in certain countries is limited. more than a routine, ritualized 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 [49]. 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 [3, 6, 39, 50]. 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.

6. Conclusions

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 [24]. 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 conservative, 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.

In our sharply divided society, the absence of social critiques does not always imply that testing and testing outcomes have been fair and reasonable, especially for student groups that are more marginalized (as is the case in Western contexts, such as the United States [6]). As with many other difficulties in education, the concerns with testing have simply the circumstances in which they happen have not been sufficiently and in-depth investigated. Further study into the nature of test production and moderation techniques is very necessary for a deeper understanding of testing procedures inside and between nations as well as the societal consequences of this type of educational judgment.

The researcher's conclusions suggest that it's critical to understand the impact of maintaining ritualistic activities. 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.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] Alderson, J. C., Clapham, C. and Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press. <http://library.uc.edu.kh/userfiles/pdf/10.Language%20Test%20construction%20and%20evaluation.pdf>
- [2] 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>
- [3] Allen, D. (2016). Japanese Cram Schools and Entrance Exam Washback. *The Asian Journal of Applied Linguistics*, 3 (1): 54–67. <https://caes.hku.hk/ajal/index.php/ajal/article/view/338/412>
- [4] 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>
- [5] Asian Development Bank. (2017). Innovative Strategies for Accelerated Human Resources Development in South Asia: Asian Development Bank. <https://doi.org/10.22617/tcs179079>
- [6] Au, W. (2009). *Unequal by Design: High-Stakes Testing and the Standardization of Inequality*. New York, NY and London: Routledge.
- [7] Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- [8] Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- [9] 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* (pp. 261–278). Oxford: Symposium Books. https://www.researchgate.net/publication/287646307_The_Global_Testing_Culture_Shaping_Education_Policy_Perceptions_and_Practice
- [10] Bajracharya, N. (2016, June 14). Big reforms ahead in Nepal's education. *Himalayantimes.Com*. <https://thehimalayantimes.com/opinion/dynamism-education-system-new-policy-needed>
- [11] 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. <https://www.nepjol.info/index.php/NELTA>
- [12] 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>
- [13] 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>

- [14] Brown, H. D., & Lee, H. (2015). *Teaching by principles: An interactive approach to language pedagogy*, (4th ed.). New York: Pearson.
- [15] 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.
- [16] Cheng, L. (2008). The Key to Success: English Language Testing in China. *Language Testing* 25 (1), 15–37. <https://doi.org/10.1177/0265532207083743>
- [17] 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.
- [18] 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=#>
- [19] 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=#>
- [20] 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=#>
- [21] 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=#>
- [22] Creswell, J. W., & Clark, V. L. P. (2011). *Designing and Conducting Mixed Method Research*. London: Sage.
- [23] Davies, D. (2015). The ‘iron gate’: high-stakes assessment at age 16 in Nepal and England. *Compare: A Journal of Comparative and International Education*, 46 (4), 582–602. <https://doi.org/10.1080/03057925.2015.1030591>
- [24] Downing, S. M., & Haladyna, T. M. (2006). *Handbook of Test Development*. Mahwah, NJ: Erlbaum.
- [25] Education Review Office, (2020). National Assessment of Student Assessment. Sanothimi, Bhaktapur. https://www.ero.gov.np/post/6_60410dcd2cc3
- [26] Fulcher, G. (2010). *Practical Language Testing*. London: Hodder Education.
- [27] 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>
- [28] 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>
- [29] Giri, R. A. (2011). Languages and language politics. *Language Problems and Language Planning*, 35 (3), 197–221. <https://doi.org/10.1075/lplp.35.3.01gir>
- [30] 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>
- [31] 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>
- [32] 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>
- [33] 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.
- [34] 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>
- [35] Harlen, W. (2009). Improving assessment of learning and for learning. *Education* 3-13, 37 (3), 247–257. <https://doi.org/10.1080/03004270802442334>
- [36] 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.
- [37] Himalayan News Service. (2018, November 3). National curriculum framework revised. *The Himalayan Times.Com*. <https://thehimalayantimes.com/kathmandu/national-curriculum-framework-revised>
- [38] Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- [39] Hursh, D. (2007). Assessing No Child Left Behind and the Rise of Neoliberal Education Policies. *American Educational Research Journal*, 44 (3), 493–518. <https://doi.org/10.3102/0002831207306764>
- [40] Hyland, K. (2006). English for academic purposes. An advanced resource book. London: Routledge.
- [41] 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.
- [42] Jilani, R. (2009). Problematising High School Certificate Exam in Pakistan: A Washback Perspective. *The Reading Matrix* 9 (2): 175–183. http://www.readingmatrix.com/articles/sept_2009/jilani.pdf
- [43] 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
- [44] Jones, M. G., Jones, B., & Hargrove, T. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman and Littlefield.

- [45] Kabir, M. H. (2008). How Validity Is Ensured in Our Language Test: A Case Study. *IIUC Studies* 5: 37–52.
- [46] Kachru, B., Kachru, Y., & Nelson, C. L. (Eds.). (2006). *The handbook of World Englishes*. Malden, MA, and Oxford, England: Blackwell.
- [47] Khaniya, T. R. (2005). *Examination for enhance learning*. Lalitpur: Millennium Publication.
- [48] 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.
- [49] 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.
- [50] 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>
- [51] Kucuk, F., & Walters, J. (2009). How good is your test? *ELT Journal*, 63 (4), 332–341. <https://doi.org/10.1093/elt/ccp001>
- [52] 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>
- [53] 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>
- [54] 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>
- [55] 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>
- [56] McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.
- [57] 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=a5cf7a-a6-3dca-4858-9f65-b88329c2a03e
- [58] 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.
- [59] Ministry of Education, (2019). *National Education Policy (in Nepali)*. Kathmandu: Ministry of Education, Government of Nepal. https://moest.gov.np/post/1_62b7036a58f86
- [60] 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>
- [61] Ministry of Education. (2016). *Flash report*. Ministry of Education: Kathmandu https://moest.gov.np/post/1_632a9b6716aa5
- [62] 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>
- [63] Ministry of Education, (1971). *National education system plan 1971-76*. Ministry of Education. <https://www.martinchautari.org.np/storage/files/thenationaleduationssystemplanfor-1971-english.pdf>
- [64] 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>
- [65] Curriculum Development Center (CDC), (2019). *National Curriculum English: Classes 9 & 10*. Curriculum Development Center. Ministry of Education. Bhaktapur. <https://bagisworischool.edu.np/wp-content/uploads/2020/09/2.-ENGLISH.pdf>
- [66] National Education Commission, 1992. *Report of the national education commission, 1992 (executive summary)*. National Education Commission, Kesharmahal, Kathmandu, Nepal. https://nepalindata.com/media/resources/items/20/bReport_of_the_National_Education_Commission_1992.pdf
- [67] NNEPC, (1956). *Education of Nepal*. Bureau of Publications, College of Education, Kathmandu. Nepal. https://himalaya.socanth.cam.ac.uk/collections/rarebooks/downloads/Education_in_Nepal.pdf
- [68] 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>
- [69] Office, E. N. R. (n.d.). *Education Review Office, Nepal*. https://www.ero.gov.np/post/6_5f16e3298b914
- [70] Organization of Economic Cooperation Development, (2016). *PISA for Development Brief 1*. <https://www.oecd.org/pisa/pisa-for-development/pisafordevelopment-documentation-briefs.htm>
- [71] 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. <https://www.bc.edu/research/nbettep/statements/V2N4.pdf>
- [72] 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>
- [73] 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>
- [74] Poudel, Lekh Nath. (2016). Reviewing the practice of national assessment of student achievement in Nepal. *Nepalese Journal of Educational Assessment*, 1 (1), 1-16. [https://www.ero.gov.np/upload_file/files/post/1595313482_598772199_1587622596_1795486007_Journal_NJEA_2\(1\)_2017\(1\).pdf](https://www.ero.gov.np/upload_file/files/post/1595313482_598772199_1587622596_1795486007_Journal_NJEA_2(1)_2017(1).pdf)

- [75] 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>
- [76] 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>
- [77] Ramanathan, H. (2008). Testing of English in India: A developing concept. *Language Testing*, 25 (1), 111–126. <https://doi.org/10.1177/0265532207083747>
- [78] Richards, K. (2003). *Qualitative Inquiry in TESOL*. Basingstoke: Palgrave Macmillan.
- [79] 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>
- [80] 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>
- [81] 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>
- [82] Roach, E. (2022, March 30). *Education in Nepal*. WENR. <https://wenr.wes.org/2018/04/education-in-nepal>
- [83] 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>
- [84] 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>
- [85] 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.
- [86] Shank, G. (2006). *Qualitative Research Methods: A Personal Skills Approach*. Upper Saddle River, NJ: Pearson.
- [87] 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. <https://www.colorado.edu/education/sites/default/files/attached-files/Effects%20of%20High-Stakes%20Testing.pdf>
- [88] 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>
- [89] Shohamy, E. (2016). Critical Language Testing. *Language Testing and Assessment*, 1–15. https://doi.org/10.1007/978-3-319-02326-7_26-1
- [90] Shrestha (2014). *Perception of Teachers towards Continuous Assessments System*. M.Ed thesis Tribhuvan University, Kirtipur. <https://elibrary.tucl.edu.np/bitstream/123456789/1948/3/12540.pdf>
- [91] Silverman, D. (2016). *Qualitative Research*. Los Angeles, CA: Sage.
- [92] 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.
- [93] 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>
- [94] 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>
- [95] THT Online. (2021, June 16). Editorial: Grading SEE students. The Himalayan Times.Com. <https://thehimalayantimes.com/opinion/editorial-grading-see-students>
- [96] 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.
- [97] Yin, R. (2014). *Case Study Research: Design and Methods*. 4th ed. Los Angeles, CA: Sage.

Biography

Rabu Ranjit (RR) is an English language educator in Nepal. Her research interests are the interdisciplinary issues in language education and applied linguistics.