The Adaptation of Language Testing Models to National Testing of School Graduates in Nepal: Processes, Problems and Emerging Issues

Ram Ashish Giri

Thesis submitted in fulfilment of the requirements for
Doctor of Philosophy

School of Education
Faculty of Arts, Education and Human Development
Victoria University of Technology
2005

The Adaptation of Language Testing Models to National Testing of School Graduates in Nepal: Processes, Problems and Emerging Issues

Ram Ashish Giri

Thesis submitted in fulfilment of the requirements for
Doctor of Philosophy

School of Education
Faculty of Arts, Education and Human Development
Victoria University of Technology
2005

ABSTRACT

The school education system in Nepal is chronically under-resourced, severely lacking in infrastructure and school teachers are often untrained and demoralised. The isolation of remote areas in the mountainous terrain, the effects of a ten-year civil war, and a culture impregnated with respect for traditional hierarchical, rule- and memory-based modes of instruction exacerbate this situation.

Nevertheless, English language is taught throughout Nepal in both private and public schools through a centralised system of education with a centrally prepared curriculum. In the mid 1990s, a new English curriculum, based on the 'communicative approach' was implemented. However, the pedagogic practices as well as the examination systems are still steeped in a traditional focus on knowledge of language, rather than usage. In addition, memorisation of information and rote learning are encouraged. The 'old' English test format of the School Leaving Certificate (SLC) Examination is still in use despite extensive critiques as to its validity, reliability and theoretical adequacy. Clearly, there is a negative 'wash-back' effect from the old test that negates the potential of the new curriculum to make a difference to the way English is taught.

This research is a study of the current SLC English test and how it may be adapted to better suit the testing and educational needs of secondary school education in Nepal. Further, it makes a contribution to knowledge about how ESL/EFL language tests may be adapted and applied in order to meet the system-wide needs for EFL testing in developing countries more generally. After analysing theories of language learning and current models of language testing, I developed a new English test package for the SLC. Developmental, operational and implementation issues of the test was addressed. The new test package consists of a test specification, a test structure, a pilot test, and a marking interpretation scheme. This test takes into account the context of (EFL) curriculum renewal, the specifics of the educational and cultural contexts of Nepal, the multiple purposes for testing, the range of available language testing models and current debates about approaches to and limitations of language testing.

The project was implemented in terms of action research in two cycles. In the first cycle, the new test was developed and in the second, fieldwork on the new test package was conducted in Nepal. Feedback was obtained from stakeholders, such as the ELT academics, teacher educators, testing personnel, test developers, examiners, teachers and students through a number of activities such as a plenary presentation, workshops, interviews and two pilots of the test at a private school and a public school.

The fieldwork produced inconclusive results. The outcome of the pilot test was satisfactory and gave consistent results, despite the fact that the low number of students who undertook the pilot test and variable conditions, meant that no statistically significant conclusions could be drawn. Other participants in the consultations felt that while the new test package was a vast improvement on the previous one, it could not be implemented in the near future because of a range of factors militating against it.

This thesis makes a contribution on three levels: a product, a process and a theoretical insight.

1. **The Product** The new English test for the SLC, which I have developed, is basically sound and appropriate for the conditions and educational needs of Nepal school education at this time. While the test may not be implemented immediately, it provides a solid basis for further development and eventually change to the current English testing system that could be implemented incrementally, in small steps.

2. **The Process:** In the course of developing the new test, I have modelled a process for developing similar tests adaptation of current language theories and testing models to local conditions and needs) in other developing countries.

3. **The Theoretical Insights:** The action research approach I have taken in carrying out my project, the subsequent field research and writing of this thesis, has helped me to conceptualise and manage it in doable stages, and to put my own learning at the centre of the development of each stage. Action research has thus enabled me to articulate some important insights regarding the broader undertaking of reforming and improving education in the context of Nepal. At base, the potential for implementation of the new test, and the positive wash-back effects this might otherwise provide will not occur in the near future because of powerful factors relating to the macro context in Nepal. Given the

harsh and unpromising macro context (political, military, geographic cultural and economic), it may be that research and development of educational innovations (such as my own 'micro' innovation) may only be effective if these are done in the course of addressing the macro situation through policy activism. In other words, the demand for implementing the new test needs to come from the field as well as from experts such as myself and be made in the context of campaigns to reform the SLC in its entirety and increase funding for school education generally.

I have modified my initial proposition about the potential for change to the system as a whole that a new testing package might bring. My awareness of the enduring influence of bureaucratic, social, economic and cultural factors, in holding back and preventing educational reform, has been heightened. I have realised (again) that educational reform is a slow process and that it sometimes may be delayed indefinitely in the sorts of conditions that Nepal is currently experiencing. Much advocacy and policy work needs to be done amongst teachers, educational bureaucrats and politicians, significant new funds from donors need to be found and the stress of civil war must be eased, before there can be significant progress in the English test of the SLC or any other aspects of education requiring reform.

Table of Contents

LIST OF FIGURES

Acronyms Used

ACER	-	Australian Council for Educational Research
ACTFL	-	American Council on the Testing Foreign Languages
AEC	-	Australian Education Council
ALIGU	-	American Language Institute of Georgetown University
ANEC	-	All-round National Education Commission
ART	-	Applied Reading test
ASLPR	-	Australian Second Language Proficiency Ratings
CDC	-	Curriculum Development Centre
CEE	-	Common Entrance Examination
CELT	-	Compulsory English language Testing
CERID	-	Centre for Educational Research, Innovation and Development
CLA	-	Communicative Language Assessment
CLT	-	Communicative language testing
CRT	-	Criterion-referenced Testing
CTEVT	-	Centre for Technical and Vocational Training
DEET	-	Department of Education, Employment and Training
DELTA	-	Diagnostic English Language Test
DEO	-	District Education Office
DFID	–	Department for International Development
EAP	-	English For Academic Purposes
EFL	-	English as a Foreign Language

EIN	-	Emailed interview
ELT	-	English Language Teaching
ESA	-	English Skills Assessment
ESL	-	English as a Second Language
ETS	-	Educational Testing Service
FN	-	Field Notes
FSI	-	Foreign Service Institute
GCSE	-	General Certificate of Secondary Education
GP	-	Gorphapatra, a Vernacular National Daily
HLNEC	-	High Level National Education Commission
HLNEC	-	High Level National Education Commission
HMG/N	-	His Majesty's Government of Nepal
IDP	-	International Development Program
IELTS	-	International English Language Testing System
ILR	-	Interagency Language Roundtable
ILR	-	Interagency Language Roundtable
IN	-	Interview
INGO	-	International Non-governmental Organisation
IPRS	-	International Postgraduate Research Scholarship
ISLPR	-	International Second Language Proficiency Rating
LAB	-	Language Assessment Battery
LTRC	-	Language Testing Research Centre
MC	-	Multiple Choice
MELAB	-	Michigan English Language Assessment Battery

MOE	-	Ministry of Education
NCED	-	National Centre for Educational Development
NCELTR	-	National Centre for English Language Teaching and Research
NDS	-	National Development Service
NEC	-	National Education Committee
NEC	-	National Education Commission
NELTA	-	Nepal English Language Teachers' Association
NEPC	-	National Education Planning Commission
NESB	-	Non-English Speaking Background
NESP	-	New Education System Plan
NLLIA	-	national Language and Literacy Institute of Australia
NRT	-	Norm-referenced Testing
NT	-	Nepali Times, a Weekly Magazine
NTA	-	Nepal Teachers' Association
OCE	-	Office of the Controller of Examinations
OPI	-	Oral Proficiency Interviews
OSS	-	Oral Structural Situational (Approach)
PABSON	-	Private and Boarding Schools' Organisation of Nepal
PAF	-	Proficiency Assessment Framework
PCL	-	Proficiency Certificate Level
RC	-	Resource Centre
RHEC	-	Royal Higher Education Commission
SAARC	-	South Asian Association for Regional Cooperation
SAQ	-	Short Answer Question

SEASSI	-	Southeast Asian Summer Studies Institute
SEDIPS	-	Society for Educational Development in Private Schools
SLC	-	School Leaving Certificate (Examination)
SLEP	-	Secondary Level English Proficiency
SPEAK	-	Speaking proficiency in English Assessment Kit
TKP	-	The Kathmandu Post, a National Daily
TL	-	Target Language
TOEFL	-	Test of English as a Foreign Language
TOURCH	-	Tests of Reading Comprehension
TRN	-	The Rising Nepal, a National Daily
TSE	-	Test of Spoken English
TWE	-	Test of Written English
UCH	-	Unitary Competence Hypothesis
UCLES	-	University of Cambridge Language Examinations Syndicate
UNESCO	-	United Nations Education and ... Children Organisation
WN	-	Written Notes (supplied by research participants)

Declaration

I, Ram Ashish Giri, declare that the PhD thesis entitled ***The Adaptation of Language Testing Models to National Testing of School Graduates in Nepal: Processes, Problems and Emerging Issues*** is no more than 100,000 words in length, exclusive of tables, figures, appendices, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my original work.

Ram Ashish Giri   **August 2005**

ACKNOWLEDGEMENTS

The Adaptation of Language Testing Models to National Testing of School Graduates in Nepal: Processes, Problems and Emerging Issues

Preamble

This is the land

Where deities abide

And fairies dance and ply

The brooks sing

And mountains kiss the sky

This is the land

Of deep valleys and peaked hills

Which the world envies

And a tourist thrills

The land of all lands

The land of Nepal

(Translated from Anonymous)

Nepal: A Land of Contrasts

Nepal, the 'Shangri-La' or the abode of gods and goddesses in the Himalayas, is famous the world over for its bio-diversity, unique geography and socio-ethnographic multiplicity. Also known as the land of *Buddha*, the enlightened one, and the land of *Gurkhas*, the brave ones, Nepal, as it can be conjectured from the poem on the previous page, is blessed with natural resources, ethnic tolerance, artistic skills and a simple attitude to living.

Divided in to three geographic and five administrative regions it has a very diverse linguistic situation with 30 different languages spoken by around 24 million *Nepalis*. *Nepali* is the national language and the *lingua franca*. English is taught as a foreign language and is a compulsory subject throughout the school and tertiary systems. However, the development of Nepal's economic and political fronts has been held back by the traditional outlook of the caste system and a feudal political system. After the fall of autocratic *Rana* regime in 1951, its development was hampered by poorly conceived, unscientifically formulated and badly implemented five-year national plans. Presently, it appears to suffer from four of Pye's six crises – the penetration crisis (where the state cannot change the prevailing values of the people), the participation crisis (where people's participation is extremely low), the integration crisis (where developmental efforts largely remain disintegrated) and distribution crisis (where resources and expertise are unevenly distributed) (Pye 1972).* Analysing the present situation of under-development and instability in her doctoral thesis, Kerr (1999) observes:

> As I left the airport precinct, I was aware that my [Western misconceptions of] 'mountains in the sky' were barely discernible through the haze of pollution. Kathmandu city was visibly over-populated and under-resourced. Roads were poorly maintained, sanitation and water were scarce (Kerr 1999:1).

Nepal is a land of incredible contrasts. It is savagely poor, and despite the recent political crisis and the fact that much of its land is barren, mountainous, rugged and dry,

---

* The two other crises are the identity crisis, which if not resolved leads to ethnic conflict, and the legitimacy crisis, which if unresolved may result in coup d'etat (Pye 1972).

it remains romantically beautiful (Gregson 2003) and one of the top ten must see destinations of the world (TKP, 20 Jan. 2004).

This study takes on educational contrasts of Nepal and describes how its aspirations to modernise education is barred by the lack of political, educational and economic commitments. 4

This study digs beneath the surface of that beauty, into the less beautiful realities of poverty and in particular, the poor state of education. It reflects my own striving towards social enlightenment and the betterment of school education in Nepal.

INTRODUCTION

Origin and Context

My Adventure in the Field of Language Testing

One of the purposes of education in general is to develop a social awareness towards family, community and the world in which learners live. The role of assessment practice in the education process should be to motivate and to stimulate further student learning. Different types of assessment practices, then, lead to different types of motivation, and therefore, different types of learning (Athansou 1997). This has precisely been my experience both as a student and an English language teaching (ELT) practitioner.

Over three decades ago, when I was preparing for my School Leaving Certificate (SLC) examination, we heard surprising news of a change in the testing scheme. We were no longer to be tested on the originality, creativity and communicability of our understanding of the subject matter in question, rather we were to be tested on how well we had learned the subjects on the basis of objective tests which consisted of multiple-choice items, fill-ins, matching and true-false items and a few short answer type questions. The SLC had changed that year following the introduction of New Education System Plan (NESP 1971).

The change in the examination scheme also brought a change in our approach to text and learning materials. Instead of developing an understanding or what is known as 'deep learning' (McKellar 2002:4), we went for 'surface learning' (op. cit.) breaking the text into discrete facts, content knowledge and basic skills. The new test did not make any sense to us, as we did not understand what it was that they were trying to measure. Many students failed that year as they were unfamiliar with the new testing system and because they were not given enough training or orientation for it. It was only later I came to know that they had changed the previous system of examination to a more objective system for the purpose of making the test more 'scientific' and 'reliable' (Khaniya 1990:83). This was indeed my first adventure in the field of educational testing and there were many more to come in the years ahead.

As I left *Sukhipur* (a small village in the eastern district of *Siraha* in the Southern plain) for the bustling and a populous city of *Birgunj* in central Nepal for my tertiary education, my perspective on secondary schools changed completely. The schools here were better equipped, had several concrete teaching blocks, a library, and other facilities we could never dream of in a rural school. So my understanding that all have an equal opportunity and life is fair began to crumble. Fortunately, I had done well in the SLC and that helped me win a place in the *Proficiency Certificate Level* (PCL) course in the Institute (now Faculty) of Education of *Tribhuvan* University.

The then new system of testing was also introduced to all tertiary and university examinations. Nearly fifty percent of the course was tested on the objective test format whereas the remaining 50 percent was on short answer type questions. In this system of examination, education was viewed as an end product, which, as the body of knowledge, could be broken down to segmentable chunks and measured separately. For those working in both government as well as university sectors, assessment was usually associated with written examinations or tests. The focus of English language assessment was on the products of learning – a limited range of relatively easy to measure competencies such as facts, content knowledge and basic skills which were tested through formal, externally imposed summative examinations such as the SLC and annual examinations.

During my Master's degree, I served as a volunteer instructor for a year in *Doti*, a remote mountainous district in the Far Western region. This was under the *National Development Service* (NDS) programme – organised by the *Tribhuvan* University college. During my volunteer year, I visited a number of primary and secondary schools in three districts of the region, most of which I found in a very bad condition with teachers mostly under-qualified and untrained. The college I was assigned to was somewhat better in that it had concrete blocks. However, educational materials, or any other teaching and learning materials for that matter were virtually non-existent. The college instructors relied on whatever reference materials they possessed. The unavailability of teaching, learning or reference materials limited the instructors' teaching ability, severely affecting the quality of education.

As for the assessment, the instructors did not have to worry much about it as yearly examinations were prepared and conducted by the Examination Section (now

Office of the Controller of Examinations) (OCE) of *Tribhuvan* University in Kathmandu who would send a superintendent to every examination centre with tests and other relevant materials. The instructors, however, did have to conduct internal assessment valued at 20 percent – a practice that was abolished some years later due to rampant irregularities. This was my first experience of teaching in Nepal.

My Career in Teaching

My career in teaching began at the same time my Master of Education degree commenced at *Tribhuvan* University. As a part-time high school teacher, I taught the SLC courses in Years 9 and 10, the latter being the final year of the secondary education in Nepal. My options as an English as a foreign language (EFL) teacher and my activities (in the classes I taught with 60 plus students in each) were limited. This was because my main task as a teacher of the SLC course was supposedly to prepare students for the national examination by providing possible questions that might appear in the examination and their answers. The students memorised these and reproduced them in the examination. There was neither time nor resources available for developing target language abilities in the students. Developing language abilities and skills required by the curriculum was ignored in order to obtain the all important pass mark through regurgitation of the learned answers.

The situation at a regional campus of *Tribhuvan* University where I worked as a lecturer in the initial years of my academic career was not much better than that of the college in the Western Nepal. The lack of educational materials was severe. The faculty staff depended on the notes they had prepared when they were students themselves. Like the teachers of the SLC, they mainly provided notes or materials to the students which were viewed to be useful for the purpose of the annual examinations, usually set by the Examination Section at the central headquarter of *Tribhuvan* University in Kathmandu. My main achievement during my six year tenure at the regional campus was the establishment of the first ever ELT resource centre in Nepal, the model of which was later replicated for *Nepal English Language Teachers' Association* (NELTA). NELTA, a professional organisation of English language teachers, has established a network of ELT resource centres in ten of its branches. Today, NELTA

has become a forum for the *Nepali*, as well as expatriate ELT practitioners in Nepal for discussion, deliberation and dissemination of matters of professional interest. Now it advocates for the amelioration of the ELT practices in Nepal through its services to governmental and nongovernmental organisations,

The purpose of the resource centre, which was established with contributions from the international agencies working in Nepal, was to improve the quality of ELT by providing teaching and learning materials and teacher training workshops. My involvement in the Centre paved the way for a *Fulbright* scholarship for my second Master's Degree in the United States.

The education in the US provided the initial training in the field of language testing. The courses and projects I did with the Afro-American children enhanced my interest in this area and the training I received became the basic foundation for my present research.

During my 10-year service at the *University* Campus of *Tribhuvan* University, I served on a number of committees, including the *Compulsory English Subject Committee* and the *Examination Moderation Committee* at the Faculty of Education, *Tribhuvan* University. In addition I also served as a member of the *English Subject Committee* of *the Curriculum Development Centre, Ministry of Education.* In the second half of the 1990s, I also worked for *PLAN International*, an international non-governmental organisation (INGO) working for community development in Nepal, where I was an education consultant in the two districts of the Southern plain. The conditions of the schools and teachers in these districts were mostly similar to that of the schools and teachers of the Western mountain districts I had visited two decades earlier.

By working with these organisations and committees, I found that there was a great interest for a change amongst education workers, particularly the ELT practitioners. However, their interest for change in the practice of English language teaching and testing was meeting strong resistance. The system prescribes a testing framework, which is rigid and restrictive and does not allow examiners to add new test types. Resistance to change had also been due to the fact that no new language assessment framework or design had so far been researched and no alternative ways of testing recommended for the ELT context and purposes of Nepal.

As I reflect on my educational and professional journey, I find that my story is not different from that of numerous other ELT practitioners who want the system to change and may be looking for a solution. However, my experiences show that *Nepali* education, after a number of experimentations, has come to a critical turning point where it must abandon its century old and largely inappropriate educational culture and adapt to a more modern and pragmatic system of education. In order to do so, Nepal must address adequately all elements of education including the examination system. Examination is an important means which drives all attempts of reform. It can either act as a retardant or as a promoter of change.

Emergence of the Research

The present research emerged out of my wondering about what role, if at all, examination could play in reforming an education system which is so tradition-bound. I therefore took up the case of English language testing, particularly the English test of the SLC as the subject of my present research. I set out with the aim of researching why and how the English test of the SLC must change in order to be compatible with the new ELT curriculum and to exert positive washback effects in ELT practices in Nepal.

This was also an action research project. In order to develop or adapt a language testing model, which I hope would eventually become a basis for the development of new tests that are appropriate in form and function for the *Nepali* context, I needed to undertake certain actions (to develop the test), then research and reflect on those actions in a cycle of development. I had hoped that this would contribute to the amelioration of the entire ELT process, including English language teacher training programmes.

By documenting, reflecting upon and analysing the pathway of development, I have produced a revised and updated test. At the same time, I have modelled a process of adaptation and development of a test for language in under-resourced and developing countries, using testing models and theories that are current in Western countries.

The Context of the Research

In the mid 1990s, His Majesty's Government of Nepal introduced a revision of the school curricula. The revision completed its cycle from primary level to secondary level in 2000. The first batch of students on the new curriculum graduated from schools in 2001.

The SLC results of the first English test on the new curriculum were 'near-disaster' (Spotlight 2001). An analysis of the new SLC English test revealed that, despite the change in the secondary English curriculum, the SLC has continued to follow the previous test format (CDC 2000; CERID 2001). In this research project, I have proposed and researched revision of the current practices of the SLC English test. Drawing from the empirical and theoretical materials, apart from my own first hand data, I have developed an argument that changing the nature of the English test could exert positive washback effects and potentially change the entire English language teaching and testing practices in Nepal.

Organisation of the Thesis

Writing this research report has been a challenging task. I have employed what is sometimes known as 'multi-methods' (or mixed methods) (Darlington and Scott 2000:167-169) consisting of qualitative analysis, action research, and quantitative evaluation of a pilot test that at times did not sit easily together.

Throughout the thesis, I engage with the question of reforming the English test of the SLC, the most important examination that happens in the life of every individual in Nepal, with a view to reforming the ELT practices. While I address this issue particularly in the ELT context of Nepal, the issues and arguments I present in this thesis may equally be applicable to a wider context of education and to other developing contexts. That is to say, I address the ESL/EFL testing issues in the context of the national examination of Nepal but I also draw lessons from other developing regions of the world. By the same token, I believe that third world countries can draw useful insights from my thesis regarding adaptation and implementation of a large scale standardised test.

The thesis, which is an account of the research journey, has been completed in two cycles of an action research project.

Thesis Structure

Cycle One   **Cycle Two**

Chapter One The Context		Chapter Seven Planning & Organising Fieldwork
Chapter Two Research Methods	Chapter Six The New Test Package	Chapter Eight Analysing and Reporting field data
Chapter Three The Problem		Chapter Nine Emerging Issues
Chapter Four Analysis of Language Testing Theories		Chapter 10 Reflection
Chapter Five ESL/EFL Tests		

Figure 1: Structure of the Thesis

Cycle one of the project, as I show in the figure above, has been reported in chapters 1, 2, 3, 4, 5 and 6. It completes a cycle of diagnosing, problematising, acting (developing a new test), observing and reflecting on the first cycle. Chapter 6 is also the planning phase of cycle two. Cycle two consists of chapters 7, 8 and 9. Chapter 10, the final chapter of the thesis, is a summary and reflection of the project and sets a context for a possible third cycle.

Cycle One

In this cycle, I addressed the following research questions:

1. What are the contextual parameters to be taken into account and the steps to be undertaken in the process of designing a test suitable for a new curriculum in a particular context, based on current theories and models of ELT testing?

2. What aspects of language testing practice, particularly ESL/EFL testing practices can be drawn upon in modelling a language testing adaptation process?

3. What issues and dilemmas emerge in relation to applying educational technologies that have been developed in the context of developed nations, to a developing context like Nepal?

Chapter 1 is the introduction to the thesis. It provides an overview of the research and the context in which the research took place. The chapter examines the changing context of education in Nepal and outlines the factors that need to be considered and the issues that need to be addressed when developing a package to reform the examination.

Chapter two outlines the reasons why action research has been an appropriate research method for the present research. Given the complexity of research tasks, I needed a methodological approach, which should allow me to investigate the prevailing conditions at the local level, work with the stakeholders in the field, and be theoretically consistent for adapting a design to develop a new English test. The chapter describes the key principles followed during the research, and it demonstrates how action research informs the theory of the field in question.

Chapter three is the statement of the problem. It explains why the current English test of the SLC is dysfunctional. It also indicates the areas to be addressed when developing a revised testing package.

Chapters four and five are the planning or adaptation of a new test design. The chapters outline the processes, aspects and theoretical underpinnings that should be

12

taken into account in the test adaptation process. They also point out what compromises are to be made in developing a process for designing a new test given the local constraints and the shift taking place in English language teaching and testing internationally.

Chapter Six is about how the new test package was actually created and how language-testing principles were addressed during the development of the package.

Cycle Two

In this cycle the main research question I addressed was:

How useful, practicable, and appropriate is the new test package?

Chapter seven of the thesis, therefore, is about my plan to collect feedback on the revised testing package. It describes how I identified people and organisations to participate in my research, how I selected them, and what activities were planned to obtain feedback. This chapter is also the planning phase of the second cycle which is basically the fieldwork and analysis of the data/feedback received. Chapter seven also tells the story from the field. It describes what activities were described to obtain feedback on the revised testing package.

Chapter eight gives an analysis of the responses obtained in Nepal. It describes how data coming from various sources and subjects were organised, how they were classified and what approach was used to analyse the data. It presents the findings of the consultation phase, the framework used to organise the diversity of views into generalised findings and finally the summary of the findings and the main issues raised during the fieldwork.

In chapter nine, I discuss some of the issues that emerged from the fieldwork and analysis of the feedback and the pilot test.

In chapter ten, the final chapter of the thesis, I reflect on my entire research journey, research activities, my observations and of course my learning in the course of the research. I explain what issues need to be resolved and why, for introducing such a change in the national examination of Nepal. In the final section of the chapter I outline

the areas to be addressed in the next cycle of research and the work to be considered for further research. Two other questions that I addressed in the final reflection on both cycles of the action research were:

1. What issues and dilemmas emerge in relation to applying language testing practices that have been developed in the context of developed nations to a developing context like Nepal?

2. Based on the experience of this project and similar projects documented in the literature, in what ways, if at all, might language tests serve as a means for educational change in the context of educational development and renewal?

CHAPTER 1

An Introduction to Education in Nepal

In the varied topography of professional practice, there is a high, hard ground where practitioners can make effective use of research-based theory and technique, and there is a swampy lowland where situations are confusing 'messes' incapable of technical solution.

- Schőn 1983 quoted in Darlington and Scott 2002:1

1.1 Introduction

Nepal is undergoing a period of rapid transition in the fields of socio-politics and education. With the dawn of the new millennium came fresh thinking and renewed commitment amongst *Nepali* education workers to ameliorate the existing education system, which they find inappropriate and incongruous with the present time. However, this transition is not working according to the expectations of the average people (Poudel 2004; Joshi 2001; Kerr 1999).

The expectations of the *Nepali* community of education have risen over the years. Schools are now expected to prepare young people with the knowledge, understanding, skills and values they need for productive and rewarding lives in an increasingly complex and competitive world. Advances in information and communication technologies, population diversity and complex environmental and social challenges are all changing the social, cultural and economic contexts in which schools are operating (Bajracharya, Thapa and Chitrakar 1998). The dominance of communication technology has increased a great deal and it is likely to keep growing in the years to come.

Nepal is geographically a remote country, yet it can no longer remain isolated from the changes taking place elsewhere. Nepal must be involved in innovation in education and modernisation of the educational practices to meet the changing needs of

individuals and growing demands of the community at large. As a consequence, there is a need for change in education in Nepal.

So far as its effort to develop education is concerned, Nepal has made huge progress in terms of numerical expansion in the second half of the 20th century, especially in the last three decades. However, the state of affairs is far from being satisfactory. While education still suffers from a lack of necessary infrastructure and resources causing a severe impact on its quality, only a small fraction of those who enrol in secondary education graduate with a *School Leaving Certificate* (SLC) diploma. About 68 per cent of those who sit for the SLC fail the expectations of the stakeholders causing a great wastage of national resources (Figure 2).

School Graduates
(128,000 or 32 % of secondary enrolment)

| | 32 % |

Secondary Level Enrolment (449296 or 40% of Lower

Secondary enrolment)

| 40% of lower sec enrolment |

Lower Secondary Level Enrolment (1058448 or 20 %
Of primary enrolment)

| 20% of Primary enrolment |

Primary Level Enrolment (5076658 or 63 %
of all school age children)

| 63% of all school age children |

School Age Children

| **8 Million** 100% |

Figure 2: The pyramid of success in the school education (based on the data from MOE 2003; PABSON 2003).

From Figure 2 it can be seen that only 128,000 students (2.5 percent of the children enrolled in the primary level) graduate from secondary school. This means that about 97.5 per cent of children either drop out or do not get through the school system. This also means that most of the human resource remains unproductive causing a huge waste of personal as well as national resources. Furthermore, the quality of these SLC graduates may be in doubt as they have failed to meet the stakeholders' expectations and

demands of the higher education system (Davies, Glendinning, and McLean 1984; Verma and Pande 1989; Joshi 1996). As such, the whole apparatus of school education, including the examination system, appears to be failing.

In the following section, I introduce the Nepali education system from a historical as well as contemporary perspective, and point out the problems that the system has faced over the years.

In the final section of the chapter, I move from general education to English language education to outline the current ELT scenario and present some of the issues and problems associated with the SLC English course that have direct bearing or implications for the practice of assessment and its revision.

1.2 The Education of the State

"... a formal education system has not been a part of either the central or political tradition, and prior to 1951, the intention was to withhold education from most Nepalese."

Vir (1988:9)

The present education system of Nepal is a reflection of the structure of the society, which is deeply founded in Karma, a Hindu philosophy of division of labour and passive acceptance. The philosophy perpetuates fatalism and the status quo through division of the community into tribes and castes, and is the basis for maintaining social and cultural division in the people, exerting restrictive effects on educational practices (Bista 1991; Shrestha 1993; Pigg 1993; Vir 1998; Macfarlane 1994; Kerr 1999). Bista (1991) and Macfarlane (1994), for example, suggest that the social structure based on the division of caste puts one caste on the top of others, creating a rift between planners and white-collar workers, and grass root level workers, which in turn affects the development process adversely. Pigg (1993) and Kerr (1999), on the other hand, argue that *Karma* is intrinsic to the Hindu religion which the so-called 'elite' uses as a strategy to strengthen their position and maintain the status quo.

Formal education, as the epigraph suggests, mainly functioned as a form of acculturation, and as a privilege restricted to the people of the upper strata of the society, until the early 1950s. After the overthrow of the autocratic *Rana* regime in the

early 1950s, Nepal embarked on the development of planned education. Since then, Nepal has developed and endeavoured to implement several successive five-year plans. In fact, nine five-year plans have been launched in the last five decades. Despite this, no significant change has taken place so far as the modernisation of education is concerned. As a consequence, education in Nepal remains more or less unchanged.

Numerous commissions on education have been instituted since then, namely, the *Education Committee* (1952), the *National Education Planning Commission (NEPC)* (1954), *All-round National Education Commission* (ANEC) (1962, in which the NESP was envisaged and was launched in 1971), the *Royal Higher Education Commission (RHEC)* (1982), *National Education Commission (NEC)* (1992), the *High Level National Education Commission (HLNEC)* (1997). These commissions have claimed to have made recommendations appropriate in the prevailing social, political and education context (Khaniya 1990; Kerr 1999). The NEC (1992) and HLNEC (1997), for example, recommended that in the changing democratic climate the goals of education should be revised creating more scientific, pragmatic and objective curricula. The curricula should (a) make education useful and meaningful; (b) make systematic provision for improved teaching and learning; and (c) be geared to the knowledge and skills needed for the citizens of the future to be functional and effective in fast changing educational, employment and professional settings. However, their recommendations have produced little impact on the educational development so far. This is so partly for the fact that education fell into the hands of people who were largely inexperienced, unskilled and under-qualified (Kerr 1999), and partly that the planners who mainly represented the elite kept it that way. The old system perpetuated their social status and authority (Macfarlance 1994). Poor governance and mismanagement by the people in authority aggravated the educational conditions.

In the next section, I present an analysis of the periodic efforts and plans made in the formal education system, which have mainly been implemented with a short-term vision producing short-term results. In the final paragraphs of this section, I reiterate the main issues and trends that emerge out of the periodic plans. I have divided the efforts and plans into four stages in terms of their goals and impacts on formal education. My main aim, as will become clear later, is to present the scenario of why educational efforts in

Nepal have failed to produce desirable results, and what implications this will have on any future reform efforts.

1.2.1 The Commencement of Planned Education

I have chosen to call the period of 1950-1970 the stage of initiation of planned education in Nepal; a period in which a planned move to expand the access to education was made. However, despite the government's intention to formalise education by creating a single system of education consisting of adequately funded vernacular schools, a dual system of public and private schools emerged under the influence of the elite clientele created by the western style English medium schools of the *Rana* regime to which the *Nepali* planners, mostly members of the elite circle, gave continuity to maintain the power structure (Vir 1988; Kerr 1999).

The private schools, usually English medium schools mainly attended by children from Anglophile families, had better facilities, better qualified teaching staff, and therefore, created better results. The public schools, initially known as the 'community schools' were managed by the local communities that were themselves, in the main, under-educated and under-resourced. Thus, a two-stream education system emerged creating a division in the school graduates, which since has become a major issue of social and economic inequity (Bajracharya et al 1998).

1.2.2 The Stage of Expansion

This stage is also known as the New Education System Plan (NESP) years (1971-1980). The NESP was regarded as a radical and revolutionary plan in its approach to planning and development (Bajracharya et al 1998). Through a centralised top-down process, the plan targeted equal access to education, linkage between education and production, production of a qualified and trained teaching force, and development of improved curriculum and curricular materials. It was an ambitious plan based on the Western model of development but with insufficient resources, an unskilled work force, a poor sustainability strategy, and above all poorly planned implementation procedures. As a result, the NESP caused resentment in the already

19

divided society, confusion in the work force at the grass roots level, and unresolved complications (NEC 1992; Kerr 1999). The education system during this stage expanded rapidly, but its decentralisation pace, local level work force, community capability and implementation mechanism were unable to keep up with the rate of expansion. Furthermore, there was a lack of communication and co-ordination within the government machinery, and a lack of cooperation from the community (Bajracharya et al 1998).

The existing private schools were nationalised. However, the poor quality of education in the state schools encouraged the people of the upper strata, who preferred educating their children in neighbouring India or elsewhere, to re-institute the dual system of public and private schools.

So far as ELT is concerned, the weightage of English in the NESP curriculum was drastically reduced from 25 per cent to 12.5 per cent. The new curriculum was based on the structural approach to ELT. It required teachers to be trained and to develop and implement instructional schemes to encourage communication, language practice in pairs and groups, manipulation, demonstration and role-plays. But conversely the government had to remove the requirement for teachers to be trained in order to meet the increasing demand of teachers to keep up with the huge influx in the school enrolment. This move placed the quality of language education at risk and as a consequence of this, ELT in the later years in Nepal became increasingly ritualistic and backward looking.

Nationalising schools brought several resource problems. When public education was declared to be free, education in the schools suffered from a quality setback. The schools were prevented from charging fees of any kind. More than 95 per cent of the government's provision for school education was spent on teachers' salary (Bajracharya et al 1998). Because of the lack of adequate funding the schools could not meet their recurrent expenses, nor could they organise improvement programmes. This resulted in a virtual absence of teaching learning materials, inadequate physical facilities, poor maintenance of physical facilities, and a scarcity of trained teachers. Furthermore, Nepal, like any other developing country, suffered from a crisis of distribution, and failed to supply books and materials throughout the country, which

contributed further to poor quality education and a high drop out rate. There was, thus, an inconsistency in planned targets and their outcomes.

1.2.3 The Stage of Privatisation

Now, I shall turn to the third stage of planned education in Nepal throughout the 1980s, 1990s and beyond, which I term 'the stage of privatisation' of school education. The inability of the government schools to meet the expectations of the community, and the government's exasperation at the deteriorating quality of the state schools, encouraged the development of an alternative system of education. Private schools mushroomed, because of a very lenient government policy which gave letters of 'no objection' to charging fees.

The growth of the private education system has never been acknowledged by the government (Kerr 1999). There is no government policy on the development of private education or on its place in the *Nepali* education system. Two decades later on, the government has attempted to regulate private education (TKP, 10 Feb. 2004). However, private education is yet to receive recognition, a situation which, I believe, has been a major obstacle to the development of and international investment in private education in Nepal. In July 2004 (GP, 12 Jul. 2004), the organisation of *Private And Boarding School of Nepal* (PABSON) demanded a separate act for private schools to end the state of confusion that exists in private education due to the lack of government policy.

The government seems to have maintained a silence regarding private education. This is perhaps to hide its inefficiency and to evade its responsibility to provide quality education for all. Private schools, however, seem to have capitalised on the situation and grown unplanned and unsupervised. The fact that private schools have a high success rate in the SLC, and that most of them teach children through the medium of English, a must for higher education, has also contributed to their popularity.

1.2.4 The Stage of Rejuvenation

Finally, the period of the1990s, which I call the 'stage of rejuvenation', is the period of a changing scenario in *Nepali* education. The changing political climate provided encouragement and new thinking for innovation of education in the state

schools. There has been a change in the educational as well as administrative structure of education in order to bring about efficiency. A change has also been introduced with a view to meeting the changing needs of the students

However, as I describe in the following section, all has not gone well in the process of the revision of education in Nepal.

To sum up, formal school education in Nepal has come a long way since its emergence in the late 1940s. The initial curriculum was nationalistic in its approach, reflecting the then mood of the people and keeping in line with the religious and social values to appease the ruling elite. A more liberal approach to education and curriculum development was adopted after the institution of parliamentary democracy as a system of government. Unfortunately the popular democratic system of government did not last long. The then king replaced the parliamentary system of government by a more autocratic one party *Panchayat* system of government, which gave absolute power to the King. Soon it was realised that the liberal approach to education was a threat to the autocratic rule. Consequently, the school curriculum was revised. The planners educated in the 'British style' of schools preferred the 'British model' of school education with a goal of producing citizenry loyal to Nepal's religion, country and the King. Any materials that offered discussions of alternatives to the government system, monarchy, and even religion were removed from the prescribed texts (Vir 1988; Seddon 1994; Kerr 1999).

The school curriculum took a gigantic leap forward with the introduction of New Education System Plan (NESP) in 1971. The curriculum was more accommodative of matters relating to different religions, culture, life styles and changes in the educational technology. It is said that the NESP curricula were the first 'scientifically' developed curricula in Nepal (Joshi 1996). However, its unstated objective remained to shore up the sagging popularity of the *Panchayat* system of government (NEC 1992). It turned the *Nepali* educational sector into a political arena, a situation that exists to this day (NEC 1992).

With the restoration of multi-party parliamentary democracy in 1990, another attempt was made to give education a catalytic role for social change and national reconstruction. However, while the social mores like caste and fatalism continue to

create inequality, innovations of the 1990s are likely to continue facing difficulties because of the uncoordinated, and unstable political situation.

The commissions' reports reveal that the periodic changes in the educational policy reflect contemporary issues and concerns. Some of these issues and concerns have changed in degree and type over time; some persistently remain as they were, some others have been resolved, while some new ones have emerged (Bajracharya et al 1998). In examining the commissions' reports, I find that there are three trends of changes in the development of education emerging. They are:

i.   changes imposed or led by political circumstances,

ii.   changes taking place as a consequence of policy intervention, and

iii. changes taking place due to changes in the international education arena.

On the ELT front, the learning and teaching of the English language in Nepal suffered a serious quality setback when the weightage of English in relation to other subjects was reduced by half in the 1970s, the requirement of training of teachers was removed, and the infrastructure could not keep up with the huge increase in secondary enrolment.

The government has since renewed its commitment to ameliorate education in the 21st century by introducing necessary changes in educational goals and curricula and innovations in the pedagogic processes. However, serious problems remain: the mismatch between the educational provisions and economic development needs, the mismatch between educational goals and the delivery process, a lack of commitment towards ELT, a lack of community participation, administrative inefficiency, an untrained workforce, a lack of infrastructure, and above all, a lack of sufficient funding.

In the next section, I give a more detailed account and look at these and some other problems that have contributed to the present state of education in Nepal. These issues and problems had to be taken into account in developing a revised English test.

1.3 The Current State of the Education

1.3.1 The Structure of Education

The structure of education (divided into the ladder of five stages[*]) is comparable to that of many South Asian countries. The pre-primary level has been added to the structure recently. Similarly, the level of higher secondary is yet to fully establish itself in the school education structure as it was introduced only a few years ago. Higher secondary education is one of the two options available for school graduates, the other one being the *Proficiency Certificate Level* (PCL) of the university education structure, which is to be phased out in a few years' time.

However, given the mountainous geography and uneven topography of Nepal, the structure creates a problem. The dropout rate from primary to lower secondary level of education, for example, is very high because the children who complete primary education in remote schools have no access to a lower secondary school, as such schools are relatively smaller in number or are generally attached to secondary schools usually situated in urban areas.

1.3.2 Lack of Funding and the Problem of Sustainability

Nepal has a slow economic growth (of less than 2.2 percent) (Gorpkapatra {GP} 2004). With limited internal revenue, Nepal is facing great difficulty in providing adequate resources to education. Lack of funding has contributed to low educational quality and administrative inefficiency, resulting in poor motivation and low morale of the teaching and administrative force, and a poor record of sustainability of innovative programmes.

The proportionate allocation of the national budget to education in Nepal has remained significantly low. While other South Asian Association for Regional Co-operation (SAARC) countries allocate 25 percent or more of their budget to education,

---

[*] pre-primary level (Nursery to KG classes), primary level (1-5 classes), lower secondary level (6-8 classes), secondary level (9 and 10 classes), and higher secondary level (11 and 12 classes).

24

Nepal's investment is significantly low, 13 per cent (UNESCO 2002) only 21 per cent of which is invested into secondary and higher secondary education. The allocation of budget, as the table below shows, remained around 10 percent in the eighties and around 13 percent in the 1990's:

Year	1985-90	91/92	92-97	1997-98	98-2001
% of budget	10.01	12.03	13.23	13.08	13.69

Figure 3: Financing education. (Source: MOE/OCE 2000)

Though it might appear from Figure 3 (above) that there is a significant increase in the budget allocation to education in 1990s, Nepal is far from meeting the demand for educational resources, as the total budgetary allocation to education is significantly low. Also there is no rationale for distribution of the allocated budget within the sectors of education. Secondary education combined with higher secondary education, for example, receives roughly around 21 per cent (compared to up to 40 per cent in some SAARC countries) of the total education budget. The rest goes to primary and higher education (UNESCO 2002).

It appears that the insufficient allocation of the national budget to education is an indication of the government's failure to view educational development as a pre-condition for the development of other sectors of the society, making educational quality and administrative efficiency virtually impossible (TRN, Jul. 2001c). Despite the government's attempt to increase the number of trained teachers, the majority of teachers are still untrained as enough training programmes cannot be launched due to lack of funding. The co-curricular and extra curricular activities in public schools are planned but are usually not implemented. The government has a target of 90% of enrolment at the primary level but the actual current enrolment is only at 62.5 per cent. The public schools cannot employ additional teachers to run enough classes so that the physical facilities of many urban schools remain under utilised. There have been insufficient female teachers recruited in primary schools to boost female enrolment. At the same time, the provision and infrastructure cannot keep up with huge increases in

enrolment. Clearly a lack of sufficient resources is a serious setback to effective implementation and sustainability of innovative programmes.

The SLC examination is an example of such a quality setback. Because of a lack of funds, the SLC Board and the *Office of the Controller of Examinations* (OCE) have not been able to run extensive training programme for the examiners and other testing personnel. This means it cannot conduct oral assessment on its own, though such a programme is viewed as necessary (OCE 2000). Each time they plan to do something new, they have to look to donors for financial support. The new SLC examination, for example, was piloted with funding from the Department for International Development (DFID), a British government donor agency which helped in piloting and implementing the oral test of the English component of the SLC. But the oral test as one of the national trainers/examiners told me during interviews, was almost certain to be stopped when the funding ceases in the end of 2003. Thus, the sustainability of educational programmes, most of which are project funded, is linked to the continuity of the funding from a variety of donors.

This means that many of Nepal's problems of educational underdevelopment are resource related. Any effort to ameliorate the current state of education, therefore, must face severe resource constraints.

1.3.3 Educational Policy and Lack of Commitment for ELT

As I mentioned in the beginning of the chapter, efforts are being made to link education to the development process with a focus on poverty alleviation and illiteracy eradication. The educational policy, now, emphasises modernising education by developing required infrastructure and expertise (CDC 1999). Secondary education is considered an important link between primary and higher secondary education. It aims at providing a foundation for prospective careers and at producing lower level human power. The secondary curriculum, for example, develops the three broad areas of foundation for higher education, pre-vocational knowledge, language, communication and academic (study) skills.

There is no official position for ELT and its role in the development of education and its modernisation process, except that it is termed as a foreign language.

However, English is viewed as an indispensable part of education at all levels (Jha 1992). This could be an elitists' approach to perpetuate their control over the *Nepali* education system. Nevertheless, having no official position is actually hampering the development of ELT. Kerr (1999) notes that no national or educational plans of Nepal make a definitive or exclusive mention about the status of ELT and the role it plays in the modernisation of *Nepali* society. No national commissions in their reports have ever commented about the need for English or the state of ELT in the country. The most recent commission, HLNEC (1997) provided only a sketchy description of ELT and failed to provide any clear directions for the future.

The lack of commitment to ELT at the policy level has created confusion on the part of the government and ELT practitioners as to what directions ELT should take in the new century, and what proportion of resources should be allocated to it. As a consequence, despite the fact that ELT is expensive, it has not been able to produce the desired results.

1.3.4 The Public vs. Private Education and the Dilemma it Creates

The so-called 'public' schools, also known as '*sarkari*' (government) schools are sponsored and managed by government agencies such as the *District Education Offices* (DEOs), and *Resource Centres* (RCs). The government exercises its control over the schools through a highly centralised system of administration and through the funding it provides to such schools. The Ministry of Education (MOE) deals with all matters relating to policy, curriculum design, textbook development, teacher selection and teacher training, implementation, supervision and monitoring of educational programmes. There are about five million children in government schools taught by about 250,000 teachers (MOE 2003). The teaching performance is far from satisfactory. Less than 20 percent of public school students successfully take the SLC examination. It is this performance of the public schools that has indirectly stimulated the growth of private schools.

The private schools, on the other hand, appear to have performed well so far as their results in the SLC are concerned. There are about 8000 private schools, which accommodate about 35 percent of the whole school population. Despite the fact that

most of the teachers are untrained (Khadka 2001), they have successfully mobilised their resources to generate good results in the SLC. However, the private schools have often been infamous as profit mongers, charging exorbitant fees to make a profit (Khadka 2001) and not delivering what they promise, i.e., quality education (Spotlight 2001). These schools are expensive and affordable to only the well-off families, yet they are popular because they occupy more than 80 percent of all SLC results (NT, 16 May 2001). Thus, such schools perpetuate class division between those who can afford education in private schools and successfully take the SLC, and of those who go to the government schools and have a high risk of failing the examination.

1.3.5 Mismatch between Secondary Education Provision and Development Needs.

The implicit goal of secondary education is to produce human resources required for the economic development of the country, especially at the grassroots level (CDC 1999). However, educational plans or programmes are not developed on the basis of any analysis of developmental or labour market needs. As a result, there is a mismatch between what the educational goals are and what secondary education actually delivers.

Schools do offer some vocational subjects as optional subjects. These range from horticulture through business accounting to computer education. However, the lack of adequate equipment defeats the very objectives of the courses aimed at developing technical/practical skills in students. As mentioned earlier, schools are unable to run innovative programmes because of a lack of adequate funding and equipment. Average SLC graduates, therefore, are not adequately trained to take up employment in their respective areas.

The *Centre for Technical Education and Vocational Training* (CTEVT) runs degree and training courses in a wide range of technical areas. However, the minimum requirement for entry into the CTEVT programmes is the SLC. This puts an additional stress on the SLC, making it the crux of all employment opportunities. And with a success rate of only approximately 2.5 percent (only about 128,000 out of 50,00,000 students who enrol at the primary level) (Figure 2), the SLC is unable to meet the developmental needs for human resources (MOE 2003).

1.3.6 Lack of Community Participation

Educational policy and planning in Nepal is a highly centralised and top-down process. All planning, implementation strategies and decisions are made at the centre by a small number of overseas educated and trained personnel who usually do not have any working knowledge of the field (Kerr 1999). The ordinary community members such as parents, teachers, school administrators, field supervisors are rarely involved in the process. As a result, communities, as reported in Koseli (Koseli, 18 Sep. 2001), do not have a sense of responsibility towards education.

The bureaucratic approach of the government toward handling education and the schools creates apathy in the people and an uncongenial sentiment that education is the responsibility of the state. The schools' responsibility is to get their wards through the SLC with the best possible marks no matter what means are employed. The lack of community participation makes it difficult for the government to meet the cost of education, and the same time, bureaucrats cannot conduct the SLC in a free, fair and peaceful manner.

1.3.7 Administrative Inefficiency and Diminishing Educational Quality

The MOE is responsible for initiating educational plans, programmes, and activities. Its role is to ensure that the educational planning, management, and service delivery systems across the country function well. Various functional units of the MOE and its constituent and autonomous bodies, working within the framework of the ministry, play an instrumental role in carrying out its functions (MOE 1996).

The CDC, as the apex body of MOE, plans, develops and monitors the curricula and textbooks, teaching-learning materials and support materials. It consists of various subject committees headed by the *Director-General* and (a) subject specialists, (b) Directors and staff, (c) members from universities, (d) practising schoolteachers, and (e) representatives from donor/assisting agencies.

The functions of the CDC include (i) developing and revising school curricula and textbooks, (ii) organising seminars/workshops for the writers on curricular improvement and textbook dissemination, (iii) conducting studies and surveys on issues and problems related to the curricular aspect of school education, (iv) preparing subject-wise handbooks for teachers, and textbooks for students, and (v) providing the scripts for textbook printing (MOE 1996).

Secondary school enrolment over the past 15 years has quadrupled, and the administrative needs have also increased considerably. However, school administration in Nepal has not undergone adequate changes to match the management needs of the schools. The school administrative system works as a replica of the total structure of the government bureaucracy, staffed with administrators and bureaucrats rather than education experts.

Verma and Pande (1989), Bajracharya et al (1998), and a report in TKP (TKP, 15 Mar. 2002a) claim that the bureaucratic structure of education is the reason for unsatisfactory educational outcomes. Departments like the Planning Division, CDC, NCED and OCE seriously lack trained personnel.

The centralised and top-down structure of education has brought a number of problems. Firstly, because of the lack of a proper communication and co-ordination mechanism, the structure has become inefficient; secondly, the under-qualification of the MOE staff makes change in education hard to bring about; thirdly, because of the lack of training, staff are unable to design, plan and sustain programmes without outside help. Finally, as a result of all this, any educational programme suffers from quality setbacks.

1.3.8   Under-qualified and Untrained Teaching-force

Educational change cannot be successful unless and until the change is in the hands of qualified, trained and skilled teachers who translate the change into practice. Enabling teachers should essentially be viewed as a precondition for effective and successful implementation of the change.

As I stated already, educational change in Nepal is a top-down process, which affects the education practitioners and their situations considerably. Given the fact that

teaching is the least preferred profession and that teacher preparation programmes in Nepal have not been very effective, the teaching profession has not proved attractive. Competent people do not take up teaching as their career as long as there are other options available. Teachers are not paid well and usually have to work with minimum facilities. Poor salary drives them into taking a second or even third teaching job. Most teachers provide private tuition for money. As a consequence, teachers, as Kerr (2000) suggests, do not spend time analysing learners' needs or designing support materials, nor do they make time to design and create practical or constructive activities or tasks for their classes.

1.4 The English Language Teaching (ELT) Context

1.4.1 The ELT Situation

Nepal has never been colonised by a western country, yet the influence of the English-speaking world is reflected in the emphasis placed on the acquisition of English and the existing structure of the education system. English is compulsory throughout school education and is viewed as a precondition for higher education, and therefore, a 'must learn' subject for educational and occupational purposes (Davies, Glendinning and McLean 1984), national and international communication (Khaniya 1990; Awasthi 1995) and for the modernisation process (Kerr 1999). English remains an inseparable part of Nepal's academic pursuits and the demand for it is growing in spite of Nepal's inability to meet it (Jha 1992; Awasthi 1995). The purpose of teaching English in Nepal is seen, as Jha suggests, as an effort to enable the learners to communicate with speakers of English and at the same time, to acquire knowledge, ideas, skills and techniques imparted formally and informally through English (Jha 1992:26). Davies et al (1984:4), commenting on Nepal's attachment to English suggest, "it has other than instrumental values, symbolic and sentimental ones". As such, English is compulsory throughout the school education, at the tertiary level and in the first two years of undergraduate education. English in public schools now starts in standard 1 (earlier it used to be introduced in standard 4 only), but in private schools, it is introduced as early as the nursery class.

31

ELT is a huge and growing profession in Nepal and is viewed as instrumental in the process of reform and modernisation. However, it seems to fall short of meeting the needs generated by the country's rapid modernisation process, particularly in the fields of economy, science and technology, tourism, and the need for increasing contact with the outside world.

1.4.2 Importance of English in Secondary Education

English has a very important place in secondary education. It is one of the six compulsory subjects and carries 12.5 per cent of the total weightage. The importance of English in the Nepali context is set out in the following terms:

> English is a major international language, one of the six languages of the United Nations, and the [most important] means of international communication in South Asia. It is a vital tool for any student to become successful in local, national and international communication. It is the only foreign language taught in all schools of Nepal and is the medium of higher education (CDC 1999:42).

There are two main goals of English language teaching at the secondary level, the development of communicative skills to enable students to communicate with people who speak English, and the development of academic skills for higher learning. English is thus regarded as the language of communication as well as 'reference', knowledge of which is essential to pursue education in any discipline at a higher level.

English was given more weightage (25 per cent) in the past than now (CDC 1999). However, there has always been an unlimited amount of unstated emphasis attached to it. A considerable amount of time and (human and material) resources are invested in it. In terms of emphasis no subject in the secondary curriculum matches its weight. To learn and to be educated, for most people is to learn English (Verma and Pande 1989).

1.4.3 Secondary ELT Curriculum and the Bases for its Renewal

The ELT curriculum is prepared by a team of experts at the *Curriculum Development Centre* (CDC) on a broad national uniform policy and principles to meet the national aspirations. The Centre takes into account a general language proficiency level, common interests and issues in idealised situations. As a result, it does not meet local needs and requirements of students. During framing of the curriculum, constraints such as under-resource, inadequate teacher training, insufficient facilities, and unfavourable environment of local schools are not considered seriously.

The current secondary ELT curriculum (introduced in 1999) has been revised on the basis of the fact that (a) the objectives of teaching English have shifted from developing an ability in various skills in the language to developing a communicative and academic ability in them, (b) the new approach views language as 'communication' and (c) changes are taking place in ELT internationally.

The curriculum acknowledges that teaching English is about the acquisition of communicative as well as academic skills. However, the pedagogic as well as the examination system continues to stress the acquisition of knowledge of the language. The whole instructional system continues to rely on the traditional approach to education in which emphasis is laid on accurate memory and memorisation of information; and rote learning is encouraged. The school-based quarterly, half-yearly and yearly examinations including the national examination, all follow a similar approach to testing in which test questions test the ability to locate and repeat answers from the texts provided, and the majority of questions are of a factual nature (Khaniya 1990; Poudel 2004).

The curriculum aims at enabling students to communicate with people of any nationality who speak or write English by exposing them to a variety of texts of spoken and written English. The two courses (one each for grades 9 and 10) are claimed to be of a practical nature with language 'functions' as the core of the curriculum; and grammatical structures and vocabulary as tools to express the functions. The CDC claims that it is applicable to society, both in Nepal and in the world outside (CDC 1999). It defines language in the following terms:

Language is seen here as a skill that allows one to get things done. The things that can be done through language are described as 'functions', such as expressing likes and dislikes, good wishes, etc. A single language function can be expressed through more than one grammatical structure or set of vocabulary items (CDC 1999:43).

The word 'function' here is used in the same sense as Wilkins (1976); Van Ek (1982); Leech and Svartvik (1994) and is regarded as the core of language instruction with grammatical structures and vocabulary as tools to express the functions.

1.4.4 The ELT Objectives and their Focus

The aim of teaching English at the secondary level can broadly be specified as follows:

- To develop an understanding of and competence in spoken English to be able to communicate fluently and accurately;

- To read, appreciate and develop competence in understanding a variety of reading texts including literary texts;

- To develop skills and ability to write English appropriately and effectively and to use simple reference materials; and,

- To develop an awareness of relevant cultural and ethical values (CDC 1999).

The figure below shows the weightage of the macro-skills in the secondary curriculum:

Macro-skills	2001 (New)	2000 (Old)
	(Weightage in percentage)	
Oral Skills (Listening and Speaking)	20	10
Reading (Vocabulary /comprehension)	45	45
Writing	35	35
Grammar	-	10

Figure 4 Weightage on macro-skills (CDC 1999)

The curriculum also lists skill-based objectives which it calls 'specific objectives'. There are 9 objectives for speaking, 11 for reading and 10 for writing.

Listening and speaking together have 12 objectives while reading and writing have 21. Thus, the curriculum shows a clear bias towards writing skills.

1.4.5 Discrepancies between the Prescribed Texts and the Curriculum Objectives

ELT in Nepal is textbook bound. The teachers receive new textbooks without being told what and how to deal with them. For the teachers, the textbooks are the *de-facto* curriculum and the single most important source of input to be provided to the students (Khaniya 1990; Kerr 1999).

The launching of the new curriculum calls for high quality teaching and learning materials that shape and support systematic, efficient and effective learning. The CDC, Department of Education, has theoretically attempted to respond to the need of teaching and learning materials by loosening its grip on material production and allowing schools to prescribe their own textbooks, written in line with the prescribed syllabus. However, in practice, it has not encouraged publishers and schools to do so. As a result, the schools remain obliged to prescribe the textbooks prepared by the government agencies.

The textbook writers working for these agencies face several problems including (a) a shortage of source or reference materials; (b) a lack of proper understanding of student learning processes; (c) a lack of understanding of the environment and constraints in which ELT is practised; (d) a lack of training in teaching and learning material development; and (e) an inability to address the need to combine traditional and 'new' learning materials effectively. As a result, most textbooks of the new generation continue to be 'knowledge-based' (Khaniya 1990).

There is a clear conflict between the curriculum aiming for communicative competence and prescribed tests which are mainly 'thematic' (Kerr 1999). This mismatch may make it difficult to achieve the aim of 'communicative skills'. Kerr further observes,

> While in a developed country, the use of the thematic (notional) approach might not exclude the development of functionality [or communicative ability], in a developing country where teachers had limited knowledge of the subject, inadequate or no training, and were conditioned by a rote learning methodology, this difference [between approach and text] was

likely to cause confusion by teaching language inappropriate for daily communicative needs (Kerr 1999:224).

Furthermore, the prescribed texts are based around the experience of urban contexts, which exacerbates the problem of relevance of teaching English in rural settings.

One other problem is, as I indicated in the previous section, the distribution of the materials. Textbook distribution in Nepal has been problematic due to difficult topography and absence of road networks in hilly and mountainous parts of the country. A recent study reported that nearly half of the total school population in mountain districts did not have a complete set of textbooks even half way through the academic year (CERID 2001). Teachers and school associations have been demanding that the government should bring an improvement in the existing textbooks and materials distribution system (TKP, 8 Jul. 2001), and in the educational standards of (remote) rural schools (Kantipur, 13 Mar. 2004).

1.4.6 Methods of Teaching English and the Teachers' Quandary

Prior to the *New Education System Plan* (NESP), the method of teaching English in practice was the 'grammar-translation' method, the premises of which were the learning of grammar, mastery of vocabulary and its use in sentence construction, translation from and into the target language, and composition. Some people today believe that the standard of English was very high in those days, higher than today's standard anyway. The introduction of NESP brought changes in ELT. As mentioned, it reduced the weightage given to English by 50 per cent, reducing it to a single paper carrying 100 marks. The curriculum was changed, shifting its orientation from literature to language making structural teaching as its basis. The structural approach to language teaching views language as a set of structures, and learning of language as gaining mastery over those structures. Language ability, then, was seen as the ability to handle discrete elements of the language system (sounds, words, grammar, etc.) and develop individual skills (listening, speaking, reading and writing). The prescribed approach to language was the 'oral structural and situational approach' and the method of teaching was the 'audio-lingual' method. However, while many teachers resented the abandoning

of literature as a resource for teaching, many others lacked training and continued to teach the new curriculum using the old ways.

'Global' or 'integrative' approaches to language teaching were never in practice in Nepal. Two decades after the introduction of structural teaching, the Communicative Language Teaching (CLT) movement finally made its way to Nepal. ELT is once again changed from structural teaching to communicative teaching. CLT, which was in practice from the early 1990s at the primary level and from 1999 at the secondary level, is affecting tens of thousands of teachers and students alike. Thus, the last seven years or so have witnessed a great change in ELT in Nepal.

CLT, which is based on the principles that 'language is communication' and 'learning a language is learning to communicate' (Widdowson 1978; Weir 1990), was introduced as a response to discontent with the structural methods of teaching which focused on grammar and structure. In spite of the fact that structural methods treated all skills equally, the teaching of reading and writing remained dominant in practice and produced unsatisfactory results, i.e., students who were unable to speak and understand English well. The existing curriculum and CLT, the official teaching method, have set communication as the goal of ELT with little or no focus on grammar, structure or translation. As a result, traditional teachers with no adequate training on CLT are sceptical as to whether CLT is really the right method of teaching English in the context of Nepal. Many cannot see the relevance of such a syllabus, which does not have grammatical items or structures as teaching points. In the present circumstances, they cannot see the possibility of non-analytical, participatory and experiential ways of language teaching and learning. In the absence of a clear understanding of what is expected of them, most teachers are confused and even frustrated, unsure of how to handle the new curriculum and the textbooks. Many of them do not believe that CLT and the new syllabus are superior to the traditional analytical approach to teaching. Most ELT practitioners in Nepal are not prepared for the departure from the traditional and/or structural language teaching approach to CLT.

1.4.7 English Language Teacher Development: A Lost Priority

'There can be no curriculum development without teacher development.' (Elliott 1991:51). This is generally accepted at least in the theory of the *Nepali* education system. However in practice, teacher development seldom gets its due consideration in the curriculum development processes or in the process of its implementation and as a consequence of this the teacher factor remains a problem.

Total no. of teachers	Number of trained teachers	Percentage of trained teachers	Teacher-student ratio
142183	30317	21.3	1.38

Figure 5: Number of (fully) trained teachers (Source – MOE 2003:87)

There is a serious problem in that English teachers, as Figure 5 illustrates, are under-qualified and inadequately trained. The teachers' qualifications and the training they receive have come under acute scrutiny lately (Kerr 2000; TKP, 2002). The in-service training programme of the MOE is merely ritualistic and therefore ineffective, usually run by under-qualified and poorly trained trainers. These trainers themselves do not have a proper understanding of the principles of language and language learning, theories and methods of language teaching, learning psychology, and so on. The pre-service training run by the Faculty of Education, *Tribhuvan* University, on the other hand, is isolated from the actual ELT situation and is idealistic in nature. Most training courses are developed on the basis of the academic knowledge of the course instructors with little or no consideration of the need, the school level courses, and the constraints in which ELT takes place. As a result, the training does not reflect the existing ELT situation, nor does it serve the need of an average teacher. It is generally said that these training programmes are more academic than professional in the sense that the participants of such programmes are trained in accumulating knowledge rather than in developing skills of quality pedagogy.

Teachers generally have a poor grasp of English and lack a proper understanding of ELT principles. Therefore, they simply tend to do what they do best, i.e. explain things in the vernacular *Nepali*, abandoning efforts to develop the target language

proficiency in students. To make the matter further complicated, nearly half of the teachers who are untrained cannot see any difference between teaching social studies and teaching language (Awasthi 1995). For them, teaching is 'preaching' whether it is history or the English language.

Furthermore, there is a misconception among EFL teachers about how languages, particularly foreign languages, are learned. They believe that a language is made of words, which are uttered, and combined, into sentences. To learn a language, therefore, one needs to learn words, how to pronounce them, and then the rules for combining them into sentences. In the technical terms of linguistics, one needs to have mastery over lexicon, the phonology, and the syntax of the target language. This folklore still has a tenacious hold on how languages like English should be taught in the foreign language context of Nepal. For most school teachers, there are lists of words and expressions to be memorised; grammar rules to be memorised and practised; and comprehension texts to be understood, and questions to be answered in order to use the words and rules memorised. These often constitute the core-activities of English class and homework assignments.

Most teachers of English have little or no idea that the teaching and learning of EFL as it is envisaged above is, in fact, an out-dated and inefficient approach to language learning. Evidence from a child learning their first language, migrants 'picking up' the language of the host community with no formal instruction and an EFL learner producing novel sentences or combining language elements in intricate ways, show that learning a language is much more than the learning of words, learning to pronounce them, or learning of grammar rules and their applications (McNamara and Quinn 1988).

However the fact that teachers hold the key to the success of any ELT reform can hardly be exaggerated. Any attempt to reform English language teaching and testing should be based or built on (a) the teachers' level of English language proficiency, (b) their knowledge about English language teaching and learning, (c) knowledge about and abilities in classroom practices, and (d) their knowledge about language assessment.

Summary and implications

My examination of the educational context reveals that secondary education in Nepal suffers from negligence, poor budgetary allocation and in general has low status of priority. There is a mismatch between the overall goals and development needs, and between the educational goals and the outcome. The lack of capability in and participation of the local community, administrative inefficiency combined with poor communication and co-ordination in the structural hierarchy of the bureaucracy tarnish any effort to improve education. An untrained teaching force, and above all insufficient resources, aggravates the problem further. As such, educational plans have failed to produce the desired results. As Weiler (quoted in Kerr 1999) observes:

> Many plans are nothing but exercises in wishful thinking or even futility....even carefully designed plans in education encounter implementation problems...[for] the development of education takes place at the base of the system. Whatever decisions may have been taken at higher hierarchical levels, the key to success or failure of a given plan lies in the hands of local teachers, administrators and parents.... greater attention must be paid to the implementation capability of the local level in the educational system, and to make sure that the community at the base of the system has the information and the resources, and most important the commitment necessary for the successful implementation of educational development programs (Weiler 1982 in Kerr 1999:115).

The current educational system of Nepal, especially the school education system is at the crossroads. As the political situation deteriorates due to ongoing fights between the rebels and the army, education is increasingly becoming a 'mess' (Nidhi 2004). There is a lack of a common approach to education: as a result, the political parties use it for their own benefit.

There is a lack of national commitment for English language teaching (ELT). As a consequence, they do not have a sense of direction for the teaching of the English language in Nepal. This may have been a cause of poor coordination among ELT agencies contributing to the deteriorating quality of its teaching and learning.

There is a slim ray of hope, however, that politicians and educationists may join hands together to develop a common approach to education. News published in the recent newspapers (GP, 12 Jul. 2004; TKP, 10 Feb. 2004) report that the Ministry of

Education (MOE), school teachers and academics are deeply concerned over the current state of education calling for a joint effort to work out a solution. This indicates that there is a renewed understanding among the government officials and education workers to get Nepal out of the 'mess' to a more scientific and modern system of education.

Nepal, however, must choose its approach to education carefully because any such move is likely to face strong resistance from the administrative bureaucracy and age old educational culture which values academic skills more than life or practical skills. Improvement in the very traditional ELT practice, for example, is likely to be plagued by, but nevertheless must be based on, the consideration of a poorly motivated and largely untrained work-force, poorly conceived curriculum, unbalanced curriculum provisions, inadequate technological support, and above all, insufficient resource provisions.

My aim in this project was to research the English test of the SLC and, against all the odds, develop a new test package, which is in line with the new (current) SLC curriculum. In so doing I initially assumed that the revised testing scheme would exert positive washback effects, ultimately improving the ELT practices in Nepal. I needed to develop an action (or intervention) to address the problem, and then to implement (or pilot) it in the actual situation. For this, I needed a methodological framework, which could allow me to consult a wide range of people, agencies and materials to analyse, explore and create an action plan. And at the same time, it should lend itself to organise a number of activities with a view to piloting the action plan and obtaining feedback.

In the next chapter, then, I outline the methodological framework I applied for the present research in analysing the problem, developing a plan of action and in piloting it in the field.

CHAPTER 2

Research Methodology

Action Research is a process as much as a research methodology ... concerned with systematic collecting of data, which is analysed and fed back to the subjects so that action plans can be systematically developed.
- Tomal 2003:9

2.1 Introduction

My educational training has been in the tradition of positivism in which end products (knowledge) are valued more than the processes or means applied to achieve them. So, initially I focused more on the potential outcome of my research than on the processes I was employing to carry out the research. However, this changed in the course of the research.

From the very inception of the present project, I was looking for a direction to address a long felt need to initiate a change and ameliorate the existing practice of English language testing at the *School Leaving Certificate* (SLC) level in Nepal. My approach, however, was not that of an outsider or an independent researcher but of an insider. As an ELT teaching and testing practitioner myself, I wanted to diagnose the English language testing scheme, investigate the context of the testing of English as a foreign language, analyse the existing models of language testing and then adopt a process to plan and implement (or pilot) a new testing scheme. I also wanted to organise a number of research activities at various levels of ELT practice and involve my fellow English language teaching and testing practitioners. In return, I hoped that the project would make the participants more aware of the situation, of the problem, and some possible solutions.

Therefore, I looked for a research methodology, which was comprehensive enough to include the multiple facets of my research but, at the same time, flexible enough to allow me to complete my project within the time and resource constraints of completing a degree.

Though I was not aware of it initially, my requirements and what I was doing as a researcher were very much in line with action research. In this chapter I describe action research, the research methodology I later formally adopted for this project, and give reasons why I adopted this, what key principles of action research I followed in carrying out the research, and the activities which I have theorised as 'cycle one' and 'cycle two' of the research project.

2.2 The Aim of the Research

As I mentioned in the preamble, from the beginning of my teaching career as an English language teaching and testing practitioner, I have been interested in investigating the English language education context, particularly the changing ELT context of Nepal. I was also keen to understand how existing language testing models and testing practices, particularly the ESL/EFL testing models and practices, could be adapted to the revision of the national testing scheme of Nepal.

My research questions, thus, grew out of the general ideas I drew from my practice of, and experience in, English language teaching and testing. The general ideas of improving the practice of English language teaching and testing were initially about intervention or change in the English language testing system of the SLC, and were directly related to ELT practitioners and stakeholders working at various levels. My main aim, when I set out to do this project, was to investigate the English language testing practice at the SLC level in Nepal, adopt a process to develop a more appropriate test in place of the current SLC, and consequently, make a theoretical and practical contribution to secondary level ELT practices in Nepal. However, I have learned during the course of this research that there are many more problems in implementing such a test than can be dealt with in the course of this research project.

The aim which has evolved during the three year course of this research is less ambitious, but nevertheless, just as useful as the original aim. This aim is to investigate and to model a *process* for developing EFL tests for the national school systems, taking into account the context of (EFL) curriculum renewal, the specifics of the *Nepali* educational and cultural contexts, the multiple purposes for testing (administrative, academic [achievement], instructional and prognostic), the range of available language

43

testing models and relevant current debates about approaches to, and limitations of, language testing.

In order to design an optimally fair, valid, reliable and practicable test, a review of current models of EFL testing was needed to be undertaken. Choice of an optimum model also entailed consideration of issues of the social character of language testing, the institutional nature of the assessment, ethics, and the broader educational policy within which the English curriculum (and hence the test itself) are set (McNamara 2000). Any test, in a sense, is a 'compromise' between the principles and established techniques of testing and the purposes and contextual limitations of testing (McNamara 2000:27). A test design, therefore, calls for a process of iteration between available models of testing and the specific contextual requirements. To apply the principles and techniques of language testing models developed for other contexts to the context of under-resourced, developing countries brings an additional layer of complexity to the challenge of designing a new test.

Thus, to develop a process of test adaptation, I took into account the shift in language teaching and testing that has taken place internationally, and 'compromises' that are occurring in the field of ESL/EFL testing. At the same time I drew strongly on the local factors and my own knowledge of the history and cultural and resource constraints of the *Nepali* education system.

I have designed the test adaptation process on the basis of an analysis of the secondary ELT curriculum, linguistic as well as communicative domains, discourses, aspects and skills, local context, current theories of testing and the most appropriate available language testing models. Resource, and developmental, operational and implementational issues have been addressed as part of the process of developing the test specifications for a new test. An analysis of issues, principles and dilemmas arising from the process of adaptation and test development was made.

2.3 The Research Questions

My main research question for this project was to investigate *how* an English language test can be revised and updated according to the best language testing knowledge available. To meet this aim, the following questions were researched:

1. What are the contextual parameters to be taken into account and the steps to be undertaken in the process of designing a test suitable for a new curriculum in a particular context, based on current theories and models of ELT testing?

2. What aspects of language testing practice, particularly ESL/EFL testing practices can be drawn upon in modelling a language testing adaptation process?

3. What issues and dilemmas emerge in relation to applying language testing practices that have been developed in the context of developed nations to a developing context like Nepal?

4. How useful, practicable and appropriate is the specific new test package I designed?

5. Based on the experience of this project and similar projects documented in the literature, in what ways, if at all, might language tests serve as a means for educational change in the context of educational development and renewal?

The research I have carried out into the *Nepali* SLC examination is broadly framed by the notion of action research: I have investigated the context, developed a 'solution' or alternative to the current SLC and have 'tested it' by seeking feedback from the stakeholders as well as by piloting the sample test with a small group of students. This thesis is a report of the research processes, its outcomes, and a reflection on the cycles of planning, investigation, action for change, and further reflection on the outcomes. The action for change or revision was the development and adaptation of a new test informed by theories of language acquisition and ESL/EFL testing, local requirements and the contextual factors.

2.4 Action Research

Action research as a research methodology is a family of activities in educational reform and system improvement in which a problem situation is diagnosed, remedial action planned, implemented and its effect monitored (Kemmis and Grundy 1981 in Burns 1998). There is an abundance of literature on action research emphasising that action research integrates research into the educational setting with a view to improving the practice. Action research, in this sense, makes a strong connection between theory and practice by helping practitioners connect their day-to-day practice with a broader educational theory. The present research aimed at exploring the existing context, as well as the practice of the English language testing in Nepal, identifying the problems and designing a method of finding a remedy for them. Action research, therefore, was viewed to be an appropriate methodology for the research.

Action research has successfully been applied for the improvement of three levels of practices – instructional practices, organisational practices and system-wide practices. Tomal (2003), Tyabane (2003), Grundy (1995), Zuber-Skerritt (1996 and Colin and Garrow (2002), for example, report successful application of action research to enhance the understanding of and to improve instructional practices. Similarly Speedy (2003), Tomal (2003), Zeichner and Noffke (2001), O'Brien (1998), and Greenwood and Levin (1998) report cases where action research was applied to bring out changes or improvements at an organisation or institutional level. According to Zeichner and Noffke (2001), O'Brien (1998), Greenwood and Kevin (1998) and Elliott (1996), action research is an effective way to explore practices and bring about systematic (system-wide) reform, professional development and cultural innovation supported by an informed theory.

In the present project, I address three levels of language testing practice at the same time. Firstly, I investigate what change or revision in the system of EFL testing can be adopted in the existing circumstances, and how it can be done; secondly, I make a critical analysis of the existing organisational structure and the processes of the *Office of the Controller of Examinations* (OCE) - the responsible organisation for the English test of the SLC with a view to recommending a more functional structure, and finally the research project envisages that the reform in the EFL testing practices at the SLC

level will bring improvement in ELT practices in Nepal in general and teaching and learning of English at the secondary level in particular.

2.4.1 The Key Principles

Action research is sometimes criticised for its subjectivity and open-endedness in its approach (Tomal 2003; Speedy 2003; O'Brien 1998) and also for the fact that it can be biased and manipulated by the researcher for a particular approach or method (Tyasbane 2003). Nevertheless. I treated action research as a research philosophy, rather than a research method, and applied its important features in the research methods of the present project. I applied the following key principles:

2.4.1.1 Practice –theory transformation

As it has been discussed in the introduction of this chapter, creating awareness, and producing 'new understanding' is a goal of the research with a view to improving the practice of ESL/EFL testing in Nepal. And action research, as a holistic approach, is appropriate in this case, where practice of EFL testing is the main concern, for several reasons. Firstly, I believe that action research is a systematic and practical strategy of action and reflection for helping practitioners to maximise learners' achievement of their education targets. By this I mean, and as will become evident in the later section of the chapter, that action research employs a systematic process of inquiry and of recommending a step-wise process of solving the problems identified. This process is potentially of great relevance to bringing about other much needed changes in the context of Nepal, not just the changes needed for the SLC.

Fundamentally, I regard improving practice as a process for developing the capacity of the practitioners and subsequently improving the practices and processes through which the outcomes are achieved. By 'practitioners' I speak of myself as well as of my teaching and academic colleagues. The research, therefore, focuses on the process of attempting to solve a particular problem, and on the theorising that has emerged from that process.

2.4.1.2 Descriptive as well as reflective

The aim of the study, as described in the previous section, was to investigate the context as well as the practice of the EFL testing in Nepal, identify the problems of the ailing testing scheme of the English component of the SLC and design a process for finding their solutions. The approach to study, therefore, has been descriptive (Long 2000) as well as reflective (Dick 1997). By descriptive, I mean that after setting the objectives of the study, as it is described in the next chapter, I describe the research journey in phases as it took place in the field and the steps I undertook to carry out the analysis and interpretation of the responses from Nepal. For instance, I describe how various stakeholders contributed to my journey, and how the process might have contributed to the development of their own understanding, what their experiences have been in this regard and what in their view should happen in order for it to improve. By reflective, I mean that I adapted a reflective approach to draw on the responses I collected during the research processes and to infer on a given issue or problem. For example, on the basis of the responses or evidence, I posit some problems, and reflect on their causes and their explanations.

2.4.1.3 Reflexive Critique

Reflexive critique is what is sometimes known as 'critical reflection' (Grundy 1995:16). By this I mean that I have taken a critical approach to my research journey in two ways. Firstly, during the reflection phase of the study, I take a critical look at the evidence I have collected and the interpretations that I have made on them in order to make sure that the conclusions based on the evidence accurately reflect the evidence. In addition I reflect on the broader social and political context that I have sketched. An important part of this is reflecting on my own beliefs, biases and predispositions, which influenced my conclusions. Secondly, the alternative views and concerns received from the participants, were also given due consideration and used to make modifications, elaborations or even alternative conclusions on a given issue. The participants involved in the research, therefore, as well as the supervisors, have been 'critical friends' (Grundy 1995:16) of the researcher.

2.4.1.4 Cyclic

In action research, actions can be thought of as cyclic rather than linear (Dick 1997). Action research recognises that reform or improvement is not a one-time affair, rather it is an on-going process of a cyclic nature with four steps which Kemmis and McTaggart (1988) term as four moments of planning, action (implementing plans), observation (systematically evaluating) and reflection, then re-planning (or revising plans for) further action to be repeated in a similar cycle. Susman (1983), however, has suggested five phases be conducted within each research cycle.

Initially, a problem is identified and data is collected for a more detailed diagnosis. This is followed by a collective postulation of several solutions, from which a single plan of action emerges and is implemented. Data on the results of the intervention are collected and analysed and the findings are interpreted in the light of how successful the action has been. At this point the problem is reassessed and the process begins another cycle (Susman, 1983 in O'Brien 1998).

Conventionally, the steps of an action research cycle are as follows:

Action Research Cycles
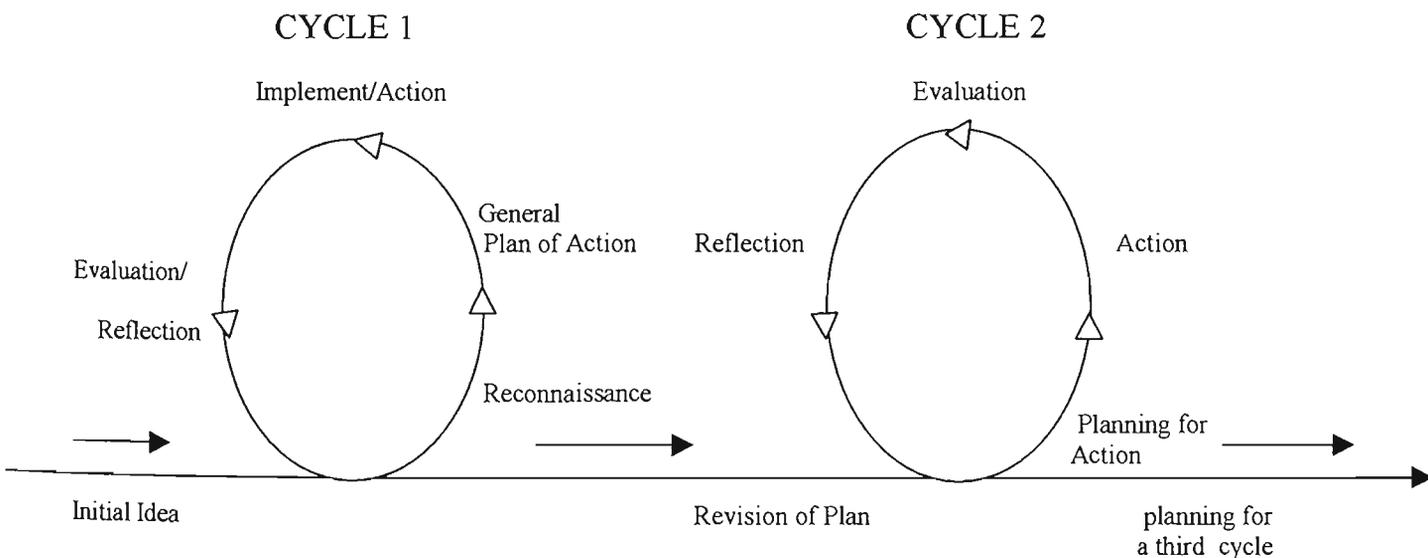
Figure 6: The Conventional Steps of an Action Research Cycle (Kemmis and McTaggart1988)

49

However, steps can be designed differently to suit particular research purposes.
As Kemmis and McTaggart say:

> Avoid being a 'slave' to the action research cycles steps, to not impose the
> cycle but let it evolve in a fluid and natural manner (Kemmis and
> McTaggart 1988:36).

In the second cycle, as I demonstrate in Figure 6, the problem is re-examined and
the plan for the second cycle revised. The cycles of planning, action, evaluation and
reflection continue again and again until the problem is resolved. For the purpose of the
present research, I have developed my own steps of the cycles. The cyclic model of
action research used in this research is presented in Figure 6.

2.4.1.5 Qualitative

The aim of the present research is not to establish any new principles of
ESL/EFL testing but to derive new understanding from the existing ones about how
improved testing practices can be implemented in Nepal. Part of this research involves
people theorising about their practice in assessing their students learning of a foreign
language – being inquisitive about the circumstances in which they work, the actions
they take and the consequences their action will bring, and coming to understand the
relationship between their practice and their professional and social life (Dick 1997;
Cohen et al 2000). This project, by analysing the English language testing practice,
aimed at making the *Nepali* testing practitioners aware of the practical and theoretical
foundation of their practice and, based on the analysis of second/foreign language
testing theories and practices, suggest ways for its improvement.

2.4.1.6 Political Process

Finally, educational reform (or change) is a political process in that it involves
awareness and understanding of the policy implications. It is 'conscientisation' (Burns
1998) or a commitment to educational improvement. Action research as a bottom up
process, then generates ideas, which link problems of understanding to policy. In other

words, action research functions as a process of advocacy and becomes a means to foster institution or system-wide reform. As Kemmis observes:

> Action research involves people in making critical analyses of the situations (classroom, school, system) in which they work – situations that are structured socially, historically and institutionally. Critical analyses aim to recover how a situation has been socially and historically constructed as a source of insight into ways in which people might be able to reconstruct it (Kemmis 1997:173).

The present research involved all stakeholders in the task of critically analysing the English language testing of the SLC with a view to raising awareness and advocating reform of the language testing system of Nepal in general and the English test of the SLC in particular. This process involves a challenge to established ways of doing things, which is also at some level a political challenge.

2.5 The Two Research Cycles of the Present Project

As I mentioned above, action research takes place in cycles. The research activities continue in a cyclic fashion until the problem is resolved or the research target is achieved. For the purpose of the present project, I completed two cycles and set a context for the third. The following is a description of the activities in the first action research cycle:

2.5.1 The First Cycle

2.5.1.1 Identifying and clarifying the general idea

My research grew out of my wondering how the English test of the SLC could be improved in the changing ELT context of Nepal, and what role it could possibly play in reforming the entire ELT practice. These ideas roughly referred to the general situation of English language testing practice. I revised the ideas constantly during the course of my study to make it more specific and researchable.

51

2.5.1.2 Reconnaissance (Analysis of the Research Context)

This activity was completed in two parts. In the first part, I analysed the educational and English language teaching and testing context of Nepal where the reform was targeted. During the critical analysis of the context, I focused on various aspects of the situation, the agents or stakeholders involved in it, the factors that affect it, and the constraints under which the system operates. Then, in the second part, I described the facts of the situation and the bearing they have on the existing situation. I employed the following methods during the activity:

a. brainstorming: I reflected on my own practice and the experience I have had in the field in the last fifteen years. Later on, I compared and assimilated the outcome of this activity with those of other subsequent activities;

b. a review of related literature: the following types of documents were critically reviewed:

  - history of education in Nepal,
  - national plans
  - educational plans
  - education commission reports
  - related research reports
  - related papers and articles

c. consultation meetings with
    - supervisors
    - language testing experts in Australia
    - language testing agencies in Australia, North America and Europe
    - Victorian Schools
    - Fellow researcher at Victoria University and the University of Melbourne

So far as the related literature is concerned, I carried out an in-depth analysis of the relevant materials and gathered information from:

- a review of some of the foundation readings of the field from North American, British, Australian and Asian writers
- agencies involved in the development and implementation of English language tests
- a review of the international tests of English as a second or foreign language
- courses on language testing
- consultation with language testing experts

2.5.1.3 In-depth Analysis of the Problem

In this phase of my research, I carried out an in-depth analysis of the English test of the SLC. I described the current state of the test and pointed out its problems. I also outlined the need for a change in the test in the changing context of ELT in Nepal.

2.5.1.4 Construction of a General Plan of Action

Having diagnosed and analysed the problem of the English test of the SLC, I outlined a plan of action. My action plans contains (a) research questions, (b) a statement of the issues or factors to be addressed, (c) development of a new test package, (d) a framework for selecting and involving people in the action phase, and (e) a statement of the resource, both human and material needed.

2.5.1.5 Action for the Change

This phase of my project consisted of developing of the new test package, which is the 'action' for the change of the current practice of the English test of the SLC. It

consists of a new test specification, a new test design, a sample test and test marking and interpretation scheme. It also consists of the process employed for developing a new English test for the SLC.

2.5.2 The Second Cycle

The second cycle of my research contains chapters on planning for the field visit, organising research activities during the field visit, analysing the feedback and reflecting on the process and the experience I gained from it.

The spiral methodological framework used in this research is summarised as follows:

CYCLE 1

CYCLE 2

Action: Development
Of Test Apparatus & Sample Test
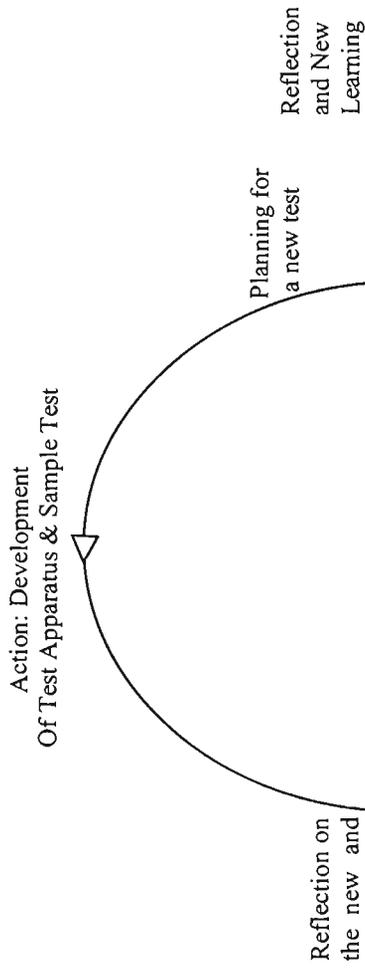
Analysis of
the feedback & Results of the pilot test

Fieldwork

Reflection on
the new and
formal planni

Planning for
a new test

Reflection
and New
Learning

Planning for
fieldwork & consultation

Indepth analysis
of the problem (the SLC)

Study & Analysis of
the context
Of research

Planning for the second cycle

planning for
the third cycle

Identification
Of research
problem
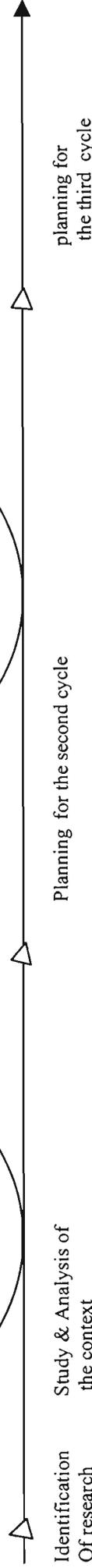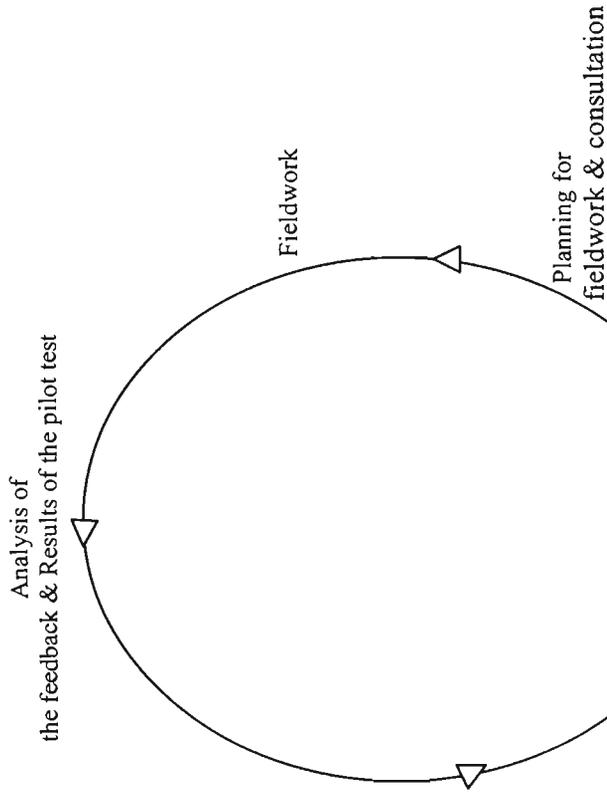
Figure 7: The Research Cycles (based on ideas from Elliot 1991, Kemmis and McTaggart1988, Colin and Garrow 1996, Greenwood and Levin 198)

The present research, thus, has been completed in two cycles of analysing the research context, identifying and diagnosing the problems in the English language test of the SLC, planning interventions (i.e., designing a process of adapting/developing a test), implementing plans (piloting the test and receiving feedback on the new test package), evaluating the results of the intervention and reflecting to what extent the problem is resolved or remains unresolved. I also evaluated any further problems that arose. The recommendations for further actions could be regarded as the starting point for a notional third cycle.

For the present project, I applied what may be termed as 'the multi-method' in the sense that it followed a combination of methods and techniques for the development and completion of the project. During the first cycle of the project, I used 'selective case study' approach (Salter 2000) to analyse cases of the English test in the national examination. I also used what is known as 'progressive focusing' (Burns 1998:330) during the analysis of language testing theories and related literature with a view to singling out factors and issues that needed to be considered when developing a new test package for the English test of the SLC.

I applied workshop and interview techniques to collect first hand data in two districts of Nepal. Feedback was sought also through a presentation at a conference and publication of journal articles. I used a testing method to obtain responses on my pilot test. Similarly, I used qualitative and quantitative approaches to analyse feedback on the new test package and responses on the pilot test.

5.2.3 Limitations of action research methodology employed in the project

Despite a number of strengths of the action research methods employed for the research (see section 2.4.1), the project has a number of methodological limitations as an action research project.

The present action research project involved the selected research participants only during the fieldwork, that is, after the new test package had been developed, and not in every phase of the project. For this reason, the project cannot claim to be a participatory action research project. Nonetheless, the fieldwork has been a significant part of the project and has contributed to the development of the thesis in two ways.

Firstly, the ideas and opinion expressed by the research participants have not only enhanced my understanding of English language teaching and testing conditions and the factors militating against the reform, they have also shaped up the outcomes of the project.

An additional limitation concerns training of testing practitioners. My project was based on the premise that training of English language teaching and testing practitioners in Nepal is a necessary condition for effective implementation of a new test package. However, given the time and resource constraints, a formal training or orientation sessions (except my presentation in the beginning of feedback sessions) for the participants before they made any comments could not be conducted. Lack of training and therefore understanding of the theoretical dimensions of the new test may limit the quality of the comments they provided on the new test package. Nonetheless, this limitation could be addressed in the next cycle of the project by the provision of some pre-training.

Ideally and in a favourable situation, I would want to work with the participants in every phase of the project. Similarly, I would train them and then allow them time to work with the new test package before they form any opinion on it. Also, giving them opportunities to work with the new test package would have helped me to explore how the new test would affect the teachers' behaviours in the classrooms; or what difference the use of the proposed marking system (and rating scale) would make in judging the value of a response against a set criteria and in arriving at a numerical score. However, owing to the resource constraints and the fact that this was only a small scale project, this was not possible.

The small-scale pilot test conducted only in two districts was another limitation of the present project. Given that the project addressed the need to reform the School Leaving Certificate Examination, a national examination for which over three hundred thousand candidates sit annually, the outcome of the limited pilot test might not be generalised. A larger-scale study would be required for that with a greater number of participants and a full-fledged new test. However, it was never my intention, and beyond the scope of my study to include a reliable, large-scale piloting of the test. However, the significance of my thesis is in the modelling and documenting the test

development process and my reflection on the constraints of the economic and political context.

In the previous chapter, I outlined the educational context of Nepal and provided an overview of its limitations and constraints. In the following chapter, which could be seen as the next step of the action research project, I analyse the research problem further.

CHAPTER 3

Problematising the School Leaving Certificate Examination

The competence of most of even those who manage to get through the SLC examination in English is questionable. Most pass the examination 'technically' because of the nature of [the] question; they can hardly function in English.

- Verma and Pande 1989:32

3.1 Introduction

With a view to elaborating the research problem, I carried out an in-depth study of the current and past *School Leaving Certificate* (SLC) (or secondary) English curricula, current and past English tests, teachers' guide, and pilot reports of the *Office of the Controller of Examinations* (OCE). In this chapter, I present a description of the SLC, the OCE and in particular the English test of the SLC. I outline issues related to the development and implementation of the test, which I shall address in the forthcoming stage, 'action for the change'. This analysis, therefore, belongs to the planning stage of my action research cycle.

The *School Leaving Certificate* (SLC) Examination is an externally administered national examination and as is the case in many Asian countries, it is administered by a separate government agency (Marsh 1999). As a measure of achievement and learning at the secondary level, the SLC is the only national examination, which every school graduate must sit in order to be eligible for tertiary level studies and for many forms of employment (Bajracharya, et al 1998). English is one of the eight subjects tested in the SLC. The education practitioners of Nepal have questioned the credibility and validity, particularly of the English test of the SLC in the recent years (Giri 1995; Khanal 1997; Kerr 1999; Poudel 2004) largely because the SLC, despite the recent changes in the SLC curriculum, still employs traditional approaches to testing. It also fails to cater to the need for a reliable source of information regarding a candidate's ability (Joshi 1996) in a given subject area.

In the first section of the chapter, I describe the status of the SLC Examination in *Nepali* society; I identify its stakeholders, and explain some of the factors, which have a bearing on the current situation. In the second section of the chapter, I explain why the English test of the SLC needs to change in order to meet the changing context of ELT in Nepal. In addition, I examine what implications revision of the test may have on ELT practice in general and testing of English in particular. I also address some of the pertinent issues and their implications on the development of test specification and new test design.

3.2 The School Leaving Certificate Examination

About 300,000 {regular (those who sit the SLC after graduating from schools annually) and exempted (those who repeat the examination after being unsuccessful in their previous attempt)} graduates from public and private schools sit the SLC held in March-April each year. As a gatekeeper for entry into the higher education, it is implemented after a school level qualifying test known as the 'send-up' test. The 'send up' test is administered by the *District Education Offices* (DEOs) for the public schools, and clusters of schools for the private schools.

The SLC is given in two language courses (*Nepali* and English), four core subjects (Maths, Science, Social Studies, and Health, Population and Environmental Studies), and two pre-vocational subjects (optional subjects I and II). Like any other external or national examinations, the SLC defines common standards of performance required to demonstrate adequate completion of a syllabus. It has status in the wider community, and is supposed to provide an objective assessment of a student's performance.

The results of the SLC are reported in two ways. Firstly, the raw scores, which are the total scores in individual subject areas, are presented. These scores are accepted at their face value only as they discriminate one group of students from another in terms of numerical scores and do not explain what abilities in the subject they possess. Secondly, an average score of all SLC subjects is presented in order to give it an interpretation. A candidate must pass in all subject areas to be eligible for a certificate. A score of 32

percent or lower is termed failure, 33 to 44 percent is awarded the third or pass division, 45 to 59 percent second division and 60 percent and above first division.

The cut-off points and the score boundaries under the current scheme are defective for two reasons. Firstly, the mark of 33 percent is too low to be acceptable as the minimum pass mark. How this number was determined as minimum pass mark in the first place remains a mystery as no MOE or CDC documents provide any justification for the selection of that number. It seems to have been done so under the influence of the education system Nepal adopted initially. In the post-colonial era, most former colonial countries, including neighbouring India and Pakistan, have revised the scheme in order to make their assessment regionally comparable. Secondly, the score boundaries in the current scheme are distributed unevenly. For example, the margin between the third and second division is 11. Between the second and first division, it is 14, but for the first division it is 40. Thus, a candidate who scores 60 or 99 gets the same class, i.e., the first division.

The SLC is a 3 hour long test in each subject area which like any other standardised test, covers a limited part of the course syllabus, and can, therefore, capture only a small sample of a student's performance even on the topics tested within the period of time allotted (Marsh 1999). I also agree with Marsh (1999) that it is biased against students who do not perform well under examination pressures, encourages a concentration of teaching only those aspects of a course which are most readily assessed (Khaniya 1990), and encourages didactic teaching and rote learning (Kerr 1999). However, the SLC is deeply rooted in *Nepali* educational tradition and is here to stay at least for the foreseeable future. It is, therefore, imperative that the problematic areas of the SLC are exposed and solutions to the problems found.

3.2.1 The High Stake Nature of the SLC and its Stakeholders

The SLC is a high stake examination, and consequently, it is accorded a great importance in the Nepali society. It plays a crucial role in the lives of most, even all individuals involved in it. As Singh (1996) observes:

The SLC examination has become a major landmark in an individual's life in the Nepalese society. It provides the ladder for one to get on to higher

61

education and also opens up to vista of making his/her own career development. Success in the SLC examination plays a decisive role in getting entrance to a campus, making the choice of subjects in higher education, taking part in scholarship competition and job competition and opting for a particular vocation. (Singh 1996:4).

Whether it is societal frontiers, admission to higher education, employment or personal achievement, the SLC results become a basis for decision making carrying a serious consequence on an individual's life. The SLC results are seen as the final arbiter of student's ability and dictate the student's future. Failure in the SLC is not only a matter of personal shame on the part of students; it also means the loss of an opportunity for a job or admission to higher education. The SLC graduates with high grades are eligible to apply for a place at prestigious institutes and faculties like Medicine, Engineering, and Science. Those who pass with an average mark seek admission in less popular areas in terms of employment prospects, such as Humanities, Education, Commerce, and Law.

The SLC affects the lives of teachers too. Failure to get students through the examination sees claims of incompetence, neglect of duty and sometimes means the loss of a job. Teachers with a high success rate in the SLC are given certificates of appreciation or monetary benefits.

Parents who can afford it make special efforts to get their children through the SLC by sending them to private tuition classes, getting them extra help during the examination period and sending them for private lessons with teachers who are potential examiners. The SLC holds the key to any walk of life in the Nepali society, so parents endeavour to get their children through the examination no matter what. They adopt any means in order to do so, because failure in the SLC means a great loss of family resources and prestige.

In Chapter 1, I described the two-stream (i.e. public and private) system of education in Nepal. I also noted that the graduates from both types of schools must sit the same SLC examination. A public school's performance, then, is judged in terms of the success rate of its students in the SLC. Schools securing high pass rate are considered to be 'prestigious' and are awarded national prizes by the Ministry of Education (MOE). Failure to secure good results is regarded as demonstrating inefficiency, poor management, and incompetence. Such schools are penalised by

withdrawing the grants-in-aid. There is no research data to show the impact of the policy of withdrawing grant-in-aid but, but like Joshi (1996), I would argue that such a policy "can have a severe and negative impact on the teachers and students' morale, on academic expectations, and on the quality of students who seek admission to such schools" (Joshi 1996).

In the case of private schools, the rate of enrolment depends on the success record in the SLC. A lower pass rate results in students/parents being less likely to enrol or continue in a school. Loss of students means loss in income for school entrepreneurs. School administrators, therefore, are mainly concerned about training the students with exam tactics and equipping them with well-prepared model answers in order to get the maximum pass rate. Numerous newspaper reports in the past (for example, Spotlight, 2001; Kantipur, 30 May 2001; TKP, 22 Apr. 2002a) alleged that some well-off schools administrators have exercised power pressure tactics or even provide monetary incentives to examiners or examination administrators to secure a higher pass rate (TKP, 22 Apr. 2001d). One example of such an irregularity was the scheme of publishing the top-ten student examinees of the SLC Board. This was abolished in 2000 because some schools used unethical practices to secure a position in the top-ten list.

The administrators of MOE, on the other hand, mainly worry about holding examinations, getting question papers to centres and collecting answer booklets from centres, getting them examined and publishing and distributing the results on time. A delay in any of these processes would be attributed to inefficiency on the part of the administrators.

As a consequence, students of the SLC develop a 'must pass' attitude and use whatever means they can to do so. Singh (1996) claims that this leads to unethical practices. He notes:

> Performance in the SLC is also considered as one of the major criteria for sanctioning the operation of schools. Success in the SLC is, thus, critically important to the students and the school. This anxiety is shared by the parents, family and community in an equal degree. For all these reasons, getting success by hook and crook has become a norm or even motivational principle for many (Singh 1996:4).

3.2.2 The OCE and its Administrative Inefficiency

The SLC examination was founded in 1933, and has been managed and conducted by a separate Board, the SLC Board with an apex body called the *Office of the Controller of Examinations* (OCE).

The Board is the highest body responsible for formulating examination policies and decisions. The OCE, as an executive body, carries out the decisions and conducts and manages the SLC examination. The Controller of Examinations assigns question setters and moderators to prepare final drafts of question papers. These are then printed and distributed through the Department of Education machinery such as Regional Directorates and district level offices, i.e., DEOs. The OCE is a highly centralised bureaucratic organisation with a very rigid power hierarchy. The functions of the OCE, which consists of Controller, Deputy Controllers, Head Examiners and Examiners, are as follows:

- carrying out the policies and decisions of the SLC Board,
- conducting SLC examinations, publishing results and awarding certificates,
- organising seminars/workshops for further improvements in the examination systems,
- maintaining records of the individual candidates of the SLC exam,
- disseminating statistical information on SLC results (MOE 2003).

The procedure for setting question papers and conducting the examination is purely administrative. Nothing is done to ensure the quality of the questions. Along with very scanty one-page guidelines, senior teachers of English are given old questions papers as model questions, and are asked to set the questions for the next SLC examination. These teachers normally follow the old model for questions and use the prescribed texts to write relevant questions. No pilot or validity checks of any kind are conducted.

A district level *Examination Co-ordination Committee* appoints a superintendent to conduct the examination. After the examination each day, the superintendent collects

all the answer-scripts and dispatches them to the regional office and /or to the centre. The OCE and the Regional Directorates distributes the answer booklets to examiners and head examiners. The examiners examine the booklets at their home, a practice which was abolished after reports of massive irregularities. This was replaced with the system of examining answers booklets at the OCE premises only (Bhandari 2002). Moderation of the examined answer booklets is mere a formality. The examined answer booklets with mark-slips are sent back to the OCE where the booklets are scrutinised and then mark slips are sent to the *National Computer Centre* for data processing. The process takes from three to six months. Upon the completion of the data processing, the results are published in the national newspapers. The OCE, the Regional Directorates and DEOs then distribute transcripts and certificates.

The OCE does not have its own administration at the regional and field levels. It relies on the staff and personnel of the education directorate and DEOs who are largely untrained in examination procedures. Furthermore, there are communication and co-ordination problems occasionally costing the confidentiality of the question papers and smooth operation of the examination. Some district level officers, who are usually those in-charge of the SLC at the district level, have been known to leak the questions to candidates from affluent families before the examinations actually take place.

3.2.3 The Resource and Expertise Constraints

The OCE, which plans, implements and manages the SLC, is poorly equipped with inadequate resources, lack of modern equipment and qualified expertise required to carry out their tasks successfully. It is generally said that, specialised examinations such as the SLC are conducted by a group of non-specialists. Most of the people who work at the OCE do not have any specialised training in testing nor do they have any experience in the field. So far as the testing of EFL is concerned, almost none of them have any knowledge or understanding of EFL testing principles, test designs and second or foreign language rating scales. As a result, testing English is 'unrealistic' and mostly 'ritualistic' (R.B. Gautam 2001a). However, the most serious problem the OCE is facing is that of lack of resources. The OCE is unable to support the SLC with adequate infrastructure, expertise, training and other facilities.

3.3 English Test of the SLC: The Focus of the Research

The English test of the *School Leaving Certificate* (SLC) Examination is at a critical stage. While it must change in the changing context of English Language Teaching (ELT) in Nepal, the test has not been able to escape from being stereotyped and reflects what (Spolsky 1979) terms as the psychometric-structuralist tradition of the past. Recently, the OCE has attempted to improve the testing (OCE 2000) by introducing some administrative reforms. However, the English test of the SLC is flawed in more than one way. This could be because the English test does not follow any officially accepted language testing framework nor does it have a well-developed test specification or any scientific test development process in place.

In this section, I give a critical analysis of the current English test and its shortcomings.

A language test is given for several formative or summative purposes, which can be categorised into three general categories – administrative, instructional and research. A test, for example, can serve the administrative purpose of placement, certification, promotion and selection. Similarly, it may also be given for instructional purposes such as diagnosis, prognosis, provision of feedback and so on. Assessments are carried for research purposes as well as to find more about language learning and teaching and learner communication strategies. In addition, a test can review the achievement of a language course or programme.

These purposes are not always mutually exclusive to each other as one form of assessment may also be used for several other purposes. The assessment given for finding evidence of progress, for example, may also be used for researching students' learning styles or communication strategies. A test given for a general purpose may also serve the purpose of motivating students towards learning. Similarly, a test given for testing achievement may also be used for certification purposes.

Thus, what to test, how to test it, when, and by whom it should be tested, is largely dependent upon, as the diagram below suggests, *why* the test is done (McNamara 2000).

FORMATIVE

T
E
S
T
I
N
G

WHY
(PURPOSE)

SUMMATIVE

WHEN: end of term, continuous.

WHAT: skills or knowledge
Use of skills/application

FORMAT: integrative/subjective
Objective/discrete point

HOW: Norm-referenced
Criterion-referenced

WHO IS END USER: teacher
administrator
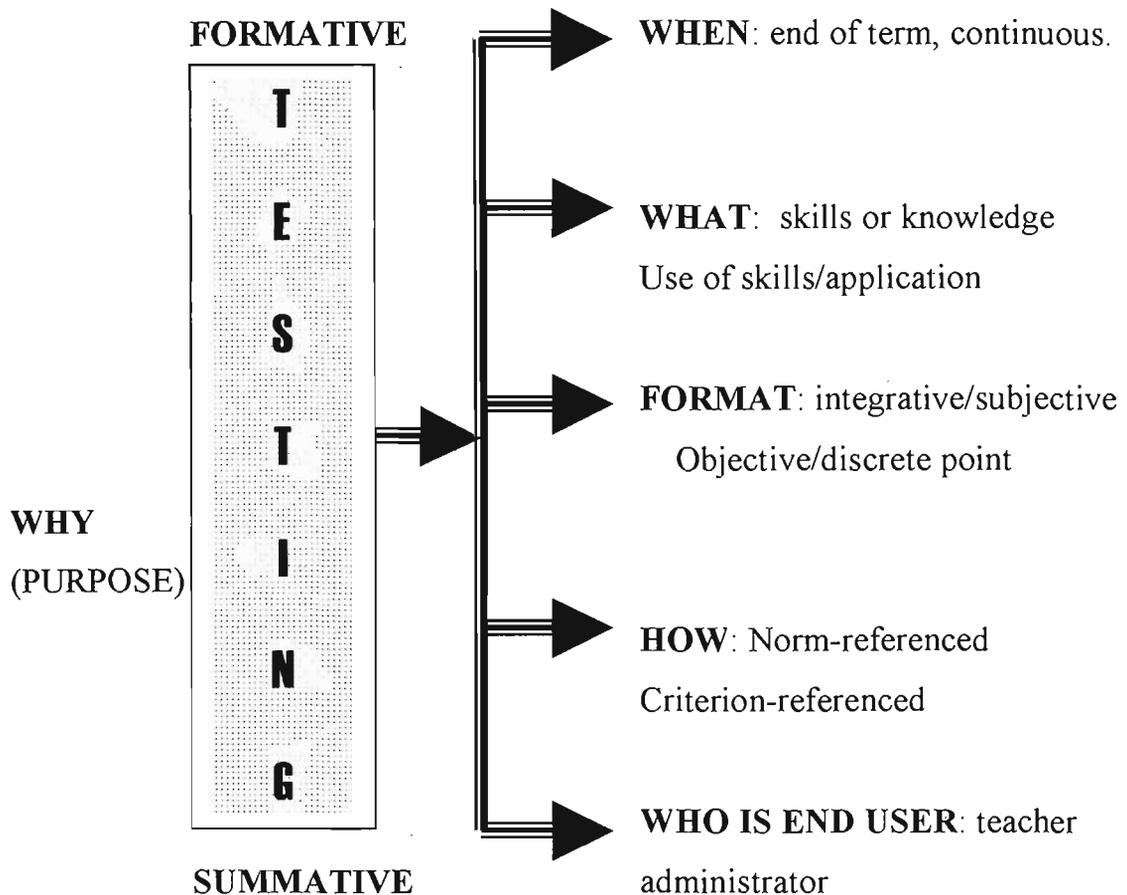
Figure 8: The WH-Paradigm of Assessment (based on Cajkler and Addleman 2000; Haines2004)

Figure 8 above demonstrates six WH- factors that influence development of a test. For example, the purpose of a test determines the format that will be used. Similarly, the subsequent decisions about what goes in the test (content), when to give the test (timing), how it will be scored (marking system) and who should use the test results largely depends on the 'why' of the test. In other words, the main alternatives in assessment, to discover how well the learners are achieving or how well they are learning, will govern all other factors in the test. These include the format/design of the test, who uses the results, the method used to interpret the results (norm-referenced or criterion-referenced), and administration of the assessment.

The English test of the SLC aims to serve two specific purposes – general proficiency and certification of achievement. The general language proficiency is indicative of the learners' linguistic competence in the target language and their ability to use this competence in actual communicative situations. As a general proficiency test,

the English test of the SLC, is, as one would expect, a test of general language skills and elements. As stated earlier, the SLC reports an aggregate of scores only, and it does not contain information on the outcomes of the test components nor does it provide descriptions of abilities. Institutes or employers, who may use the test results for selection purposes, make their decisions on the basis of the aggregates only

Decisions about certification are usually made on the basis of the achievement of the particular course. What type of certification (first, second or third division) is issued depends on what has been achieved or learned from what has been taught in the SLC course. The English test of the SLC, in this sense, is an achievement test. In other words, the SLC examination is a measure of achievement, the results of which are sometimes used for screening or selection purposes. A score given on a subject is an indication of how well a candidate has taken the examination. It is less an indication of how well he has learned during the course.

3.3.1 The Content of the SLC English Test

The English test of the *School Leaving Certificate* (SLC) examination given at the end of secondary education has been a subject of acute dissatisfaction in recent years partly because its design is faulty (McGrath 1996), and partly because it does not comply with the current theories of language testing (Khaniya 1990; Awasthi 1995; Kerr 1999). For example, language testing in Nepal is based on what is termed as 'the psychometric-structuralist approach' while, given the recent change of approach of the SLC English curriculum, it should follow what is called the 'psycho-sociolinguistic' approach (Spolsky 1975 in Howard 1980). The examination sometimes follows the pre-scientific or 'elitist' approach in the sense that the onus is on the teachers/examiners to judge the proficiency of students solely at their own discretion (Spolsky 1975). It follows the psychometric-structuralist approach because it imposes, what Lado (1961) terms as, a scientific view on language, assuming that language ability consists of elements of the language systems and development of individual skills of grammar, vocabulary, listening, speaking, reading and writing; and that learning should be tested at a number of discrete levels. As a result, in terms of current language teaching and

68

testing theories the current English test of the SLC is invalid as it does not portray students' real proficiency in the language (Khaniya 1990; Giri 1995).

The SLC, being the national examination, is not designed to assist teachers in diagnosis of pupils' weakness or abilities. The form and content of the examination has a profound limiting effect on ELT practices as the requirements of the SLC dominate all levels of teaching as well as learning practices. What the SLC requires becomes far more important than the goals of teaching. In effect, the SLC tests becomes the de facto curriculum for Year 10, controlling what is taught and learned during the secondary years and determining overall educational process (Kerr, 1999:231).

Joshi (1996), for instance, suggests that the SLC has become the most influential factor in determining what the students learn and what the teachers focus on for teaching. As the examination mostly demands answers, which are outcomes of recall, memorisation of facts and information has become a major thrust of the teaching and learning process. This has a negative impact on the curricular process and as Khaniya (1990) suggests, is certainly not congenial for developing the desired level of language proficiency in students. The students usually suffer from low intrinsic motivation and near negative attitude towards learning English, as they do not see any practical reasons for learning it other than obtaining a satisfactory score in the subject. Their sole aim is to obtain an SLC certificate.

About two decades earlier, Davies et al (1984), published a similar critique of the SLC. His team suggested that without a response based examination, more flexible teaching methods will not be attempted as long as the teachers remain SLC bound, and that as long as the SLC examinations remain unchanged, it will be difficult to improve the teaching of English in the Secondary schools (Davies, et al 1984:10). Two decades later, the situation remains unchanged.

A language test or examination should be developed on the basis of the specification of the targeted language behaviour and the test design adapted out of the specification (Bachman and Palmer 1996). In the case of the SLC, however, no such specification of the targeted behaviours has been developed. The SLC English test, therefore, continues to follow the previous test format. The testing for reading, vocabulary and grammar, for example, continue to test lower level knowledge using objective tests such as multiple choice item, true-false item, matching, fill-in and short-

answer questions. In writing, the same stereotyped questions are used for letter writing, job applications, developing an outline into a text, completing a conversation and essay writing. I present some specific cases below to support my argument.

The oral test is actually a test of a 'verbal essay' only. Candidates usually memorise answers to possible questions, and reproduce them during the oral test. Candidates are asked to choose one of the five questions, such as,

If you were the chief of your school/village, what would you do to improve it (OCE 2000)?

To answer such a question which is easily guessed, a candidate reproduces what s/he prepares or memorises prior to the test.

Some reading questions are superficial and do not require the candidates to read the text to answer them. For instance:

Match the words in column A with the words in column B

A	*B*
Universal	*difficult*
Rapidly	*international*
Drew near	*quickly*
Complicated	*came closer*

(SLC 2001)

In writing, questions are often poorly worded which confuses the candidates as to what is expected. The question below, which is a repeat of the question from an old SLC question paper, for example, asks candidates to rewrite in their own words and expand the given outline in several paragraphs. However, the way the question appears on the paper does not state this very clearly. As a result, a candidate may simply complete the story by supplying six (or seven) grammatical words and make it readable. In this case, the test becomes a grammar test rather than a writing test.

Write a readable story with the help of the guidelines given below:

A king [was] disappointed [because] his people [were] lazy. To teach them a lesson, he had a big stone put in the middle of the road one night. Next day, merchants pass[ed] and go [went] round it. An officer driving his car did the same. A young soldier came cycling [and] did the same. All cursed the stone and blamed the government for not removing it. Then the king had the stone removed. Many people watched it. Under it was an iron box marked 'For the man who moves away the stone'. Inside, [there] was a purse full of money. The people were ashamed.

(SLC 2001)

The parenthesis in the story contains possible answers, which I have supplied to illustrate that by supplying six grammatical words, the story may be rewritten into a readable story, which is not the intention of the test.

The grammar question below, despite the fact that the new SLC English curriculum does not subscribe the testing of grammar as a separate entity, is similar to that of the previous English test of the SLC:

Rewrite the following sentences choosing a suitable word from the brackets:

i.	*"What's wrong .... you?" She asked me.*	*(in, for, with)*
ii.	*Who is .... best player of your team?*	*(a, an, the)*
iii.	*Sushila met him .... the way to work.*	*(in, on, at)*
iv	*New Zealand is .... island?*	*(a, an, the)*
v.	*It is .... one rupee note.*	*(a, an, the)*
vi.	*That is .... book I like most.*	*(a, an, the)*
vii.	*She is ....the sari today.*	*(at, in, by)*

(SLC 2002)

As it can be inferred, grammar (preposition and article) is being tested on the discrete-point format of test.

As for reading comprehension, two types of texts are included in the test – seen and unseen. Type one or so called 'seen' texts are those which are selected from the prescribed texts and are, therefore, seen or read by the candidates prior to the test and are, thus, familiar to them. Two seen texts are included in the test. An example of a reading comprehension test on a seen text is given below:

Read the following passage carefully and do the exercises given below:

What are the parts of a computer? The word 'hardware' refers to the physical parts of a computer. The main are the monitor, the keyboard, the mouse and the CPU or Central Processing Unit, which contains the hard disk.

The monitor is also known as the screen or VDU (Visual Display Unit). It looks like a television screen and displays what we do on the computer. It comes in different sizes.

The actual compute, i.e., that part that performs tasks we ask it to do is in a unit holding the CPU, memory and hard and floppy disks. It comes in a rectangular box.

CPU stands for Central Processing Unit. This could be called the brain of the computer. It performs calculations and manipulates data. The keyboard functions like a typewriter. The computer is given commands by typing them on the keyboard. Information then is displayed on the computer.

Questions

1. *What does the monitor look like?*
2. *What does 'hardware refer to?*
3. *Why is the CPU called the brain of the computer?*
4. *What does the monitor do?*
5. *How does the keyboard function? (SLC 2001)*

By the time the candidates sit for the SLC, they will have gone through the 'seen' texts and questions such as the ones given in the question paper several times. Given the fact that almost all questions are information-type questions based on the surface (local) level information on the text, the candidates can easily memorise answers of such questions and reproduce them in the examination.

Type two or 'unseen' texts are supposedly selected from other sources than the prescribed texts and therefore, are new for the candidates. The following text, which is a formal letter, is an example of reading comprehension on an unseen text.

Read the following letter and answer the questions given below:

3/95 Fareway Street
WA 6539 Perth
Australia

Feb. 25, 2001

The Manager
Brown's Sport Store
164 Victoria parade
Sydney, Australia

Dear Sir /Madam,

Recently I turned fifteen and currently I am seeking part time work. This year I am completing class 10 at St. Catherine's college. The course includes English, Japanese, science, mathematics, computer studies, physical education, history and geography.

Eventually, I wish to take up a career in physical education and I am therefore interested in part time work involving equipment in the field. Should there be a vacancy for part time work in your store, I would appreciate the opportunity of an interview at a time convenient to you.? I can be contacted at the above address or by telephone on 3569872.

Yours faithfully,

Peter Smith

Questions

1. *Who is the letter addressed to?*
2. *What is the qualification of the candidate?*
3. *How old is he?*
4. *How can he be contacted?*
5. *What languages are included in the course?*
6. *Where is the letter sent from? (SLC 2001)*

The given type two or unseen text, which is seemingly artificially crated, is a job application, the questions on this text, like the ones on the seen text above are memory level questions based on the surface level information only and does not involve much interaction from the candidates.

The questions to test writing, as the example below demonstrates, are no better than the questions on oral skills.

Write several paragraphs about how people celebrate the Dashain festival in your town or village. (SLC 2002)

This is an essay type question intending to test candidates' originality or creativity in the writing skill. However, there are three problems with the question. Firstly, the question is framed in such a way that candidates could easily prepare the

answer in advance. Secondly, the question is a repeat from previous tests. Therefore, candidates can guess a handful of such questions and memorise their answers only to reproduce during examinations. Secondly the festival of Dashain is celebrated by Hindus only. Those who are not Hindus are likely to be disadvantaged. This raises an issue of equity and fairness which I take up in Chapter 9.

These questions are only a few taken from numerous other examples which appear on the SLC question papers every year. However, they are demonstrative of the fact that these questions do not provide opportunity to the candidates to demonstrate what ability in English they possess or what they can or cannot do in the target language. In other words, such questions do not test candidates' ability in the language; rather they encourage rote learning and memorisations of answers which can be mechanically reproduced during examinations.

The OCE claims that the new SLC English test is a departure from the old one in that it measures practical abilities and language functions on the basis of the curriculum objectives, and is more valid and reliable than the previous test (OCE 2000). However, it does not explain why and how the new test is more valid and reliable. Furthermore, the curriculum prescribes that practical abilities are (a) ability to communicate effectively in varying social, educational and professional situations; (b) ability to deduce and select the appropriate meaning of an unfamiliar lexical item from its particular context; (c) ability to recognise inter- and intra-sentential relations, and to recognise the organisation of a text; (d) ability to select implicitly and explicitly stated information; (e) ability to determine conceptual meanings of a text, to answer specific to the text (direct reference) and, to answer questions going outside the text (inference); (f) ability to map the main points, take notes, summarise, transfer information into non-text forms, and to use reference skills, and (g) have knowledge of literary devices used in the text. The current testing format neither subscribes to these abilities nor tests them. Skills such as choosing an item from a number of choices, saying whether a statement is true or false, matching items or filling in are not practical skills. While such a test may test one's knowledge of a language element or skill, they do not measure one's communicative ability in the language area in question.

An analysis of the current English tests (2001-2003), thus, reveals that:

- both old and current English tests of the SLC follow the discrete-point approach to language testing in that knowledge about a particular skill or element of language is tested by breaking them into discrete segments ;
- test items are based on surface level information;
- the previous oral test tested candidates' knowledge of the sound system of the target language; the current oral test is a verbal essay test and it does not test their conversational ability, an important aim of the SLC English curriculum;
- reading tests follow more or less the same format in both tests. Most questions do not require candidates to produce answers in their own language. They do NOT test the candidates' functional/ communicative competence,
- writing tests are of a similar type which encourage prepared answers. This does not necessarily encourage a candidate's originality, creativity or thinking;
- The curriculum does not envisage teaching and testing of grammar as a separate entity. However, the English test of the SLC continues to test grammar in a traditional fashion.
- most tests are stand alone tests in that they are written separately without establishing any association among them. As a result, these tests measure, if at all, language abilities in parts. And in language learning, the sum of the parts does not constitute the whole of language ability.
- there is no integration sought among the tests.

It can be said that the defects of the old English test of the SLC are the inherent defects of the new SLC.

Commenting on the old (previous) English test of the SLC, a study conducted by the University of Cambridge Local Examination Syndicate (UCLES) reported that it was

> ... stereotypical, traditional and the one which does ... not test students' abilities in functional English. [The test components] are unsuitable and require a complete overhaul. The question papers are prepared to test a range of skills but concentrate mainly on factual recall of textbook information and sometimes encourage repetition of learned model answers based on textbook exercises (UCLES 1985 cited in Awasthi 1995:35).

To sum up, despite the recent curriculum renewal shifting ELT to communicative teaching and aiming towards developing communicative competence as well as academic competence, the SLC test of English more or less remains the same. The questions are traditional in their form testing the candidates' knowledge of the language at each of the discrete levels. Also, there is a discrepancy between the mark allocation in the curriculum and the distribution of marks on the tasks.

A scheme of oral examination for 20 marks has been introduced. Under this scheme, a candidate is asked to describe a topic. The description is rated by a couple of raters, usually teachers of English, who have little or no training. A similar scheme was in practice during the NESP years. However, the NESP system of internal evaluation was abolished in 1980 calling it 'impractical' in the existing situation. A more formal scheme of assessment was put in place, which evaluated students' language proficiency on the basis of their performance in writing.

The English test of the SLC is based on the prescribed texts. Comprehension questions as well as tests of speech, vocabulary and grammar are traditional and of a factual nature, and they only test students' memory or ability to recall. In other words, the English test of the SLC encourages rote learning and guesswork and does not foster language learning. It is therefore unable to reflect the students' actual proficiency in the English language. The test items are stereotypical in nature and the same model of questions is recycled and repeated year after year. Consequently, students, teachers, school administrators and even parents abandon the aim of developing the desired level of language proficiency, and concentrate on preparing the students for the SLC by equipping them with answers to the potential examination questions, and by tutoring

examination tactics. As a result of all this, Year 10 lessons become practice sessions for the examination. What is not examined is not taught.

3.3.2 The Efficacy of the English test of the SLC

As discussed, recent studies on the English test of the SLC reveal that it is invalid and unreliable (Davies et al 1984; Khaniya 1990; CERID 2001), and unrealistic (Giri 1995; Poudel 2004). It only tests students' ability to reproduce memorised written responses to questions, and does not provide a complete picture of the learner's ability to function in the language in either speaking, reading and writing (Kerr 1999). McGrath (1996) claimed that it does more harm than good, and is a defective and inaccurate reflection of what has been taught. As a result, the examination has lost its credibility (Khaniya 1990; Awasthi 1995; Giri 1995; CERID 2001). Research on the performances of the SLC graduates enrolled in tertiary education found that 76.43% of university entrants failed in English (Bhadra and Yadav 1988), that they had difficulty in coping with tertiary education because of their poor grasp of English, and therefore, 10% or fewer passed the compulsory English paper at tertiary level education (Davies et al. 1984).

The OCE has recently, introduced some reforms such as – coding and decoding to check irregularities, parallel sets of questions to check cheating, decentralisation of marking answer scripts, and central (at one location) checking of answer scripts (OCE 2000). This shows their intention to improve the test. However, these reforms are administrative rather than academic or qualitative in that the reforms do not help improve the quality of the English test as such.

As the current English test of SLC is based on the old format/scheme, it is, like the previous one defective. This is caused in part by the selection of the test format, content of the test, organisation of the test, procedures of marking, and partly for the system of examination itself. The test does not measure what it is meant to measure. It does not match the syllabus and course materials as it is supposed to. It does not yield consistently dependable results, and above all, the development of the test is not backed by any pilot testing or research.

3.4 The Potential Role of the SLC English test in Reforming Existing ELT Practice

I am a strong believer in the power of testing. Testing has the potential to change practices of teaching and learning. As I have indicated in this chapter, the English test of the SLC exerts negative washback effects. In the same way, an improved SLC could exert positive washback effects. Heyneman and Ransom (1992) suggests that in order for a test to provide positive washback effects it must have an appropriate distribution of recall, application, synthesis and evaluation type questions. The test scheme should be able to improve the content of examinations and set up good feedback mechanisms.

The role that examination plays in educational reform has led to the 'examination-improvement approach' to educational reform (Kellaghan 1992:102). According to this approach, examination plays a major role in raising educational standards and provides justifications to bring about educational reform (Noah and Eckstein 1992). It also influences educational processes such as setting up learning goals, determining the teaching syllabus, selecting course materials and organising classroom processes (Taylor 2004).

In a country like Nepal where education is constrained by severe resource limitation, an examination can be a low cost means to improvement. Emphasising the importance of examination in an under-developed country, Heyneman and Ransom (1992) write:

> Especially in the context of scarce resources with declining educational standards and the ever increasing demands for better qualified manpower, education officials are looking for low-cost ways to improve their education system...examination can be powerful, low cost means of influencing the quality of what teachers teach and what students learn in school (Heyneman and Ransom 1992:109).

So, where should reform commence? Davies (1985) suggests that examination is the most sensitive, most controllable and most certain change-producing factor in the total educational innovation process. So, if one has to choose an agent of educational change, to begin with, one should always choose examination first because 'creative and

innovative testing ... successfully attracts to a syllabus change or a new syllabus' (Davies 1985:8). I illustrate this point further with a few examples.

Davies (1985), reporting a Malaysian case of ELT reform, describes the problem it can create when examination is not given its due consideration. The Malaysian CDC introduced a new communicative syllabus into the secondary education aimed at developing in school graduates the ability to communicate in English. The lack of coordination between the CDC and the examination syndicate, a separate government agency with responsibility to conduct national examinations, led the examination to be incongruent with the syllabus resulting in a fiasco. The main reason for this disaster was, as Davies pointed out, the mismatch between the syllabus goals and the examination.

After an evaluation of the situation, a further reform package was introduced in Malaysia in the mid 1990s which considered examination as one of the primary components of the reform process, and because of this consideration, the ELT practice made tremendous progress towards the desired goals (Ahmed 1997).

The Alderson Report (1986) on the *National Certificate of English* of Sri Lanka provides another example of how a national examination can exert a positive effect on classroom practices. The national examination in Sri Lanka was changed along with other sister elements such as the course objectives, text materials and teaching methods. All support materials such as teacher's guide, test materials, learning materials were prepared on the basis of the skills, sub-skills and activities as specified in the test specifications. The underlying principle of this practice was, as Alderson indicates in his report, classroom practices developed the language skills and abilities, which the course objectives targeted, and which were assessed in the examination. Alderson further points out that tests had a strong impact on teaching and learning. Thus, whether the effect of a test is negative or positive depends on the 'nature, ingredients and use' of the test (Alderson 1986:104).

Lee (2000), who studied the effect of testing on freshmen in Canada, found that testing methods results in a significant change in the students' performance in the target language.

The closest example, however, comes from the Caribbean country of Trinidad and Tobago. Essay writing, despite being a part of the school curriculum for so long, had

not been taught simply because it was not tested in the national examination. However, with its inclusion in the *Common Entrance Examination* (CEE), the national examination in Trinidad and Tobago, the change in the school instructional practice was sudden, immediate and direct. As London (1997) writes:

> Government policy to include essay writing in the Common Entrance Examination in Trinidad and Tobago has changed the content of teaching. How to write an essay is now deliberately taught, and both teachers and government officials believe that the change in the test represented an improvement in the primary school curriculum (London, 1997:144).

London further argues that the reform in the CEE, a high-stake examination in Trinidad and Tobago, benefited the country's education system in several ways. Firstly, the reform impacted on the instructional practices. Essay writing, previously marginalized, surged into prominence as soon as it was included in the national test. Secondly, the reform enhanced school education. Signs of students using language creatively began to appear. Thirdly, the reform boosted the credibility of the examination. The reform was viewed as a refinement and modernisation of the traditional examination. Fourthly, the reform also raised public awareness about the national examination, which served as a basis to secure popular support for the intended curriculum reform. The most important gain of the reform, however, was the change in the teachers and the classroom practices where teachers, as the most direct agents, designed, planned and implemented the instructional programmes and evaluated student progress.

Kellaghan (1992) expresses a similar view:

> If the quality of examination is changed, then because of high stakes associated with examination in terms of student opportunities and teacher accountability, the educational experience of students will also change (Kellaghan 1992:102)

Khaniya (1990), who researched the washback effects of the SLC, also found that:

> if the people responsible for the design of the exam can make explicit what exactly the students are expected to have achieved and if the ingredients [of the examination] are in accordance with purpose of the

whole teaching programmes, the teachers and students can work towards achieving that (Khaniya 1990:327)

In summary, it can be said that examination reform may facilitate more general reform in education. Change in education can effectively be implemented through examination. However, change in other elements of the educational process does not necessarily change the examination process. A good examination does not only monitor learning, but, as Heyneman and Ransom (1992) observe below, it also provides directions for reforms:

> A well designed examination system can monitor and measure achievement and occasionally aptitude, provide performance feedback to individual districts, schools and students, inform education officials about the overall strengths and weaknesses of their educational system and suggests directions for change and improvement (Heyneman and Ransom 1992:108).

Summary and Implications

The SLC is regarded as a measure of learning during the school years, and is considered to be the minimum requirement for lower level employment and higher education. The SLC, thus, serves the functions of (a) a measure of achievement (b) a qualification for employment, (c) a requirement for higher education, and (d) an asset for social prestige as a person without a SLC degree is not considered to be educated. As a result, it becomes a controlling and dominating factor in the total education process. Hence, it has, in its present form, an undesirable effect on schools and constrains curriculum innovation.

The new curriculum acknowledges that the teaching of English is about acquisition of communicative skills and aims at enabling students to communicate with English speakers by exposing them to a variety of texts. The secondary level course claims to be of practical nature with functions as the core of the curriculum and grammatical structures and vocabulary as tools to express them. There has been a drastic change in the way language is viewed and in the way it is expected to be taught and learned. But the testing scheme has not changed to reflect the changed view.

The current English test of the SLC remains traditional and is unsuitable, as it does not assess on the new curriculum objectives.

The SLC must change the way it measures language proficiency. Unless the SLC changes to a response based examination, more flexible teaching methods will not be attempted, and teachers will remain 'SLC bound'. Furthermore, as long as the SLC examination remains unchanged, it will be difficult to improve the teaching of English at the secondary level.

Any language test or examination should be developed on the basis of the targeted language behaviour. It should have a well-defined test design and elaborately developed test specifications. In particular, revision of a high stake national test like the SLC must be well planned, well researched, and above all, well prepared. Revising the current procedure or scheme of the English test of the SLC would hopefully entail that:

- the OCE is reorganised to allow a separate cell for research and training;
- the capability of the OCE and its personnel has to be enhanced by training staff and developing necessary infrastructure;
- the English test of the SLC addresses both purposes it is supposed to serve (proficiency and achievement);
- the English test revises its design as well as its content;
- the resource constraints of the OCE are addressed.

In the following chapters, I suggest what design the SLC English test might adopt in order to develop a valid as well as reliable test to meet the current goals of the curriculum. Finally, I outline the structural change that the OCE needs to undergo as well as the expertise and infrastructure requirements for successful implementation of such a test (Appendix J).

CHAPTER 4

Planning for the Change:

What do we know about Language Testing?

What one chooses to put into a test or what one samples to put into a test depends not only on the model of the test, but also on the purpose of the test for which it is intended.

- Davies 1997, MMW Video Series 1

4.1 Introduction

In chapter 1, I outlined the context of Nepali education and pointed out the inherent problems and constraints the system is facing today along with their impact on educational quality and the modernisation process. I also sketched the English language teaching (ELT) scenario in Nepal and pointed out some of the issues associated with it. In Chapter 3, I discussed the English test of the *School Leaving Certificate* (SLC) and the reasons why this test must be revised in order to meet the changing need of ELT learners and practitioners. I also suggested what implications the revision of the test might have on school level ELT practice.

In order to produce a new test that reflects recent advances in language testing theory, I embarked on a comprehensive review of linguistic theories and language testing models. This review, along with Chapter 5, can be seen as a part of planning phase of the action research cycle. It can also be seen as the literature review of the thesis.

Our understanding of the nature of language, how it is learned and how best to teach it, has changed over the years. These changes, according to Ingram (1985), have inevitably been accompanied by changes in the ways and means of language testing. In other words, the trends and the resulting approaches not only influence the way we view language and language learning, they also affect the way we do language teaching and testing.

Trend, in its generic sense, is a general direction, course or tendency of thoughts, events or acts which emerges as a result of or as a reaction to an earlier trend. In language teaching and testing, particularly English language teaching, a trend is a general belief or view about language and how it should be taught and tested. Approaches, on the other hand, are methods used or steps taken in setting a task or problem and solving it. In other words, approaches are designed in line with the general view (i.e., a trend) of language to achieve the desired goal. For example, 'communicative language teaching' is a trend in which language is viewed as communication, and therefore, should be taught and learned through communicative acts. A number of communicative approaches, therefore, can potentially be used to realise one single set of goals.

In this chapter, I describe how different writers/mentors of different linguistic traditions have defined language ability and how their understanding of this ability reflects on subsequent language testing theories and models. In the first section, I introduce some of the changes that have taken place in the last century and discuss the trends that have been prevalent in the field of second or foreign language testing. Then I analyse the resultant language testing models, particularly the second or foreign language testing models and describe the way the models reflect different view of language, language learning, teaching, and more importantly, testing. I describe what is common in their discussions on language ability and where they differ.

My aim in this chapter is to report how I selected a number of tests for review, and point out what issues and aspects could be drawn upon when designing a process for the development of a new SLC test for Nepal.

4.2 Selection and Sampling of Language Tests for Review

It was a privilege for me to be in a situation in which I could benefit from all available opportunities and get materials on the most recent development in the field. However, overwhelmed initially by the vast amount of literature, data bases, websites e-journals, and the sophistication of research done in the field of language testing, particularly the second or foreign language testing, I began to update myself by –

- searching and sampling tests and other materials,

- consulting with relevant people and agencies,
- taking up a postgraduate language testing course,
- participating in research seminars, and
- net-working with like-minded people.

In order to select and review available tests, I first sampled from the most recently published books, reports and materials along with the ones most projects on language testing referred to, and then categorised them according to their origin or reference; for example, North America, United Kingdom and Europe (particularly the Council of Europe), Australia, and Asia. For each of the four categories, I selected foundation readings for language testing and later on added readings on research methodology, action research, global education, education and curriculum reform and so on. In addition, I contacted several agencies and organisations for material support including:

- Australian Council of Educational Research (ACER)
- Language Testing Research Centre (LTRC)
- National English Language Training and Research (NELTR)
- Department of Education, Employment and Training (DEET), Victoria
- Board of Studies, Victoria
- Education Testing Services, Princeton, USA
- International English Language Testing System (IELTS), Cambridge University, UK
- The British Council, Manchester, UK

These consultations were important in the sense that the agencies provided valuable information about ongoing or completed projects, which could provide necessary input to my research. They also let me use their website for further information and test materials for review. The people concerned also provided me insights by telling me about the test development process within their own organisations.

As I mentioned earlier, after my initial confusion about where to begin given the overwhelming amount of literature, I began to look at some of the major works of the four continents; namely, North America, Europe (UK in particular), Asia and Australia. I analysed the major works on language testing models or approaches on the basis of the following questions:

- How do they view language, language learning and language proficiency?
- What are the salient features of the model or approach?
- How useful are they from the adaptation point of view in a developing context?
- What features or aspects of the models can be replicated or adapted in the context of under-resourced, developing countries?
- What issues or problems may I have to deal with in the process of adaptation?

In the following section, I discuss the trends in language testing that emerged out of the analysis of language testing models or the approaches that have developed out of these trends.

4.3 Trends in Language Testing

Modelling in second or foreign language testing is not a recent concept. In fact, language testing theorists have always emphasised the need to base the development and use of language tests on a theory of language and language proficiency. More recently, they have called for the incorporation of a theoretical framework of what language proficiency is, with methods and technology, involved in measuring it. Consequently, language teaching theorists and methodologists have developed and used language testing models or methods that matched the language teaching models in practice at a particular point of time. In other words, trends in language testing have always followed the trends in language teaching.

These different definitions of language ability focus on different aspects of ability, so developing a single language testing design to accommodate all of these

aspects is a complex task. Furthermore, the test instruments, or methods used to elicit language abilities, are themselves based on the assumed level of abilities, making it uncertain whether the test measures will elicit data to characterise an individual's language performance in non-testing situations. Bachman (1990) describes the complexity of the problem of defining and measuring ability in the following terms:

> All language tests must be based on a clear definition of language abilities, whether this derives from a language teaching syllabus or a general theory of language ability, and must utilise some procedure for eliciting language performance (Bachman 1990:9).

Language testers, therefore, face a dilemma of attempting to measure abilities that are not precisely defined, and of using methods to elicit language abilities that themselves are manifestations of the very same language abilities. In other words, the test methods or data elicitation procedures one uses to measure language abilities are characteristically related to the way one views these abilities. The most important task of a language tester, therefore, is to define language abilities in such a way that the test methods or elicitation procedures applied elicit language test performance that is characteristic of language performance in non-testing situations. The different ways of looking at language ability at different points of time, and consequently, proposing and using different procedures for measuring it, have been marked as trends in language testing. In the following section, then, I present an account of different trends in language testing in examining the ways these trends view language learning and testing.

4.3.1 The Pre-scientific Trend

I indicated above that one cannot separate language testing from an understanding of the nature of language, of language abilities and of its learning and teaching. I also noted that different language testing theorists view language and language abilities differently. In consequence, this has become the basis of different language testing trends in different, but sometimes, overlapping periods. Prior to 1960s, for example, language was viewed as a means to approach the target culture and language learning was a way of training the mind (McGarrell 1981). The target

87

language, therefore, was an object of study, which could be separated from its context for the purpose of teaching and learning. Literature was the main source of learning materials. Language learning, therefore, was viewed as a process of learning grammatical rules and vocabulary, and applying those rules and words in writing and translation exercises. In this 'separatist' view (Davies 1968:216), the relationship between society and language was not considered important. Language testing was, therefore, testing the learning of words and grammatical accuracy through writing exercises.

The pre-scientific trend is also known as what Madsen (1983) calls the 'intuitive' stage or in Spolsky's (1978) term the 'pre-scientific era' in that decisions about teaching and testing mainly depended on the personal discretion or subjective judgement of the teachers. It embraced the traditional grammar - translation approach in which learners were required to apply the rules of grammar taught to them deductively and the vocabulary they learned by memorisation. The grammar translation course was basically a grammatical progression, which included grammatical structures and the vocabulary needed to express them. Both were selected arbitrarily on the basis of principles of easy to difficult, familiar to unfamiliar and concrete to abstract. The items that were viewed as simple, easy and familiar were taught before the ones that were less frequent, or of increasing abstraction, complexity or difficulty.

The merit of this trend can be seen in the creativity and accuracy it encouraged in students' writing, and in the way language proficiency was assessed. To facilitate control over formal accuracy, grammar practice tasks such as analysis of sentence structure, labelling its parts, or synthesising parts into larger units, were commonly practised. According to Ingram (1985), in order to reinforce the mastery over grammatical accuracy further, abundant writing activities such as translation, essay, letter writing, précis, and open-ended answers based on reading and comprehension, were organised. As for assessment, the testers made a global or holistic evaluation of the learners' ability in the target language through their composition and writing activities, which required them to synthesise their knowledge of the rules and vocabulary to produce syntactically acceptable utterances. The testers, according to Madsen (1983), made their judgement about candidates' ability in the target language on the basis of their knowledge of grammar and vocabulary exemplified in writing.

So far as its demerits are concerned, this trend is characterised by a lack of testers' concern for statistical analysis, objectivity, or test reliability (Howard 1980; Spolsky 1993). In this approach, the onus was placed on subjective judgements of the learners' language proficiency. Therefore, a language test within the framework of this trend was the responsibility of the individual teacher who set his/her own standards or criteria. A typical language test, thus, comprised translation, framing of question answer and composition exercises. Oral tests, however, were rarely given.

4.3.2 The Psychometric - Structuralist Trend

The behaviourist language learning theories evolved in the early 1960s into what Spolsky (1978) calls a psychometric- structuralist trend, which was mainly inspired by structural linguistics (Bloomfield 1933; Firth 1953) and behaviourism (Skinner 1957). Also known as the era of scientific language testing, this trend, as Madsen (1983) notes, viewed language as a multitude of discrete point patterns, which could be learned by a mechanical, repetitive stimulus-response habit formation process. Language ability was seen as the ability to handle these discrete elements of the language system, and to develop aspects of individual language skills (Ingram 1985).

During this period, several approaches to language teaching appeared, among these were the audiolingual approach, oral-aural approach and oral-structural approach. These structural approaches to language teaching regarded language as a set of structures with language learning as gaining mastery over these structures through the process of repetition and practice. As Carroll (1966) observes, language learning was basically:

> ... a process of acquiring conscious control of the phonological, grammatical, and lexical patterns of a second language, largely through study and analysis of these patterns as a body of knowledge (Carroll 1966:16).

The discrete-point approach to testing measured language proficiency by testing learners' knowledge of the discrete items of grammar and vocabulary and of language skills by testing one item or aspect at a time.

In the 1960s and 1970s, discrete-point tests became the most widely used tests worldwide and are still practised in many parts of the world. So much so that the psychometric-structuralist trend in second or foreign language testing became synonymous with objective testing. In subsequent years, Lado's (1961) work, which was seminal in introducing objective testing to second or foreign language testing, claimed a universal appeal and received a great deal of support from methodologists, teachers, course designers, and test developers. As Spolsky notes:

> ...(it) has correctly pointed to the desirability of testing for very specific items of language knowledge and skills, judiciously sampled from the usually enormous pool of possible items. This procedure makes it a highly reliable and valid testing (Spolsky 1978, quoted in McGarrell 1981:24).

In the late 1970s and early 1980s, this approach to testing became controversial. The majority of its opponents criticised the way it views language and language learning and the procedures it employs to testing language proficiency. This criticism paved the way for integrated and communicative testing.

4.3.3 The Psycholinguistic - Sociolinguistic Trend

Learners of a second or foreign language are exposed to the target language through its practising/observing its skills (listening, speaking, reading and writing) and elements (sounds, words, sentences). However, a language is not learned as separate skills or elements, rather it is learned in its entirety. That is to say, language skills and elements are learned as integrated parts, not as isolated components of a system. This holistic or integrative approach to language and language learning implies that learning of language is a unitary or holistic process in that it goes beyond the mastery of individual skills and elements. The sociolinguistic view of language posits that every utterance occurs in a culturally determined but identifiable context in which individuals engage as participants to perform various roles. What they say and how they say it is determined by cultural as well as contextual factors.

The psycholinguistic-sociolinguistic trend, which is also known as the integrative-sociolinguistic trend (McGarrell 1981), emerged from opposition to the

psychometric-structuralist trend or the discrete - point/atomic approach to language testing. Language proficiency, for this trend, is not an accumulation of discrete elements of language, rather, it is more than gaining control of a set of habits or structures. It is an ability demonstrated in performance in an actual situation. Language testing, therefore, should be viewed as a measure of communicative competence revealed by a candidate's performance in a given social situation (Howard 1980; Weir 1988).

There are, therefore, three aspects to this trend. They are (1) the psycholinguistic aspect, (2) the integrative aspect, and (3) the sociolinguistic aspect.

4.3.3.1 The Psycholinguistic Aspect

Psycholinguistics, which draws upon cognitive psychology and cognitive learning theories, views the human mind as central to all learning processes. Unlike the behaviouristic approach, which regards language learning as an external mechanistic process shaped by environmental factors, psycholinguistics considers language learning as a complex internal and abstract mental process underlying the planning, production, perception and comprehension of language. This aspect of language learning, which later led to theorising language acquisition and subsequently second language acquisition, draws considerably on Chomskyan linguistics (Chomsky 1961). Chomskyan linguistics is fundamentally different from Saussurean linguistics (Saussure 1966) in the sense that while the former claims the existence of "innate ability" (i.e., predetermined language learning ability which exists in every normal human being), the latter describes it as a physical process of learning language structures and systems.

Language testing, under psycholinguistic theory, shifts its emphasis from linguistic accuracy to functional ability. Language tests reflect on problem-solving approaches and are supposed to reveal what underlying rules the learners have internalised. It is during this trend that learner errors in the target language received a positive attitude and were considered to be beneficial as they could be an indication of the learners' level of 'transitional competence' (Corder 1967) or 'interlanguage' (Selinker 1972), rather than a problem.

91

4.3.3.2 The Integrative Aspect

The integrative aspect of this trend emphasised the assessment of a combination of language abilities at the same time. According to Banerjee (2000), integrative testing, unlike discrete point testing, closely reflects a candidate's real language use. Oller (1979), the mentor of what is popularly known as the Unitary Competence Hypothesis (UCH), proposed that there is a single, unitary factor that underlies language proficiency. He argued that language competence is indivisible and that the four macro-skills of listening, speaking reading and writing are so closely interrelated that it is appropriate to consider a general language proficiency factor rather than to distinguish finely between them. Oller (1979) writes:

> ...the concept of an integrative test was born in contrast with the definition of a discrete point test. If discrete items take languages skill apart, integrative tests put it back together. Whereas discrete items attempt to test knowledge of language one bit at a time, integrative tests attempt to assess a learner's capacity to use many bits all at the same time (Oller 1979:37).

Oller suggests that integrative tests such as cloze and dictation assess several elements and/or skills simultaneously and are a better guide to a learners' aptitude and potential communicative ability, as they require them to perform under real life constraints, such as time pressure and pragmatics. Weir (1988) suggests that they are also practical to administer, economical to set and mark, and have respectable degree of reliability.

The main defect of this approach, as McNamara (2000) points out, is that it cannot measure the communicative language ability of a candidate, and therefore, it cannot serve all the purposes of a language test.

4.3.3.3 The Sociolinguistic Aspect

The sociolinguistic aspect is based on the premise that knowing a language involves more than the knowledge of language elements and skills. It involves, alongside the knowledge of language code, the ability to use the language in specific contexts. There are two aspects of language ability, (a) knowledge of the language code (linguistic competence), and (b) knowledge of socio-linguistic factors which include

knowledge of the culture and social rules. A language test should measure ability to use language in both types of knowledge because a test measuring the linguistic aspect only, cannot account for how well the user can use it to communicate in given social contexts.

So far as language testing is concerned, this aspect calls for a testing scheme that reveals the learners' ability in using the language in communicative situations. Language tests should evaluate not only the learners' knowledge of the elements and skills but also their ability to comprehend and produce utterances that are both situationally and contextually appropriate.

4.3.4 The Trend of Communicative Language Testing

Communicative language testing belongs to the psycholinguistic-sociolinguistic trend. However, following the works of several language testing methodologists, particularly of Canale and Swain (1980); Bachman (1990); Weir (1990); Bachman and Palmer (1996); McNamara (1996), I find that there is enough ground for communicative language teaching to establish itself as a trend. After Hymes' (1972) two-dimensional model of communicative competence consisting of linguistic and sociolinguistic elements, several communicative language teaching and testing approaches (and models) appeared in Europe as well as in North America. These approaches shifted testers' attention from grammatical accuracy, objectivity and reliability, to functional and contextual appropriateness, language use and validity. During the 1980s, and early 1990s, language testing models such as those proposed by Morrow (1979), Canale and Swain (1980); Bachman (1990); McNamara (1996) advocated that language testing should measure both competence and performance, i.e., (a) what the learner knows about the form and about how to use it appropriately; and (b) the extent to which s/he actually demonstrates language knowledge in meaningful performance.

Bachman and Palmer (1996) added additional features to Hymes (1972) model of communicative competence. They suggested that a communicative language test has four important dimensions, namely (i) specificity of context, (ii) authenticity of materials, (iii) authenticity of test tasks, and (iv) simulation of real life situations, and must operationalise all four types of competencies – linguistic, sociolinguistic,

discourse and strategic. In other words, in a communicative language test, communicative performance tasks should be representative of the types of tasks and materials the candidates are likely to encounter in real-life situations and at the same time, correspond to normal language use.

From the discussion above, the following issues emerge. Firstly, the new SLC English test should be an opportunity for the candidates to synthesise their linguistic knowledge (interlanguage) with that of the communicative context in order to produce language that is not only syntactically acceptable but at the same time contextually appropriate. Secondly, as a communicative language test, it should be integrative in design. Thirdly, it should have the characteristics of specificity of context, authenticity of materials and test tasks, and simulation of real life situations. And finally, the SLC English test should be based on the holistic view of language proficiency and should maintain objectivity and reliability in marking and minimise subjectivity.

In the section below, I analyse some of the language testing models and approaches that have been most prevalent at different times to see how they address these issues.

4.4 Approaches to Second or Foreign Language Testing[*]

4.4.1 The Discrete point approaches

Discrete-point approaches to testing language attempt to test knowledge of a language in decontextualised segments. Such approaches, as I discussed in the previous section, are based on the theory that language consists of different parts (grammar, vocabulary, sounds, and so on) and different skills (for example, listening, speaking, reading and writing), and these are made up of elements that can be tested separately. A language test, according to this approach, comprises of a large number of discrete test items, which, as a group, cover all aspects of language ability. Language testing, thus, is viewed as testing the ability of handling different elements of language systems and development of individual skills separately and independently. According to Banerjee (2000), the approach assumes that each item gives precise information about a

---

[*] A version of this section was published in Journal of NELTA, (2002) Vol. 7, No. 1-2, Pp-11-25

94

candidate's mastery of that element or skill and that the total score describes his or her language ability as a whole.

A typical discrete-point test consists of (a) a test of a specific component of language, such as a syntactic point or phonological item or a vocabulary item, (b) a reading test in which comprehension of a text is tested through testing the understanding of specific items contained in the text, and (c) a writing test in which a candidate is required to write on, or about, a specific item or information already provided. The focus, obviously, is on the achievement of accuracy through reproduction of isolated language elements independent of their context, rather than on fluency and construction of meaning.

The discrete point approach to testing became the centre of controversy during the latter part of 1970 and in the 1980s. Davies (1978), for example, criticises the way language is viewed under the approach and says that whole language development cannot and should not be equated with the separate development of its constituent parts because the whole is always greater than any one of its constituents. Ingram (1985) makes a similar comment that language is not just the sum of its parts, but the parts have to be mobilised and integrated together to carry out particular tasks in particular situations. Ingram's remark is a reflection of Oller's point on the defects of the discrete point approach to testing. He says that it suffers from several deficiencies and that it depends on language proficiency being neatly quantifiable. He outlines the deficiency in the following terms:

> Discrete point analysis necessarily breaks the elements of language apart and tries to teach them (or test them) separately with little or no attention to the way those elements interact in a larger context of communication. What makes it ineffective as a basis for teaching or testing languages is that crucial properties of language are lost when its elements are separated. [because] the whole is greater than the sum of its parts (Oller 1979: 212).

Language knowledge, therefore, must be tested in language contexts in order to see if a person can communicate appropriately in a given context. Testing formal knowledge of language (i.e., linguistic competence) is necessary, but not sufficient to predict that the person can use the language effectively in a social situation. A person taking a driving test or music test, for example, must demonstrate their driving or

performing ability as a whole. The knowledge of engine parts or the keyboard does not necessarily make him or her a good driver or performer.

According to Davies (1990), discrete point tests, therefore, though they are highly reliable, are not valid tests because they do not give information about how the candidate uses language in real communicative situations, and are, therefore, as Weir (1988) points out, irrelevant and artificial. Weir goes on to suggest that grammatical (linguistic) competence is not a good indicator of one's communicative skills at all. He concludes that atomistic types of tests, which assess language in isolation, have no practical value because they give only limited information about a candidate's language ability.

The discussion in this section reflects some of the issues I mentioned in the previous section. This analysis re-emphasises the fact that a language test should be a test of linguistic as well as communicative ability. It also reiterates that the test should take a holistic approach, rather than an atomistic approach, to language proficiency. The analysis in this chapter also spells out an additional feature, the feature of validity, that I should consider when developing a new test for the SLC.

4.4.2 The Integrative Approaches

Integrative approaches to testing, based mainly on the work of Oller (1979), are seen as in opposition to the discrete point approaches, and promote the notion that language ability is holistic or unitary rather than a divisible construct. According to the Unitary Competence Hypothesis (UCH), underlying language behaviours can be specified on the basis of a candidate's Pragmatic Expectancy Grammar. The Pragmatic Expectancy Grammar, which, according to Shohamy (1996), is constructed in the course of language acquisition, is a psychological representation of the language user's ability to expect utterances in a given context. It is the chief mechanism underlying the skills of thinking, understanding, speaking, reading and writing. Language testing, then, is primarily a task of assessing the efficiency of the pragmatic expectancy grammar the learner is in the process of constructing.

The integrative approaches to language testing are based on the principle of component analysis, in that, as Banerjee (2000) suggests, scores from tests of vocabulary, grammar, phonology, reading, dictation, cloze and composition are

analysed and correlated to determine the candidate's language proficiency. Integrative tests are commonly used in proficiency tests because they reflect real language usage more closely.

Oller's UCH has been questioned for the fact that the data for his research, as Cummins (1979) claims, came from academic tests which focus on conscious and explicit features of language rather than the functional and sociolinguistic features, and that the data, according to Bachman and Palmer (1982), is not conclusive to support the existence of unitary competence. The approach implies that based on unitary competence, a language learner should be able to use the target language equally well in all its manifestations. However, this is not usually the case. The experience of ESL/EFL practitioners is, as Bachman (1989) notes, that non-English speaking background (NESB) learners are good at one skill or aspect (writing, for instance), but have difficulties in some other aspects (e.g., speaking). One other problem of the integrative approach to testing is that the results are reported as a single score though the tasks in the tests combine different components of language ability. It is, therefore, difficult to interpret the score as to what it actually means.

This section implies that use of a cloze test increases validity of a new test because the test assesses a candidates linguistic, textual and pragmatic ability to process thinking and the target language in a given context, a point for me to consider when developing a revised test. Other issues that emerge from the discussion above are that componential analysis and co-relation among the sub-tests and the whole test are important procedures to establish validity of tests. Furthermore, reporting of a test outcome should be more descriptive and informative than it normally is with this kind of test.

4.4.3 The Proficiency Approaches

The term 'second or foreign language proficiency' is used in two different ways. Firstly, it is related directly to 'ability' and is defined as the degree of competence or capability in a given language demonstrated by an individual at a given point of time independent of a specific course, textbooks or teaching methods. Secondly, it relates to the extent and adequacy of an individual's control (or mastery) of the target language in

all kinds of social or interactive situations including work settings. The former meaning of proficiency denotes competence whereas the later one specifies performance.

All language tests, in this sense of the term, are proficiency tests. The discrete-point approach (Lado 1961), integrative approaches (Carroll 1961, 1972), and Oller 1979), the pragmatic approach (Davies 1965), the theoretical frameworks (Cummins 1984), rating-scale approaches such as the American Council on the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines and the Australian Second Language Proficiency Ratings (ASLPR), and Proficiency Assessment Frameworks (PAF) (Hinofotis, Bailey and Stern 1981) are approaches to proficiency assessment. The differences in these approaches lie in the ways they view language and the focus they attach while developing language tests.

The Proficiency Movement in North America began in the 1980s with the publication of ACTFL Provisional Proficiency Guidelines in 1982. The ACTFL Proficiency Guidelines aimed at assessing a learner's ability in functional and communicative language. Oral Proficiency Interviews (OPI), which were face to face interviews could be rated on the ACTFL Proficiency Guidelines. Cummins (1979 and 1984) also provided a theoretical framework of proficiency, focusing on its cognitive and academic aspects. Like Davies (1965), Cummins considers proficiency to be of abstract, global and general nature, and performance to be a contextually embedded activity. The Proficiency Assessment Framework (PAF) (Hinofotis et al 1981), on which Bachman's (1990) model appears to be drawn, is a measure of componential language ability which focuses on context specific, conscious and explicit features of language. The context specific PAF needs a special mention because it is an exemplary case of how a tester can take a theoretical model of language testing and adapt it to a specific context, immediate purpose, testing conditions and other constraints.

The PAF was developed to assess the ESL oral proficiency of foreign teaching assistants (FTAs). The raters, under this empirical research, were required to provide two types of ratings, firstly their overall impression of the FTA's performance; and then their justifications in written comments.

Three main categories of the FTA's performance were identified for further elaboration, namely language proficiency, delivery, and communication. Language proficiency included ability in vocabulary, grammar, pronunciation, and flow of speech.

Delivery comprised personality and non-linguistic factors such as eye contact, gestures, confidence in manner and enthusiasm, whereas communication included features like explanation, use of supporting evidence, clarity of expression and ability to relate to students.

A rating scale was provided to the raters during the second round of viewing the tapes. A series of regression analysis studies (study of test relatedness) of the PAF revealed that all main categories and sub-categories contributed to the evaluation of FTA's oral performance. The PAF's categories can be related to the corresponding components of Bachman's (1990) Communicative Language Assessment (CLA) model. For example, the grammar component and the communication categories can relate to grammatical competence and illocutionary competence of the CLA model respectively, though in much curtailed forms.

Similarly, Chaulhoub-Deville (1997) adapted the PAF for assessing students' oral proficiency for everyday communication. He confirmed the conclusion of Hinofotis et al (1981) study that there is a correspondence of PAF categories with the components of CLA model. Thus, the PAF studies provide empirical justification for adapting specific aspects of the more general theoretical models to a specific purpose in a specific context.

In the discussion above two very important issues emerge. The first one is related with the process of test development. Features are added and/or deleted from an existing model to make it adaptable to a specific context. And secondly, rating scales are effective means for measuring general language proficiency.

4.4.4 The Rating Scale Approaches

Rating scales as a basis for assessing language proficiency have become widespread in recent times because of people's interest in transparency in educational systems. According to North (2000), rating scales are useful because they describe what the attainment of a given level of language proficiency means in practice. Rating scale approaches are becoming more and more popular today because rating scales (a) provide a basis for the learner to compare him/herself, (b) establish a framework of reference to describe achievement in terms which are meaningful and user-friendly for all stakeholders, and (c) increase the reliability of subjectively judged ratings, especially

of productive language skills. They provide common bands and meaning for such judgements. Rating scales also set learning goals and descriptions of proficiency for learners and allow the test results to be measured against the targets. They also offer a profile of a candidate's ability in the target language to the prospective employment organisations and training or education institutes. In effect, they may become a basis for selection, recruitment or remediation.

Today, the most popular rating scales are the Australian Second Language Proficiency Ratings (ASLPR) (now ISLPR, Wylie 1997), the Interagency Language Roundtable (ILR) scales, and the American Council on the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines. They are all derived from the same source namely the rating scale developed by the US Foreign Service Institute (FSI). The ILR scales are for use with US government employees, so, it is excluded from the discussion here.

4.4.4.1 The American Council on the Teaching of Foreign Languages (1982)

The American Council on the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines (1982, later revised in 1986) describes a candidate's language proficiency on a 9-point scale ranging from novice to superior. The descriptors represent degrees of proficiency in real world communication and describe how language learners typically function along the range of competence. The tri-sectional framework of the scale consists of (a) functions which tell what a candidate can do at what level, (b) content, which says what contents or topics can serve as a basis for assessment at the level in question, and (c) accuracy, which describes what quality or feature the candidate's language will contain at a given level.

The ACTFL Proficiency Guidelines are sometimes criticised for (a) lacking notions of communicative competence (Liskin-Gasparro 2000) (b) promoting an artificial context where the interactional skills and the set of roles that a candidate can play are limited (Bachman 1989), and (c) not corresponding to findings in L2 acquisition because they take a generic, educated native speaker as a criterion against which to measure the proficiency of NESB learners (Brindley 1991).

4.4.4.2 Australian Second Language Proficiency Ratings (1984)

Australian Second Language Proficiency Ratings (ASLPR) is an elaborate procedure used to measure general proficiency in a language learned as a second or foreign language. It is a descriptive scale in the sense that it consists of descriptions of language behaviour at nine levels, from zero to nine. According to Ingram (1985), ASLPR consists of general descriptions of observable behaviour with 'examples of specific tasks' that exemplify the sorts of behaviour that are observable. ASLPR has been developed in English, French, Italian, Japanese and Spanish, and its use has expanded rapidly in Australia. In 1997, the name of ASLPR was change to International Second Language Proficiency Rating (ISLPR) (Wylie 1997) in order to give it an international appeal.

According to Ingram (1984), each macro-skill in the ASLPR (or ISLPR) is described separately and a candidate's proficiency is stated in a profile. Each profile has a numerical value, which is accompanied by descriptions. The descriptions are statements of what the candidate's behaviour is in terms of the tasks s/he can carry out and how s/he can carry them out. The numerical scores are interpreted directly from the descriptions so that the reader knows what sorts of things the candidate can do in the target language.

However, the ASLPR is not a test instrument but a rating scale, which is based on, as I mentioned earlier on, the IRL oral interview. The problems are that the raters are expected to develop their own test tasks but, the scale provides no guidance for that, nor does it say how the new test tasks can be validated.

The numerous divisions and subtle differences between the divisions they recommend in the ACTFL Proficiency Guidelines and ASLPR Guidelines, (comprehensive though they are in terms of their coverage and application) are 'blurred' and difficult to comprehend for the largely untrained Nepali teaching force, and therefore, there is a possibility for them to be unreliable. Also, rating scales are potentially inefficient in terms of the time it takes to score on them. With careful

planning and development, these problems, however, can be surmounted (Bachman and Palmer 1996).

For the revision of the SLC, I can draw upon two aspects of the rating scale approach. Firstly, the use of a rating scale will provide a framework for reference to proficiency levels, and secondly it will increase the reliability of the new test.

4.4.5 The Communicative Approaches

The communicative approaches to language testing are based on the assumption that language is learned to communicate different kinds of situations. They emphasise the need for testing language by using processes of communication. These processes include using language appropriately to interact with people in different types of situations and perform different types of tasks. Such tests emphasise authenticity of test methods, authenticity of test materials and authenticity of responses so that the test outcomes give a real picture of a candidate's communicative competence. Thus, the focus in this type of test, as Savignon (2000) suggests, is on qualitative evaluation of the candidate's language knowledge rather than the quantitative assessment of linguistic features.

The communicative approaches became a common practice during the late 1980's and early 1990's, particularly after the much publicised works of Canale and Swain (1980), and Bachman (1990) who, according to Shahomy (1996), responded to the concern of the 1970s that language produced on non-communicative tests was artificial and non-representative of the candidate's actual communicative competence. Thus, the communicative approaches were a response to the call for direct tests in which both the test format and procedure duplicated as closely as possible the real life situation.

The tradition of communicative language testing began much earlier though, i.e., from ideas of Hymes (1972) in which he makes a distinction between linguistic competence and communicative competence. Hymes's work includes formal or linguistic competence or knowledge, and ability to use language appropriately in social situations. In addition, he tested performance, which is the actual use of the language. However, the model, as Bachman (1990) claims, does not distinguish between

competence and skill, a deficiency that gave way to, and resulted in, the development of subsequent models of communicative language testing. Hymes's model has had a lasting effect on many recent models of language teaching and testing.

Language testers have attempted to define the construct of language knowledge, isolate its variables and examine the relationships among these variables to provide a basis for what to test and how it can be validly tested. It is, therefore, important to see how competence and performance have been addressed in the theoretical works of language testing. Some notable works are Halliday (1976); Morrow (1977 & 1981); Munby (1978); Widdowson (1978); Canale and Swain (1980); Canale (1983); Davies (1985); Bachman (1990) and Bachman and Palmer (1996). But the models that attracted widespread attention are those of Canale and Swain (1980), Davies (1989), and Bachman (1990). A brief discussion of these models is given below:

4.4.5.1 The Communicative Competence Model (Canale and Swain 1980)

Canale and Swain's (1980) model has been influential because of its convincing treatment of competence or knowledge which includes (a) grammatical competence (i.e., knowledge of the language elements), (b) communicative competence (i.e., knowledge of socio-cultural rules) and (c) strategic competence (i.e., coping strategies in a communicative situation). Canale (1983), amending his and Swain's previous classification of competencies, makes a distinction between knowledge and skill as components of communicative competence. He added a fourth aspect of knowledge, "discourse competence" which is mastery of how to combine grammatical forms and meanings to achieve a unified spoken or written text.

However, this model, as Bachman (1989) suggests, is descriptive rather than a working model because it focuses on competence and ignores the implementation of that competence in language, a criticism which reflected on an earlier comment made earlier by Cziko (1984). Cziko argued that as a descriptive model, the communicative competence model, defines all the knowledge and skills that learners need to use effectively. However, the differentiation between communicative and discourse

competence, as Cziko (1984) claims, is not the result of any empirical study, and is separate or independent rather than inclusive.

The model seems rich enough to conceptualise the issues involved in a second language performance situation, and comprehensive enough to incorporate every major aspect of the language ability; and as an evolving framework, it provides a map rather than a prescription of language ability, setting out those areas of ability that need to be taken into account in second language assessment.

The problem with this approach is multifold. Firstly, it is not easy to identify test performance with trait or ability because a performance can only be associated with ability if the test task is 'direct'. Furthermore, language tests, according to Davies (1995), like all mental measures, are indirect. Therefore, the task of identifying performance with trait is problematic. Secondly, the approach is primarily concerned with construct validity and disregards face, content, and predictive validity (these terms are defined later in Chapter 6). The communicative approach suggests coverage in terms of language components skills and abilities, and, as Bachman (1990) recommends, the content of communicative language tests should be motivating, substantive, integrated and interactive. This may ensure coverage and authenticity but it still does not ensure that the test will have content validity, a necessary quality of a test. In order for such a test to have content validity the test construction should be based on the test specification, which, according to Davies (1995) and Davies et al (1999), is prepared after a thorough analysis of the target domain.

4.4.5.2 The Pragmatic Model (Davies 1989)

According to Davies (1989), what makes a language test communicative or non-communicative is an interplay between problems, tasks and the manipulations of language skills. To Davies, the purpose and context of testing are the most relevant features of communicative language testing. He, in a similar fashion to Hymes's dichotomy of knowledge and ability for use, distinguishes two aspects of communicative competence – 'knowledge *that*', which refers to knowledge in the Hymes's dichotomy, and 'knowledge *how*', which refers to ability for use. Davies

retains the term 'knowledge', as Figure 9 below shows, to refer to competence but uses the term 'proficiency' or 'control' to refer to performance (Davies 1989:162).

Proficiency for Davies is observable behaviour or the product of behaviour valued in itself. By control, he means that a candidate needs to demonstrate not only knowledge of language but also skill in the use of that knowledge in settings, which are, to some extent, communicative. The figure below summarises his theory of language and language testing:

Competence	Performance
Knowledge *that*	knowledge *how*
General	specific
Indirect tests	direct tests
Proficiency tests	achievement tests
Construct, predictive validity	content validity
Test of formal competence	context specific, work place
TOEFL, IELTS	Free Interviews, SLC
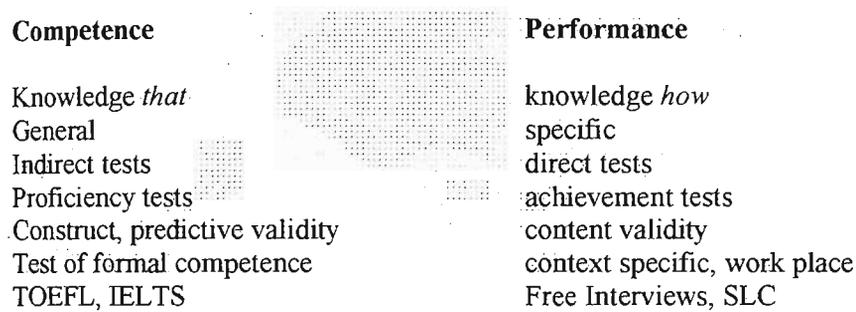
Figure 9: Davies model of language proficiency and proficiency testing (based on the ideas of Davies 1989 & 1991)

Davies views 'competence' as the figure above shows, as a general, abstract ability, which therefore, cannot be tested directly. Proficiency tests such as TOEFL and IELTS can only provide an indirect approximation of a candidate's competence. All proficiency tests in this sense are indirect tests. Performance, on the other hand, is context specific, and can be measured directly though performance tests such as free interviews and through semi-direct tests such as simulations and role-plays, which combine general and specific testing contexts.

4.4.5.3 Bachman's multi-trait, multi-method Model (1990)

Bachman's model (1990) is sometimes termed as the 'latent-trait model' or multi method, multi trait model (Davies 1995; Chalhoub–Deville 1997). This model addresses the traits or personal characteristics of a candidate that are latent, or invisible, but, which affect his or her test performance, and may be detrimental to the total outcome. The model views language as a combination of knowledge and competencies coupled with the ability to implement that knowledge appropriately in any given context. There are three main parts in the definition –

105

(a)   combination of knowledge and competencies. Bachman (1990) views them in combination with each other, and therefore, language tests based on this view must test knowledge as well as competence,

(b)   ability to implement. This implies that being competent or having knowledge in a language is not enough. One must also have ability to implement, use or put to practice. The term 'ability' includes linguistic ability, social ability, affective ability, and communicative ability. In fact., all types of abilities that come into interplay during an interaction or communication, and

(c)   language appropriate in a context. One should not just have ability to use language, but be able to use it appropriately in a given social context. Whether or not the language produced is appropriate is judged by the continuation or success of the interaction and the fulfilment of the task.

Bachman's model (1990) is an elaboration of Canale and Swain's (1980) model. This elaborated view provides a broad basis for the development and use of language tests. In his model, Bachman makes a distinction between organisational competence (which includes grammatical competence and textual competence) and pragmatic competence, (which consists of illocutionary competence and sociolinguistic competence). Organisational competence has tended to be the focus of most language tests because it is quantifiable. However, it is not easy to assess pragmatic competence because it is more difficult to elicit an utterance indicative of pragmatic competence than to measure it. Pragmatic competence involves an understanding of the social rules governing the use of language. The users must know what is socio-linguistically appropriate in a given context and how to be sensitive to varieties of language, register, and native-like naturalness, and to cultural references.

Bachman suggests that performance on a language test is affected by a wide variety of factors, and an understanding of these factors and how they affect test scores is fundamental to the development and use of language tests. His model reflects the conviction that if one is to develop and use language tests appropriately, i.e., for the purpose for which they are intended, one must base them on clear definitions of both the

abilities to be measured and the means by which the abilities are observed and measured.

According to McNamara (2000), Bachman and Palmer (1996) revised Bachman's earlier framework by adding an affective component, re-conceptualising strategic competence as a set of metacognitive strategies, and relabelling knowledge structure as topical knowledge.

Bachman and Palmer investigated grammatical, sociolinguistic and pragmatic competence using a modified version of ACTFL interview, a writing sample, a multiple-choice test and a self-rating scale. They claim that their model is based on both theoretical and empirical work. However, a test based on the multi trait, multi method model, is essentially a large test because it accommodates a large number of variables. According to Chalhoub-Deville (1997), while the length of a test may affect the candidate's performance adversely, the operationalisation of different variables may become a problem resulting in a low level of reliability.

4.4.6 The Performance Approaches

The performance-based approaches to language testing developed out of the uncertainties that existed in the language testing field in 1980s mainly because of the non-existence of a well-developed theoretical model, in spite of the fact that they have a long history and widespread usage in other fields. So, pragmatic considerations and operational definitions guided the development and use of language tests. As a result, as Shohamy (1996:145) suggests, language tests were 'theory-free', performance based, task-driven, while considered to be communicative, functional, direct and authentic.

Performance approaches to language testing were developed in response to practical needs in specific-purpose contexts drawing on the theory and practice of performance assessment in vocational fields (McNamara 1996). Language tests on this approach were developed and used on the basis of (a) purpose of testing, (b) testing context, (c) sampling of test task, (d) simulation of a situation, and (e) coverage. Performance tests are oral or written. They may be direct, reflecting an actual life situation (e.g., tourist guides in action), or semi-direct (role-plays and interviews). Performance on these tests is analysed by means of a rating scale, which provides a score and a description.

The reasons why performance based approaches became very popular are that:

(a) they are free from complications of theoretical terms,

(b) language described in behavioural/function terms is easy to comprehend,

(c) rating scales describe language in a communicable way

(d) rating scales are generic, descriptive and the hierarchical criteria in them are taken from everyday life situations.

Performance based tests have, thus, high face validity, high washback effects and high client acceptability.

According to Shohamy (1996), the performance approaches to language assessment are simplistic, narrow and are not based on a theory of construct. The very simplistic approach to tests becomes a weakness of performance approaches. As I noted earlier, performance tests focus on tasks rather than constructs. Constructs, such as theories of competence, knowledge, communication, and skills, are overlooked therefore, causing lack of construct validity. However, it is the constructs that should drive the development, scoring, and interpretation of performance assessment.

Summary

Approaches	Views on Language Proficiency	Considerations for Test Development	Comments
Discrete point Approaches Lado 1961	Formal knowledge grammatical competence Language a system of habits of communications	Indirect reference to performance Overall Control Cross cultural understanding	No test of interactive ability No consideration of authenticity of materials/ materials response
Proficiency Approaches Pragmatic Model Davies (1965&1985)	Componential Overall - control A combination of skills and work sample domains	Learning approach (linguistic) Performance approach (language at work) Authenticity of materials performance	Componential analysis Reliable, valid Componential analysis of test
CALP/BICS Cummins 1979 & 1984	Cognitive and academic aspects of language proficiency, Communication fluency Abstract, general, global	Surface fluency Academic and communicative performance Active and cognitive participation in a task context embedded activity	The extent to which measure of language proficiency should relate to academic achievement and language proficiency Context specific
Proficiency Assessment Framework Hinofotis, Bailey and Stern 1981, Chalhoub-Deville (1997)	General language proficiency Componential ability	Focus on conscious and explicit features of language Language proficiency Delivery Communication	
Integrative Approaches			
Oller 1979	Unitary ability Abstract, general	Focus on academic conscious and explicit features of language	Lacks functional social aspect Componential analysis of test performance
Carroll (1961, 1972 & 1985)	Componential Abilities	Domains, skills Integration through test specification and test modes	Componential analysis of test performance

Communicative Approaches

Morrow ((1979) Canale and Swain 1980) Bachman 1990	Componential Abilities	Real life approach Interactional approach Contextualisation of language Test methods should mirror aspects of performance under test Domains and skills	Too elaborate/ comprehensive resulting in large tests
Performance Approaches McNamara 1996	Capabilities involved in language use Language knowledge underlying performance	Specific purpose testing Direct tests Analysis of skills and components of performance Work sample approach Quality of execution of performance	Focus on task fulfilment and quality of task execution
Rating-scales Approaches ACTFL/ ASLPR	General Componential ability	Language domains Language components Skills, Descriptions of levels	Scales, not tests lack of guidelines for test development

Figure 10: Summary of language testing approaches

Summary and Implications

In this chapter, I have discussed the most important trends and approaches to language testing and the views of language proficiency and language learning reflected on each. I have referred to issues that I should address when developing a new English test for the SLC. The SLC English test, as a proficiency test should have elements of competence (construct validity, predictive validity and tasks of formal competence). Similarly, as an achievement test, it should have elements of performance (i.e., content validity and use of interviews). It should also test all four types of competence – formal, sociolinguistics, strategic and discourse. I have summarised these approaches in the figure above.

The trends or approaches to second or foreign language testing, as I have described in this chapter, may be described in summary as analytical, global and communicative. The analytical approach to language testing follows the description of language, which suggests that language consists of several discrete systems and sub-systems, and that language learning is gaining mastery over the systems one by one. Testing, then, is employing objective tests, which could be scored consistently. The global approach to testing, according to Ingram (1985), advocates combining various language sub-skills and elements in testing because they are interrelated and interdependent, and for integrating various test items into a total language event. The communicative approach, on the other hand, views language as communication, and language learning as developing communicative competence, which, as Read (1981) says, is essential for enabling learners to use language in the multiple functions it serves in real life. In communicative testing, the best examinations, according to Madsen (1983), are those which make the candidates use their learning of the target language to respond naturally as one does when exchanging ideas orally or in writing.

Despite the progress in communicative language testing, only a small number of communicative language tests in second or foreign language have been published. The UCLES's Certificates in Communicative Skills in English (UCLES 1990), and the Australian Educational Council's (AEC) ESL Scales (AEC 1994) which I used as a reference for the development of the proposed rating scale) and the National Language and Literacy Institute of Australia's (NLLIA) ESL Development: Language and

Literacy in Schools (Lumley 1994), are currently in use in the British and Australian education systems.

Most modern language testing models or approaches favour a componential approach to language proficiency though the question regarding the nature of the components remains debatable to this date. Furthermore, many models have a strong theoretical foundation but scanty empirical support. Some models are based on empirical data but suffer from poor statistical support while other models are too abstract and generic or too elaborate and comprehensive to be of immediate practical value. Therefore, adaptation of a model (or approach) or models (or approaches) to suit the specific characteristics and purpose of testing without altering the theoretical foundation is possibly the best option.

From the perspective of creating a new test design, a synthesis of discrete-point approach and integrative procedures was viewed necessary in the 1970s. Today, when the psycholinguistic-sociolinguistic trend is still in vogue and the communicative approach to language testing is becoming increasingly popular, it is assumed that no single language testing model or test 'can accommodate the wide variety of test scenarios' (Weir 1990:11). What appeals to ELT practitioners and present day language testers most, in the wider context of second or foreign language testing, according to Davies (1995), is the combination of two or more testing frameworks (or approaches) in order to make it more adaptable for a particular purpose in a particular context. I developed a plan to frame my revised version of the SLC on a combination of holistic, integrated and communicative testing frameworks (see Figure 12, page 140).

The planning phase of cycle 1 of the present project continues into the next chapter in which, with a view to bringing out issues and aspects on which a new test could be built, I analyse some of the most common ESL/EFL tests.

CHAPTER 5

Planning for the Change:

What do we Learn from Second or Foreign Language Tests?

... the competing requirements of test validity and financial practicality will maintain the distinction between tests which can be administered reliably to large numbers of students, and more holistic tests which can potentially reveal all aspects of the candidates' language proficiency.

- Clapham 2000 (in Byram 2000:51)

5.1 Introduction

In this chapter, I review some of the British, North American and Australian ESL or EFL tests, along with some tests of specific language skills and elements. I also outline the implications they might have on the development of a new test design. The in-depth analysis of the tests, their contents, task types, the characteristic features, and more importantly of the features of the test development process is a part of my search. Consistency in rating test performance and the holistic approach to testing are two of the main concerns on which I further build upon in this chapter. Like the previous chapter, this chapter is a part of the planning stage of the first cycle of my action research project, the implications of which I have carried over to the next stage, the stage of action for change.

5.2 International Tests of English as a Second or Foreign Language[*]

5.2.1 The British Tests

Testing English as a second language (ESL) started in Britain in the second half of the 19th century. However, testing English as a Foreign Language (EFL) formally started in 1913 when Britain instituted the Examination for the Certificate of Proficiency in English, a test for candidates of non-English background given by the

---

University of Cambridge. This examination was modelled on the traditional native speaker language syllabus with an essay as the key feature. In addition to a paper in phonetics, there was a literature paper and a paper involving translation. Oral examinations, writing questions, such as letter writing and composition (either on a narrative, descriptive or an imaginative topic), reading and writing main points, and a summary of a story, were added in the Certificate Examination in subsequent years. The examination was recognised as the equivalent of the standard of English required for all students, British or foreign, before their entrance to a university. With the growing demands of certification in English, different levels of the Certificate were introduced and later conducted through the British Council (Howatt 1984).

Cambridge University examinations put emphasis on literature-based curriculum and there was absolute reliance on subjective grading. Thus, these examinations were, as Spolsky (1993) says, more concerned about what to test (i.e., curriculum), rather than about how to test (i.e., methods of testing).

The Cambridge University ESL or EFL examinations have developed considerably since their commencement. The University's ESL/EFL tests of *General Certificate of Secondary Education* (GCSE) today are administered to thousands of non-English speaking background (NESB) students enrolled or seeking enrolment in British schools. At the higher level, the University conducts O (Ordinary) and A (Advanced) level examinations and also ESL/EFL proficiency tests. The *International English Language Testing System* (IELTS), a test for students of Year 11 onward is a measure of English language proficiency for students seeking admission for higher education or training in British or Australian universities. It was jointly developed in the mid-1980s by the *University of Cambridge Language Examination Syndicate* (UCLES), The British Council (BC) and the International Development Program (IDP) of Australian Universities (Ingram 1991; Buell 1992; Clapham 1995). The test reflects the ideas of communicative language testing and was probably the first communicative language test administered over a large population. A brief description of IELTS follows:

5.2.1.1 IELTS

The IELTS (UCLES 2000) contains both language skills (receptive and productive) and language elements (grammar and vocabulary). The productive skills are

tested by employing direct assessment tasks, whereas the receptive skills are tested through indirect measures. The present IELTS assesses both the general English proficiency and English for Academic Purposes (EAP) to establish whether the candidates seeking admission in an academic or training programme have sufficient English to undertake the programme without their study being unduly inhibited by their English skills. It also provides a basis on which the nature and length of a professional English course, if required, is determined.

According to the IELTS Handbook (UCLES 2000), the IELTS has three components, the first, general, consisting of listening and speaking; the second, modular, which consists of English for academic purposes; and the third, general training, containing reading and writing.

The listening test, basically a paper-and-pencil test, assesses a candidate's ability to process spoken language and to comprehend it. The reading and writing tests are designed with specific academic populations in mind, whereas the modular components of the test focus on a specific sub-population in the academic field. Lately, the modular component, according to Davies et al (1999), has been abandoned in favour of a more general proficiency test because of the difficulty in establishing equivalence of performance across domains. Reading justifiably remains a dominant skill in the IELTS, which consists of six basic elements – recognition skills (recognising, selecting and manipulating information), knowledge of vocabulary and structure (making use of lexical and structural markers), knowledge of formal discourse structure (making use of cohesive devices), background (content/world) knowledge (inferring conceptual meaning of texts), synthesis and evaluation strategies (making inferences and interpretations) and metacognitive awareness and skill monitoring (relating academic texts to diagrams).

The reading test, as Grabe (1991) suggests, assesses the candidates' ability to handle texts, comprehending texts and reproducing information in one's given field. The tests consist of skimming, scanning, making inferences, recognising paraphrase, recognising logical organisation, discerning main ideas and supporting details, recognising synonymy, and recognising grammatical patterns on the following test formats, such as, (a) matching, (b) summary cloze, (c) text interpretation, (d) multiple-

choice items, (e) fill-ins, (f) completing a chart, diagram or a table (g) labelling a diagram, (h) classification, and (f) comprehension questions.

The IELTS employs both a direct grading scheme (with some degree of subjective marking procedure) and indirect marking scheme. It uses '0' to '9' band specifications, '0' being no proficiency or failure to take the test, and '9' the highest level or native-like of proficiency.

According to Buell (1992), the IELTS test, as an integrative/ communicative test of language abilities, mirrors real life tasks that language learners may have to perform, and therefore, has a high degree of validity.

5.2.2 The North American Tests

Tests of English as a second or foreign language in the US were initially developed out of the Department of Education's 1975 guidelines for schools, which called for assessment of the English language proficiency of students with a linguistic background involving a language other than English. Several ESL proficiency tests emerged as a result, including the *Language Assessment Battery* (LAB) (Abbott 1982), which assessed receptive as well as productive skills, and *Secondary Level English Proficiency* (SLEP) (ETS 1984), which only assessed receptive skills.

For higher level education, candidates seeking enrolment in the US and Canadian universities were required to take one of the following four tests - the *Comprehensive English Language Test* (CELT) (Harris and Palmer 1986), the *Michigan English Language Assessment Battery* (MELAB) (Jenks 1987), the *American Language Institute of Georgetown university* (ALIGU) (Alderson et al 1987) test or the *Test of English as a Foreign Language* (TOEFL) (ETS 2001). Both the CELT and MELAB were based on the previously used versions of placement tests. The ALIGU test, a test of academic English was given to applicants for scholarships awarded by various US government agencies. TOEFL has been considered to be the most reliable evidence of English language proficiency by most North American universities.

About the same time, Interagency Language Roundtable (ILR) procedure, which emerged as a parallel development, has been increasingly popular. The ILR procedures, influential both in and outside the US, have been a guiding resource for the development of some of the most popular rating scales such as the ACTFL Proficiency

Guidelines (1986) and ASLPR (Ingram 1984), which I described in the previous section. A brief introduction of TOEFL and the Speaking Proficiency in English Assessment Kit (SPEAK), an oral performance test is given below:

5.2.2.1 TOEFL

The *Test of English as a Foreign Language* (TOEFL) (ETS 2001) is a standardised test of English proficiency conducted by the *Educational Testing Service*, a US testing organisation. The test is used widely by candidates seeking admission in American and Canadian universities. In addition, the test is also taken by many people to produce evidence of English language proficiency for career decisions – seeking employment in international companies or advancement in a career (Bachman, Davidson, Ryan and Choi, 1995).

The TOEFL is a prototypical 'psychometric/structuralist' language test, and consists of three sections – listening comprehension, structure and written expression, and reading comprehension, all of which are designed on a multiple-choice format (Davies et al, 1999). The format of the test allows a high degree of reliability and low cost test administration.

As complements to the general TOEFL, the *Test of Spoken English* (TSE) and the *Test of Written English* (TWE) have been developed and are administered as options. The candidates of general TOEFL may or may not take these. It depends on their requirement. The TSE is a verbal essay. The TWE is an essay writing test on a given topic.

The TOEFL is a norm-referenced test in which raw scores from each section are converted to an aggregate score, which then is used for ranking.

5.2.2.2 SPEAK

The *Speaking Proficiency in English Assessment Kit* (SPEAK) (ETS 2001), a semi-direct test of oral performance, is the institutional counterpart of the TSE. It consists of forms of the TSE and is available for institutions to use locally.

117

The format of the SPEAK is non-reciprocal in that candidates do not participate in a face-to-face interaction, rather, they listen to taped prompts, read stimulus materials and supply responses which are taped along with the prompts for analysis and ratings.

There are four aspects of the SPEAK – pronunciation, grammar, fluency and comprehensibility, which are rated against a four-point SPEAK rating scale, '0' meaning frequent phonemic error, no grammatical control, fragmented speech or incomprehensibility, and 3 representing native-like language proficiency (Bachman et al 1995).

5.2.3 The Australian Tests

The Australian system of education follows the British one and the tests of ESL or EFL are more or less similar. Institutes like the *Australian Council for Educational Research* (ACER), the *Language Testing Research Centre* (LTRC) and the *National Centre for English Language Teaching and Research* (NCELTR) have developed tests of English language proficiency for specific levels, tasks and situations. Most of these tests are either for lower level students (Years 1-9) or tertiary level students. ACER, for example, has developed *Applied Reading Test* (ART) (ACER 2000), *Tests of Reading Comprehension* (TORCH) (ACER 2000), *English Skills Assessment* (ESA) (ACER 2000) and *Diagnostic English Language Tests* (DELTA) (ACER 2000). ART is particularly designed for selection purposes. TORCH, on the other hand, is based on the cloze model. ESA and DELTA are both diagnostic tests, but only DELTA is designed for the students of NESB background studying in Australian institutes.

The *Australian Second Language Proficiency Ratings* (ASLPR) (Ingram 1984), a structured interview procedure, as I described in the previous chapter, has been designed to measure general proficiency in a language learned as a second or foreign language. Candidates are rated on a descriptive (criterion-referenced) scale in which a number, title and behavioural description specify the language proficiency in the four basic communicative skills. The scale consists of descriptions of language behaviour at nine levels - 0 to 5, five being native-like proficiency. The ASLPR provides three kinds of information about each level - (1) a description of the language behaviour appropriate to the level, 2) a series of examples of observed behaviour and (3) a series of comments explaining the key features in a given area.

The analysis of the British, North American and Australian tests points out issues I discussed in the previous chapter. Two points, however, can be singled out, namely the psychometric features used in the test to achieve objectivity and reliability, and the use of test specification to ensure validity.

5.3 Classification of Language Tests

Tests are designed to serve different purposes, and therefore, measure different aspects of language proficiency. Given below is a classification of ESL/EFL tests, their goals and some of their characteristics:

5.3.1 Goal base tests

Every test is constructed for a purpose (or a goal). Tests designed for different purposes are named differently, e.g., to place new students into groups for the purpose of teaching (placement tests), to check on learners progress in learning particular components of a course (diagnostic tests), to assess the candidates' language ability in social as well as academic situations (proficiency tests), to screen out candidates for a specific purpose (screening test), and to show the candidates' standards attained in the course for the purpose of certification, comparison or summation (achievement tests).

According to Harrison (1981), placements and proficiency tests are generic and are not based on any particular course. Both of them have future orientations i.e. reference forward to future learning. However, they differ in their purpose and also in the fact that placement tests refer to the general situation whereas proficiency tests refer to particular situations such as specific area of study. A wide variety of tests can be used as placement or proficiency tests - skill based tests, cloze, dictation, role plays, conversation, interviews, to name a few. A brief description of these tests is given in the subsequent section.

Diagnostic tests (which are also known as formative or progress tests) and achievement tests (often also termed as attainment or summative tests), on the other hand, as Cohen (1984) suggests, have backward reference to the prescribed course of study. The achievement tests, however, serve long-term objectives and cover a wide

119

range of material rather than the short-term objectives of finding out learning problems and developing a remedial programme. Achievement tests and diagnostic tests cover a wide variety of tests skill tests, grammar tests, vocabulary tests, cloze, dictation, role plays, interviews and so on. However, an achievement test has a wider range of sampling of representative items as it follows a course closely than a diagnostic test which is based on general and broad language skills (Hughes 1989:10-13).

Placement tests, proficiency tests, diagnostic tests and achievement tests serve, as I mentioned earlier, different purposes. However, the division among the test types is often blurred rather than sharp, and in practice, one test type serves several different purposes. An achievement test, for example, has a backward reference and is concerned with a past course, but a candidate, as Harrison (1991) argues, takes the test to convince someone else of his future potential, either as a student or employee. Similarly, a proficiency test, which in theory is not based one any course, may provide reference on how well the candidate prepared in order to secure a desired outcome.

5.3.2 Criterion-referenced and Norm-referenced Tests

Learning materials are divided into learnable/achievable objectives, each of which is measurable. According to McNamara (2000), a test, in which a task is securely tied up with a specific target or objective, i.e., a criterion, is called a criterion-referenced test (CRT). Success is measured in terms of the defined objectives. However, a test, in which the elicited sample of performance is judged in terms of a numerical value and by ranking it in a normal population of test takers, is a norm referenced test (NRT). In other words, in a CRT, judgment about a candidate's performance, as Khaniya (1990) suggests, is made by comparing it with a pre-specified criterion, whereas in a NRT, the performance is compared against the performance of others who sit the same test.

Whether a test is criterion-referenced or norm-referenced, therefore, depends on the way the test scores or test results are used. Criterion referencing, however, is more popular than norm referencing and is regarded as more suitable because, it tests an externally defined objective required by a syllabus or content rather than on a rank order. Also, criterion-referenced tests can be used for diagnostic purposes because 'they test all relevant behaviour, not just samples of it' (Davies 1990:18).

Criterion referencing is not as easy as it is popular. Determining a cut-off point requires a value judgement to establish how much of language learning is enough. Davies (1990) claims that,

> there is no finite inventory of learning points or items, there are very many behavioural objectives, there are variable external criteria of success, of fluency, and of intelligibility, and there is no obvious way of establishing adequate knowledge of saying how much of language is enough (Davies 1990:18).

Norm referencing has traditionally been linked to the psychometric models or discrete point approach to testing and analytical grading or interpretation of test results. Conversely, criterion referencing is linked to the integrative, global or communicative approach to testing.

In language testing, it is not the case of either CRT or NRT. While the usefulness of NRT is debatable in the language testing literature (Davies 1978; Cziko 1984; Khaniya 1990), the testers over the years have often used both at the same time. Cziko (1984) for example, suggests that it is possible and often useful to make norm-referenced interpretations of scores obtained on criterion/domain-referenced tests. Rudner (1989) agreeing with Cziko says that:

> ... by themselves, these labels [CRT and NRT] are unimportant. Many tests and testing programs properly incorporate aspects of both types of tests. For instance, many criterion-referenced tests have been normed and many norm-referenced tests permit content-based interpretations (Rudner 1989:9)

5.3.3 Discrete-point and Integrative Tests

A test that is designed to measure knowledge or performance in discrete areas of the target language is known as a discrete point test. It focuses, as I discussed in the previous section, on specific, segmentable elements, which can be learned and tested in isolation (Henning 1987; Davies et al 1999).

The terms 'integrative tests' and 'integrated tests' are often used interchangeably to refer to tests which require candidates to combine various language skills in answering test items (Davies et al 1999).

According to Oller (1979), integrative tests such as cloze and dictation operationalise the candidate's language ability, which includes linguistic as well as extra-linguistic knowledge to reconstitute the meaning of a text. That is to say, such tests integrate a number of language sub-skills, such as grammatical knowledge, the ability to extract information, lexical knowledge and the ability to construct a discourse.

Integrated tests, according to Banerjee (2000), are a sub-set of integrative tests in the sense that the latter term can be applied to any test that assesses two or more language skills. The former, however, refers to "an assessment of language ability in which two or more language skills are combined" (Banerjee 2000:291).

A cloze test, for example, tests grammatical knowledge and ability to extract and predict meaning from a text. A dictation test assesses a candidate's listening ability, ability to decode in writing and therefore their ability of writing skills and at the same time their ability in receptive phonology. A writing test, as Banerjee (2000) suggests, can test their knowledge of grammar, ability to construct a discourse and ability to manage their language resources. Similarly, a reading test can test a candidate's comprehension, ability to extract, construct, establish and guess meaning, knowledge of grammar, and lexical knowledge.

The dichotomy between discrete point test and integrative test, as discussed in the previous section, was first highlighted by Carroll (1971) who recommended the use of discrete point tests for diagnostic purposes because such tests consist of a large number of discrete items which as a group cover many aspects of language ability, and therefore, can reveal the language areas or aspects in which learners are facing difficulties. Integrative tests, on the other hand, are regarded as a measure of overall language proficiency because they 'tap a greater variety of language abilities' (Henning 1987:5). Such tests therefore have less diagnostic or remedial guidance value.

5.3.4 Subjective and Objective Tests

Differentiation between subjective and objective tests is made on the basis of the manner in which they are graded or scored. An objective test, according to Henning (1987) is objectively scored, i.e., scoring of such a test is free from the examiners' interpretations or judgement. No matter whether the answer is examined by one or

many examiners, by a human or machine, the score does not change. A subjective test, on the other hand, requires opinion-based judgements from the examiners, which, based on an individuals' perception, expertise, experience and training, may differ from one examiner to another. Multiple-choice tests are examples of objective tests, whereas free compositions are subjective tests. Scoring objective tests is easy and economical, but scoring a subjective test is expertise-extensive and time consuming (op. cit.).

Objective tests, according to Harrison (1991), are reliable because scores of such tests are free from subjective judgements. However, subjective tests are valid tests because they allow freedom for the candidates' originality, organisation of ideas, and creativity of style. The construction of subjective tests is relatively easy as there is only a small number of test items, but an objective test has greater coverage because the number of test items in an objective test is usually large.

5.3.5 Tests of Language Elements

5.3.5.1 Pronunciation Tests

Teaching and testing of pronunciation has always been an important consideration in a second or foreign language. In the tradition of structural teaching, this evolved as a broad category of production, or a prominent area in its own right requiring the language practitioners to develop discrete teaching and testing strategies. Testing of pronunciation takes such forms as modelling and drilling, phonemic discrimination, pair or group interaction on provided prompts, visual representation of sound patterns, and phonetic transcription.

In communicative language teaching, though intelligible pronunciation is viewed to be crucially important, the focus on pronunciation diverts teachers as well as learners' attention away from language forms and pragmatic competence (Seidlhofer 2000). However, the goal of teaching language is to help learners achieve meaning in the context of language use through production and perception of the target language. In communicative testing approaches, as Hughes (1989) suggests, testing pronunciation is not mentioned explicitly for the reason that validity of pronunciation as an assessment category is problematic, and that there is no linguistic criteria against which to judge the

pronunciation performance (Davies et al 1999), and therefore, tests of pronunciations, as Hughes (1989) claims, are not viewed as an effective measure of real communication.

5.3.5.2 Vocabulary Tests

There are three major concerns while testing the knowledge of vocabulary. Firstly, as vocabulary refers to the total number of words of a target language that a candidate knows, a tester often has to consider what vocabulary to test, i.e., general vocabulary or vocabulary of a particular domain. According to Hughes (1989), when a tester decides to test either general or specialised vocabulary, then s/he confronts the problem of sampling as the total number of words in any given field is large. Furthermore, s/he must decide what aspects of vocabulary should be tested, its form, meaning or use. In addition, and independent of the first two, is how the vocabulary should be tested.

Davies et al (1999) suggests that vocabulary items can be tested explicitly and in isolation following the discrete point procedure. Integrative procedures for testing vocabulary items test vocabulary knowledge needed to complete a task. This involves the ability to choose and use it precisely from appropriate registers.

5.3.5.3 Grammar tests

A grammar test is designed to measure the knowledge or control of grammatical elements of a language. Teaching and testing of grammar as the core of language ability, according to Hughes (1989), has declined considerably in the recent years to give way to the communicative language teaching and testing because the former lacks a communicative focus, an essential aspect of language ability. A grammar test, nonetheless, still popular and often an indispensable part of large-scale achievement, placement or diagnostic tests, is believed to measure ability or knowledge that underlies other language skills. It is also a popular test because it is 'relatively easy to construct and is objectively scored' (Davies et al 1999:69). However, a grammar test is an indirect test of language ability and there is little evidence to support that the knowledge of grammar facilitates the use of language in communicative settings.

5.3.6 Tests of Language Skill

5.3.6.1 Tests of Listening

Listening is vital for teaching as well as for learning. For teaching, it is a tool or a means for providing prompts and drawing learners' attention to language forms. For learning, it is a valuable source of receiving comprehensible input required for interaction and comprehension (Nunan 1999).

According to Alderson and Lynch (1988), there are two types of processes involved in listening. One is the bottom up process, which as a linear process takes a tape recorder view of listening and decodes sounds, units, and texts. The other process is termed as the top-down process, a reconstruction process, in which listeners construct meaning supported by the linguistic clues they encounter. These are shaped by the listener's prior knowledge of the context. This phenomenon, according to Nunan (1999), is also known as the schema theory. According to this theory, past experiences lead to the creation of mental frameworks that help listeners make sense of new experiences. The skills involved in the former process are sometimes termed as lower-level skills as the listener's attention is limited to the local clues available in the listening. The skills involved in the latter process are called high-level processes because these are cognitive skills and are dependent on what the listeners already know about the content and/or context of the material.

The four major components of listening, as described in Brown and Yule (1983) and Nunan (1987) are processes, materials, tasks and learners. Test developers must consider who is to listen, what type of materials are to be used, and at what stage of listening (pre-, while-, or post-listening) the candidates are required to engage in. They also need to determine what sorts of processes (top-down or bottom-up) should be performed and what types of tasks are most appropriate.

Madsen (1983) suggests that tests of listening incorporate either tests of language elements (such as grammar, vocabulary and pronunciation) or tests of listening comprehension. The difference between these two types of tests is often blurred. While the sub-skill tests of grammar, vocabulary and pronunciation focus on the linguistic elements (lower level processing), the comprehension tests incorporate

broader communication interacting with linguistic as well as non-linguistic knowledge (macro-comprehension). Listening then, as Buck (1991) suggests, is an interaction between linguistic knowledge or lower-level skills (i.e. knowledge of phonetics, lexis, syntax, and discourse), and non-linguistic knowledge (i.e. situational, content-related and general world knowledge). Thus, listening comprehension is an inferential process in which a listener constructs meaning through interaction, and his interpretation of the text is guided and influenced by the context and purpose of listening.

Buck (1991) makes a distinction between propositional meaning, which is understanding of the stated information, and inferential interpretation, a higher level cognitive skill. He suggests that a test constructor (or tester) can establish clear criteria against which to judge a candidate's propositional interpretation because most competent listeners agree on what the text says. On the other hand, however, listening is also a very individual and personal process of constructing (inferring) interpretations in the light of background knowledge, the purpose of listening and one's own assessment of the context of situation. In such a case, constructing objective tests of listening comprehension, i.e. of communicative listening ability, seems rather difficult, as there are no clear criteria to discriminate between levels of ability.

Listening tests in a second or foreign language are usually non-interactive in that listeners do not interact with the speaker or tester. Although, attempts have recently been made in the state level or national examinations to include interactive listening tests, the emphasis in such a listening test, however, remains confined to testing transactional language, e.g. language to convey information.

5.3.6.2 Tests of Speaking

There are four aspects of speaking ability – competence (linguistic as well as communicative), performance, environment and purpose. Linguistic ability is just one aspect of the total process, and is in itself insufficient to enable an individual to speak appropriately in a given situation. The process of speaking involves an interlocutor, a purpose for speaking, a topic, and a situational context. The process of testing speaking, therefore, involves (a) types of speaking ability involved, (b) tasks, topics or materials on which the speakers are required to perform and (c) the environment in which the

performance is to take place. However, creating a testing environment for speaking is a difficult task as eliciting sufficient, authentic oral response is expertise intensive, expensive and time consuming. Fulcher (1997) observes:

> Eliciting a large enough [spoken] language sample for adequate assessment is time consuming and expensive while scoring will, for the foreseeable future, depend on the use of expert human judges (Fulcher 1997:75).

Fulcher's statement outlines some of the most pertinent issues in the testing of speaking in a second or foreign language. These include eliciting a large enough sample, time needed, cost involved, and dependence on human judgement. Weir (1990), expressing a similar view, writes that while eliciting an adequate and qualitatively rich sample of spoken language is crucial, the process of doing so is often difficult because the very nature of speaking is transitional. It is crude and rudimentary on the one hand, and on the other, the tasks developed for this purpose need to be authentic, purposive, interesting and motivating in order for the test to have a positive washback effect on teaching that precedes the test.

The process of constructing an oral test, piloting, administering it, tape-recording the candidate's performance, transcribing and marking it, is time consuming and very expensive. This poses severe resource constraints on any examination board. Oral tests, many formats of which are direct tests, can meet the criteria of validity with regard to the communicative ability in real life situation. However, because marking on various aspects of oral ability, such as appropriateness, fluency, accuracy, pronunciation, stress and intonation, is subject to a high level of subjectivity, there is a high possibility for the tests of oral ability to be unreliable.

According to Davies (1995), more and more examination boards today are adapting some form of free or controlled interview. The paired interview, for example, is becoming increasingly popular because it is natural and authentic, and the situation is relaxed, and the candidates' anxiety level decreases if they can talk freely. Bachman and Palmer (1996), for example, writes that if the informality of the interview situation can be maintained, which is difficult and largely depends on the interlocutors and testers, the interview as a measure of oral ability has a high degree of face and content validity. However, in order to maintain intra-rater reliability and inter-rater reliability, a

formulised, standard procedure for eliciting language performance and training of a large number of raters is essential. Even with intensive training and set procedures for conducting interviews, it is difficult for the interviewers to create a real life communicative situation with appropriate motivation, purpose and roles. Also, if a candidate, during an interview, is paired with an unknown person, the unfamiliarity can affect their performance. Furthermore, extrovert partners can determine the conversation and receive better grades than an introvert person.

5.3.6.3 Tests of Reading and Comprehension[*]

The importance of reading in foreign language learning cannot be exaggerated. Its value in an EFL course, as Howatt (1984) suggests, has long been recognised. Similarly, Weir (1997) suggests that there has always been a significant amount of interest in comprehension exercises on texts requiring careful reading. In fact, in most EFL programmes, learning English is equivalent to learning to read, and write on the basis of it. Slow silent reading, for example, was a main feature of many British second/ foreign language tests in the 20th century, particularly in the first half, whereas the American interest was in the speed of silent reading. Such features of the British and American traditions were, and in some ways still are common in international tests such as, the IELTS and TOEFL (for example, see the previous section).

Over the years, language practitioners' approaches to reading have changed from traditional, unidirectional to multi-dimensional. Traditionally, reading was regarded as giving information from the text to readers. Teachers of a second or foreign language, according to Ushiro (2000), conducted their classes in such a way that engaged learners in what is known as the bottom-up process to help them with the linguistic knowledge and literal understanding of the text. However, according to schema theory (McNamara 2000), a theory of the organisation of background knowledge and of its use in comprehension, readers' background knowledge, their motivation, and purpose influence the way they read and the strategy they develop to comprehend a text. Therefore, the ESL/EFL practitioners have acknowledged the role of the readers' schema in reading, and have recommended both types of processes: top-

---

[*] A version of this section appeared in Young Voices in ELT (2003) Vol. 3, No. 1. Pp. 5-12

down and bottom-up, for effective reading. Reliance on the bottom up process alone may help develop lower level skills but not overall structure of the text. Similarly, with over-reliance on the top-down processes, there is a danger of being too subjective and of making idiosyncratic predictions. Gray (1960 quoted in Alderson 1989:12), for example, suggests three types of reading, 'reading the lines', 'reading between the lines' and 'reading beyond the lines'. 'Reading the lines' is a bottom up process whereas 'reading beyond the lines' is purely a top-down process, but 'reading between the lines' is a combination of the two processes.

The process of fluent reading is rapid, purposeful, interactive and flexible. This means that a reader uses a range of strategies to interact with the text at a reasonable speed and purpose in mind. The skills and knowledge involved in reading may be divided into nine groups - knowledge of lexis, knowledge of syntax, knowledge of content, knowledge of concept, knowledge of literary devices, knowledge of linguistic devices, knowledge of discourse, assessment skills (ability to appreciate, critique and argue, and to take one's own argumentative position), and information transfer skills (ability to map the mains points, take notes, summarise, transfer information into non-text forms to use reference skills) (Grabe 1991).

The three most popular principled methods for testing reading comprehension are multiple choice questions, short answer questions, and cloze. A brief description of these techniques is given below:

The Multiple - Choice Tests

Multiple-Choice (MC) testing is regarded as a procedure, which is used as a testing technique for testing listening comprehension, reading comprehension, sentence structure and vocabulary. This technique, as the main feature of the discrete point format, is perhaps the most widely used technique in the world for testing a second or foreign language. According to Lado (1961), an MC test usually consists of a stem and a number of options consisting of distractors and a correct choice. A candidate is required to listen or to read the text or the stem and the distracters and select the answer of his or her choice. In the case of listening comprehension and reading comprehension, the stem follows the text, which may be in the form of a narrative, interactive or

expository piece. The form of the stem itself may be declarative sentence with a blank or an interrogative sentence. In the former case, the choice made of options completes the sentence, whereas in the latter's case, the choice answers the question. For a structure and vocabulary, the stem is usually a statement with a blank, though a question form is also often used.

Commenting on the test, Bachman (1989) writes that the MC items are context - reduced, non-reciprocal and the ones in which there is no potential for feedback. Selecting the contents and setting items as well as making decision about the correct answer are subjective, however, the marking process is objective and mechanical. Thus, there is a complete inter-rater, intra-rater reliability, as marking is free from idiosyncrasies. There is a validity problem too. Weir (1993) found that there was a low correlation between MC reading test and other measures of reading.

Short-Answer Questions

Short answer questions (SAQs) are extremely popular with the ESL/EFL teachers almost everywhere because of ease of construction, brevity in answering and the wider range of purposes they serve. Questions are called short answer question because answers to such questions are specific and brief. A candidate is required to answer a number of questions based on reading or listening. Such questions, as Weir (1990) suggests, are useful tools to check a candidate's comprehension and ability to perform a task such as inference, recognition of ideas, and they provide reliable data about his or her ability in the target language.

Recently, the SAQ format has gained currency. Weir (1990), for example writes that the format now incorporates mind mapping of main ideas and support details a reader is expected to extract from a text. Some SAQs also include introspection analysis of what is being tested to ensure that the test fulfils its purpose. In addition SAQs require the writing of answers in one's own words, along with the requirement of answering questions in a non-text form such as labelling a diagram, completing a chart, numbering a sequence of events, paraphrasing, and translation into the first language.

The Cloze Test

As one of the most common integrative tests, the cloze test has become increasingly popular with ESL/EFL teachers in recent years. The cloze procedure involves the deletion of every 'n'th (usually every fifth or seventh) word from a passage. The reduced redundancy that results from this process forces a candidate to rely upon his or her knowledge of underlying linguistic rules and also to maintain the meaning of the passage to fill in the blanks. In order to do so, the test-takers not only have to comprehend the text well, but they also have to find the items that are syntactically and contextually appropriate. The test is, thus, an overall index of the test-taker's ability to process their syntactic and pragmatic knowledge, the latter being a necessary quality for making it a pragmatic test.

In 1990s, the use of the cloze technique entered an innovative phase in which teachers of English put the technique to various practical applications. The cloze was used in listening, (Viswat and Jackson 1994) and conversation (Jonz 1990; Brett 1997). Various studies on the interpretations of the cloze test scores were also conducted (Abrahan and Chapelle 1992; Gibson 1997). Swaffer, Arens and Byrnes (1991), for example, suggest that the cloze procedure is a product-oriented test and is more text-based than any discrete point tests. Oller (1992) reaffirms the properties of the cloze by saying that being a pragmatic testing technique the cloze reflects the theories behind the more recent concepts such as text-linguistics discourse analysis and pragmatics. McNamara (2000) suggests that a cloze test is an appropriate substitute for a test of productive skills because it requires readers to integrate grammatical, lexical, contextual and pragmatic knowledge in order to be able to perform on the cloze, such tests are easy to construct and relatively easy to score.

The cloze technique was also used in various types of tests such as diagnostic (Fotos 1991) aptitude and prognostic (Aweiss 1993, Storey 1997). But the most important development of the cloze technique was its use in educational technology for both-teaching and testing purposes (Grant 1991; Katchen 1994; El Koumy 1997; Coniam 1997 and 1998).

In an EFL context, the cloze test has been used as a test of EFL proficiency in Southeast Asia (Brown et al. 1991); Japan (Brown 1995); China (Li 1996); South Korea

(Lee 1996) and the Middle East (El-Koumy 1997) all of which established the cloze as a measure of overall language proficiency.

The cloze procedure has been praised for its efficiency and reliability, but at the same time, concerns have been expressed over its validity as a device for testing the global comprehension of the text.

According to Alderson and Lukmani (1981), it is important to test higher skills of comprehension as well as lower level skills. They adapted the six components of the famous Bloomian (Bloom 1972) taxonomy of knowledge, comprehension, application, analysis, synthesis and evaluation simultaneously. They then tested reading at Bombay University, India; and Lancaster University, England, by breaking it into eight skills, namely, recognition, identification, discrimination, analysis, interpretation, inference, synthesis, and evaluation. They maintained that the first three skills are lower level skills. The remaining five are higher level skills, which measure cognitive skills, logic, and reasoning ability alongside language skills. They confirmed earlier findings that lower level skills and higher-level skills correlate with each other, and that the lower level skills facilitated the mastery of higher-level skills. The implication of this research for teachers and testers is that these two types of skills should be viewed as complementary to each other rather than supplementary in that they are integral elements of the same thing and mastery of one type of skills facilitates the mastery of another type of skill. More recently, Weir (1993), who also recommends testing of both levels of reading skills, presented a four-level breakdown of reading for testing purposes – two of which are recommended for global comprehension (high level reading), and two for local comprehension (lower level skills).

5.3.6.4 Tests of Writing

Writing is a complex mental process intimately related with critical thinking. Writing in a second or foreign language is essentially different from writing in one's first language (L1). The second or foreign language writers, unlike the L1 writers, have to worry about the language and mechanics alongside the fluency and accuracy of the contents of their writing. Therefore, it provides indicators of candidates' understanding of the purpose of writing, content, fluency, audience, organisation, accuracy and

vocabulary. While second or foreign language testers need to maintain a balance between fluency and accuracy during assessment, they often encounter another problem - that of assessing contents that are expressed in the language versus the language that is used to express the content. In other words, the testers, as Hamp-Lyons (1991) suggests, often have to deal with the problem of allotting appropriate or reasonable weightage between accuracy and fluency, and between content and language. They often face candidates who have fluency without accuracy and those with accuracy but little fluency. Similarly, they have to grade writing that shows very strong content while grammatical or textual competence lag far behind or vice versa.

There are three levels of writing exercises used in secondary level classes in an ESL/EFL context. They are controlled writing in which control is exercised through an imposing structural and lexical limitations, guided writing for which guidelines in the form of points to be covered or paragraph outlines to be developed are provided, and free composition in which the writers are provided with a context or topic, and are free to use their own ideas, organisation and style, and develop their own argument.

Writing involves many processes such as (a) accumulating information/data through a literature review, brainstorming, discussion, observation, and research. The writer then has the task of organising them, creating an outline for writing; drafting, receiving feedback, and revising and repeatedly rewriting drafts. For assessing writing, this means that a timed, single-session essay test or a single writing sample alone is not a valid method for evaluating writing performance. Writing tests should reflect the conditions and processes through which the final product of writing is achieved.

So far as approaches to assessment of writing are concerned, Weir (1993) provides a detailed framework for assessing writing. In his framework, he outlines that any approach to assessing writing should be based on the considerations of (1) operations, which consider what skills and tasks are in question, and whether the text type required is interactional or informational (academic), (2) performance conditions which take into account, the purposes of writing, demands on the writer (organisational, propositional, illocutionary), topic, channel of presentation (visuals, layout, typeface) addressee, setting (where it is written), time available, size of output, amount of support given (e.g. dictionary, calculator), or response mode, (3) formats for written production, whether they are indirect (gap filling, form completion, editing, responding to given

information) or direct (open-ended essay test, integrated tasks), and (4) quality of assessment criteria such as relevance and adequacy of content, organisation, cohesion, adequacy of vocabulary for specified purpose) grammar, mechanics, and appropriateness of language (such as register, and style).

Approaches to writing assessment are either global (or holistic), or analytical. In the global approach, a rater gives a single score on a piece of writing on the basis of their total impression of the writing as a whole. This 'impressionistic' method of marking suffers from considerable discrepancies between marks awarded due to the differences in the marking standards employed by various raters (Hamp-Lyons 1991; Weir 1990 and 1993). Cooper and Odell (1997) criticise the holistic approach by saying that it is unsuitable for informed decisions about composition. It is not an overall indicator of writing and that this method distorts the raters' ability to make a sound judgement. However, in spite of the criticism there are merits of this approach also. According to Spandel (1997), fallible as it may be, the holistic approach to assessing writing can be reliable and useful if it is followed up by an in-depth individual assessment such as interview and/or review of a portfolio to confirm results.

The analytical approach to assessing writing "refers to a method whereby each separate criterion in the mark scheme is awarded a separate mark and the final mark is a composite of these individual estimates" (Weir 1990:69). Though this approach is blamed for not assessing any writing ability because assessment is done on language segments or components when the writers write continuous prose; this is, as Spandel (1997) suggests, the most preferred form of assessment for the end of a course, as well as for external examinations because analytical assessment facilitates the writing process through correction, feedback, and diagnosis (see also Section 4.4.4 in Chapter 4).

One other form of writing assessment, which is becoming increasingly popular, is the portfolio assessment. In this form of assessment several pieces written over a given course of study are evaluated. The portfolio brings together several pieces of writing collected at intervals over time, which may include early drafts. This method of assessment, according to Brand (1992), emphasises assessing writing as a process, not a product, and provides a more naturalistic approach to teaching and assessment. The opportunities for multiple drafting, self reflection and receiving and responding to

feedback implied by the portfolio, presents the reality of the writing ability of a candidate. Questions are often raised about the reliability and validity of readers' responses in portfolio assessment. However, as Hamp-Lyons (1991) suggests, incorporation of portfolio materials, as well as impromptu writings for assessment can, to some extent, minimise the problem of reliability and validity.

Summary: Classification of Language Tests

The table below summarises the tests discussed in this chapter and includes a few other. The table also shows what they focus on, the test format, type of language proficiency they measure, or the way they interpret the test outcomes.

Basis	Test Type	Characterising Features
Goal	Proficiency Achievement Diagnostic Prognostic Screening Placement	- general, for a specific purpose, forward reference - general, backward reference to a course, for certification, promotion - backward reference to a course, for motivation and remedial work. - a form of aptitude test, general, for predictive purposes - general, screen out candidates for a specific/remedial purpose - general, for grouping or instructional purpose
Medium	Oral Written	- oral elicitation/response format - written elicitation/response format
Mode of scoring	Objective Subjective	- responses examined with an established set of responses, MC - free composition may be objectified in scoring through the use of rating scales
Skills	Listening Speaking Reading Writing Communication/interaction	- receptive ability via phonic signals, crucial for input for language work and comprehension - productive ability, using language through vocal sounds to perform a language task - receptive ability via graphic symbols - productive ability via graphic symbols - ability to use language in all its manifestations in real-life, communicative situations
Elements	Pronunciation Vocabulary	- abilities in receiving and comprehending spoken tests - aspects of words, general or specialised vocabulary

136

Category	Type	Description
Grammar		- knowledge and control of grammatical knowledge of language.
Approach to Design	Discrete point	- a verity of diagnostic test, measure of proficiency, or knowledge in restricted area, total score reflects overall language proficiency, has remedial or guidance value
	Integrative	- taps variety of language ability concurrently, e.g., cloze, dictation
Replicability in real life situation	Direct	- replicable in real/uncontrived communicative situations, valid measure of language proficiency, interview, group interaction, conceptualised vocabulary test
	Indirect	-not replicable in real situation, MC, cloze, synonym matching
Reference	Criterion - referenced	- target behaviours, criteria set helpful in clarifying objectives in advance
	Domain - referenced	- target language domains prescribed contents = clarify contents
	Objective - referenced	- items are selected to match objectives without reference to domain or target behaviour
	Norm- referenced	- acceptable standards pre determined by administering the test to sample population
	Standardised	- large scale, same as norm fairer and less arbitrary norm changes from group to group
Speed	Speed	- easy items, speed of performance under time pressure
	Power	- difficult items, sufficient time allowed focus on knowledge rather than speed

Figure 11: Classification of tests (based on ideas from Cohen 1980, Harrison 1984, Hughes 1989, Heaton 1988)

137

Summary and Implications

Language tests in the past tended to measure memorised rules and words, and their application in writing exercises, without regard to their language context. And as marking in such tests depended on the sole discretion of individual testers, such tests were neither valid nor reliable because 'knowledge of rules and memorisation of words bear no relation to being able to speak [or use] the language' in social situations (Lado, 1957:4). Lado's work, influenced by structural linguistics and reinforced by psychometrics, facilitated the emergence of discrete-point testing procedure, which viewed language knowledge as amalgamation of discrete systems. This procedure tended to decontextualise language knowledge and test language features in isolation (McNamara 2000). Though such tests had a high degree of reliability, and it was easy to find out the item difficulty level and item discrimination index through the process of item analysis, the tests themselves were 'stand alone' tests and seriously lacked validity because they did not reveal the true level of a candidate's language proficiency required in specific disciplines. That is to say, the test tasks were not representative of the skills or abilities the candidates required in their academic field. Attempts have been made in recent years to rectify the deficiency by incorporating integrative tasks, and moving from general proficiency to specialised proficiency testing. The IELTS, for example, emphasises validity at the expense of reliability. The TOEFL for the purpose of broadening its appeal, is under review, and is moving from being too specialised to a more general proficiency test.

The foregoing examination of some of the ESL/EFL tests demonstrates that in general, British and (North) American EFL proficiency tests reflect different approaches to language test development. The North American tradition, for example, is heavily based on the psychometric properties of tests. Issues such as reliability and concurrent and predictive validity are of particular interest in this tradition. Hence objectivity of scoring and generalisibility of results play a dominant role in the development of test methods. Multiple-choice items are often used in testing receptive skills to gain the desired internal consistency even if the test is expected to measure communicative competence. Moreover, this tradition highly recommends the use of trained scorers and

detailed specific instructions in conducting interviews for testing productive skills in order to achieve high inter-rate reliability.

The British tradition, on the other hand, puts emphasis on the specification of test content and expert judgement. While reliability receives less attention in this tradition, content and face validity are the major concerns of the test designers. The British tests, therefore, enjoy more variability in their formats and include various communicative activities.

Thus, most international tests of English as a foreign language are making compromises between (a) general language proficiency and specialised language proficiency, (b) between objective testing and integrative testing, and (c) between reliability and validity.

The shift towards the communicative approach to language testing in the last decade has encouraged testers to make language tests more integrated and less discrete. The move has prompted them to use more authentic test tasks in terms of content, purpose and response, to turn their attention to validity, and to develop and use direct tests (McNamara 1996). The Australian research institutes such as NLLIA, LTRC, ACER, NCELTR and AEC, for example, have developed authentic or direct ESL test instruments and ESL scales following Bachman's (1990) communicative competence model of language testing, which many examination boards use as an integrative, purposive, authentic and valid way of testing ESL in Australia.

Most English language tests I described in this chapter are of a general nature or they are context- or course specific. Most international tests of the English language, for example, are generic in type and do not consider the candidates' background, nature and amount of input and the context in which learners learn the language. Disregarding these may, as Skehan (1989) suggests, make a test impractical and 'may not give a true picture of the ability of the examinee' (Khaniya 1990:122). The international ESL/EFL tests, in this sense, are all constructed towards a particular purpose and they would not measure what a local test like the SLC does. They, therefore, do not serve the specific purpose of English language teaching and testing in Nepal directly.

In a context like Nepal, where there is a move towards communicative language teaching, but at the same time, an attitude bias towards reading and writing as dominant language skills, international tests of ESL and EFL and other 'stand alone' tests cannot

be adopted for practical as well as resource and curricular reasons. A compromise, therefore, must be found between test content, test method, and authenticity of response. In a test setting, testers often attempt to create conditions, under which the candidates' interactions with the test tasks or contents are produced in reality. In other words, a 'principled compromise' (McNamara 2000:27) is essential in order for a testing approach to be suitable under the prevailing conditions of a given educational context.

From the discussions in this and previous chapters, the implication I draw upon is that the English test of the SLC has to be a holistic test and a synthesis of all four types of competence, namely grammatical (linguistic) competence, communicative competence, strategic competence and discourse competence. This in turn implies abandoning the existing test and adapting communicative testing on an integrative test structure, which should be based on an elaborated specification of the target goals and use. The theoretical underpinnings and desired test features for the new SLC English test along with rationale for their inclusion in the new test are summarised in the table below:

Theoretical Issues / desired test features	Rationale
Holistic	The design of the new SLC English test should provide a picture of a candidate's overall English language proficiency, not of proficiency of different segments of the language.
Integrative Test Structure	The structure of the test should be integrative in that tests of different types of competence and different language skills are tied up together by a common theme, language point or context
Communicative Competence plus other three types of competence	The test should be an opportunity to synthesise all four types of competence with that of communicative context to produce language that is syntactically acceptable and at the same time contextually appropriate.
Competence vs. performance	The SLC test should not just reveal what language ability a candidate possesses, but it should also show if they can accomplish a practical (academic, occupational and/or professional) task using the language.
Criterion-referenced test with norm-referenced interpretation	The new SLC test should be basically a criterion-referenced test in the sense that a criterion as a standard or bench mark of learning as described in the test specification is used to rate test performance. The test scores arrived in this way are then norm-referenced to report the test outcomes.
Subjective vs. objective	Objectives tests are reliable so far as marking is concerned. Subjective tests, on the other hand, are valid as they assess a candidate's ability to use the language creatively. In the new test, therefore, with a view to ensuring reliability and validity, a proportional balance between the two should be maintained.
Validity, reliability and practicability	Validity, reliability and practicability are essential characteristics of a test. The new test should be valid (i.e., it should assess what it intends to assess), reliable (be consistence in producing results) and practicable (implementable in the context for which it is developed.) In Chapter 9, I

	describe what steps I undertook to ensure validity, reliability and practicability of the new test.
Cloze procedures	Cloze tests reveal an overall index of a candidate's ability to process linguistic, textual and pragmatic knowledge. Inclusion of cloze procedures in the new test will increase its validity and reliability.
Planning Component in writing test	Introduction of a planning component in a writing test will reveal a candidates' ability to gather, organise and create an outline before writing a draft. It may also reveal whether a candidate generates writing from his learning or s/he reproduces a memorised answer.
Forward reference vs. backward reference	As a proficiency test, the new test should have forward reference and include elements or skills required at a higher level of study or occupational settings. Similarly, as an achievement test, it should have backward reference to prescribed test materials
Curriculum based	The new test should be based on the SLC curriculum. That is to say, it should be developed by specifying the abilities, skills and sub-skills aimed by the curriculum.
Test specification	A test specification will provide details of the target language abilities and skills which are under test. It will also specify test objectives and provide information about test types, test tasks and marking and scoring schemes. It will ensure that the new test is relevant to the aim and is appropriate in form and function.
Rating scales	In order to make the rating of test performance transparent and to establish a common point of reference, a rating scale is a useful device. Its use will ensure reliability in marking.

Figure 12: Theoretical issues and desired test features

In the following chapter, I describe the creation of the new test package and how these features are incorporated in it.

CHAPTER 6

Action for the Change

Adapting the Old to Create the New

There are different models of language proficiency each of which supersedes the other by adding new elements or by taking away some elements or even by recombining different elements of different models together in different ways.

- Elder 1997, MMW Video Series 1

6.1 The Challenge

In the foregoing chapters of the thesis, I have given a detailed description of the context of SLC English language test, and some limitations and constraints in developing and implementing a new test. I have also outlined theoretical underpinnings that should be taken into consideration in order to be able to develop a useful test for a developing context like Nepal. Given below once again, then, are the theoretical underpinnings (see also Figure 12, page 140) that needed to be addressed when designing a new test and the sections of the thesis in which they have been discussed:

- It should be a test of communicative competence as well as of linguistic competence (section 4.3.3.3; 4.4.5);

- It should provide a holistic view of the actual language proficiency of a candidate rather than segmental scores of a candidate's achievement in different skills and elements (sections 4.3.3.1; 4.3.3.2; 4.4.3);

- The structure of the test should be integrative, i.e., tests of different skills and elements should be integrated together through content, thematisation or context, so that the tests are not 'stand alone' tests (sections 4.3.3.2 and 4.4.2);

- The test should be criterion-referenced with a provision for norm-reference interpretations (section 5.3.2);

- A balance between validity, reliability and practicality should be maintained (sections 4.4.1; 4.4.2 and 4.4.5);

- There should be a balance between subjective and objective tests (section 5.3.4);

- All of the elements of a curriculum should be treated adequately and reflect the requirement of the curriculum (sections 3.3; 4.4.6); and

- There should be a provision for a use of rating scales (section 4.4.4).

The challenge I faced was to develop a testing package, which not only addresses these theoretical issues, but also is practical in the context for which it was intended. The diagram below summarises how these theoretical factors (as discussed in Figure 12 and above) influenced the design of the new test package:

```
┌─────────────────┐      ┌─────────────────┐      ┌─────────────────┐
│  Communicative  │      │Holistic approach│      │   Integrative   │
│   competence    │      │  to language    │      │    structure    │
│                 │      │                 │      │                 │
└─────────────────┘      └─────────────────┘      └─────────────────┘
```

┌─────────────────┐   ┌──────────────────┐   ┌─────────────────┐
│   │   │ New SLC Test   │   │   │
│   │   │   Package: │   │ Criterion referencing│
│ Competence &   │ ───▶ │   │ ◀─── │ plus norm referenced │
│ performance   │   │ 1. Specification │   │ interpretation │
│   │   │ 2. Design   │   │   │
│   │   │ 3. Sample Test   │   │   │
│   │   │ 4. Marking,   │   │   │
│   │   │   Reporting   │   │   │
│   │   │   Scheme   │   │   │
│   │   │ 5. A Process for Test│ │   │
│   │   │   Development   │   │   │
└─────────────────┘   └──────────────────┘   └─────────────────┘

┌─────────────────┐   ┌──────────────────┐   ┌─────────────────┐
│ Subjective and │   │Validity, reliability,│ │Reflection of curriculum│
│ objective tests│   │   practicality   │   │   requirements │
│   │   │   │   │   │
└─────────────────┘   └──────────────────┘   └─────────────────┘

Figure 13: Considerations in designing a new English test for the SLC.

Figure 13 above depicts the theoretical considerations that I needed to bear in mind when developing a test specification for the English test of the SLC, designing a test structure, writing test tasks (sample test), and in formulating a marking and reporting scheme. I did this in stages. In the following section, then, I describe the creation of the test specification as the first stage. Then, in the second stage, I designed a test structure. In the third stage, I wrote sample test tasks, and in the fourth stage I developed a test outcome marking and reporting scheme. In the final section of this chapter I describe the overall process of developing the new test package.

The present chapter, therefore, can be taken as the action phase of the first cycle of the project.

6.2 The Test Specification

A test specification is a document which specifies every bit of information on a prescribed course that a test developer or examiner or a teacher needs to have in order to be able to develop an appropriate and valid test. It consists of information on the target language use or outcome, test purpose(s), materials to be used in tests, the test input (the test itself), expected response, marking or scoring scheme and any additional information that may be useful.

In Nepal, like in other developing countries, English courses are taught with a view to achieving specific academic and communicative goals. That is to say, the courses are designed to meet a particular set of objectives that are specific to the context of Nepal. The test specification I developed is based on an analysis of the new ELT curriculum, prescribed texts and teachers' guides. I broke the target macro-skills into micro-skills, and their sub-skills and specified the aspects and abilities involved in them. I also set target outcomes (standards or bench marks) for each of the skills and domains in which these skills and abilities could be assessed. Given below is the description of the new test specification, its components and the processes I employed to develop it. I followed the Bachman and Palmer's (1996) multi-trait, multi-method model, though, with a view to keeping the specification simple for the target teachers' population, I did not use terms such as pragmatic and discourse competence in the new

test specification because these technical terms may confuse the ordinary teachers. In the section about how I developed the Pilot Test (section 6.4), I show which tests of the pilot test assess which of the four types of competence.

6.2.1 The Format of the Test Specification

Different authors recommend different formats for a test specification. Donaldson and Lynch (2002), who adapted their model from Popham (1978), for example, propose a five-component model, which consists of general description, prompt attributes, response attributes, sample items, and specification supplement. Bachman and Palmer (1996) use a different set of terminology to name the components of a test specification. Their terminology roughly corresponds to what the five components mentioned above indicate. However, there is an important addition of 'scoring method' (Bachman and Palmer 1996:51-52). Alderson, Clapham and Wall (1995) suggest yet another set of terminology which again overlaps what the other two formats generally cover, but add a component of 'Test Battery' (Alderson, Clapham and Wall 1995:14). This component includes information about the target language use domains, sub-skills and the aspects of the sub-skills to be tested. On the basis of the ideas from these writers, I developed a seven-point test as set out below:

Target Language Use/outcomes	General Description	Prompts (stimulus)	Response Characteristics	Sample Tasks	Scoring scheme	Supplementary Information

Figure 14: Test Specification Format (based on the ideas from Donaldson and Lynch 2002, Bachman and Palmer 1996, Alderson, Clapham and Hall 1995) (see Appendix A)

The test specification which I developed based on the model is in Appendix A. I will now describe each of the seven points (or contents) of the new test specification:

6.2.2 The Contents of the Test Specification

6.2.2.1 Target Language Use/Outcomes

In this section of test specification, I described target skills and sub-skills. After an analysis of the English courses (Years 9 and 10) and related documents, I:

- prepared an elaborated list of target skills, their sub-skills along with their aspects; and,
- identified domains in which these abilities and skills were to be tested (for complete lists of skills, aspects and domains, see Test Specification, Appendix A).

The secondary English curriculum does not provide an exhaustive list of domains. The domains explicitly mentioned in the document are home and family, school and education, work and employment, and games and sports. However, an analysis of the prescribed texts revealed that additional repertoire of social domains including everyday life, entertainment (TV, movie and games), public places (e.g., restaurant, shopping centres, bus stops), media, fitness, friendship and transport were required. The documents also implied social domains and everyday communication domains for listening and speaking, and more formal and academic domains for reading and writing.

6.2.2.2 General Description

This section contains:

- the purposes of testing of the concerned abilities and skills and sub-skills to be tested,

- the test types,

- the sources of materials

- the general description of the context in which the test should be carried out, and

- the inferences of test scores to be made to reflect the purposes of a test.

It recommends that

- the materials to be used for developing tests could be derived from a number of different sources including newspapers radio news casts, TV news casts, journals, information leaflets, brochures, bills and menus,

- the difficulty level of such materials should be the same as that of the prescribed text studied by the students,

- the difficulty level of the materials can be determined by carrying out an analysis of the selected texts on the criteria of lexical complexity, index of syntactic complexity, discourse variety, and pragmatics involved.

6.2.2.3 Prompts

Prompts are descriptions of characteristics of input or stimulus that are presented to candidates.

Based on the analysis of the prescribed texts and exercises, I:

- provided a description of instructions, test types, test tasks, number of items and arrangement of these items in a test paper,

- recommended as many test skills as there are skills (sub-skills) under test. For example, I recommended eight types of task for listening because there are eight listening skills or sub-skills under-test;

- determined the number of items in an individual test on the basis of their nature. For example, a task based on explicitly stated information in a

text has more test items than the one which is based on implicit information; and

- estimated time required for each task against number of items in the task, and the nature of response needed.

6.2.2.4 Response Characteristic

This section consists of characteristics of expected response to a test task. These include:

- a description of the ways in which candidates would interact and respond to tasks (prompt or input) (or a description of what the candidates do during a test), and
- details of expected style of interview (reciprocal or non-reciprocal); format of writing, language style, and register, and so on.

6.2.2.5 Sample Tasks

Sample tasks comprised

- illustrative examples of test items that the test specification should generate, and
- a description of the context of the task, the task itself, alternatives if given, task layout and the actual instruction.

6.2.2.6 Scoring Scheme

This section provides:

- description of how candidates' responses should be assessed or evaluated,
- description of the criteria to be used during rating, time to be allotted and marks to be awarded for each criterion,
- description of the procedures to be used for arriving at a score.

- allotment of marks and a scheme for scoring on different task type,
- allotment of half a point to each task in an eight-item test, 1 point for each task in a four-item test, 2 points each for items that required understanding of the implicit meaning of the text,
- schemes for scoring different task types, and,
- time to complete a task (from ½ a minute to 2 minutes each depending on the level of complexity of a test item).

6.2.2.7 Supplementary Information

The section of supplementary information contains further information for test developers, test implementers and evaluators (examiners) of test responses. These include:

- information needed for the construction of a test,
- any further references that might be useful for the test developers,
- description of what material and human resources are involved in the implementation of the test of a particular skill,
- the number of audio cassette players and trained invigilators that would be needed during the implementation of the test,
- for the speaking test, advice on how assessors and interlocutors of the test could make candidates feel comfortable and how they could maintain informality in the situation in order to get authentic responses, and,
- the need for some form of training for testers and raters.

6.3 The Test Structure

The next stage was to integrate the considerations, the specification and subject materials in order to develop an integrative test design and individual test tasks. As the figure below shows, the tests of different skills were integrated in a single set of materials. Similarly, the test tasks and skills they assessed were the ones specified in the test specifications.

6.3.1 Test Structure 1: The Process of Producing Integrative Tests
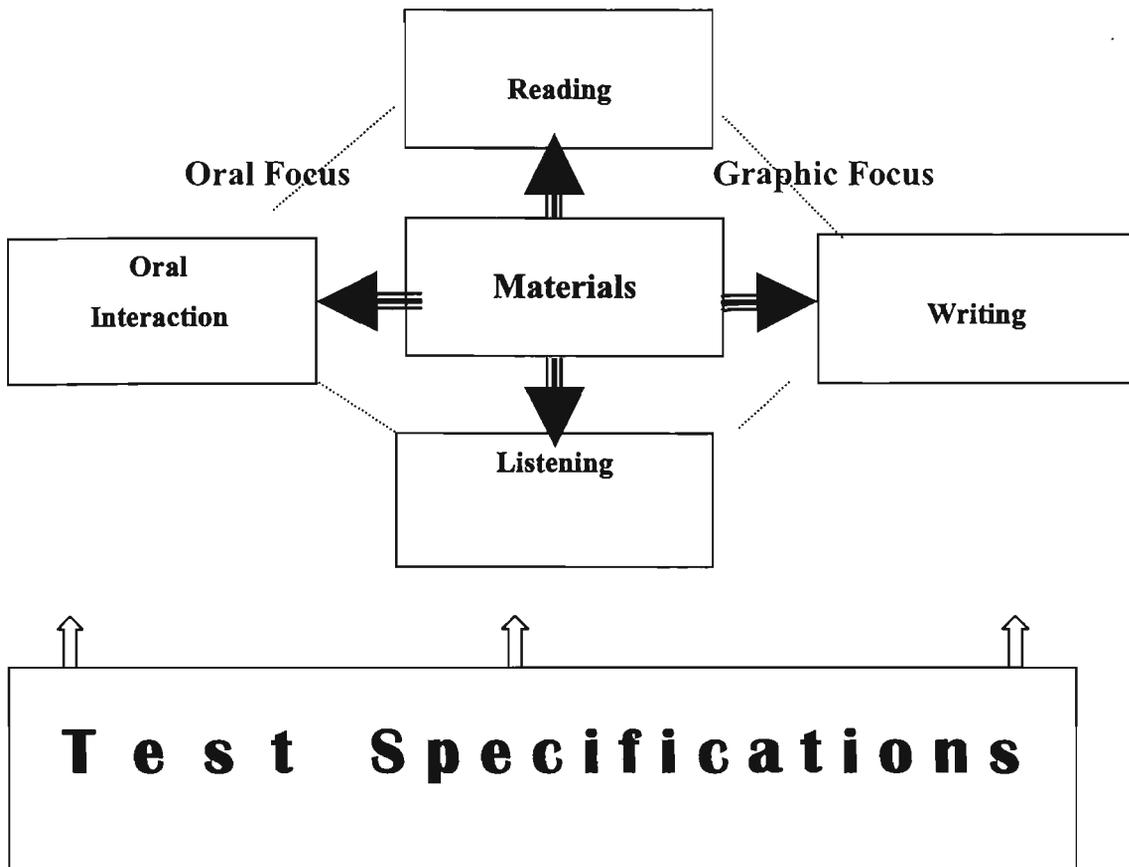
Figure 15: Integrative Test Structure (adapted from Carroll and Hall 1985)

In the structure above, integration was achieved through a common context presented either by thematisation (in this case, "an embarrassing situation"), contextualisation or a language feature. A common theme or context was included in all or most tests and test tasks. For example, a single set of materials, as the figure above demonstrates, became the basis for the test of all language skills or elements. Listening or speaking on or about the theme of the material would become a prelude for reading, and writing tests, and the writing test comes as a follow up of the reading test.

In an ideal situation, communicative as well as academic abilities in all four major skills of language should be tested. It needs, however, adequate provision of resources, equipment and facilities, and sufficient preparation in terms of training and materials. In a situation in which testing is severely constrained by limited resources, I

have suggested that an alternative and more realistic structure such as the following be considered:

6.3.2 Test Structure 2: Alternate Integrative Test
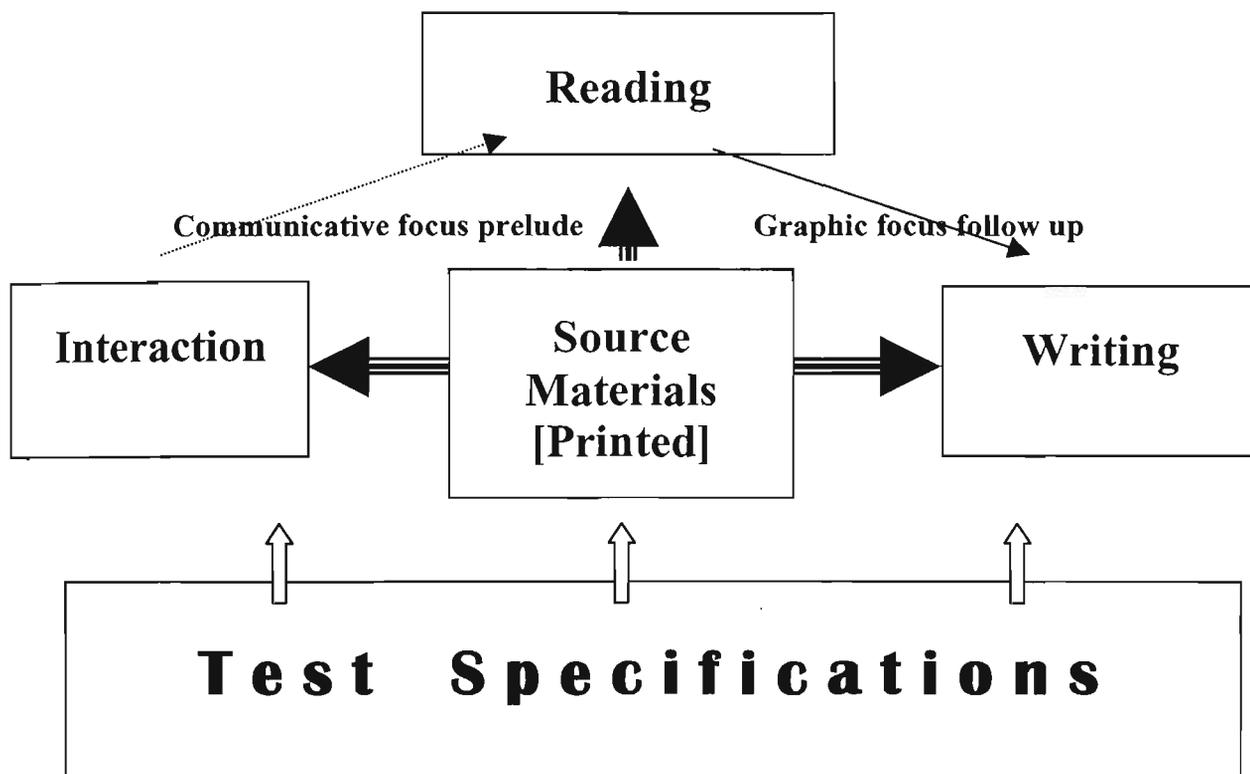
Figure 16: AlternativeTtest Design of the Integrative Test. (adapted from Carroll and Hall, 1985: 8)

In the alternative test structure, as the diagram above shows, I replaced listening and oral tests with tests of interaction skills. I recommended this structure for a period of five years during which the concerned authority should develop the necessary resources (both human and material) and infrastructure for testing listening and speaking. This alternative test structure was based on the following assumptions:

- The SLC had, as an achievement test a backward reference (i.e., to the course), but at the same time, as a proficiency test, it had a forward reference, i.e., to the kind of ability and skills candidates should be required to demonstrate in higher education and employment settings. The selection of materials and test types, therefore, was based on the prescribed texts as well as on the intended abilities and skills.

- A test should match the curriculum and not vice-versa.

- There were expertise and resource constraints for conducting oral tests.

- There was a lack of adequate resources and equipment to administer oral tests.

- The SLC test of English was independent of any school based assessment.

- Incorporation of proficiency based creative tasks required authentic communicative language and situations.

- Adding a cloze test into the design would improve predictive validity of students' proficiency.

6.4 The Pilot Test

The next stage was to develop the sample or pilot test which I did by compiling the samples of tasks given in the test specifications (see Appendix B). In the sample test, I included some seen materials and familiar test types with a view to giving a backward reference to the course. However, I designed new tasks using these materials in line with the course objectives, but at the same time ensured that while responding to the tasks, the candidates make use of their understanding of the text and of their own language to perform the tasks rather than produce answers from memorisation and rote learning.

To ensure that the test has a forward reference, I included unseen materials and tasks, which should provide some indications of what abilities in the target language a candidate possessed and whether or not s/he met language requirements for a higher course. During the construction of the sample test, I ensured:

- that the objectives of the course were covered and the target skills had adequate representation in the test;

- that the test had adequate sampling of the skills and abilities expected from a tertiary level entrant. For this purpose, I carried out an analysis of the 10+2

and PCL English courses to determine the target skills, sub-skills and abilities (see Appendix I);

- that the test had positive washback effects. It was, for example, ensured that the test is communicative in nature and that the marking scheme includes the features which the teachers should take up during their classroom practices.

In principle, I followed the multi-trait, multi-method model (Bachman 1990) in developing the test tasks in which I covered the four types of competencies the model proposes. Test tasks A, B, E, F, H, I, J, L and N of the pilot test aimed at linguistic competence in the sense that they measured phonological, lexical and syntactic accuracy. Similarly tasks A, B, C, D, G, K, O, P and Q targeted candidates' communicative ability as they tested how appropriately and successfully a candidate could use the target language to respond to a given situation. Tasks C, D, G, M, and N also provided evidence of their strategic competence in that they related to the candidates' ability to contextualise their response.

(a) *'Have you (or your friends) been in an embarrassing situation? How did you (or they) react when you (or they) were embarrassed?*

Or

(b) *'What would you do if you found one million rupees? Speak in English for five minutes what you would do to spend the money'*

Tasks such as the ones given above provided evidence that the candidates understood the situation and made a response they thought was appropriate. Similarly, task H. I, J, K and Q were the tests of discourse competence in that they tested their ability to establish references and co-references, and to respond to a task cohesively and coherently. In the figure below I show what type of competence different tasks of the pilot test assessed:

Skills\Tasks	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Listening	X	X															
Speaking			X	X													
Reading					X	X	X	X	X	X	X	X	X	X			
Writing							X	X				X			X	X	X
Competence	Lc	Lc	Cc	Cc	Lc	Lc	Cc	Lc	Lc	Lc	Cc		Sc	Lc	Cc	Cc	Cc
Under-test	Cc	Cc	Sc	Sc			Sc	Dc	Dc	Dc	Dc	Lc		Sc			Dc

Lc = Linguistic competence
Cc = Communicative competence
Sc = Strategic competence
Dc = Discourse competence

Figure 17: Test tasks and the type of competence they test

The individual tasks of the test are tasks of communicative competence in that they ensure that the students interacted with the test tasks and used their own words to respond to them. In other words, the candidates were made to use their linguistic, communicative, discourse and strategic competence to perform the tasks of the sample test.

A number of issues were addressed during the development of the new test specification, test structure and the sample test including:

- authenticity of materials
- authenticity of tasks
- authenticity of responses
- criterion-referenced and norm-referenced interpretation
- reliability and validity
- practicality

A brief description of these issues follows:

6.4.1 Authenticity of materials

Authenticity of materials is the degree to which language materials have the qualities of natural speech or writing. Authentic materials refer to texts that are genuine, undoctored and appropriate to the situation, not specially prepared for pedagogical purposes (Richards 2001). For the pilot test, I selected materials from newspapers, cookery books, magazines, information leaflets, conversion tapes, and TV. While selecting the materials, I applied the following criteria to maintain their authenticity:

Familiarity to the context or concept

I avoided materials which were 'foreign' to the students in terms of concept, terminology, and cultural context. I did this by making an analysis of language domains recommended in the curriculum and the prescribed texts and then by comparing them with those of the selected texts. I ensured that the concept and context of the selected texts fell within those domains (for a complete inventory of domains are listed under each skill, please see Test Specifications, Appendix A).

Length

The SLC English curriculum, teachers' guide and prescribed textbooks did not give any indication about how long a text should be. However, a cursory analysis of the texts given in textbooks and teacher's guide revealed that for listening, the texts were one to two minutes long, and for reading the length was between 150 to 300 words. Language teaching and testing experts also support the view that the length of a reading text for this level of students should be somewhere between 150-300 words (Heaton 1988; Khaniya 1990). The texts I selected were within this range.

Interest

The text selected should be novel, and at the same time interesting. Interest is linked to comprehension and reading speed. Grellet (1981) suggested that the students

of this age like to read texts dealing with adventure, things of national and regional importance, food, fun, festivals, music, TV and so on. During the selection of texts, I took care that texts were from these categories and fell within the domains specified in the secondary curriculum.

Difficulty level of the text

Neither the curriculum nor the prescribed textbooks mention anything about the target range of vocabulary and sentence structure. However, based on the information passed on to seminars and workshops organised by the *Curriculum Development Centre* (CDC), I gathered that the range of the target vocabulary for this level is 8,000. My analysis of the sentence structures used in the prescribed texts revealed that the average length of sentence was 12 words and the sentence structure was one or two clause-sentences. There were a few instances of three-clause sentences also. I ensured that the selected texts fell within this range. With a view to maintaining anonymity of the materials, and meeting the criteria mentioned above, specific names were at times, altered and texts edited.

6.4.2 Authenticity of Tasks

Authenticity of test tasks refers to the extent to which a test task reflects a real life situation. A language test is said to be authentic when it closely mirrors the content and skills being tested. A test of listening, for example, is authentic when it tests a candidate's academic or social listening ability in an academic and/or social situation. Similarly a conversation test designed to measure a person's conversation skills is an authentic task.

Complete authenticity in a test, however, is unachievable because a test in itself is a special occasion. Anything that is special or specially prepared is not authentic. Authenticity in a test is, however, approximated by simulating reality through the use of authentic materials, authentic tasks and authentic situation (Davies et al 1999).

Authenticity of texts and tasks in the pilot test was, thus, maintained in two ways. Firstly, only authentic texts from sources such as information leaflets, brochures,

newspapers, recipes and menus or bills, were selected for the test. Secondly, tasks were designed to reflect the contents and abilities specified in the test specification.

6.4.3 Authenticity of Response

Authenticity of response is the extent to which candidates' interactions with the test materials replicate the real world processes. In this test, attempts were made to create tasks or conditions so that the candidates' interactions with or response to the tasks were produced in the criterion setting.

In (writing) task P, for example, the candidates were asked to assume that they were the victims of a particular set of circumstances and to respond to the task as a victim. Similarly, in the task Q, they were given the role of a tourism development officer of their village who had to prepare attractive information leaflets to lure tourists to visit their place. However, as most test tasks are simulated, and done in an artificial environment, it is rather unreasonable to expect a high degree of authenticity in response. Therefore, a scorable performance elicited in this way was a 'principled compromise' so far as the authenticity of response is concerned (McNamara 2000:27).

6.4.4 Criterion-referenced testing vs. norm-referenced testing

A criterion, as discussed in the previous chapter is a standard of learning. A test is a simulated environment created to elicit approximations or samples of the criterion. It is the way these approximations or samples of performances are interpreted that makes a test a criterion-referenced test or norm-referenced test.

The sample test used in this research is a criterion-referenced test in the sense that I have closely tied up the tasks with the target language uses or behaviours. That is to say, examiners grade candidates' performance against the criteria set in the rating scale. At the same time, I have proposed a numerical scale to interpret the performance into a numerical score, which is then compared with all other scores to assign a class or division.

158

The proposed testing package is, therefore, criterion-referenced testing with norm-referenced interpretation of the grades arrived at by using a criterion.

6.4.5 Validity and Reliability

Validity is defined as appropriateness, meaningfulness and usefulness of the specific inferences or interpretations made from test scores, and is directly related to the purpose of the test (Athanasou 1997). For example, if the purpose of a test is to discriminate fast readers from slow readers, a test will be valid if it gives an accurate indication of the students who read fast and who read slowly.

When a test has a high degree of validity, the results from the test may be reliable. Reliability is associated with the questions of stability and consistency of results. It is the degree to which the scores are consistent and dependable (Athanasou 1997) which are inferred from a statistical procedure known as a correlational coefficient (Burns 1998).

For the purpose of the revised testing package, my intention was to establish face validity, content validity, construct validity and predictive validity of the sample test. Similarly, in order for establishing reliability, I established consistency of item difficulty, consistency of scores, consistency of performance and inter-rater reliability, which I discuss in Chapter 8.

6.4.6 Practicability

A test must be practicable to administer in a specific situation. By practicable it is meant that the test should be of appropriate length, it should be easy to administer, and there should be adequate equipment and infrastructure for its administration. Furthermore, it should be easy to mark, and adequate resources both human and material resources (such as interviewer, raters, examiners, funds, etc.) should be available.

The field of language testing in Nepal lacks the necessary infrastructure and resources, so that, realistically, listening and speaking tests cannot yet be implemented. In order to ensure practicability, I recommended that an alternate test structure be

159

practised for an interim period of five years. This period should allow time for the concerned agencies to develop the necessary infrastructure and human and material resources needed to administer a test which includes listening and speaking. I, therefore, have recommended a phased implementation of the new test package so that funding could be sorted out step by step.

6.5 Marking and Interpretation of the Test Performances

6.5.1 The Rating Scale

For marking and interpretation of test scores, I have adopted a five-point analytical scale for scoring on the four skills of language. The five points of the scale roughly correspond to the labels used by the *Office of the Controller of Examinations* (OCE) as the five levels of performances. The five levels of performances are as follows:

- Failure
- Pass (Third Division)
- Second Division
- First Division
- Distinction

While many scales use seven or more points (or levels) of performance, I developed five levels with a view to making the transition to the new test easier for the Nepali users. The differences between two levels in a seven- or nine-point scale are very subtle and often blurred which would not help an average Nepali user. In the figure below, I define the proposed levels and give their descriptors:

160

Level	General Descriptors	Language Descriptors
0 (Zero)	No evidence of knowledge Unattempted Poor Failure Incoherent Lack of basic knowledge	No comprehension of spoken language Limited oral production of memorised stock phrase No communicative ability Unable to comprehend running texts No accuracy Inadequate vocabulary and structure to most tasks Poor hand writing
1 (One)	Limited but satisfactory Accuracy confusing Serious grammatical errors Inconsistency in use of TL Pass	Can comprehend short conversations on familiar materials and tasks Can ask and respond to simple questions Can hold simple conversations on familiar topics Limited comprehension of explicitly stated information Can answer questions on surface level information Limited range, frequent lexical inappropriateness Hesitation, repetition and inconsistencies in writing Lexical and grammatical errors
2 (Two)	Moderate or good Medium target language use range Some inconsistencies Simple/accurate realisation of tasks Second class (division)	Can comprehend most simple conversations in social situations Limited comprehension in academic settings Can describe familiar things, personal feelings, events and processes Can read and comprehend most texts written in simple language Can read and enjoy graded prose and poetry Can answer questions on surface level information Medium range use of cohesive devices Can perform on familiar written tasks Fluency and accuracy inconsistencies
3 (Three)	High, very good or excellent Large target language use range Minimum, non-basic gram. Errors Complete realisation of the task First class (division)	Can understand without paraphrase or repetition Can comprehend and speak to English speakers Can hold conversations on most topics with reasonable fluency Can read, understand and respond to questions on most texts within the structure and vocabulary range Can perform on most written task confidently and without hesitation Occasional grammatical flaws
4 (Four)	Excellent Effortless and smooth language Near-native proficiency Almost flawless language Distinction	Can use language appropriate to context, function and orientation Can understand both formal and colloquial speech Can initiate, sustain and conclude a wide variety of communicative tasks Can read at a reasonable speed and comprehend implicit meaning of texts Can convey meaning precisely and accurately Can perform a wide variety of written tasks confidently

Figure 18: The five points of the proposed Analytical Scales and their general and language descriptors (based on the ideas from ASLPR 1984; ACTFL 1986; North 2000)

161

The proposed rating scale with calibrated descriptors is given in Appendix D

6.6 Reporting of the Test Outcomes

6.6.1 The cut-off points

The marks required to obtain a 'pass' certificate vary between countries. In most European countries, for example, the satisfactory mark is 50 percent or higher (West, Edge and Strokes 1999:88). Similarly, in many states in neighbouring India and Pakistan, and in many Asian countries, the minimum mark to pass a test is 40 or higher (Keeves 1997). In Nepal, however, it is 33 percent. This figure seems to be entirely arbitrary as I have been unable to find any explanation for it. It appears to be a traditional cut-off rate of the education system which many of the colonial countries adopted.

Many researchers have expressed their concerns about 33 being the pass mark (Giri 1995; Khanal 1997). When school boards in other developing and developed countries set the 'pass park' much higher, it is difficult to accept 33 as the minimum required level of language proficiency.

An ideal situation is the one in which users of the test results set their own pass mark depending on the purpose of the use of results and the level of language proficiency required for that purpose. Examiners or raters in such a case report scores only without categorising or labelling them. TOEFL or IELTS results are the examples of such a practice.

In a situation like Nepal, however, examiners or raters are provided with cut-off points and label categories. The examination board, in this case the OCE, sets the minimum satisfactory level (or pass mark) at 33 percent and labels categories of scores such as the third division, second division, first division, and reports the scores as follows:

Score boundaries	Class or division
60+	First division
45-59	Second division
33-44	Third or pass division
32 or below	Failure

Figure 19: The existing cut-off points and score boundaries

As the figure above shows, the score boundaries are not scientifically determined. The range between the score boundaries for the third division is 11, for the second division, it is 15, but for the first division, it is 40. Thus, whether a student scores 61 or 100 percent, they are awarded with the same class, i.e., the first class. Clearly, this needed to be adjusted after the current test is revised. Therefore, for the new test package, I have taken a more scientific approach to determining the cut-off points and score boundaries:

Score boundaries	Class or division
86-100	Distinction
71-85	First division
56-70	Second division
40-55	Third or pass division
39 or below	Failure

Figure 20: The Proposed cut-off points and score boundaries

As the figure above suggests, the minimum cut-off point is 40. This figure is also arbitrarily selected. However, this classification makes it comparable to the neighbouring countries, and it makes sense to award a pass certificate when some one achieves at least more than one third of the course. The score boundaries, as it is given in the table above, have a regular interval of 15 and a new class of 'distinction' has been added.

For labelling grades, different countries use different set of terms. Some of the common terms used are - pass, credit, merit, distinction, and honours (West, Edge and Strokes 1999). These terms appear to have been chosen arbitrarily. The concerned examination boards have given no explicit justification for choosing the terms.

For the new test package, with a view to causing minimum diversion, I used the same terms as the OCE uses in the existing scheme but added the term of distinction to it.

6.6.2 The Reporting of Test Outcomes

With a view to developing a simple yet comprehensive and functional mechanism for reporting, making the reporting of the SLC results widely acceptable, and establishing a common language across all stakeholders including funding agencies, I have developed a new scheme for reporting the SLC results. The new scheme provides information on many aspects of the outcomes. Firstly, by using a rectangular shape, it shows what the score range of the examinees has been in English in a particular year. Secondly, by using a dark rectangle (■), it shows what the average score is. Thirdly, a diamond sign (◊) inside the rectangle shows where in the range a particular candidate stands. And finally, it provides descriptors of the abilities a candidate with a particular class (division) should possess.

The proposed reporting scheme, with an even score boundaries, descriptors of abilities for each division, and a display of score range, average score and a candidate's standing in the score range is a scientific one.

A sample of the test outcomes reporting scheme, based on Figures 18, 19 and 20, is likely to look like the one given below:

Alphabetical Grade	Candidate's score (75)	Score Range	Division	Level and their Descriptors
A+ A		86-100	Distinc-tion	4 - native-like proficiency - can understand both formal and colloquial speech - can initiate, sustain and conclude a wide variety of communicative tasks - can use language appropriate to context, function and orientation - can read at a reasonable speed and comprehend implicit meaning of texts - can convey meaning precisely and accurately - can perform a wide variety of written tasks confidently
A- B+	◊	71-85	First Div	3 - can understand utterance without paraphrase or repetition - can comprehend and speak to English speakers - can hold conversations on most topics with reasonable fluency - can read, understand and respond to questions on most texts within the structure and vocabulary range, - can perform on most written tasks confidently and without hesitation - occasional grammatical flaws,
B B- C+		56-70	Second Div	2 - can comprehend most simple conversations in social situations - limited comprehension in academic settings - can describe familiar things, personal feelings, events and processes - can read and comprehend most texts written in simple language - can read and enjoy simple, graded prose and poetry - can answer questions on surface level information - medium range of use of cohesive devices - can perform on familiar written tasks - fluency and accuracy inconsistencies
C C-	■	40-55	Pass Div	1 - can comprehend short conversations on familiar materials and tasks - can ask and respond to simple questions - can hold simple conversations on familiar topics - limited comprehension of explicitly stated information - can answer questions on surface level information - limited range and frequent lexical inappropriateness - hesitation, repetition and inconsistencies in writing - lexical and grammatical errors
D		39 or less	Fail	0 - unattempted - no comprehension of spoken language - limited oral production of memorised stock phrase - no communicative ability - unable to comprehend running texts - no accuracy - inadequate vocabulary and structure to most tasks - poor hand writing

◊ = A candidate's achievement in the skill
■ = Average performance
☐ = National score range

Figure 21: Proposed result reporting format

165

The sample format of the result-reporting scheme above is a hypothetical case of a candidate who achieves 75 in English. A star sign in the middle of the first division column shows that his or her achievement has been in the first division. The rectangle shows that the range of score is somewhere between 15 and 90. The average score, shown by a dark line is 45. The figure also provides a comparison between alphabetic grades and the numerical grades. In the right column, it also provides descriptors for a level.

So far in this chapter, I have outlined the issues and factors that I considered when developing a new test package for the English test of the SLC. I have described how I developed the elements of the package (test specification, test structure, sample test, test marking and interpretation scheme). I have also pointed out, during discussion of the package, how I dealt with the issues or factors in question. Given below is a summary description of the process I employed for the development of the new test package.

6.7 The Process of Test Development

6.7.1 The Process

Looking back, I used a five-stage process for developing the new test. The process may be represented as follows:

Stage 1	Theoretical and practical conceptualisation

$$\Downarrow$$

Stage 2	Bench Marking and Designing Test Structure

$$\Downarrow$$

Stage 3	Test Development

$$\Downarrow$$

Stage 4	Marking and Interpretation of Test Performance

$$\Downarrow$$

Stage 5	Reporting Test Outcomes
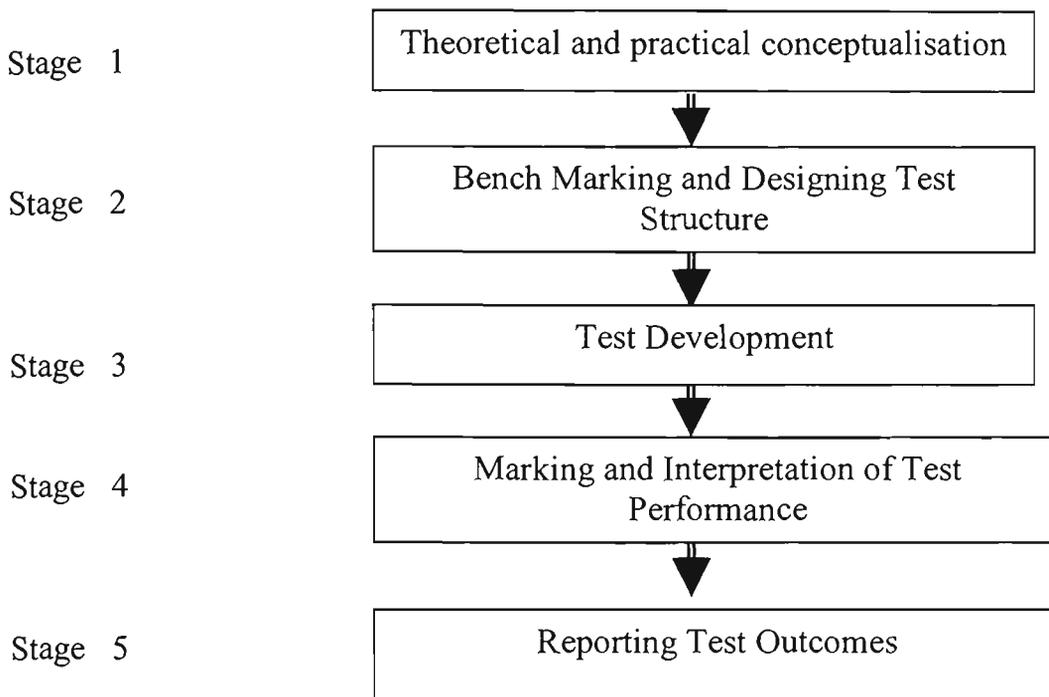
Figure 22: The Process of Test Development

In using this five-stage approach, I have followed Canale and Swain (1980), and Bachman (1990). However, I have elaborated or modified their three stages model in the course of developing and piloting the new test which I have designed for a particular purpose of Nepal. The first stage of the process is about theoretical and practical conceptualisation, which encompasses various theories and approaches to language assessment, curriculum provisions and resource constraints, the broader educational context and the prevailing English language teaching and testing situation of Nepal in particular.

The second stage of the process is setting benchmarks for desired language learning standards, developing test specifications and designing a test structure.

The third stage is what Bachman and Palmer (1996) calls test operationalisation, i.e., development of test tasks and blue prints. This stage consists of steps such as construction of test tasks, producing a pilot test and evaluating them before recommending them for use.

The fourth stage is the development of a rating scale for marking and rating test performances. Similarly in the fifth stage, I developed a scheme for interpreting and reporting test outcomes. The diagrammatic representation of the test development process is given below:

6.7.2 Model of Test Development Process

The Process   **The Components of the Process**

Stage 1
Theoretical
Conceptualisation

Stage 2
Benchmarking
Designing test

Stage 3
Test tasks
Development

Stage 4 Marking
Test Performance

Stage 5
Reporting Test
Outcomes

Context of Testing

Theories and
Approach to
Testing

Desired
Performance
Outcomes

Conditions/
Environment of
Testing

Target Language Levels

Construction

Piloting

Evaluation

Development of
Test Tasks

Marking and Rating of
Performance

Reporting Performance
Outcomes

Figure 23: Model for Development of a New Language Test (based on ideas from Oller 1979, Carroll 1972, Carroll and Hall 1985, Canale and Swain (1980), Bachman 1990, Bachman and Palmer 1996)

There are two parts of the model. The first part of the model represents the five-stage process and the second part indicates their components. The stage of conceptualisation, for example, consists of components such as theories and approaches to language testing, desired goals, context of testing, and conditions of testing. The first stage of the process provides an orientation to the test developers by raising their understanding and awareness of the theories and approaches to language testing, and the educational context in which the test is to be used. In addition, it provides understanding of the nature and provision of the (SLC) English course, and the most appropriate conditions and constraints in which the test is to be adapted.

The second stage is of setting of learning standards or benchmarking. In this stage, a tester, based on the awareness in the first stage, identifies the main domains of the language course, which are envisaged by the (SLC) curriculum and prescribed textbook(s). The test developer then establishes aspects and skills of language proficiency that characterise a proficient (SLC) graduate. Subsequently, a test structure (specification), which is actually a test blue print, is developed on the basis of the above. The test blueprint developed in the second stage helps the test developer in the third stage of construction of the test tasks, piloting them, and evaluating them before they are included in the test paper.

In the fourth and fifth stages, approaches to grading or rating performance are determined and a mechanism for reporting performance outcomes is prepared.

The stages of the process (or the components within a stage) are processes in themselves and are recursive rather than linear in the sense that each component influences and in turn is influenced by what has gone before and what can be expected in the next phase. Desired performance outcomes, for example, are determined by context of assessment (educational and physical environment, course and assessment requirements), approaches to foreign or second language testing (whether the approach or format is integrative or communicative), and context of performance (environment in which the performance is carried out and personal characteristics with which a candidate perform on a task). Desired performance outcomes, along with others, influence the identification and establishment of target language domains, aspects, and skills, which become the basis for the test structure or test specification. Similarly,

identification of test tasks is dependent on the test structure and must influence how the tasks are to be rated or graded.

Completion of the test specification (see Appendix A) and the pilot test (Appendix B) meant that I was ready to embark on the next stage – to travel to Nepal in order to present the test specification and pilot test to teachers and colleagues, and to carry out a small-scale pilot amongst students.

CHAPTER 7

Action and Reflection

Planning, Organising and Conducting Fieldwork

The excitement resides not so much in reaching the destination ... but in the voyage, and what might be found on the way.

- Darlington and Scott 2002:20

7.1 Introduction

In chapters 3-6, I described some of the theoretical as well as the practical issues relating to *Nepali* education and discussed the likely effects of efforts to improve or revise it. I indicated that underdevelopment of education in Nepal is largely the result of the social, economic and caste system which views education as a means to maintain the power structure. The lack of vision and a clear policy for education in general and English Language Teaching (ELT) in particular has held back the growth of educational institutions and systems.

I also demonstrated that language testing in Nepal, especially English language testing at the SLC level, is poorly conceived against the new curriculum, and in consequence, it is poorly formatted, badly developed and inefficiently implemented, mostly by under-qualified and untrained personnel.

In developing the new test package, I kept an average test ESL/EFL test developer in mind. I based my package on the view that an average test developer required (a) a clear understanding of the context of assessment, test purpose and language testing approaches (conceptual clarity), (b) standards or bench marks for assessment, (c) processes involved in test development and (d) schemes for marking and rating performances, and (d) schemes for interpreting and reporting test outcomes (test interpretation). It is with these factors in mind, I came up with a five-stage test designing process, which I believe every new test should go through. It was now time to find out how my colleagues in Nepal would find the new test package.

There are two main sections of this chapter. In the next section I describe how I planned to get feedback on my new test package and to conduct a test pilot in Nepal. In the following section I describe what actually happened in the field.

7.2 Planning to Get Feedback

This section documents the planning phase of the second cycle of the research project, how I identified particular stakeholders, and planned to include them in providing feedback and assisting with a pilot study.

Having developed a methodological framework (Chapter 2) and a revised testing package (Chapter 6), I planned a fieldwork trip with the goal of piloting the sample test and receiving feedback from ELT practitioners of all levels including those involved in the testing of English at the secondary level. To do this, I identified and selected research participants, and gathered their feedback through a number of activities. I followed the procedures described below:

- Identification and selection of people and organisations
- Selecting and sampling of the research participants
- Selection of research sites
- Planning data collection activities

These procedures are described below:

7.2.1 Identification and selection of people and organisations

In order to gain sufficient quality, depth and authority in the feedback about the revised test package, I viewed it necessary to approach the following stakeholders:

- **The Office of the Controller of Examinations (OCE) Officials**

The Office is responsible for developing tests in different subject areas of the SLC including English, implementing them, getting the answers examined

172

and performances rated, and publishing the results. Two officials, who oversaw the implementation of the English test, were invited to participate in the research.

- **The Members of the Curriculum Development Centre (CDC)**

The Centre, which is a constituent agency of the Ministry of Education, develops the policy and curricula and curriculum materials in all SLC subject areas. It is now under the *Department of Education* in the new organisational structure of the Ministry. The centre comprises several subject committees with subject specialists working in each of them. I requested the members of the English Subject Committee to participate.

- **The Trainers at National Centre for Educational Development (NCED)**

The Centre is responsible for training of teachers and for developing necessary training materials for the development of expertise in the SLC subject areas. It comprises national trainers in different subject areas who train teacher-trainers to train teachers at the district and local levels. I invited the national trainers of English to take part in the project.

- **Teacher Trainers**

I approached English language teacher trainers particularly, at the *Faculty of Education, Tribhuvan University*, and *School of Education, Kathmandu University*. The *Faculty of Education* is a major teacher-training institute with more than 10 campuses throughout the country, which train secondary level teachers. The *School of Education* at *Kathmandu University* has recently started to develop and launch programmes in secondary and post secondary teacher education. There are private run teacher training colleges as well. However, they are mostly affiliated to *Tribhuvan University*.

- **Test Developers and Examiners**

I selected test developers and examiners and performance raters from among experienced teachers, teacher educators and subject specialists, most of them from Kathmandu Valley.

- **English Language Teachers**

Teachers teaching SLC English in both the public (now called community) and private schools were asked to take part in the research.

- **School Graduates**

I wanted to involve those students of the selected schools who had graduated from the schools by having taken the 'send-up' (or SLC qualifying) tests successfully, but were yet to sit the SLC. The qualifying tests are usually administered by the local education office (DEO) in the case of the public schools; and school networks such as *Private and Boarding School Organisation of Nepal* (PABSON), in the case of private schools. The school graduates would be used to pilot the test.

7.2.2 The Sampling Procedures

I used voluntary 'opportunistic random sampling' in order to locate individuals in each of these categories to elicit feedback. According to Burns (1998), 'opportunity sampling' is carried out on conveniently accessible groups when there is no possibility of generalisation in a wider population (Burns 1998:86). The selection of stakeholders and organisations was, therefore, based on:

1. recommendations from my supervisors and colleagues;
2. my own knowledge of networks and how the *Nepali* ELT system operates; and

3. the feasibility of visiting them in terms of geographic location and time available for the fieldwork.

Through my colleagues in Kathmandu, I had already contacted the schools, the organisations and the people I had selected to participate in my research. I had already sent a formal request for their participation. I had contacted:

- 60 teachers from Kathmandu and *Chitwan* districts (from both public and private schools)
- 7 ELT experts/academics form *Tribhuvan* University and Kathmandu University
- 2 teacher trainers at NCED
- 5 curriculum specialists at CDC
- 2 officials of the Department of Education, MOE
- 2 OCE officials
- 3 English test developers
- 3 examiners and performance raters, and
- school graduates of the two selected schools.

While selecting these groups, I took care to ensure that the people selected represented the wider context in which the proposed testing package was to be considered. The school graduates were selected from a rural public school and a private school in *Chitwan* for the pilots of the sample test. I had sent a formal letter to the heads of the schools requesting their permission to conduct the test and to use the test scores from the SLC examinations of the concerned students for test validation purposes. Similarly, the concerned parents were contacted for permission for their children to participate in the sample test.

7.2.3 Selection of Research Sites

Kathmandu, as the educational hub of the country with most of the people and agencies involved in the SLC residing there, was an obvious choice. I expected to receive useful feedback as the English language testing academics and experts had working experience in the area in question.

I chose *Chitwan*, a district in the central region for piloting the test because of its geographic proximity and its socio-economic situation. The schools of the plain of *Chitwan*, which stretches as far in the South as the border with India, were representative schools of all regions of the country in terms of infrastructure, students' enrolment and number and quality of teaching staff.

Compared to Kathmandu, *Chitwan* has a different economic make up. Most people in *Chitwan*, (unlike Kathmandu Valley where people are either businessmen or office workers) are farmers. A significant number of people are labourers and must work everyday to maintain their livelihood. In terms of human resources, facilities and infrastructure, a big hiatus exists between the schools of these two districts. The inequality manifests itself in the poorly equipped schools, and school buildings, and in the standard of teaching in public schools. The teaching force also, like the rest of the country, is largely under-qualified and untrained. *Chitwan* has the poorest and most disadvantaged schools, while Kathmandu has both relatively rich and poor schools. I did the pilot in two schools in *Chitwan*: a poor public school and a relatively rich private school.

7.2.4 Planning the Feedback and Piloting Activities

As indicated above, I planned to conduct a number of research activities addressed to (a) members of NELTA, (b) English language teachers, (c) examination administrators, (d) ELT academia, and (e) SLC candidates with a view to obtaining sufficient and representative views on the revised testing package.

I planned to do a plenary session in the international conference of *English Language Teachers' Association* (NELTA) in 2003 with a view to raising the level of general ELT practitioners' awareness of the SLC English test and its development and

implementation. I also hoped to receive feedback on my proposed revised testing package.

I planned several rounds of workshops and interviews with ELT academics and ELT practitioners. The former included teachers, teacher trainers, test developers, examiners, and examination administrators, whereas the latter included teacher educators and ELT experts.

I planned to conduct the sample test on the selected school graduates with a view to finding out their performance on the test and seeking their overall reaction to it.

Thus, I sought feedback on my revised testing package through the following activities:

- Conference presentation
- Workshops
- Interviews
- Email interviews, and
- Test pilots.

In addition, I kept a diary and wrote research journals regularly on my activities, progress and thoughts as they developed through the research process. The research journals are the basis for the accounts given in the thesis and the final reflection on my journey as a researcher.

7.3 Trying out the New Test Package

As an ELT teacher, I believe in the value of the means and processes for achieving an outcome or end, rather than in the end product itself. As a researcher, I have found that the process of research may be as valuable as the findings in the sense that the challenge lies not so much in what useful and relevant knowledge may be found, but in how best to seek that knowledge.

In this section, I describe the actions and processes I carried out in order to seek information and feedback. Carrying out this phase of the research (the consultations or

fieldwork) can also be viewed as the action phase of the second cycle of the action research.

7.3.1 The Fieldwork

In February 2003, I returned to Nepal in order to get systematic feedback about the revised SLC English examination from my colleagues and other stakeholders and to run a small pilot examination with Nepali SLC students. My visit coincided with the Ninth International Conference of *Nepal English Language Teachers' Association* (NELTA) in Kathmandu. This enabled me to present my research and my proposals for change to the SLC examination (at a conference plenary) and hence to gain informal feedback from colleagues in addition to the formal data gathering that I had planned.

As I left Australia for Nepal, the site of my fieldwork, I had very mixed feelings about what might happen. I was excited because the fieldwork was going to be an important opportunity for me to present the ELT leaders and academics with my testing reform proposals. I was also uneasy because I was not sure how they would receive the new test package. I was worried too in that although I had already planned and arranged the data collection schemes and activities, I was unsure whether, in the turbulent political scenario, everything would go as I had planned. As is revealed in this section, while the political scene was peaceful at the time, the feedback and results of the pilot were very mixed and forced me to rethink my original project in a fundamental way and revise the test development process I had used for the present research.

The fieldwork ran for two months and was completed in March of 2003. In addition to the NELTA Conference in February, this time of the year was appropriate for the research for the fact that, in February-March, being the end of the academic semester, most of teaching and preparation for the examination had been completed so that the stakeholders could participate in the research. This was also an appropriate time for the pilot test because I wanted the test to be tried out on those who had successfully completed the Year 10 course, having taken the qualifying test, but were yet to sit the SLC.

As I arrived in Kathmandu, a truce between the army and the rebels had been declared, and peace talks were being planned.

7.3.2 Gaining Systematic Feedback

I carried out the field visit with the following objectives in mind:

- To gain feedback from ELT practitioners, ELT academics, teacher trainers and English testing personnel on the test structure and test specifications aimed to reform the current English test of the SLC;

- To try out the pilot test; and,

- To learn what issues and factors might influence the implementation of the reform package.

In addition, I had two covert aims, which were important to my argument about the role of examination as an agent of reform. They were:

- To find out the participants' views, in the changing context of education in Nepal, about the role of examinations in general and the SLC examination in particular; and,

- To find out what kind of reform the stakeholders thought the SLC English test required.

To achieve these aims, I planned to consult all the stakeholders involved in the SLC Examination, particularly the English test.

7.3.3 Data Collection Techniques

The following activities were organised to obtain feedback:
- A plenary presentation at the Ninth International Conference of NELTA
- Two teachers' Workshops
- A workshop with a range of people (academics, teachers, administrators)

179

- Five interviews with ELT academics
- Five email interviews with ELT academics
- Two test piloting sessions

The data collection techniques allowed access to ELT practitioners at all levels, from the grassroots teachers in rural areas to policy makers at the centre.

7.3.3.1 The Presentation

Nepal English Language Teachers' Association (NELTA), which invited me to do the plenary, is the only professional organisation of ELT practitioners in Nepal. It has 10 regional branches with more than two thousand members. It organises international conferences every year in which around 700 ELT practitioners from throughout the country and 30 to 40 international/regional ELT experts participate. The keynote speakers usually come from the UK or the US and several publishers exhibit their ELT materials.

The plenary presentation at the NELTA conference was my first attempt to unpack the proposed reform package to my colleagues and ELT professionals. The plenary was an opportunity for me to raise the participants' awareness regarding the reform of the SLC, and gain informal feedback on the proposal.

My 50-minute long presentation was well received by the conference participants. However, I did not have enough time for an extensive question and answer session, which would enable me to receive feedback and comments. I fielded a couple of questions and talked to some of the participants during the tea and lunch breaks. I recorded their comments on Field Contact Summary Sheets (for a sample Field Contact Summary Sheet, see Appendix F). My presentation seemingly raised interest and questions about the existing English language testing practice. When some conference participants were keen to know more about my proposals, many others seemed confused.

7.3.3.2 Teachers' Workshop One

The first of the two workshops with the Nepali ELT practitioners was held in *Bharatpur*, the district headquarters of *Chitwan*, in which 26 secondary English

teachers, about one-third from private schools, participated. The workshop was conducted in English in the following manner.

- presentation of the new test package by the researcher
- question answer session
- group work
- group-reports
- post-reporting discussion

During my presentation, having outlined the context of English language teaching and testing at the secondary level, I outlined some of the pertinent issues and then, I presented my proposal for a new English test. The question answer session gave me an opportunity to clarify some of the concepts or terms that might have been too technical for the participants. After the question-answer session, I divided the participants into five groups and gave them separate tasks. The groups were formed as follows:

- Listening group
- Speaking group
- Reading group
- Writing group
- Marking/Reporting group

Each group was assigned a relevant section of the test package to read and then work on the assigned task. I worked with the groups in turns and helped them understand the task by giving examples. I allowed them half an hour to go through the relevant section and prepare comments.

I began to make notes extensively on their presentations when the power supply was interrupted immediately after the first group reporting began. The presentations were brief and to the point. The groups also submitted written comments. After the presentations, I conducted a follow up session in which I asked for their clarifications

on some of the issues they had raised during the presentations. The outcome of the follow up sessions was recorded on Group Contact Summary Sheet.

The teacher participants, in general, liked the new test package. However, they pointed out some issues and problems, which had to be addressed if the package was to be implemented. They also suggested a few revisions to the package. A few of them appeared indifferent, an issue which I later discuss under 'teacher resistance'.

7.3.3.3 Teachers' Workshop Two

The second workshop took place at the Faculty of Education, *Tribhuvan* University *Kirtipur*, Kathmandu in which 30 teachers, most of them from the public schools of Kathmandu, attended. It ran in a similar fashion to that of workshop one, except the fact that the participants expressed their opinion more assertively than the participants in the first group. This was something I had expected because some of the participants were also students in the in-service training programme of the Faculty.

The participants, like those of the first workshop, commended the package but at the same time, they pointed out problems that could hamper its implementation. They include incapability of the examination body, lack of training in the teachers and testing personnel, impracticability of some aspects of the new test and lack of funding that may be required to implement the new test.

7.3.3.4 The Workshop of the ELT Academics and Test Developers

I organised this workshop at the central office of NELTA in Kathmandu with 18 participants from different organisations. The participating organisation and the number of participants are given below:

- Faculty of Education, *Tribhuvan* University   3
- Curriculum Development Centre, MOE   3
- School of Education, Kathmandu University   1
- Department of Education, MOE   1
- NELTA   2
- National Centre of Educational Development (NCED)   1
- Administrator, Private School   1
- Administrator, Public School   1

- Test developers who had developed the SLC English tests   3
- Examiners who mark/rate candidates' performances   2

The *Office of the Controller of Examinations* (OCE) officials were also invited to the workshop, but they did not turn up. I took this as their resistance to the revision and unwillingness on their part to share the criticism of the current practice of English language testing at the SLC level. The presentation, which I completed with the help of a whiteboard and OHP, lasted for about 40 minutes. Following the question-answer session, I divided the participants in 5 groups and explained the tasks. The groups were asked to read the materials of the revised testing package. I allowed half an hour to read and another half an hour for group discussions. The groups were asked to look at the materials from the following perspectives:

- Administrators of the Examination
- Test Developers
- Examiners
- Teacher Trainers
- School administrators

The group presentations were lively: one group often cross-questioned another and at times contradicted the opinion expressed on a particular issue. I made extensive notes on their presentations, which I later transcribed on Group Contact Summary Sheets. Their responses concentrated around 5 issues, which could make the reform difficult in the existing educational circumstances. These issues included the following:

- the centralised structure of the OCE
- bureaucratic mentality of the concerned authority
- training of the concerned people
- the capability of the OCE, and
- resources

7.3.3.5 Interviews

I conducted five face to face (unstructured) interviews with ELT academics and personnel from professional organisations, who could not participate in the workshops, but whose input I viewed as important. People drawn from the following organisations were interviewed:

- Faculty of Education, *Tribhuvan* University
- National Centre for Educational Development (NCED)
- Private and Boarding Schools' Organisation of Nepal (PABSON)
- Nepal Teachers' Association (NTA)

Email interviews were conducted with those ELT experts who were either out of town at the time of data collection or were not available for a face-to-face interview. I emailed them the reform package and a set of questions, and asked them not to restrict their opinions to the questions. Participants for the email interview were drawn from the following organisations:

- School of Education, Kathmandu University
- *Buddhanilkantha* school, a special secondary school in Kathmandu run jointly by the HMG/N and the British Government
- English Subject Committee CDC/MOE
- Curriculum Development Centre/MOE
- Education Campus, a private teacher training institute in Kathmandu

Extensive notes on the face-to-face interviews were recorded while email responses were printed out and used as data sheets during analysis. Most interviewees, like the ELT academics and administrators, blamed the current education system for the existing state of the SLC English test. They pointed out the need for reform and raised issues similar to those of the workshop participants.

7.3.3.6 The Pilot Tests

I conducted the **first pilot test** at a private school, supposedly a better one in *Chitwan* because of its 100 per cent success rate in the SLC. The Secondary Wing of the School ran in a new complex, about 5 km from the city centre. After an introduction to my colleague, whose company I had sought for assistance, I told the class the purpose of my visit. With the help of the class teacher, I announced the names of the randomly selected students. I made them aware that their participation was entirely voluntary, that they did not have to participate in the research if they did not want to, and that they could withdraw from the research at any stage, a decision which would not jeopardise their position in the school in anyway. Only nine of the selected students volunteered to participate in the pilot test.

After I set up the cassette player in a separate room, I called the selected students, explained the tasks and test procedures and began the listening test. After the tape was played once, my colleague and I noticed that most students were unable to complete the tasks, especially those tasks that required longer responses. So, after consulting my colleague, I played the tape again. Playing of the tape twice, instead of once, is one of the several issues I would reconsider when revising my new test package in the third cycle.

For the speaking test, I called one of the students, explained the procedure of the speaking test and asked him to select a folded slip of paper on which the task was written. After he selected his task, I allowed him about 2 minutes to prepare mentally (they were not allowed to make written notes on this test) and then asked him to speak about 5 minutes. Both my colleague and I did the rating on the test on the basis of the provided analytical rating scale. This process was repeated nine times and the sessions were recorded on tape.

I called them for reading and writing tests which ran smoothly as the students seemed familiar with what they were expected to do. The planning component of the writing task was entirely new for the students. Many of them did not attempt any planning for writing. They simply wrote the answers.

The noise coming from the neighbouring classes was occasionally intrusive. After the test, I conducted a post-test group interview in which the students expressed their impression of the test and test tasks. On the whole they seemed to like the test.

Some of them said that they did not attempt the planning component of the writing tasks as they were not clear what they were expected to do. Others ignored it simply because they were worried about time and thought that they would only waste their time. So they just started writing on the task. Some other students said that the tasks were new and made them think.

The **second pilot** of the sample test took place at a public school, an average one, situated in the periphery of the city. Only five students volunteered to participate in the research. The tests of listening, speaking, reading and writing went in a similar fashion to that of the previous school. I noted that the students here struggled to complete most tasks in time. After the test, the students told me that they did not have any listening and speaking practice at all, and some of the reading and writing tasks were not familiar to them.

The outcomes of each of the pilots are further discussed in Chapter 9.

Summary and Implications

In general, the fieldwork went well and produced data which showed strengths and shortcomings of the new test package. The data also pointed out factors which militated against the implementation of the new test. They raised issues which made me reconsider some aspects of the new test package and my position on the role of the SLC English test in the amelioration of the ELT practice in Nepal.

Some participants were critical of the fact that I had allowed insufficient time for them to prepare themselves for the workshop/interview sessions. Some of them said that the material (the revised testing package) was too long for them to go through in one session, and that it should have been made available to them in advance so that they could prepare for the workshop or interview in advance. There was some truth in these comments. However, given the sensitivity of the material, and the fact that the new test package was still under review, I gave them the package during the workshops and asked them to read the relevant sections. While emailing the package, I specifically mentioned that after the response to the research questions was made, the participants were to delete the copyright material immediately. I emailed the package only when they agreed to do so.

Time was always a constraint. Some of the participants who preferred to put their responses in a wider context rather than to make specific points, complained about not having enough time to complete their comments. Others made concise comments specific to the issue in question. This made it difficult for me to integrate the two lots of feedback in a balanced way. During analysis of the responses, I incorporated as much of their diversified views as the scope of the research allowed.

Using electrical equipment was also a problem. I often had to run around to find the right power point, a plug that would fit or for an extension cord to put the equipment in an appropriate place. Frequent power interruptions and lack of enough substitute batteries also made it impossible to record all sessions in full and compelled me to take written notes. These and several more problems like these are related to underdevelopment of the required infrastructure and part of the context of Nepal as discussed.

In this chapter, I have described the action research second cycle of planning and organising fieldwork, trying out the new test package and obtaining feedback. In the chapter that follows, I document the feedback and analysis of the feedback and pilot test outcomes. Then in the final chapters, I discuss the issues noted in this chapter at length and their likely impact on implementation of the new test package.

CHAPTER 8

Feedback and Analysis

A Mixed Result

In qualitative research, we are particularly interested in how others see and experience the world. This requires us to be very aware of the lens which we bring to the task ... it is a perpetual impossibility to look at one's lens at the same time as one is looking through it, but this is one of the many challenges of qualitative research.

- Darlington and Scott 2002:20

8.1. Introduction

The proposed new test package produced mixed reaction from the participants. While many of them commented positively on the package, some of the participants were sceptical about its practicality in the prevailing circumstances.

In this chapter, I present the voices of the wide range of ELT practitioners (i.e., the teachers, academics and testing personnel) who took part in the research, the problems they saw in the current system of education, and the responses they made to the proposed new test. There are two main sections in this chapter. In the following section, I describe the framework used to analyse the data and after that I present the findings. The section on findings consists of four sub-sections. Firstly, I describe the positive findings on the new test package, secondly, I explain their concerns, thirdly, I outline their suggestions for the improvement of the package, and finally, I describe findings on the pilot test.

8.2 Analysis of Data – Organisation of the Findings

Initially, the prospect of putting together a diverse range of viewpoints from the varying groups of research participants and making coherent meaning out of them

188

(while keeping my own perspectives) was daunting. My own thoughts and experiences, my tasks as a researcher to present the voices of the ELT experts and the need to evaluate the feedback about how appropriate and acceptable my proposed revision of the SLC was, often mingled with one another. However, in order to maintain a degree of objectivity in the analysis of the viewpoints I tried to maintain a distance whilst making coherent and objective meaning out of the data, and retaining the diverse voices of the participants as far as it was practicable.

I needed a framework, i.e., a way of organising the diversity of views into generalised findings whilst acknowledging their common as well as their contradictory voices. I applied data reduction techniques (coding, transcribing, classifying and categorising data) and data display techniques (extended text, graphs and charts) as described in Miles and Huberman (1994). I carried out axial coding on the different and at times contradictory responses, which, as it is described in the forthcoming section, yielded themes and issues. For the pilot test, I established an overall performance of the students who took the sample test, and then calculated their skill level as well as school performance separately. Finally, I established validation of the pilot test.

8.2.1 Coding of the Data

For the purpose of the present project, an open codification of data (Cohen, Manion and Morrison 2000; Miles and Huberman 1994) was applied to analyse the field notes (FN), written notes from the research participants (WN), (taped) interviews (IN) and emailed interview (EIN). I applied colour coding which is basically a way of organising data under different categories, each one associated with a particular colour (Delamont 2002). Having colour-coded the field notes, written responses and interview notes, an axial coding (Strauss and Corbin 1990, in Skinner 2001) was undertaken which involved employing set procedures to put the data back together into categories. The data under a particular category was then organised into sub-categories. The six categories created for the initial themes were:

Categories	Colour
The Educational context	Yellow
The ELT context	Blue
The ELT practices	Orange
The SLC Examination	Pink
The Existing English Test of the SLC	Red
The Proposed Testing package	Green

Figure 24: Colour Coding Scheme

I then made multiple copies of the total data and filed them under each theme or category. This helped me to keep all relevant data readily available avoiding confusion between segments, and the need to constantly sift through the pages of the data. The organisation of data in this way also helped in the reflection stage – a core component of this research by allowing me to verify some of the implications of the data.

A sample set of codes under a colour category with a name and definition is given below (see Appendix G for details):

Names of codes	Codes	Definitions
1. The Educational Context	EdCon	The Educational Context
1.1 Expectation to Education	EdCon/Expect	The context in which education is
1.2 Attitude	EdCon/Attitude	provided in Nepal. It includes
1.3 Social Issues	EdCon/Social	factors like general educational
1.4 Political Issues	EdCon/Political	goals, policy, organisational
1.5 Economic Issues	EdCon/Economic	structure, provisions and practices.
1.6 Constraints	EdCon/Constraints	It also includes the constraints that
1.7 Discrepancies	EdCon/Discrepancies	the education in Nepal faces and
1.8 Exam Disease	EdCon/Exam	the discrepancies that may be in the system, in the planning, implementation, provision and in the examination.

Figure 25:Sample of the codes used in the analysis of data

8,2.2 Triangulation of the Data

The views and opinions of the different groups of stakeholders (ELT practitioners, academics, testing personnel) were in agreement on most issues. The differences in opinion between teachers (who were largely untrained) and experts could mainly be explained by the difference in understanding of the theoretical and technical

issues of testing. Likewise, the workshops, interviews and email interviews produce the same sort of views and opinions. Each method of data collection appeared to produce the same findings. The different methods could therefore be said to triangulate (Flick 1992) each other.

8.3. Presentation of the Findings

The findings on the new test package are presented in four parts. The first three parts summarise the findings on the new test package, which include the research participants' positive response, critical comments and their suggestions for improvement. And the fourth part is an analysis of the participating students' performance on the pilot test. A conclusion summarising the main points of this section is given in the end of the chapter.

8.3.1 The Positive Response

Most participants of all three groups found the new test package:

- well done and comprehensive (FN: 7)
- relevant to the new English curriculum and useful (FN: 15)
- informative (FN: 26), and
- an improvement on the current test (IN: 5)

In particular, they appreciated the proposed -

- test structure (FN: 7)
- test specification (IN: 5)
- marking criteria (FN: 26)
- analytical scale (FN: 15) and
- results reporting scheme (WN: 2).

The participants of the workshops said that the reform package was comprehensive in the sense that it incorporated all aspects of test development, and that, as a group report showed, all skills and elements of the SLC English course were adequately stressed.

Comprehensive as expected. All skills are clearly and adequately addressed. The mark-sheet format is attractive (FN: 7).

The proposed test structure and test specification received widespread attention. The participants expressed their comments on the test format and various parts of the test specification. Many of them found that the test structure and specification were workable, and an improvement over the existing test. A national teacher trainer made his comments as follows:

> *Based on our experiences and the knowledge that we have of testing theories, the proposed test schemes/test specifications seem to be better and more useful than the existing one* (IN: 5).

They also found the marking scheme, marking criteria and reporting schemes systematic, transparent and an improvement on the current schemes. A teachers' group report said:

> *... it is much more informative. The inclusion of grading criteria, score boundaries are very appreciable [sic][appreciated.* (FN: 26).

This view was shared by the academics as well:

> *... the grading scheme is appropriate. It will make the answer checking systematic and transparent. It will help teachers and students see what is expected of them at each level* (FN: 15).

The majority of participants in all three categories recommended the package for implementation. One of the group reports commends the revised testing package in the following words:

> *The proposed scheme is very appreciable [sic][much appreciated]. It has to be recommended. The government should show its intention to implement and materialise of this scheme into practice* (FN: 15).

There was consensus among teachers, testing personnel and examiners on the sample questions, their layout, and mark and time distribution. For example, they said that some tasks on listening, reading and writing were new and innovative. They felt that the tasks were skill-based, creative and required the students to make independent interpretations of the text and write answers in their own words. Some of them were of

the opinion that the reform package could help reduce the negative washback effects, which were detrimental to classroom teaching and learning. In some of the group reports, they said:

> *We find the sample questions good and skill based. Though some questions are new and unfamiliar to the students, they will encourage the students to think, be creative in their answers and write in their own words* (WN: 3).

> *I think the tasks are well written and will require students to be imaginative in responding to them. This in turn will improve the teaching practices as well* (WN: 2).

However, as the section below reveals, they also pointed out areas which needed reconsideration.

8.3.2 The Critical Comments

The analysis of the critical comments showed that their comments were mainly around five areas of the test package:

- its practicability in the existing circumstances (FN: 27)
- cut-off points (FN: 7)
- score boundaries (FN: 6)
- the grading scheme (FN: 6), and
- the sample questions (FN: 2).

The teachers in both workshops expressed the view that the reform package, particularly the marking and reporting scheme, good though it was, was not practicable in the existing circumstances, as the required support mechanisms were non-existent. During group report presentations, one of the groups said that:

> *... the marking scheme that you have presented is good. However, it is not likely to be put to practice in the context of Nepal. It is so because everything in this country lies in theory, and there is no consistency between theory and practice. So, the scheme is not practicable in the theoretical context of Nepal* (FN: 27).

The statement does not clarify what actually the disagreement is about. The group probably raised their doubt about the government's capability in this regard and questioned if the package would be implemented successfully. It should be noted that Nepal as a bureaucratic society is considered to be weak at the practice level (Vir 1998).

Many teachers and testing personnel were concerned about the raising of the pass mark from 33 to 40. They thought that, in the situation where the majority of students did not even secure 33 to pass the exam, and that a very large number of candidates obtain marks between 33 and 39, raising it to 40 would only result in failing more students. Many of them recommended that cut-off points and score boundary be 'improved' (WN: 8), 'revised' (WN: 12) or 'modified' (WN: 18). As one of the teacher's reports said:

> *The proposed reporting scheme is attractive. But I think, the cut-off points and score boundaries should be revised keeping the Nepali context in mind.* (FN: 7).

The testing personnel agreed with the teachers' concern. The commented that:

> *... the cut-off points and score boundaries suggested in the proposed package are not satisfactory. The score boundary suggested for the second division is 56-70. The average students cannot achieve this level. The level of second division should be achievable to an average student. I think the existing scheme of cut-off points and score boundary is suitable enough* (FN: 6).

The teachers and testing personnel seemed to have missed the point. They were, in fact, comparing the proposed score boundaries with the outcomes of the current tests in which the average score of the SLC examinees was in the mid to high forties. They thought that raising of the minimum cut-off point to 40 and subsequently other score boundaries would end up failing more students. Their concern, however, is without a theoretical basis because the proposed new test is a different test and the proposed marking scheme is based on well-defined criteria and an analytical scale which means that if they were followed closely, they were likely to improve the examiners and raters' performance. The analysis of students' performance on the pilot test supported the view. The ELT experts and academics were in agreement that an improved marking/scoring

scheme would make the practice of marking more 'transparent', consistent' and 'reliable' (IN: 1).

Commenting on the test items on listening and speaking, the participants of the workshops said that the tape should be played twice. This could be easily accommodated in the next revision of the new test. Some teachers said that questions on speaking, especially the one in which the students were asked to interpret a pie chart verbally, would not be practicable in the rural areas where, as the quotation below suggests, students did not practice speaking on such materials at all.

Some aspects/parts may not be feasible for the rural areas, specially the task which has a pie chart to be interpreted as such materials are not practised in such areas (FN: 2).

These comments suggested that the test should be developed in a way that allows some variation from region to region. The problem of feasibility of some questions in some areas is associated with the practice of ELT at the classroom level. If certain parts of the curriculum or prescribed texts are not implemented in some areas of Nepal, it should be viewed as a teacher's problem or a problem of the teacher training agencies which are responsible for training teachers and providing them with materials to be able to teach them. The students should not therefore be penalised for what they have not been taught. A test uses a representative sample of the target curriculum and prescribed texts on the assumption that all teachers know the curriculum and what is expected of them. The teachers are also expected to provide practice on all types of materials prescribed by the curriculum.

Here there is an issue of equity and fairness. In Chapter 1, I noted that there were regional discrepancies in teaching standards and in resource distribution. However, the SLC as a national test assesses candidates on the basis of the national standards set by the secondary curriculum. It is, therefore, not supposed to be based on the regional variations in teaching and learning. This raises the issues of equity and fairness to the candidates of the five different regions of the country. I would like to argue that the national test of SLC English can be divided into two parts – (a) oral, and (b) reading and writing. The oral test, which consists of listening and speaking tests, may be regionalised. This means that five different versions of the oral tests, each developed on

the basis of the themes and concepts appropriate for an individual regional context, would have to be developed and administered separately in the concerned regions. These five versions of the oral test would be different but equivalent so far as their objectives and standards are concerned. This, in my opinion, would address, to some extent, the issue and equity and fairness.

On the reading test items, some of the teachers said that seen (prescribed) texts should not be included in the test, that some reading texts (especially 300 word long texts) were too long and questions on establishing co-references or locating cohesive devices in the text were not practicable, as students were not familiar with such tasks. A participant expressed his concern in the following terms:

> *I wonder what the aim of giving a seen passage to candidates is. In my opinion, seen passages (or texts) should not be given. If the students are to be tested on reading and writing skills, seen passages do not measure what they ought to measure* (FN: 1).

This participant was concerned that the new test could become a stereotype. As I noted in Chapter 6, some prescribed texts were included in the new SLC test to give the pilot test a backward reference. However, new test tasks were suggested to reduce the possibility of the new test being stereotypical reflection of 'seen' texts.

In the test of writing, they indicated that the students were not familiar with planning processes, and that could be a problem especially in the initial years. They also thought that there were too many long-answer questions. I agree that asking students to provide written planning of the writing they do during a test would be a problem in the initial years of implementation. However, this problem would gradually diminish once the task become familiar to the students. Their comment about too many long answer questions would be considered in the next revision.

Speaking of the need for equipment recommended in the package to implement listening and speaking tests, the participants said that:

> *... some recommendations, such as having a tape recorder for every examination hall and recording every candidate's answers on the speaking test, do not seem to be practicable in the present circumstances* (FN: 5).

The concern is genuine given the fact that the OCE is under-resourced. However, a test of listening and/or speaking cannot be effectively implemented if a

tester does not have the support of adequate equipment. Tape recorders would be needed to implement the listening tests and to record students' responses on the speaking test. As is suggested in the test specification, one tape player would be needed for every examination hall. This might not be feasible for the OCE in the initial years. I have, therefore, suggested a strategy for a phased implementation of the new test package in order to resolve this problem in the final chapter of the thesis.

Commenting on the marking scheme and analytical scale, the teachers and teacher trainers said, as the excerpt from a group report below shows, that the criteria for marking different skills and elements of language needed sharpening and should be more explicit because the boundaries between two levels of performance were not clearly stated.

The criteria given in the analytical scale should be sharper and more explicit showing clearly the expected level of performance. An ordinary examiner cannot see the difference between the levels of performance criteria (FN: 5).

The academics/experts, however, did not see any ambiguity in the criteria of the proposed rating scale. For them, the scale was 'well-done' (IN: 3). The teachers seemed unfamiliar with the idea of a rating scale. This demonstrates a need for training with a view to familiarising the teachers and examiners with the schemes of the new test package including the rating scale. In fact, training of teachers, test developers and examiners should be a cyclic process in which new training modules are formulated and implemented on the basis of findings from an ongoing research process.

8.3.3 Useful Feedback for Improvement

A participant, during the interview, said:

A separate training package for each of – examination setters/ test developers/examiners performance raters and teachers should be developed and implemented (IN: 3).

As I discuss in the final chapter, training of teachers and the testing personnel should be an in-built programme in the new test package because it would require the

197

ELT practitioners to be familiar with the proposed new schemes. Making some other specific comments on the test package, the teachers, agreeing with testing personnel, said:

> *The pass mark should be lowered from 40 to 35. The cut-off points and score boundaries should be revised in keeping with the pass mark being 35.* (FN: 5)

This teacher is concerned that raising of the minimum pass mark to 40 would end up failing more students than is the case in the existing scheme because they would be unable to achieve the raised (proposed new) level of scores. However, as I discuss later in the chapter, this concern is unsubstantiated.

> *The rating scale should be a six-point scale rather than a five-point scale* (FN: 5).

The teacher did not elaborate on how a six-point rating scale, rather than a five-point rating scale, would be more useful. I do not think adding another point in the rating scale would make any significant difference in rating outcomes. As discussed in Chapter 6, I took a simplistic approach to propose a five-point rating scale with a view to keeping it simple for the prospective users who are largely untrained.

> *More varieties of writing test items should be considered. Short answer questions should be included cutting down the number of long answer type questions* (FN: 4).

I thought this was a very useful suggestion which could be accommodated in the next revision of the new test.

> *Separate sets of tests on the same text should be given to check cheating* (FN: 1).

It should be noted here that during the SLC examination, two or three candidates sit on the same desk which can make it possible for one candidate to copy answers from another. To prevent them from copying from one another, the OCE has trialled parallel sets of questions under which different candidates sitting on the same desk receive different sets of question papers. However, as I have pointed out earlier, these parallel sets are not based on any elaborate test specification. As a result, they are not always of

the same standard. My preference would be to use one test only and to find alternative methods to prevent copying.

On the listening and speaking test, the teacher participants said:

The tape for the listening test should be played twice (FN: 1).

The tasks of speaking test should be simple and preferably familiar (FN: 2).

As indicated on the previous pages, with a view to addressing the problem of regional discrepancy in the teaching of listening and speaking, a room for some sort of regional variation would be considered during the next revision of the test.

In general, the comments are useful for the improvement of the new test package and, where possible, will be incorporated in the revision of the package in the third cycle. Having varieties in the writing test items, parallel sets of the questions to check cheating during examinations, playing the listening text twice, and developing speaking tests on familiar concepts would be considered. However, I doubt that lowering the proposed minimum cut-off point from 40 to 35 would make any difference so far as the success rate on the new test is concerned.

However, some of the comments given in section 8.3.2 and in this section (8.3.3), especially the comments about raising the minimum cut-off points, the concern that the new scoring scheme would fail more students, and the comment about the inclusion of the seen (or prescribed) texts in the test, indicate that the participants missed the point in question. In their comments they implied that the proposed schemes would not improve the test outcomes. In fact, they looked at the proposed schemes from the angle of the existing test in which there is a very low success rate. These concerns and comments were not so helpful, firstly because the effect of the proposed test should be judged in the light of the outcome of the new test, and secondly, the outcomes of the pilot test, as the section below reveals, showed that, at least in this limited pilot test, a higher cut-off point is possible.

8.4 Outcomes of the Pilot Test

8.4.1 Overall Outcome of the Pilot Test

As discussed in previous chapters, the focus of this research was on getting qualitative evidence on the new test package of which the pilot test was a part. A small scale pilot was conducted on 14 students from public and private schools in order to assess its strengths and weaknesses. Despite my attempt to work with 60 (public and private school) students, only a small number of students volunteered to participate in the pilot test. They had initially agreed to take part in the test. However, when told that their participation was entirely optional, they chose not to.

On the whole, the pilot test yielded better results with the success rate of 85.72 percent, than the current English test of the SLC of which the success rate hovered around only 32 percent (OCE 2003). The overall average score on the new English test was 54 percent. While it was pleasing to see that the pilot test produced the desired success rate, the test will need to be administered on a greater number of students in order for its validity, reliability and practicability to be properly assessed.

8.4.2 Performance of Language Skills on the Pilot Test

The students who took the pilot test, as the figure below demonstrates, performed best in reading with an average score of 64 percent. Writing, listening, and speaking respectively ranked second, third and fourth. The students performed better in reading and writing than in listening and speaking. This was something I expected because the majority of the students usually do not have adequate practice in listening and speaking. However, they did not perform as well as expected in writing. This could be the result of the introduction of a new component of 'planning' in the writing task.

Skill	Average Score
	(of both private and public schools combined)
Listening	52
Speaking	47
Reading (+cloze)	64
Writing	55

Figure 26: Average skill-based scores

8.4.3 Performance of Public schools vs. Private Schools

The Private school students performed better in the test with an average of 65 percent than those of the public schools whose average performance was 41 percent. The average score of the private school students fell in the second division whereas the average score of the public school students was in the third (or pass) division. Only 11 per cent (or 1 student) of the private school students failed. The majority of the students, 67 percent (or 6 students), got the second division and 22 percent (or 2 students) the first. On the other hand, 20 percent of the public school students failed and the rest scored the third division marks only. This suggests that teaching and learning of English in the public schools is not as good as that of the private schools.

The graph below presents the performance of public and private school students on the four language skills:
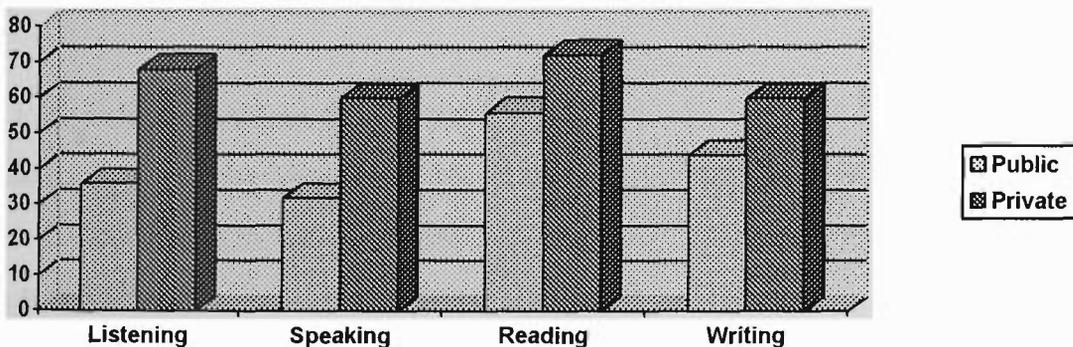
Figure 27: Skill-based performance of the public and private schools

The students of private schools, as the chart above shows, performed better in all four skills of the language than those of the public schools, as would be expected. The

students of private schools performed best in reading, followed by listening, speaking, and writing. The performance of the public school students was also better in reading and writing compared to their performance in speaking and listening.

8.4.4 Validation of the Pilot Test

In language testing, a tester looks for authentic information to make inferences regarding the performance of a candidate on a given task. To do this, a tester asks the candidate to perform a language task. The validation of the test then is the process of confirming whether the construct, design or content of the test is good enough to elicit genuine information.

Validity is defined as the appropriateness, meaningfulness and usefulness of the specific inferences made from test scores. It is, therefore, concerned with a candidate's ability to perform, and is directly related with the purpose of a test. Despite the fact that validation was not a part of the aim of the study, I conducted various validation checks on the pilot test and established different types of validity.

As I mentioned earlier, the pilot test was administered only on 14 candidates. Though the results of the validity checks are favourable, I cannot make any definitive claim about the validation, for which the test needs to be tried on a greater number (600 or more) of candidates.

Face Validity

Face validity is about the physical appearance of a test and relates to the general impression it gives to the concerned stakeholders. In other words, face validity is to find out how the test looks to the candidates, teachers and others and whether or not it has surface acceptability.

During post-test discussion, when I asked the students participating in the pilot test how the test looked, they said that they generally liked the overall appearance and the layout of the test contents. Similarly, participants of the workshops pointed out that the test was well designed. As a group report says,

... the proposed test appears to be scientifically designed, and the layout is user friendly' (FN: 10).

I was satisfied that the test had face validity.

Construct validity

A construct is an ability, a trait or a skill under test, which is reflected on a test performance. Inferences about the performance are made on the basis of the achieved test scores. Construct validity, then, indicates the degree to which a test reflect the theory on which the test is based, or how the theory is operationalised during a test. Construct validity may be established in two ways, firstly, by correlating the candidates' performances on different sections of the test (sub-tests) and then performance in each section of the test with the whole test, and secondly by making inferences from the content validity (Banerjee 2000) as explained in the next section below.

The pilot test, as I noted earlier, is based on the theory of communicative competence and academic language proficiency. As the Figure 29 on page 200 shows, the sub-tests of the pilot test correlated highly with one another and with the whole test. As I noted in Chapter 6, the pilot test was created by taking the sample test tasks from the test specification. This ensured that the pilot test assessed the ability or skill, which is stated in the test specification. I was, therefore, satisfied that the test has construct validity.

Content Validity

Content validity examines the extent to which the test is a good representation of the material that needs to be tested. In other words, a test has content validity if its items constitute a representative sample of the skills, tasks or domains as specified in the test specification.

I established content validity of the pilot test in two ways, firstly by comparing the test contents with the test specification and then by feedback from the stakeholders.

I analysed the language traits that the pilot test proposed to assess and compared it with the language traits or abilities given in the test specification.

A summary of traits/abilities/skills tested by the pilot test is given below:

Skills	Tasks		Traits	Task Type
Listening		A	Academic Listening (Instruction)	Fill in parts of a sentence Complete incomplete sentences
		B	Informal/social listening	Find missing words in informal conversations
Speaking		C	Interpreting non-text materials	Interpreting a pie chart
		D	Contextual Response	Responding to a given context, scenario or visual prompt
	Text 1	E	Meaning from context	Locating a word in text for a given meaning
		F	Skimming for main ideas	Arranging main points of a text in order
		G	Comprehension	Answering comprehension questions
Reading		H	Understanding grammatical function	Framing/completing sentences on a given syntactic pattern
	Text 2	I	Linguistic Competence	Cloze test (every n^{th} word missing)
	Text 3	J	Scanning	Fill in spaces (particular grammatical units
		K	Establishing co-references	Locating co-references(e.g., pronouns)
		L	Summary writing	Summarising a passage
	Text 4	M	Scanning/ inferring from a non-text material	Locating/writing information
		N	Comprehension of non-Material	True-False
		O	Imaginative/Descriptive	Describing a character of a story story writing completing a story
Writing		P	Letter writing	Writing a letter of complaint
		Q	Preparing non-text writing	Preparing a tourist information sheet

Figure 28: Language abilities tested by the pilot test

The abilities and skills stated in the figure above closely reflect the abilities and skills detailed in the proposed test specification which I prepared on the basis of my analysis of the curriculum and the prescribed texts.

Content validity is also established by asking for an expert opinion or feedback from the concerned stakeholders. During the field trip, I asked the participating ELT

academics and the language testing personnel if the pilot test had representative samples of the skills specified in the test specification. Their response showed that the pilot test had high content validity.

Predictive Validity

Predictive validity is the extent to which test scores predict future performances. This is achieved by comparing the results of the test with the performance (test scores) of the candidates in a future test.

For the present project, I compared the candidates' performance in the pilot test with that of the English test of the SLC, which was conducted about 1.5 months later. I calculated a correlation coefficient following the Pearson Product Moment Correlation Coefficient formula (see Appendix E for details). The result of the statistical analysis showed that the co-relationship between the pilot test and the English test of the SLC rated .9, which means that they had a high correlation. In other words, the correlation coefficient of .9 means that the candidates who perform well on the pilot test were likely to perform well in any future. English language tests. Thus, the pilot test seems to provide reliable information on the candidates' abilities in the target language, and has predictive validity.

Reliability

When a test has a high degree of validity, the result from the test must be reliable. Reliability is associated with stability and consistency of results. It is the degree to which test scores are consistent and dependable (Athanasou 1997). Consistency is inferred from statistical procedures known as item analysis and correlation coefficient. For the pilot test, I carried out three types of consistency checks.

a. Consistency of difficulty of items

In order to ensure that items of the pilot test are moderate in terms of their difficulty level, and that they have discriminability, I calculated two types of statistical procedures – an item analysis and a discrimination index.

Item Analysis

Item analysis is a statistical practice used to derive the level of difficulty in a test item and its discrimination ability. These two statistical procedures are known as item facility and item discrimination index.

Item facility (or p-value) is the proportion to which examinees answer an item correctly. It is the mean score usually expressed as a percentage or a decimal of all examinees tested on a particular test item (Oller 1979). The p-value ranges from the lowest value of .00 (or 0 percent) to the highest value of 1.00 (100%). The easier a test item, the higher the p-value. Similarly, the harder the item; the lower the p-value.

For the present research project, p-value of all 17-test items was calculated. The p-value of the test tasks is given in the figure below:

Item Difficulty Level

Test Item	Full Mark	Total Score	Mean of score	p-value
A	7	41	2.9	.41
B	3	33	2.3	.78
C	5	32.5	2.3	.46
D	5	33.5	2.4	.48
E	5	50.5	3.6	.72
F	3.5	31	2.2	.62
G	5	33	2.4	.48
H	5	53.5	3.8	.76
I	10	91	6.5	.65
J	2.5	30	2.1	.84
K	2.5	23.5	1.7	.68
L	5	26.5	1.9	.38
M	2.5	24.5	1.8	.72
N	4	37	2.6	.65
O	10	73.5	5.2	.52
P	10	67.5	4.8	.48
Q	15	105.5	7.5	.50

Total number of candidates - 14

Figure 29: P-value of the test items

The p-value of the test items, as the figure above shows, ranges from .38 to .84. According to the figure, the most difficult item was item L (a reading test) with the P value of .38 whereas the easiest test item was item J (also a reading test) with the p-value of .84. My analysis of the test items showed that the items of the pilot test fell reasonably well within the acceptable range and that they have discriminability of an appropriate level.

So far as the acceptance range of the p-value is concerned, experts vary in their suggestions. Henning (1997), for example, suggests .33 to .67 to be acceptable whereas McNamara (2000) recommends .33 to .66. For my project, however, I adapted Oller's (1979:247) range of .15 to .85, which means that items falling in this range can be accepted without revision. Items with p-value of .15 or lower and .85 or higher should be revised or discarded. In the new test, all items fell between .15 and .85. Although the pilot test was numerically inadequate, it did indicate that the items of the pilot test if conducted with an adequate number of participants would probably pass the item analysis test.

Item Discrimination Index

Item discrimination index is a statistical procedure to calculate a value or degree to which an item of a test discriminates high performing examinees from low performing examinees. An index of discrimination values (or D values) was prepared for all items of the test by identifying two contrasting groups of high achievers and low achievers. 27.5 percent of high achievers and 27.5 percent of low achievers on a particular test item were chosen and then a correlation between the scores of high and low achievers and the total score of the test were computed.

The discrimination values for an item range from −1.00 to +1.00. When an item has high positive value, it is considered as an item with effective discrimination power. Crocker (1992) recommends that if the discrimination value of an item is greater than or equal to .40, the item is quite effective, if it is in the range of .30 to .39, then it is satisfactory, a range of .20 and .29 means that some revision is needed, and a discrimination value of lower than .19 means that the item needs substantial revision or may be eliminated (Crocker 1992:654).

Most items of the pilot test had a D value of .30 or higher (see Appendix E, page 390 for details). This means that most test items had a satisfactory level of discriminability. Only three items (items L, M and O) needed to be revised to improve or to avoid ambiguity.

b. Consistency of scores

In order to check the consistency of test scores, I carried out two types of statistical procedures (1) internal consistency (or inter-component consistency) and (2) inter-rater reliability.

Internal Consistency

For the purpose of establishing inter-component consistency, the scores on the five sub-tests (listening, speaking, reading, writing, and the cloze) and the whole pilot were correlated:

	Listening	Speaking	Reading	Cloze	Writing
Pilot test	**.92**	**.93**	**.83**	**.80**	**.93**
Listening		.92	.74	.76	.86
Speaking			.77	.83	.85
Reading				.70	.80
Cloze					.66
Writing					

Figure 30: Correlation of scores on the pilot test and its subtests

As the Figure 29 shows, the pilot test correlated highly with its subtests. Also, the correlation between the scores of the sub-tests and the whole pilot test ranged from .66 to .92, which means that the subtest correlated reasonably highly with each other and that all sub-tests contributed reasonably well in measuring English language proficiency of the examinees.

Inter-rater reliability

As I mentioned earlier, the answers were marked by a colleague in Kathmandu, (henceforth Rater 1) and by me in Melbourne, (henceforth Rater 2). We both followed the Analytical Rating Scale given in the proposed revised test package. The scale was developed on the basis of ASLPR (Ingram 1984), ACTFL Guidelines (1986), ESL Composition Profile, ESL Scale (Mackay 1992), and Multiple-Trait Score (Shohamy 1985). (See Appendix J for details). An average of the two numerical scores was obtained and reported. For making answers on the cloze test, contextually appropriate method (Oller 1979) was used, which means that answers which were appropriate in the linguistic and communicative context of the text were awarded credit (Oller 1979).

I established a correlation coefficient of the two sets of scores obtained from Rater 1 and Rater 2. The correlation coefficient between the two sets of scores was .9, which is very high. This means that the two sets of scores correlated highly. It suggested, therefore, that the use of the proposed Analytical Rating Scale helped maintain reliability.

c. Consistency of Performance

I established consistency of performance in two ways. Firstly I compared the candidates' scores in the subtests (listening, speaking, reading, cloze and writing) of the pilot test with the overall score on the test. I found that candidates who did well overall also did well in other subtests and vice-versa.

Secondly, I compared the candidates' performance in the pilot test with that of the English test of the SLC. I found that most students who did well in the pilot test also did well in the SLC, and candidates who did not do very well in the pilot test also performed badly in the SLC (details of scores are given in Appendix E).

In this chapter, I described the framework I used to analyse the feedback and data and presented the findings in two parts. In the first part, I discussed positive as well as critical comments and the issues the participants raised on the new test package. The second part outlines the results of the pilot test (which produced satisfactory results). However, given the fact that it was only a very small scale test conducted on a small number of self-selected students, it is not possible to make any conclusive claims about the outcomes. Chapter 9 takes on some major issues that emerged from the analysis of the feedback and the pilot tests.

CHAPTER 9

Views about Problems of Context

In working with people with varied experience, we value the experience of all, we accept their experience as valid and we find ways to elicit their experience, for their voices to be heard.

- Tomal 2003:25

9.1 Introduction

Almost all the participants spoke of the contextual factors which would inhibit or prevent a revision of the School Leaving Certificate (SLC) examination, such as the one I was proposing, from taking place in the near future. The most significant factors to emerge were:

- Bureaucratic inertia
- Organisational incapability
- Lack of commitment to ELT
- Cultural problems
- Resource constraints

The following section is a brief analysis of these factors, as discussed by the participants:

9.2 The Influencing Factors

9.2.1 Bureaucratic inertia

I had several in-depth discussions about *Nepali* education in general and the SLC Examination in particular with lecturers and professors of *Tribhuvan* University who had a significant amount of experience in the school education system either as researchers, practice teaching supervisors or parents with children either in private or public (community) schools. The discussions confirmed or elaborated upon my findings

from the teachers' workshops. To the academics, the current system of education is under-developed and is based on rote learning, memorisation, and receiving knowledge as facts to be reproduced, all of which contribute an educational culture which is, as discussed in Chapter 1, dominated by examinations.

The university academics supported a common view that education in Nepal is unproductive and that it is 'product bound', that is that education in Nepal is about what the students learn rather than why and how they learn it. Teachers, most of whom have surrendered to the current tradition, applied product-based approaches to teaching rather than process-based approaches in order to get students through the system. These traditional old-fashioned ways of education hampered ELT practices rather than foster it. As an ELT academic wrote:

> *We all know that the SLC encourages rote learning, but an education system requires students to learn critical thinking, express opinions, make judgements on their own and not rely solely on memorisation. We would need to bring changes in the three areas of applications to achieve these objectives: one, to use textbooks with more social relevance, two, to introduce test items that require independent thinking or intelligent responses, and three, to bring corresponding changes in the marking and grading system (EIN: 3).*

I found that there was a tremendous amount of frustration amongst the ELT practitioners regarding the current system's failure to change and modernise. They blamed the uncertainties prevailing in the country's politics for the last two decades, and the top-level mentality, which is, to a large extent, unresponsive to change. Many participants, especially the ELT academics, agreed with the teachers in saying that the traditional education system and bureaucratic inertia might make the implementation of the proposed test revision difficult, if not impossible.

9.2.2 Organisational incapability

The ELT academia spoke of the present organisational structure of education to a large extent as responsible for its under-development, a view that was shared by the teachers as well. The centralised system of bureaucratic administration not only creates regional imbalance, inequalities in educational resources, facilities and opportunities between the urban and rural schools, but it also perpetuates a lack of coordination

among the governmental agencies and the private education sector. This contributes to the gap between the two streams and to the two-tier education system in the country.

When discussing the issue of organisation with the ELT academics and testing personnel, I asked them why the present structure was a problem. They made the following two points:

- that the bureaucratic structure is a problem because it prevents building capability at the local level.

> *The top-down nature of administration has mainly been a means to shift responsibility and accountability. This, to my mind, is a major hindrance to implementation of a project successfully and to capability building of the local level workers* (IN: 2).

- that the discrepancy between the public and private schools in outcomes and resources, especially when the students of both types of schools sit the same SLC, promotes a tendency in the parents to send children to private schools. A teacher from one public school wrote:

> *There is a difference between what the students of a private school learn and what the students of a public school learn though they follow the same curriculum. Therefore, the same test for these two groups of students is not appropriate. There should be different tests for them* (WN: 8).

This teacher's comment may sound sensible on the surface. However, the proposition of having separate examinations for the students of public and private schools could raise further problems. Firstly, it might entrench inequity and facilitate the two-tier system I discussed in Chapter 1, and secondly, it would be difficult to justify two national examinations, and therefore two standards, on one single curriculum.

- that the centralised system of education brings problems in resource distribution, creating regional variations in delivery and outcome standards.

The teachers expressed their strong dissatisfaction over the resource distribution mechanism. Most of the resources, including teaching materials, they said, remain at the centre. Even if they are distributed, they are either insufficient or not available in time. As a result of the lack of appropriate support and adequate resources, teaching and learning in rural and remote areas is, what they termed 'sub-standard' (FN: 2). The distribution crisis (Pye 1972) is ongoing and it was felt that this would in itself prevent the success of implementation of a reform.

9.2.3 Lack of Commitment and Coordination

The university academics also expressed their strong resentment over the government's approach to ELT and pointed out that ELT in Nepal suffers from a lack of commitment on the part of the government. As a consequence, the 5-year plans of the government do not even mention ELT, nor do the reports of the numerous commissions on education. There is no mechanism or agency in place to regulate, manage and monitor ELT practices. Also, due to the lack of a clear policy on ELT, there is little or no coordination among the organisations or agencies involved in the teaching and testing of the English language namely – *District Education Office* (DEO), *Curriculum Development Centre* (CDC), *National Centre for Educational Development* (NCED), *Office of the Controller of Examination* (OCE), *Faculty of Education, Tribhuvan University* and ELT experts. A university professor expressed his frustration in the following terms:

The government has failed to show its commitment to ELT. As a result, ELT is not given its due importance at the top level. For them, it is just one of the SLC subjects. In consequence, ELT cannot be treated adequately, as a special undertaking. There is no central level agency to determine the policy on ELT and to manage and monitor the implementation of the policy. There are different agencies involved in the teaching and testing of English. However, in the lack of a clear policy, there isn't a required level of co-ordination among them (IN: 1).

The testing personnel supported the view that the proposed reform of the SLC test, which advocate a combination of bottom-up and top-down processes (Chapter 10) to implement the revision, could not be monitored in a proper manner, as the agencies whose joint involvement was crucial for its success, largely remain uncoordinated.

214

9.2.4 Disruption of School and Classes through Political Actions

All three groups of participants (teachers, language testing personnel and ELT academics) were critical of what they called 'the culture of politicisation of education' in Nepal (FN: 5). By this they meant that the political parties use students as their weapons and mobilise them to exercise their influence. This practice affects the teaching/learning and testing of English badly. Many classes are missed due to strikes and *bundhs* (closures and picketing)). Unruly behaviour on the part of students could affect the testing processes also. In July 2002, for example, the students of *Tribhuvan* University set the Central Department of English ablaze simply because the majority of them had failed in the annual examination. In July 2003, the Faculty of Education office was set alight for a very similar reason. They were protesting the examination results when they should have really protested for better teaching and learning, and against the loss of teaching/learning days due to strikes and *bundhs*.

9.2.5 Cultural Problems

The dominance of the examination culture has a very damaging influence on the ELT practices, according to the participants. Some academics and some members of the CDC were very critical about the effect of the examination culture on secondary education. One said:

> *School education, particularly the secondary level education equals examination taking. Secondary schools throughout the country become oriented around preparing students to take up the examinations successfully. This means that students are burdened with homework, pressures to succeed, supplementary classes, study aids and practice examinations. They are not encouraged to be creative or solve problems, but rather rote-learn or memorise answers of model questions* (IN: 4).

The participants pointed out over and over that Nepali education, particularly school education, was dominated by the SLC examination. The SLC was what may be termed as the 'baton' by which students, teachers, schools and parents were 'conducted'. The SLC certificate was the key to upward, social and educational

mobility, and the stepping-stone into desired professional fields. A public school teacher who was also a parent expressed his opinion as follows:

I have paid a 100,000 rupee-donation to get my son into the school, which is in addition to 5,000 rupees I pay every month. I have done this so that my son can study in this school, which has the reputation of being the best in the city and gets the most students through the SLC in the higher first division. It is really important to get a first division certificate to get into a good college, to get into the course of one's choice or to get a good job (IN: 4).

The story illustrates the degree to which the SLC dominates education and is the test which determines a possibility of a life of affluence, or of economic marginalisation. The better the results, the better chance one had to get into college or a course of one's choice. The students, and all others involved in this, employ every possible means, fair and unfair, to get a high result. Bribery (in the form of donation) or the use of political influence are part and parcel of this, such is the desperation of families to ensure a secure life for their offspring.

The academics were highly critical of the meritocratic approach (a tendency to see merit in a diploma) to the SLC that creates what is sometimes known as the 'diploma disease' (Simon 2000:54) in the sense that every school graduate wanted to get the SLC, regardless of what they did for it. The meritocratic approach also exacerbated the negative 'washback' effect where all teaching and learning became increasingly oriented towards preparing students to examination tactics. As a consequence of this, the school and teachers' performance was measured by the proportion of successful candidates in the SLC.

Discontent was also quite high among the English language teachers, particularly those of a new generation and others seeking change. During the workshops, a significantly high number of teachers condemned the way they had to teach. They disliked how the examination dominated system forced them to concentrate solely on the textbooks disregarding oral skills and students' creativity. During group presentations, one of the teachers described his experience in the following terms:

I hate to have to teach with the pressure to finish the textbooks on time, to teach students exam tactics, and to write model answers for them so that they can memorise and reproduce them in the examinations, rather than develop their practical skills in the target language (FN: 4).

To reform such a firmly established tradition is, as the teachers pointed out, a Herculean task, and needs a properly coordinated, adequately supported and funded effort. Given the fact that the political situation was fragile and unpredictable, it seems rather unlikely that the current government would support reform of this kind.

9.2.6 The Resource Constraints

In response to my question on the constraints that might make the implementation of the reform package difficult, the ELT academics agreed with the representatives from CDC and NCED in saying that the OCE is severely under-funded. Because of the lack of sufficient funds, the OCE is not able to develop necessary infrastructure, nor could it develop enough human resources. As such, as an ELT academic pointed out, any reform proposal, which required support from modern technology and trained people, was difficult to implement and sustain unless necessary resources were provided:

There is a great deal of cost involved if the proposed reform is implemented in a proper manner. For example, money will be needed to develop necessary expertise at all level, to develop schemes and infrastructure for them, for training of teachers, for equipment, and so on and so forth. At the moment, the concerned agency does not simple(sic) have that kind of resource available (IN: 4).

So far as human resources are concerned, the group of ELT personnel pointed out that the proposed package could not be implemented successfully and effectively, because the people who would be involved in that process are not appropriately qualified or adequately trained. Without sufficient training, the present testing practitioners would be unable to manage it. One claimed:

In order to handle the reform well, the testing personnel (test developers, administrators and examiners) have to be trained. So, I suggest that training of the testing personnel should be an integral part of the package (FN: 8).

The participants noted that the lack of resource is very much responsible for the present deplorable state of the SLC. Adequate funding must be obtained, required equipment and materials accumulated and necessary expertise developed in order to

improve the situation. When asked that given the fact that resources – both human and material – were so scarce, what chances there were for the proposed package to be adapted, one of the NCED personnel said:

Chances are that your package, in the existing circumstances, will go unattended or remain as an academic exercise only. However, there is a fair chance also that this would become a basis for seeking funding with donor agencies (IN: 5).

In the section above, I have detailed the circumstances that the research participants said are problems or factors militating against the implementation of the proposed revision. I have also outlined the growing discontent among ELT practitioners regarding the current uncertain state of ELT, which is so dominated by 'the examination culture'.

There was a consensus among the research participants that, despite the circumstances preventing a proper reform or revision of the English test of the SLC, it is imperative that such a revision takes place because:

- the current structure of the OCE is very rigidly centre-oriented. Every decision regarding the SLC is taken at the centre and then is handed over to the next level in the structure for implementation. This is obstructive in that the actual ELT practitioners do not have any involvement in the decision making. The ELT practitioners and academics alike strongly criticised this practice and suggested that the OCE should be decentralised, creating a mechanism to effectively implement what is agreed upon at a given level. One participant seemed to sum up the general view:

... for any reform to be feasible and effective, it must move from bottom to top. It is the 'top-down' attitude 'which makes the system weak and the implementation of a reform difficult and ineffective (IN: 3).

- the structure or format of the current (existing) test of English is like a 'new paint on the old wall' (IN: 1) in that they (the OCE) have changed the outward appearance, not the inner nature of the test.

There is no relationship between the curriculum and otherwise a very sketchy [existing] specification grid. It does not specify the skills and sub-

skills that should be tested. Consequently, the existing English test of the SLC resembles the old one, at least in its content any way (WN: 9).

- the existing test lacks validity and systematisation in that it did not reveal a true picture of the candidate's language proficiency (FN: 1, FN: 5). And as a teacher from a private school notes below, there is no practice of piloting or pre-testing test items:

> *The massive failure in the examination may have been due to the fact that tests are used in the national tests without having their authenticity and validity assessed* (IN: 2).

- the current testing scheme, as another workshop participant pointed out, was not transparent as it was subjective and did not follow any performance rating scale:

> *There aren't well defined, pre-determined criteria for testing speaking and listening. As a result, their assessment is highly subjective, solely done on the assessor's discretion* (FN: 23).

- the stereotypical nature of the test encouraged 'cheating' (test questions could be easily guessed and 'guess papers' could be bought prior to the examination). The system enabled candidates to sit the examination with readymade answers in their pocket.

- that the provision of parallel examinations, introduced recently, was not working, because it was not based on any properly developed test structure or test specification. A group report said:

> *The scheme of parallel questions introduced recently to check cheating is not working because the parallel sets are not of equal difficulty. As a result, some students get easy questions while others get difficult ones* (FN: 15).

Conclusions

There are profound institutional as well as systemic problems which contribute to the deplorable state of the English test of the SLC in Nepal. The ELT practitioners

want change in the existing teaching and testing practices. The new test package, despite its merits, could not be implemented because of the factors discussed.

Clearly the implications are that although reform in the existing English test of the SLC is needed, it is not possible at this time for the proposed test reform package to be implemented in the current situation. The participants in the study identified four pre-requisites as necessary to ensure effective implementation of the reform. They were (a) the commitment of the government to reform, (b) adequate resources, both human and monetary, (c) sufficient preparation in terms of (material) development and training, and (d) administrative provision for monitoring and supervision.

The participants expressed the view that in order to implement the new test package, it is imperative that an awareness or orientation campaign involving all stakeholders at all levels is organised. This could be done, as the excerpts below suggest, by organising discussions, debates, orientation meetings and publications at the central, regional and local levels:

> *Revision of the SLC should be a bottom-up process. The ELT practitioners' mentality should be prepared for the change. Awareness in the people concerned is an important step to achieve this. This will make people realise why, how, what change is necessary* (IN: 1).

> *A general public awareness about the proposed revision needs to be created. For these debates, orientation programmes, discussions should be held at all levels – administrators, test developers, examination implementers /moderators, assessors/examiners, and those involved in reporting test results, teaching and teacher training* (IN: 4).

The participants also expressed the opinion that the concerned agencies should do necessary homework before introducing the new test. They should, for example, develop required human resources, train the concerned teachers, acquire and develop necessary equipment and materials and create a favourable atmosphere for the revision. A mechanism for reviewing the revision process should be in-built and on going.

While I was gratified by the response from my colleagues at one level, at another I was disappointed in that it seemed that my reform package could not be implemented any time in the near future.

The implications leave three options for the stakeholders:

a. Given the complexity of the constraints, do not implement the new test package, and let the current English test continue. In consequence, this will lead to a further downward spiral of inappropriate teaching and learning of English. It also means that the aim of improving the quality of teaching and learning of English will remain largely unachieved.

b. Attempt to resolve the constraints and implement the new test package. Changing the SLC English test will also mean changing the ways of its teaching and learning. However, given the fact that the constraints are real, this is totally unrealistic.

c. Implement the proposed test reform in phases. Given the intensity of the current constraints, phased implementation, with continual revision of the concerned proposal, may also reduce some of the problems incrementally. This, however, means that the optimal range of positive outcomes will not be achieved at one stage, but they will emerge in successive stages.

These then are the findings of my consultations and fieldwork in Nepal. I am left with a patchwork of critiques and suggestions that would have been woven into the next round of test development, had the circumstances in Nepal been more propitious. However, the poverty, under-development, corruption and bureaucratic dysfunction that is typical of so many 'developing' countries is a fact of life that cannot be easily changed and must be addressed in the context of targeted educational reforms such as this one.

In the final chapter, I look back on the research process and the findings, reflect on my journey as a researcher and consider some of the issues emerging from the research.

CHAPTER 10

A Final Reflection

Initiation of change never occurs without an advocate.

- Fullan 1991: 54

10.1 Introduction

In this final chapter, I reiterate the main outcomes and findings and reflect back on the research journey I have undertaken over the last four years: I summarise the research project as a whole; and its constraints and limitations; I present the model for the process of test development that I have developed; I describe my learning as a researcher and action researcher, I comment on the issues that have arisen out of my attempts to reform and revise the *School Leaving Certificate* (SLC) Examination; I outline possibilities for a what would be a 3rd cycle of the action research; I briefly reflect on the SLC and how I now see its potential role in the long-term project of reforming ELT in Nepal; and finally, I make some suggestions as to what more could be done in order to build an English language education culture and research culture in Nepal.

10.2 Recapitulating the Research

English language education in Nepal is old-fashioned, product based and unproductive. ELT is in a bad shape and lacks a clear state policy. ELT suffers from a lack of a trained teaching force, mismanagement, under-development of basic infrastructure and inadequate resources. Moreover, the current English test of the SLC is outmoded, poorly conceived and badly developed, using a weak language test format. It does not produce credible and consistent results. More importantly, it does not reveal

222

the actual language proficiency of a candidate. The test exerts negative washback effects on teaching and learning of English at the SLC level

This project was aimed at producing a new and improved SLC test, while at the same time modelling a process for developing such a test.

My research project began with a general idea about addressing the problems the English test of the SLC faced. I framed the idea into a researchable question – how could I develop an alternative English test that would be appropriate and acceptable in Nepal. The project broadly followed an action research process which consisted of two full cycles or reflection, planning, action, observation and planning for new action. It was accomplished in 12 steps from the identification and analysis of the problem through reflection on the research journey, and finally to my conceptualising a third possible cycle moving further towards implementation of the revised test and addressing the problem of the SLC on a number of other levels. These 12 steps are illustrated in the flow char given below

Steps of my action research project

Cycle One   **Cycle Two**
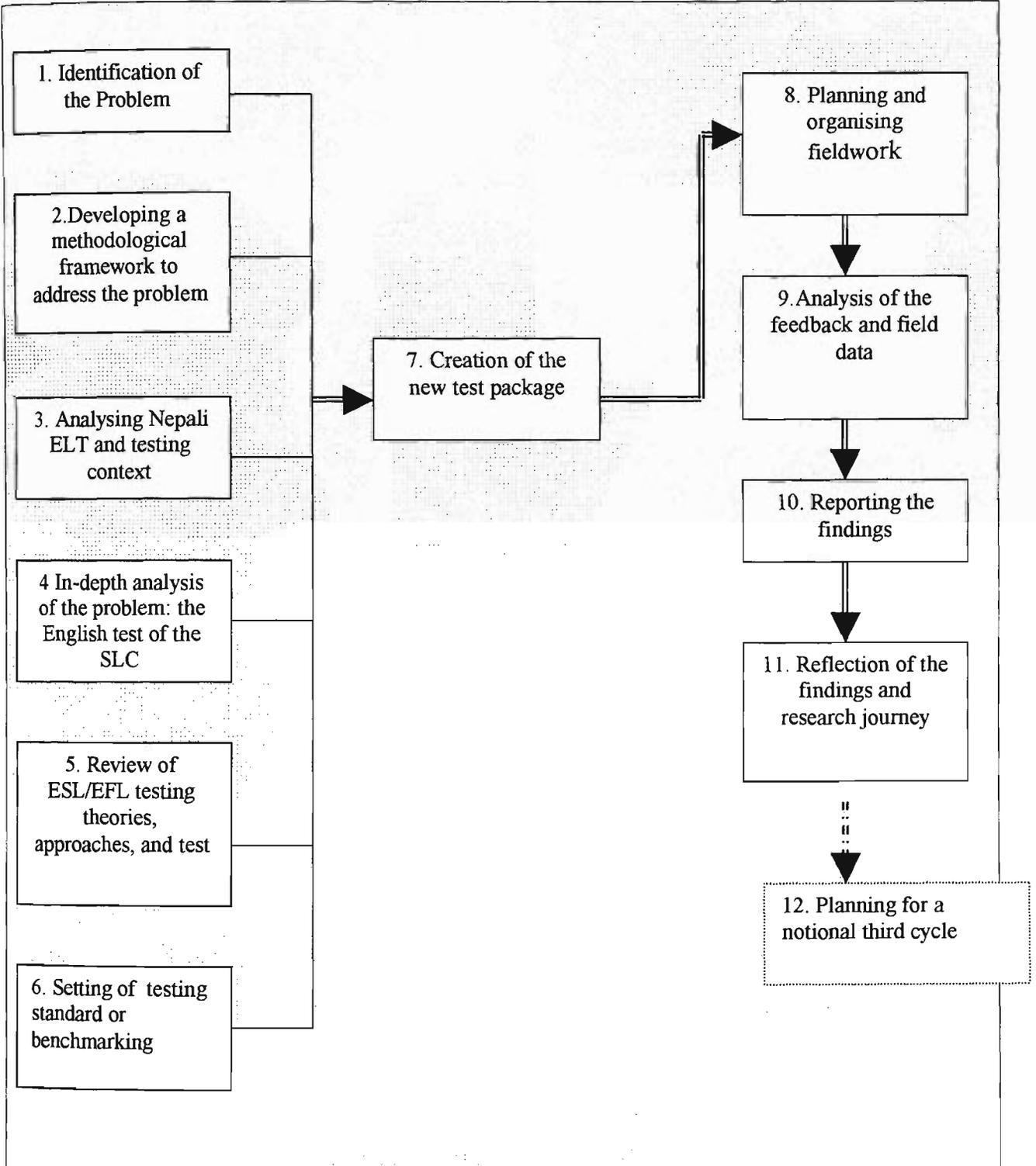
Figure 31: Steps of the present research project

Eleven of the 12 steps of the research were completed in two cycles. Cycle one consisted of steps from 1 to 7 (chapters 1-6), and cycle 2 steps 8-11 (chapters 7-10). Step 12 is a notional third cycle, which can be built upon as a follow up of the present project. A brief accounts of the steps is given below:

10.3 Modelling the Test Development Process

I first reviewed some of the most common North American, British and Australian language (especially ESL/EFL) testing approaches and models. I analysed some of the widely used ESL/EFL tests in order to learn what aspects of these approaches and tests could be adapted to designing a new English test.

I aimed at maintaining a balance between reliability and validity which I achieved by detailing the English language teaching and testing purposes, developing an elaborate test specification, using an analytical rating scale and by carefully selecting materials for the test tasks. I also resolved the issue of maintaining a balance between tests of academic ability and tests that assess communicative abilities by including tests of study skills and communicative skills. In my new test, I had to include tests of proficiency as well as tests of achievement because the SLC English test is a proficiency as well as an achievement test. I also had to maintain a balance between tests of productive and receptive skills.

The new test package consisted of (a) a test specification, (b) a test structure, (c) a pilot test, (d) a marking/interpretation scheme, and (e) a process for developing a new test. I employed a five-stage process to developing the new test package:

1. The first stage of conceptualisation provided a theoretical foundation necessary to enhance or modify the 'belief' or understanding of second or foreign language testing, the context of assessment and its constituent parts. In other words, the first stage of the test adaptation/development process offered not only the theoretical underpinnings of language testing (chapters 4 and 5) but also made me (as a tester) aware of what was to be tested (curriculum goals) before determining why it was to be tested (purpose of testing), and for whom it was to be tested (interpretation of the test outcomes) (chapters 1 and 3).

225

2. In the second stage, learning standards or benchmarks were set on the basis of the analysis of the secondary English curriculum and prescribed texts. This became the basis of setting assessment goals in the test specification (For details, see Appendix A).

3. The third stage was the development of the test and test tasks and consists of processes such as construction of test tasks, piloting and evaluating them. The new test types and test tasks now closely reflected the discourse of the academic and social settings. A structural linkage or coordination among the tests of various skills was maintained through a common theme or context. The test of listening, for example, served as prelude for the speaking test whereas the writing test was a follow up and a logical extension of the theme, concept and/or language items under test (Chapter 6).

4. In the fourth stage, I developed an analytical marking scheme based on a five-point rating scale.

5. The fifth stage was the development of a test outcome reporting mechanism for which I designed a new scheme to report test outcomes with information and descriptors to help users to better understand the scores (Chapter 6).

The issues of balancing test types, test tasks, tests of productive and receptive skills and in validity and reliability were addressed in stages 2 and 3. In this way, I completed the first cycle of the research. A diagrammatic representation of the test development process is given below:

```
┌─────────────────────────────────────────────────────┐
│      Stage 1: Theoretical Conceptualisation          │
└─────────────────────────────────────────────────────┘
                          ▼
┌─────────────────────────────────────────────────────┐
│      Stage 2: Setting Testing Standards or           │
│      Benchmarking                                     │
└─────────────────────────────────────────────────────┘
                          ▼
┌─────────────────────────────────────────────────────┐
│         Stage 3: Test Development                     │
└─────────────────────────────────────────────────────┘
                          ▼
┌─────────────────────────────────────────────────────┐
│    Stage 4: Marking/Rating Test Performance          │
└─────────────────────────────────────────────────────┘
                          ▼
┌─────────────────────────────────────────────────────┐
│      Stage 5: Interpreting Test Outcomes             │
└─────────────────────────────────────────────────────┘
```
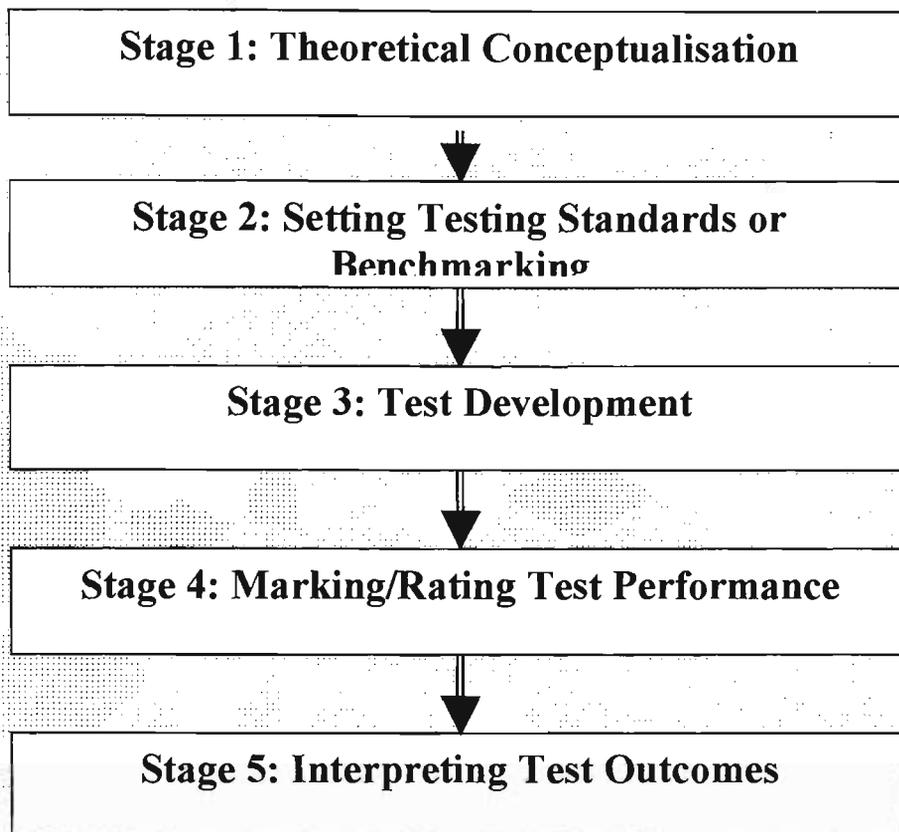
Figure 32: The Test Development Process

10.4 Outcomes of the Consultation and Pilot Test

I took the new test package to Nepal where I involved a number of stakeholders – ELT academics, teacher educators, testing personnel, (test developers and examiners), and teachers. I organised a number of activities to obtain feedback. I gave the pilot test to a small number of students and engaged them in a post-test group-interview (chapter 7).

The findings from the consultations with the stakeholders and pilot test are summarised below:

I conducted a series of workshops and interviews in two districts of Nepal. Feedback on the new test package was also obtained from a plenary presentation at an international conference and five articles on various aspects of the research.

Feedback from these consultations produced a mixed result:

- the new test package was useful and in line with the recent change in the ELT curriculum in Nepal,

- the pilot test was an improvement over the existing English test of the SLC,

- the proposed marking and interpretation scheme was not practicable, it needed improvement,

- proposed minimum cut-off point and score boundaries were impractical in the context of Nepal, they needed revision,

- the new test package could not be implemented in the existing circumstances as a number of educational, political and economic factors were militating against it,

- an extensive training and professional development programme would be needed,

- the structural provisions and capability of the OCE would have to be improved,

- the political and economic issues would have to be resolved before the new package could be put to practice.

The pilot test (whilst limited) produced satisfactory results:

- the success rate on the pilot test was 86 percent,

- despite the teachers' concern that raising of the minimum cut-off points would end up failing more students on the proposed testing scheme than the existing scheme, the pilot test produced better results than the existing test,

- students of both public and private schools performed better overall,

- students' performance was better in reading and writing than in listening and speaking,

- the pilot test had face validity, construct validity and content validity,

- it had a high correlation with the English test of the SLC (2003), which was administered about one and a half months later than the pilot test, revealing that it had a predictive validity as well,

228

- most test items of the pilot test had an acceptable level of difficulty and were able to discriminate between high and low achievers,

- the scores on the sub-tests were consistent and there was a high inter-rater reliability (see chapter 9).

10.5 Reflection on the Research Journey

I chose to work on the English test of the *School Leaving Certificate* (SLC) examination because of its high stake nature and the fact that it determined the social, occupational and academic future of every individual, because it is fatally flawed, and because it exerts profound washback effects on what goes on in a language classroom. I therefore felt that changing and revising the SLC might bring about other changes in classroom practices and contribute to educational equity.

Initially, I focused on obtaining quantitative evidence from the pilot implementation of the test to prove its effectiveness. I came to realise that the complex dimensions of developing a new test suitable for the challenging conditions of Nepal could not be measured by quantitative analyses. I also realised that a new up-to-date test alone was unlikely to solve the problem, as the problem was not only the test but also the process and factors or issues associated with the process of developing it. I revised my project accordingly.

In 2000, I left Nepal amidst the chaos of civil violence and political turmoil. Several rounds of ceasefires, peace talks, all-out wars and negotiation meetings have taken place during this period. The annihilation of the royal family also took place in this period in which the king, queen, and crown prince were massacred. The donor agencies, which finance school education in Nepal, threatened to withdraw their support if the situation did not improve. The trauma of these events have affected the way I think about the role of the research and at times have introduced a note of despair that education will ever really change for the better in Nepal.

Recently (May 2004), a new government has been put in place and the political parties have shown their support. Both the government and the rebels have shown signs of compromise in their approach to resolving the conflict. The donors have temporarily

offered to continue their support. It is a sorry situation and the hope for reform and indeed survival of public education in Nepal is reliant on foreign aid.

In terms of action research, I have assumed a dual role of 'insider' (as past English teacher and present teacher trainer and advocate for change) and 'outsider' (as expert researcher applying 'Western' methods of research and test development). In the initial stages of the project, I worked as an insider. I acquainted myself with the wide range of inconsistencies and shortcomings that existed in the English language teaching and testing policy and practices in Nepal. Having defined the problem and designed a plan of action, I became an outsider seeking the opinion of others and feedback from other 'insiders' to see what they had to say about the action plan.

Throughout the process of analysis and writing of the findings, I have struggled to represent the different voices and to give each a fair share in the conclusions that I make. My 'insider' self and 'outsider' self at times contradicted each other. Though I maintained an academic way of looking at things, I found, especially when I was back in Nepal, that the force of that culture (which is my own) led me to take up conservative and traditional standpoints. My perspectives were conflicting, shifting and changing throughout the whole process, until I achieved the balance that is represented in this thesis.

Normally, a large number and wide range of stakeholders would participate in such a complex and politically fraught project as the revision of the SLC. However, this was obviously impossible given my timeframe and limitations of a PhD study. For the consultation, I therefore narrowed my focus to key stakeholders living in Kathmandu and Chitwan districts, both in the central region.

Given the focus of this thesis on the process of test development, it was not possible to carry out a pilot test that would be statistically significant (a pilot of about 600 students would be needed for that). I therefore sought to test the new test on 60 students (30 from a private school, and the same number from a public school). However, my expectation did not materialise for the reasons I outlined in Chapter 7. These reasons are part of the context of underdevelopment that is the underlying problem in the first place. The small number of participants in the pilot test does weaken any claims regarding the validity and reliability of the test. Nonetheless, the

positive outcomes that the test revealed show that the new test package has potential merit and validity.

I have tried my level best, within the limitation of the methodological framework; to give the range of research activities described in the thesis a fair and complete analysis. However, the lack of participants from the *Officer for Control of Examinations* (OCE) is a weakness in this regard. Again, their reluctance to participate can be seen in the context of other environmental factors I have discussed.

Despite these hitches, my research journey has been a fulfilling one in the sense that I have achieved most of my research goals: I have developed a vastly improved test for the specific conditions of the SLC in Nepal. From the consultations with the ESL/EFL testing academia in and beyond Melbourne during this period, and the courses and training I have had in the field of language testing. I have learned much. At the same time, I have developed my approach and capability as a researcher by orienting the whole project as an action research project and learning skills of qualitative research and evaluation.

10.6 . The Development of the New English Test as Action Research

The whole project can be conceived as a "spiral" consisting of two action research cycles. In the first cycle, I analysed the problem and explored the context (the 'reconnaissance' phase), analysed the tests, models and approaches from which I could learn (the planning phase), designed test specifications and developed a new test package for the English test of the SLC (the action phase) (Actually, this process also consisted of several much smaller cycles of action and reflection as I went back and forth from theory, knowledge of context and the curriculum to design an optimum model.). The final phase of the first cycle (the observation phase) led into the 2nd cycle. This consisted of assessing what kind of feedback was needed and possible ('reconnaissance' developing the methodology and logistics of my field trip to Nepal (planning), carrying out the interviews, workshops and pilot testing (the 'action') and reflection on and analysis of the results (observation/reflection). A hypothetical third stage might entail a revision of the test specifications and sample test, and a statistically valid test pilot. However, as my discussion in Chapter 9 demonstrates, more needs to be

done to advocate and organise for a demand for change from the 'bottom up', as well as from the 'top-down' before any real change can take place. The third cycle, therefore, might consist of my activating with my colleagues at the grass roots level, rather than, or as a prelude to, attempting to introduce the test from the top.

In summary, I oriented the new test towards the notion of 'communicative competence'. The individual tasks of the test ensured that the students interacted with the test tasks and used their own words to respond to them. In other words, the candidates were made to use their linguistic, communicative, discourse and strategic competence to perform the tasks of the pilot test.

10.7 The Emerging Issues

10.7.1 Change is Desired

The majority of ELT teachers and practitioners in Nepal are frustrated with the current state of the existing ELT practice and want change. They want the English test of the SLC to change despite the administrative resistance and resource constraints. They see the proposed package as a likely intervention, with the potential of changing English language teaching practices. The majority view was that the concerned authority should find ways and means to implement it.

10.7.2 Need for Training and Professional Development

Training and professional development in all aspects of testing is an essential condition for success. Developing new skills in the people concerned and training them how to use new ideas should be a part of a reform process.

Training and professional development of the change agents, in this case, the teachers and testing personnel, was a major concern of the stakeholders. I believe that the process of any reform begins with the development of its agents. The present project, as I describe in the next section, envisages training and professional development of the working force necessary to ensure effective implementation of the proposed schemes. As an extension of the proposed project, short term training

packages and professional development programmes for the concerned teachers and testing personnel can be developed. These training packages may be more effectively carried out as an integral part of the existing training, rather than as a separate training programme.

10.7.3 Need for Funding and Infrastructure

Much of the resistance I faced during my fieldwork was related to the fact that funds to implement the process were simply not available. Many participants said that implementing the new package would require a huge amount of money and that the needed infrastructure was not in place. However, as I discussed earlier, a reform usually requires new skills and new actions. This means that, in order for the OCE to be able to implement the new test package, it would require new expertise, new organisational provisions and above all new resources for training and professional development. An adequate level of funding, therefore, must be obtained. The current donors of education could be involved in the planning process to work out a scheme to accumulate funding for the project in a gradual manner. In Appendix J, I put the case for the reform of the SLC to take place in five stages of implementation. Even this gradual and cautious implementation, however, would, require a stable political situation.

During the fieldwork, I experienced some degree of resistance from the language testing personnel and teachers. Their resistance was apparently about the practicability of some of the proposed schemes especially in the context of the present teaching force largely being untrained. I acknowledge the need of training. However, there seemed to be a subtext to their resistance: perhaps, as mentioned in the last chapter, fear of the unknown, 'upsetting the apple cart', or simply, of having their own inadequacies exposed by the new system (Fullan 1993).

I would say the same about the bureaucratic resistance. However, this could be mitigated if the changes took place as part of the development of a new educational and organisational culture involving all stakeholders

10.8 Facing up to the Issues

Reforming the SLC test should be viewed as a continuous, constantly evolving process. In the process, problems should be expected, detected and dealt with, rather than ignored, overlooked or denied (Fullan 1993:37; Brady 1995; Elliott 1996). Dealing with problems and issues should be continuous and should be viewed as a learning exercise, rather than an opportunity for shifting responsibility or accountability.

The fieldwork in this research picks up on an extensive literature on educational reform in Nepal, arguing against the top-down bureaucratic system. The system exercises excessive control over the local districts (Fullan 1993:37). The periodic successes obtained through centralised policy are often short-lived because implementation gradually becomes an exercise of power or authority. On the other hand, the bottom-up process, which is often viewed as the most effective process of reform (Rijal 2001), may sound good in principle but it becomes chaotic if there is no control or coordination from above. Therefore, both the top-down and bottom-up strategies are necessary. Reform of the SLC in Nepal would be no exception to this. As Fullan writes:

> What is required is a different two-way relationship of pressure, support and continuous relationship. It amounts to simultaneous top-down bottom up influence. Individuals and groups, who cannot manage this paradox, become whipsawed by the crosscutting force of change (Fullan 1993:39).

An example of this can be found in the Bihar Education Project in North India, which I visited in 1997 as a member of Plan International study team. During the planning meetings of this UNICEF funded project, education bureaucrats, donors, international consultants and grassroots level workers all sat together to set educational goals, formulate programmes and develop implementation strategies. Together they worked out plans to deal with constraints and limitations, a practice which could rarely be conceived in the highly bureaucratic society of Nepal.

A developing country like Nepal, in order to avoid the high cost of different layers of the bureaucratic processes, should adopt a more practical and working level process in which education workers of all levels, from education minister to primary level

workers, along with international consultants, sit together to deliberate on and plan a reform programme. In the same manner, implementation strategies can be developed, solutions of problems can be proposed, and support mechanisms can be designed. The authority would tell the meeting what they wanted, the workers would suggest what might happen and how it would happen. This, in my view, is an important condition for any reform programme to succeed in the sense that workers of all levels build a consensus together, and initiate change with a common motivation and understanding. For this to happen, there needs to be a new political, educational and organisational culture. There needs to be a change in the political will of educational workers and bureaucrats at the top. However, the prevailing political and economic conditions are not conducive for such consensus formation to take place. The problems or issues raised by participants do not necessarily negate the merit, or for that matter, the need for the new test package. What it means is that these issues should be addressed as a priority if this or any other test package is to be implemented.

Educational reform in Nepal should therefore be regarded as both a negotiated process between the 'top' and 'the bottom' and, as a phased process in which one phase is planned and implemented in one area before moving on to another. The lessons learned from one phase should become the guiding principles for the next.

If the new test package were to be implemented, it is crucial that the issues discussed earlier in this chapter, i.e. the need for consensus, readiness in the people to be involved (training and orientation), structural improvement of the OCE and infrastructure and resources must be resolved. It would not be possible to resolve these issues all at once. Therefore, I am proposing a phased implementation plan. The proposed phased plan is in Appendix J.

There are some encouraging signs showing improvement in the people's understanding about practice of English language testing. A debate, for example, has begun as to how to link the SLC English test with school based assessment practices (KOL, 27 Mar. 2004). Secondly, the *Nepal English Language Teachers' Association* (NELTA) has asked my permission to use some part of the project material for their training purposes. And thirdly, an international consultant currently working for the private schools of Nepal has shown interest in using some of the materials of the present project for her teacher training programmes.

In the course of the project, I have demonstrated a method and a process for applying state-of-the-art language testing models and techniques to the particular context of a poor, third world country, namely Nepal. I have demonstrated that language test development is a cyclic process (test construction, test piloting, test evaluation and then test revision) which lends itself to action research as a way of framing that process. I have learned in the course of the project that no real change can occur without adequate resources being made available for implementation and training. I have become more convinced than ever that reform of the SLC should be based on the empowerment of and participation of all stakeholders; and that the new test package I have developed could be implemented if the resources needed were to be made available and phased in, in manageable stages

10.9 Reflecting on the Role of the SLC as a Possible Vehicle for Change

This thesis documents a number of studies which focus on the role of a national examination, and demonstrate to what extent and in what way a test can become a basis for improving instructional practices (see Chapter 3). As Brady's work (1995) indicates, examination is a powerful instrument and strong catalyst for change. Brady suggests that in the process model of educational reform, the content is relatively less important than the process and the purpose. The goals of instruction in this model are derived from current theories of the discipline. However tests and examinations shape the nature and ingredients of goals, content and process of instruction. In turn, the test or examination is shaped by the above factors in a dialectical process. The effect of a test on modes of instruction is, thus, inevitable. How it affects the instruction, however, depends on its nature, and the skills and levels of training of teachers.

Similarly, Davies (1985), Alderson (1986), Noah and Eckstein (1992), London (1997) and Taylor (2004) reported on the advantage of using a high stake, national examination as a change agent. Davies (1984) and Khaniya (1990) suggested that the SLC could be used as a basis for instructional amelioration in Nepal.

Clearly, the new SLC test has potential to reform and improve ELT in Nepal. It specifies language skills and abilities to be tested, outlines the procedures for testing them, and reflects a required level of proficiency. In order to meet the requirement of

236

the test, it can be expected that the teachers would change their teaching strategies in ways which would improve English language instruction.

In a country like Nepal where education is constrained by severe resource limitations, examination can be a low cost means to improvement. As Heyneman and Ransom (1992) writes:

> Especially in the context of scarce resources with declining educational standards and the ever increasing demands for better qualified manpower, education officials are looking for low-cost ways to improve their education system...examination can be powerful, low cost means of influencing the quality of what teachers teach and what students learn in school (Heyneman and Ransom 1992:109).

10.10 The Way Forward?

Nepal introduced an educational reform in the mid 1990's in the form of curriculum revision, new set of textbooks and limited teacher training. However, classroom practices have remained unchanged as confirmed by Singh 1996; CERID 1996; Khanal 1997; CDC 1999; OCE 2000; CERID 2001 and Poudel 2004.

Nepal's experiment with the objective and interactive models of educational revision was unsuccessful because it had been isolated and piecemeal practice and it largely ignored examination (Khaniya 1990). It is time that Nepal adopted what is known as "the examination improvement approach" to educational reform (Kellaghan 1992:102). According to this approach, ingredients and make up of a high-stake test like the SLC will provide a basis for improving the instructional practices. Therefore the plan for improving an instructional scheme begins with a plan for improving its examination.

This project began with the hypothesis that a new test package would have the potential of changing or improving ELT practice in Nepal. However, given the limited scope, time and resource, it was obviously not possible to provide empirical evidence to support this idea. Therefore, the next step forward or **the third cycle** of the project could be to work with a school or school network, link the new test package with the school-based (quarterly, half-yearly and yearly) tests, work with the teachers concerned and see whether and how change in the assessment schemes change their ELT practice.

On my return to Nepal, I propose to find funding and gain support to expand the pilot test into a full-fledged 3 hour long test, administered on a sizeable population of the school graduates.

I will attempt to raise funds and to develop training and professional development programmes for teachers and other testing personnel. For this I would work with governmental and non-governmental agencies supporting secondary education to work out modules for short term, medium term and long term training programmes.

10.11 Building an ELT Research Culture

In order to build a better English language teaching system, Nepali educators need to work on four areas – ELT policy development, improvement of teaching and testing methods, gaining resources, and developing a research culture. For me, one of the main outcomes of this research is that it demonstrates the need for the development of an appropriate, national ELT policy for Nepal. This must provide input in the area of language testing policy. A new policy should not only provide a direction for future ELT practices, but it should also provide a basis for structural change, accumulation and distribution of resources, and for developing necessary infrastructure. In addition, it should provide directives to the ELT practitioners for developing appropriate goals for improving classroom practices. For example, a central level institute for ELT can be established which, with its regional and local units throughout the country, will be responsible for its policy, development and management of teaching/learning and teacher training programmes, implementation of such programmes, supervision and monitoring. Tests of English at all levels can be developed and administered in collaboration with agencies like the OCE and the institute.

Furthermore, improvement in the existing ELT practice should be a constant focus of concern and action. In order for the improvement to be a continuous and effective process, it is important that we build a research culture as a basis for informed decisions. Various aspects of current ELT practice should be researched. Based on the present research, I suggest the following:

1. Ironically nearly fifty percent of English language teachers in Nepali public schools are trained whereas only a handful of private school English teachers have received formal training. However, it is the private school students who are fluent in English and secure most places in the SLC results. So, a research project might explore what differences exist between the English teachers of these two types of schools in their methods of teaching English. This project can address questions like what are the factors that influence students' success (or failure) in the SLC. The findings from such a study can provide valuable input for developing training programmes for the teachers of the public schools.

2. A further project could be to investigate how teachers in the private schools organise school-based language assessment; how the assessment reflects what they do in the language classroom and to what extent it contributes to the development of English language proficiency in the students. Again the outcome of this study could be used as a basis for developing programmes for the public schools.

3. In the introductory chapter, I noted out that Nepal has a two stream system of education in which students of private schools take most of the educational opportunities by claiming 80 per cent of the SLC results. As the private school students are mostly from well-to-do families, it is clear that the present system of education creates social injustice and that the children of the majority of poor *Nepalis* are being unfairly discriminated against. Furthermore, the SLC as a high stake examination creates a 'diploma disease' (Simon 2000:54) or 'examination mania' (Foster 1992:123) and puts the health and life of many children at risk. Therefore, an inter-disciplinary project could be to investigate how different groups of people view the SLC, whether and how the SLC examination creates division in the Nepali population, and how it affects the health and life of the examination takers and/or stakeholders.

4. Analysis of the pilot test outcomes reveals that the use of the proposed analytical rating scale improved the marking performances of the raters in two ways. Firstly, it provided more information on the candidates' performances (what a successful candidate at a particular level can or cannot do in the target language). Secondly, the high level of inter-rater reliability between the performances of the raters showed that the use of the rating scale could provide consistent scores. However, it should be further investigated to what extent the performance of the raters who use an analytical rating scale differs from those who do not.

5. In this age of communication technology, use of the computer in teaching ESL and EFL is becoming more and more commonplace every day even in a developing country. It can be expected that in the next five years, the computer will be extensively used for teaching and testing purposes in Nepal. It should, therefore, be explored further what part of the package could be tested directly on the computer.

6. If the proposed new test package is implemented, an important project could be to research its wash-back effects. Such a project can address questions about the effect of a healthy relationship between teaching and testing, how educationally beneficial (i.e. having positive wash-back effects) a test can be, how a test can exert beneficial effects in general, and what can be done to ensure that a test has beneficial effects.

In this project, I have attempted to connect the theories of second or foreign language testing (as developed in and for Western contexts) to the practice of the English language testing at the SLC level in Nepal. The project, on one level is a contribution to raising the understanding and awareness of the ELT practitioners. To what extent this attempt has been successful is perhaps for the reader to judge, or else, may be the topic for an evaluation study further down the track, an additional cycle of reflection, planning, action and observation.

Finally, the epigraph given in the beginning of the chapter is that 'change needs an advocate'. This project can be used as a basis to advocate for change through reforming the English test of the SLC. Change is needed to reform the entire school level ELT in Nepal. My research, and the writing of this thesis, is perhaps a small step in that direction.

Bibliography

Abbott, M. M. (1982). Language Assessment Battery. Paper Presented to Arizona State Department of Education. Phoenix, Arizona, USA.

Abidine, J. B. T. (1991). "Oral Testing of English in Malian Secondary Education." Language Testing Update(9): 23-25.

Abraham, R. G. and C. A. Chappelle (1992). "The Meaning of Close Test Scores: An Item Difficulty Perspective." ERIC DIGEST 33(6).

ACTFL (1986). ACTFL Proficiency Guidelines. Hastings-on-Hudson, NY, ACTFL.

Adhikari, D. (2004). Causes of Nepal's Underdevelopment. The Kathmandu Post, 28 April 2004. Available: www.kantipuronline.com. Kathmandu.

Adhikary, M. and D. R. Sharma (2001). Find Ways to Improve [the] SLC Exam System. The Rising Nepal, 26 April 2001. Available: www.gorkhapatra.org.np. Kathmandu.

AEC. (1994). ESL Scales. Canberra, Curriculum Corporation, Australian Education Council.

Aghbar, A. and H. Tang (1991). "Partial Credit Scoring of Cloze Type Items." ERIC DIGEST 32. Available: http://www.ericir.edu/plweb-cgi/

Ahmed, H. H. (1997). Assessment of Student Learning: The Malaysian Experience. Re-engineering Education for Change: Educational Innovation for Development. UNESCO. Bangkok, Thailand, UNESCO - ACEID.

Aitken, K. G. (1977). "Using Cloze Procedure as an Overall Language Proficiency Test." TESOL Quarterly 11(1): 59-67.

Alderson, J., C. Clapham, et al. (1995). Language Test Construction and Evaluation. Cambridge, Cambridge University Press.

Alderson, J. C. (1978). "The Effect on the Cloze Test of Changes in Deletion Frequency." Journal of Research in Reading 2(2): 108-119.

Alderson, J. C. (1979). "The Cloze Procedures as a Measure of Proficiency in English as a Foreign Language." TESOL Quarterly 13(2): 219-228.

Alderson, J. C. (1980). "Native and Non-native Proficiency on Cloze Tests." Language Learning 3(2): 219-223.

Alderson, J. C. (1986). Innovations in Language Testing. <u>Innovations in Language Testing</u>. M. Portal. London, Nelson: 93-105.

Alderson, J. C. (1987). An overview of ESL/EFL Testing in Britain. <u>Reviews of English Language Proficiency Tests</u>. J. C. Alderson, K. Krahnke and C. Stansfield. Virginia, TESOL: 3-4.

Alderson, J. C. (1989). <u>Bands and Scores</u>. Paper Presented at the Language Testing Colloquium, Bournemouth, UK, Language Testing Colloquium.

Alderson, J. C. (1990a). "Testing reading comprehension skills (Part One)." <u>Reading in a Foreign Language</u> **6**(2): 425-438.

Alderson, J. C. (1990b). "Testing reading comprehension skills (Part Two)." <u>Reading in a Foreign Language</u> **7**(1): 465-503.

Alderson, J. C. (1991). Language Testing in the 1990s: How Far have we Come? How Much Further have we to Go? <u>Current Development in Language Testing</u>. S. Anivan. Singapore, SEAMEO Regional Language Centre: 2-26.

Alderson, J. C., K. J. Krahnke, et al. (1987). <u>Reviews of English Language Proficiency Tests</u>. Washington DC, TESOL.

Alderson, J. C. and Y. Lukmani (1987). "Cognition and Reading: Cognitive Levels as Embodied in Test Questions." <u>Reading in a Foreign Language</u> **5**(2): 253-270.

Alderson, J. C. and B. North (1991). <u>Language Testing in the 1990s</u>. London, Modern English Publications and the British Council.

Alderson, J. C. and D. Wall (1993). "Does Washback Exist?" <u>Applied Linguistics</u> **14**(2): 115-129.

Alexander, D. (1997). Problems of Educational Reform in Zambia. England, SCUTREA. **2004**.

Allen, J. P. B. and A. Davies, Eds. (1977). <u>Testing and Experimental Methods</u>. Edinburgh Course in Applied Linguistics. Oxford, Oxford University Press.

Amatya, K. (2001). Textbooks Trauma. <u>The Kathmandu Post, 6 November 2001. Available: www.kantipuronline.com</u>. Kathmandu.

Anderson, J. (1976). <u>Psycholinguistic Experiments in Foreign Language Testing</u>. England, University of Queensland Press.

Angwin, J., Ed. (1997). <u>The Essence of Action Research</u>. Geelong, Australia, Deakin Centre for Education and Change, Deakin University Press.

Anivan, S., Ed. (1991). Current Development in Language Testing. Singapore, SEAMEO Regional Language Centre.

Ardovino, J., Hollingsworth, J. and Ybarra, S. (2000). Multiple Measures. Thousand Oaks, CA, Corwin Press Inc.

Arena, L. A., Ed. (1990). Language Proficiency: Defining, Teaching and Testing. London, Plenum Press.

Aryal, K. R. (1970). Education for the Development of Nepal. Kathmandu, Ratna Pustak Bhandar.

Aryal, V. (2001). Private Schools. The Rising Nepal, 5 June 2001. Available: www.gorkhapatra.gov.np. Kathmandu.

Athanasou, B. (1997). Introduction to Educational Testing. Australia, Social Science Press.

Awasthi, J. R. (1995). A Linguistic Analysis of Errors Committed by Nepali Learners of English. Unpublished PhD Thesis. Department of Linguistics. Hyderbad, University of Hyderabad.

Aweisis, S. (1993). Comprehension Assessment Measures in Second and Foreign Language Teaching, ERIC. Available: http://www.ericir.edu/plweb-cgi/.

Bachman, L. (1989). An Examination of Some Language Proficiency Tests from a Communicative View Point, ERIC Digest. Available: http://www.ericir.edu/plweb-cgi/.

Bachman, L. and A. Palmer (1996). Language Testing in Practice. Designing and Developing Useful Language Tests. Oxford, Oxford University Press.

Bachman, L. F. (1985). "Performance on Cloze Tests with Fixed Ratio and Rational Deletions." TESOL Quarterly 19(3): 535-555.

Bachman, L. F. (1986). The Test of English as a Foreign Language as a Measure of Communicative Competence. Technology and Language Testing: A Collection of Papers from the Annual Colloquium on Language Testing Research. C. Stansfield. Princeton, New Jersey, ELTS. 7: 69-88.

Bachman, L. F. (1988). "Problems in Examining the validity of the ACTFL Oral Proficiency Interview." Studies in Second Language Acquisition 10(2): 149-164.

Bachman, L. F. (1990). Fundamental Considerations in Language Testing. Oxford, Oxford University Press.

Bachman, L. F. (1999). Designing and Developing Useful Language Tests. Studies in Language Testing. A. A. B. C. Davies, Edler, N. Iwashita, T. McNamara. Cambridge, Cambridge University Press. **109-118**.

Bachman, L. F., F. Davidson, et al. (1995(a)). An Investigation into the Comparability of two Tests of English as a Foreign Language. Studies in Language Testing. L. F. Bachman, Davidson, F., Ryan, K. and Choi, I . Cambridge, Cambridge University Press. 1.

Bachman, L. F., F. Davidson, et al., Eds. (1995b). Studies in Language Testing. Cambridge, Cambridge University Press.

Bajracharya, H. (2001). School Education and Community Participation. The Kathmandu Post, 24 July 2001. Available: www.kantipuronline.com. Kathmandu.

Bajracharya, H., Thapa, B. and Chitrakar, R. (1998). Trends, Issues and Policies of Education in Nepal. Kathmandu, Research Centre Educational Innovation and Development, Tribhuvan University.

Baldauf, R. and I. Propst (1979). Measuring ESL Reading Achievement with Matching Cloze, ERIC Digest. Available: http://www.ericir.edu/plweb-cgi/.

Banerjee, J. (2000). Integrative Tests. Routledge Encyclopaedia of Language Teaching and Learning. M. Byram. London, Routledge: 291-292.

Barrett-Lennard, S. (1997). "Encouraging Autonomy and Oreoaring for IELTS: Mutually Exclusive Goals." PROSPECT **12**(3): 29-40.

Baumel, S. D. (2003). "Teaching English in Israeli Haredi Schools." Language Policy **2**: 47-67.

Belbase, L. N. (2001). Parents should Monitor Schools. The Rising Nepal. 12 November 2001. Available; http://www.gorkhapatra.gov.np/ Kathmandu.

Bensaoussan, M. and R. R. (1984). "Testing EFL Reading Comprehension Using a Multiple-choice Rational Cloze." The Modern Language Journal **63**(3): 230-239.

Berkoff, N. A. (1985). Testing Oral proficiency: A New Approach. New Directions in Language Testing. Y. P. Lee, C. Fok, R. Lord and G. Low. Oxford, Pergamon Press.

Bhadra, S. and Y. Yadav (1988). Causes of Failure in the Proficiency Certificate. Kathmandu, CEEID, Tribhuvan University.

Bhandari, A. (1996). Present Practices, Problems and Reforms Activities of OCE Related to SLC Examination. SLC Examination in Nepal. G. B. Singh. Kathmandu, CERID, Tribhuvan University.

Bhandari, S. (2002). [Development of Education System and National Education Goals]. Kantipur, 13 November 2002. Available: www.kantipuronline.com. Kathmandu.

Bhattarai, G. (2001a). Private Tuition Mania. The Rising Nepal, 3 September 2001. Available: www.gorkhapatra.gov.np. Kathmandu.

Bhattarai, G. R. (2001b). "Is the Practical Examination Really Practicable?" Journal of NELTA 6(1).

Bianco, J. L. (1987). National Policy on Languages. Canberra, Commonwealth Department of Education.

Bista, D. B. (1991). Fatalism and Development. Delhi, Orient Longman.

Bloom, B. S. (1972). Taxonomy of Educational Objectives. London, Longman.

Bloomfield, L. (1933). Language. New York, Holt, Rinehart and Winston.

Board of Studies, V. (1996). ESL Companion to the English CSF. Carlton, Victoria, Board of Studies.

Bona, J. (1989). Critical Perspectives on Indian Education. New Delhi, Bahri Publications.

Bond, L. (1995a). Critical Issues: Rethinking Assessment and its Role in Supporting Educational Reforms. Washington: NCREL.

Bond, L. (1995b). Norm and Criterion Referenced Testing. Washington: NCREL.

Borbora, S. (2001). Education, an Ongoing Process. The Kathmandu Post, 30 July 2001. Available: www.kantipuronline.com. Kathmandu.

Bormuth, J. R. (1968). Experimental Application of Cloze Tests. International Reading Association Conference.

BOS (2000). Guide to the VCE 2000. Melbourne, Australia, Board of Studies, Victoria.

Brady, L. (1995). Curriculum Development. Sydney, Prentice Hall.

Brand, A. G. (1992). "Writing Assessment at College Level." Washington DC: ERIC Clearinghouse on Reading and Communication.

Brett, P. (1997). A Comparative Study of the Effects of he Use of Multimedia on Listening Comprehension System, ERIC Digest 37. Available: http://www.ericir.edu/plweb-cgi/

246

Brindley, G. (1986). The Assessment of Second Language Proficiency: Issues and Approaches. Adelaide, Australia, National Curriculum Resource Centre.

Brindley, G. (1989). Assessing Achievement in the Learner-Centred Curriculum. Sydney, National Centre for English Language Teaching and Research, Macquarie University.

Brindley, G. (1991). Defining Language Ability: The Criteria for Criteria. Current Development in Language Testing. S. Anivan. Singapore, SEAMEO Regional Language Centre: 139 -159.

Brindley, G. (1995a). Language Assessment in Action. Sydney, National Centre for English Language Teaching and Research. Macquarie University.

Brindley, G. (1995b). Competency-based Assessment in Second Language Programs: Some Issues and Questions. Language assessment in action. G. Brindley. Sydney, National Centre for English Language Testing and Research, Macquarie University: 145-164.

Brindley, G. (1998a). "Outcomes-based Assessment and Reporting in Language Learning Programs: A Review of the Issues." Language Testing **15**: 45-85.

Brindley, G. (1998b). Assessing Listening Abilities. Annual Review of Applied Linguistics. B. Grabe. New York, Cambridge University Press. **8**: 171-191.

Brindley, G. (1999). Language Assessment and Professional Development. Studies in Language Testing. A. Davies. Cambridge, Cambridge University Press.

Brindley, G. (2000). Studies in Immigrant English Language Assessment. Sydney, NCELTR.

Broadfoot, P. (1992). Towards Profiles of Achievement. Examinations: Comparative and International Studies. M. Eckstein and H. Noah. Oxford, Pergamon.

Brown, B. L. (2000). New Learning Strategies for Generation X, ERIC Digest, Available: http://ericir.edu/plweb-cgi/.

Brown, G., K. Malmkjaer and J. Williams (1996). Performance and Competence in Second Language Acquisition. Cambridge, Cambridge University Press.

Brown, G. and G. Yule (1983). Discourse Analysis. Cambridge, Cambridge University Press.

Brown, J. D. (1980). "Relative Merits of Four Methods for Scoring Cloze Tests." Modern Language Journal 64(3): 311-317.

Brown, J. D. (1989). Short-cut Estimates of Criterion-referenced Test Reliability. Paper presented at the San Antonio 11th Annual Language Testing. Research Colloquium. San Antonio.

Brown, J. D. (1996). Testing in Language Programs. Upper Saddle, NJ, Prentice Hall Regents.

Brown, J. D., Ed. (1998). New Ways of Classroom Assessment. Alexandria, VA, TESOL.

Brown, J. D., H. G. Cook, et al. (1991). Southeast Asian Languages Proficiency Examinations, ERIC Digest. Available: http://ericir.edu/plweb-cgi/.

Brown, J. D. and S. O. Yamashita (1995). Language Testing in Japan., ERIC Digest. Available: http://ericir.edu/plweb-cgi/.

Brown, R. (2001). Educational Values and Summative Evaluation: A View Across Three Educational Systems. Annual Conference, England, University of Leeds.

Brumfit, C. J. (1995). Language Education in the National Curriculum. Oxford, Blackwell.

Buck, G. (1991). "The Testing of Listening: An Introspective Study." Language Testing 8(1): 67-91.

Buck, G. (1997). The Testing of Listening in a Second Language. Encyclopaedia of Language and Education. C. Clapham and D. Corson. Netherlands, Kluwer Academic Publishers. 7: 65-74.

Buell, J. G. (1992). TOEFL ad IELTS as Measures of Academic Reading Ability: An Exploratory Study, ERIC. 2002.

Burghart, R. (1994). The Political Culture of Panchayat Democracy. Nepal in the Nineties. M. Hutt. Delhi, Oxford University: 1-13.

Burns, R. B. (1998). Introduction to Research Methods. Cheshire, Longman.

Businessworld (1999). New English Testing System Offered. Businessworld. Manila.

Butler, J. (1991). "Cloze Procedures and Concordances: The Advantages of Discourse Level Authenticity in Testing Expectancy Grammar System." ERIC Journal.

Byram, M., Ed. (2000). Routledge Encyclopaedia of Language Teaching and Learning. London, Routledge.

Byrnes, H. and M. Canale (1987). Defining and Developing Proficiency: Guidelines, Implementations, and Concepts. Illinois, National Textbook Company.

Cajkler, W. and R. Addleman (2000). The Practice of Foreign Language Teaching. London, David Fulton.

Canale, M. (1983a). On Some Dimensions of Language Proficiency. Issues in Language Testing Research. J. Oller. Rowley: Newbury House: 102-115.

Canale, M. (1983b). From Communicative Competence to Communicative Language Pedagogy. Language and communication. J. C. Richards and R. Schmidt. London, Longman: 2-27.

Canale, M. (1994). On Some Theoretical Framework for Language Proficiency. Language Proficiency and Academic Achievement. C. Rivera. Clevedon, Avon, Multilingual Matters.

Canale, M. and M. Swain (1980). "Theoretical Bases for Communicative Approaches to Second Language Teaching and Testing." Applied Linguistics 1(1): 1-47.

Carrell, P. (1993). "First and Second Language Reading Strategies: Evidence from Cloze." ERIC Journal(EJ 516241).

Carroll, B. J. (1961). Fundamental Considerations in Testing English Language Proficiency in Foreign Students. Teaching English as a Second Language. H. B. Allen, R. N. Campbell and (1972). New York, McGraw Hill: 313-320.

Carroll, B. J. (1972). Fundamental Considerations in Testing for English Language Proficiency in Foreign Students. Teaching English as a Second Language: a Book of Readings. H. B. Allen and R. N. Campbell. New York, McGraw-Hill: 313-321.

Carroll, B. J. and P. J. Hall (1985). Make your Own Language Tests. New York, Pergamon Press.

Carroll, J. B. (1972). Defining Language Comprehension: Some Speculation. Language Comprehension and the Acquisition of Knowledge. R. O. Freedle and J. B. Carroll. New York, John Wiley.

Carroll, J. B. (1983). Psychometric Theory and Language Testing. Issues in language Testing Research. J. Oller. Rowley, Newbury House: 195-207.

Carroll, J. B., Carton, S. and Wilds, C. P. (1979). An Investigation of Cloze Items in the Measurement of Achievement in Foreign Languages, ERIC Digest. Available: http://ericir.edu/plweb-cgi/.

Caufield, J. and W. C. Smith (1981). "The Reduced Redundancy Test and the Cloze Procedure as Measure of Global Language Proficiency." ERIC Journal EJ. 521174.

CDC (1999). [Secondary Curriculum]. Kathmandu, CDC/MOE Nepal.

CDC (2000). Secondary Education Curriculum. Kathmandu, Curriculum Development Centre, Ministry of Education, HMG/N.

CERID (1987). A Study of the Critical Aspects of Secondary Education in Nepal. Kathmandu, Research Centre for Educational Innovation and Development, Tribhuvan University.

CERID (1988). Secondary Education in Nepal. Kathmandu, CERID, Tribhuvan University.

CERID (1989a). Factors Affecting Science Teaching/Learning at Secondary Level Schools in Nepal. Kathmandu, Research Centre for Educational Innovation and Development, Tribhuvan University.

CERID (1989b). Causes of Failure in English in the SLC Examination. Kathmandu:, CERID, Tribhuvan University.

CERID (1991). Raising the Quality of Learning of Secondary School Students. Kathmandu, Research Centre for Educational Innovation and Development, Tribhuvan University.

CERID (1996). Secondary Education and SLC Examination: An Employment Perspective. Kathmandu, Research Centre for Educational Innovation and Development, Tribhuvan University.

CERID (2001a). Revitalizing Secondary Education: School Teacher Training. Kathmandu, Research Centre for Educational Innovation and Development, Tribhuvan University.

CERID (2001b). A Study on the Feasibility of Parallel Question Papers for the SLC Examinations. Kathmandu, Research Centre for Educational Innovation and Development, Tribhuvan University.

CERID (2001c). Secondary Education: A Need for Diversification. Kathmandu, Research Centre for Educational Innovation and Development, Tribhuvan University.

Chalhoub-Deville, M. (1997). "Theoretical Models, Assessment Frameworks and Test Construction." Language Testing 14(1): 3-22.

Chapman, C. (1990). Authentic Writing Assessment. Practical Assessment, Research and Evaluation. Washington, D.C., ERIC Clearing House on Assessment and Evaluation. **2001**.

Cheng, L. (1997). "How does Washback Influence Teaching? Implications for Hong Kong." Language in Education **11**(1): 38-54.

China Daily (2001). New Exam Designed for Students. China Daily. 10 December 2001. New York.

Chomsky, N. (1957). Syntactic Structure. The Hague, Moulton.

Clapham, C. (1995). The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension. Studies in Language Testing. Studies in Language Testing. M. Milanovic. Cambridge, Cambridge University Press. **4**.

Clapham, C. (2000). Assessment and Testing. Routledge Encyclopaedia of Language Teaching and Learning. M. Byram. London, Routledge.

Clapham, C. and D. Wall, Eds. (1996). Language Testing Update. Lancaster, Lancaster University.

Clark, C. J. (1985). "Curriculum Review in Second Language Learning." The Canadian Modern Language Review **42**(2): 342-360.

Clark, J. L. D. (1978). Psychometric Consideration in Language Testing. Advances in Language Testing. B. Spolsky. Washington, DC, Centre for Applied Linguistics: 15-30.

Coates, S., L. Fitzpatrick, et al. (1995). National Reporting System. Australia, Australian National Training Authority.

Cohen, A. (1980). Testing Language Ability in the Classroom. Rowley, Newbury House Publishers, Inc.

Cohen, A. (1994). Assessing Language Ability in the Classroom. Rowley, MA, Heinle & Heinle.

Cohen, L. and Morrison, K. (2000). Research Methods in Education. London, Routledge Falmer.

Colin, T. and A. Garrow (2002). Thinking, Listening, Looking, Understanding and Acting as You Go Along. Action-in-Research Symposium, Melbourne, Graduate School, Swineburne University.

Coniam, D. (1993). Co Text or No text? A Study of an Adapted Cloze Technique for Testing Comprehension/Summary Skills, ERIC Digest. Available: http://ericir.edu/plweb-cgi/.

Coniam, D. (1995). "Towards a Common Ability Scale for Hong Kong English Secondary Schools Forms." Language Testing 12(2): 182-193.

Coniam, D. (1997). A Computerized English Language Proofing Cloze Program, ERIC Digest. Available: http://ericir.edu/plweb-cgi/.

Coniam, D. (1998). From Text to Test, Automatically -- An evaluation of a Computer Cloze-Test Generator, ERIC Digest. Available: http://ericir.edu/plweb-cgi/.

Conrad, C. and J. Oller (1971). "The Cloze Test and ESL Proficiency." Language Learning 21(2): 183-195.

Coombe, C. (1998). Current Trends in English Language Testing, ERIC Digest. Available: http://ericir.edu/plweb-cgi/.

Cooper, C. R. and L. Odell (1997). Holistic Evaluation of Writing. Evaluating Writing. C. R. Cooper and L. Odell. Buffalo, State University of New York: 83-92.

Corder, S. P. (1973). Introducing Applied Linguistics. England, Penguin.

Corson, D. and C. V. E. Clapham (1997). Language Testing and Assessment: Encyclopaedia of Language and Education. Dordrecht, Netherlands, Kluwer.

Courchene, R. (1995). An Alternative Method for Teaching and Testing Reading Comprehension, ERIC Digest. Available: http://ericir.edu/plweb-cgi/.

CRLI (1999). Secondary Screening Profiles. Edinburgh, UK, Hodder and Stoughton.

Crocker, L. (1992). Item Analysis. Encyclopaedia of Educational Research. M. C. Alkins. New York, MacMillan. 4.

Crook, N., Ed. (1996). The Transition of Knowledge in South Asia. Delhi, Oxford University Press.

Cross, D. (1992). A Practical Handbook of Language Teaching. London, Prentice Hall ELT.

Cross, L. H. (1995). Grading Students. Washington, D.C., ERIC Clearing House on Assessment and Evaluation. 2001.

Crossley, M. (1999). "Reconceptualising, Comparative and International Education. Compare." A Journal of Comparative Education 29(3): 249-258.

Crossley, M. and P. Jarvis (2000). " Introduction: Continuity, Challenge and Change in Comparative and International Education." Comparative Education 36(3): 261-265.

CTSDC (1972). Secondary Education Curriculum. Kathmandu, Ministry of Education, HMG/N.

Cumming, A. (1996). The Concept of Validation in Language Testing. Validation in language testing. A. Cumming and R. Berwick. Clevedon, Avon, Multilingual Matters: 1-14.

Cumming, A. (1997). The Testing of Writing in Second Languages. Language testing and Assessment: Encyclopaedia of language in education. C. Clapham. Dordrecht, Kluwer. 7: 51-63.

Cumming, J. (1979). "Cognitive/Academic Language Proficiency, Linguistic Interdependence, the Optimum Age Question and Some other Matters." Working Paper on Bilingualism 19: 121-129.

Cummins, J. (1994). Wanted: A Theoretical Framework for Relating Proficiency to Achievement among Bilingual Students. Language Proficiency and Academic Achievement. C. Rivera. Clevedon, Avon, Multilingual Matters: 2-19.

Cziko, G. A. (1984). "Improving the Psychometric, Criterion-referenced, and Practical Qualities in Integrative Language Tests." TESOL Quarterly 16(3): 367-379.

Dahal, P. (2001). Testing Teachers. The Rising Nepal. The National English Daily Newspaper. Available: http://www.gorkhapatra.gov.np. Kathmandu.

Darlington, Y. and D. Scott (2002). Qualitative Research in Practice: Stories from the Field. Sydney, Allen and Unwin.

Darnell, D. K. (1970). "Clozenthropy: A Procedure for Testing English Language of Foreign Students." Speech Monograph 37: 36-46.

Davey, L. (1992). "A Case for a National Testing System." ERIC Practical Assessment, Research and Evaluation.

Davey, L. and M. Monty (1991). "A Case Against a National Test." ERIC Practical Assessment, Research and Evaluation.

Davidson, F. and L. Bachman (1990). "The Cambridge TOEFL Comparability Study: An Example of the Cross-national Comparison of Language Tests." AILA Review 7: 24-45.

Davidson, F. and B. Lynch (1996). Criterion-referenced Language Test Development: a Prolegomenon. Language Testing: New Openings. A.

Huhta, K. Sajavaara and S. Takala. Jyvaskyala, Finland, University of Jyvaskyla Institute for Educational Research: 73-89.

Davidson, F. and B. Lynch (2002). Testcraft. London, Yale University Press.

Davies, A. (1965). Proficiency in English as a Second Language. Unpublished PhD Thesis. Department of Linguistics. Birmingham, University of Birmingham.

Davies, A. (1978). Language Testing. Survey 1: Eight State of the Art Articles on Key Areas of Language Teaching. V. Kinsella. Cambridge, Cambridge University Press.

Davies, A. (1983). The Validity of Concurrent Validation. Current Developments in Language Testing. A. Hughes and D. Porter. London, Academic Press: 121-128.

Davies, A. (1985). Follow My Leader: Is that What Language Tests do? New Directions in Language Testing. Y. P. Lee, C. Fok, R. Lord and G. Low. Oxford, Pergamon Press: 3-14.

Davies, A. (1988). "Operationalising Uncertainty in Language Testing: An Argument in Favour of Content Validity." Language Testing. 5(1): 32-48.

Davies, A. (1989). Testing Reading Speed through Text Retrieval. Language Learning and Community. C. N. Candlin and T. F. McNamara. Sydney, NCEL: 115-125.

Davies, A. (1990a). Principles of Language Testing. Oxford, Black Basil.

Davies, A. (1990b). An Evaluation Model for ELT Projects in South India: The Policy of Change. Annual Meeting, Applied Linguistic Association of Australia, Sydney.

Davies, A. (1991a). Is Proficiency Always Achievement? Fourth ELICOS Association Educational Conference, Melbourne, Monash University.

Davies, A. (1991b). Issues in English Language Testing. ESLD/LLS Project Seminar, Melbourne, University of Melbourne.

Davies, A. (1991c). Can Test Bias be Claimed When Language Proficiency is Inadequate? International Conference on Language Education: Interaction and Development, Ho Chi Minh City, Vietnam.

Davies, A. (1995). "Testing Communicative Language or Testing Language Communicatively: What? How?" Melbourne Papers in Language Testing 4(1): 1-20.

Davies, A. (1996). Australian Immigrant Gatekeeping through English Language Tests: How Important is Proficiency?, Tampere, Finland.

Davies, A. (1998). "Immigrant Gatekeeping and the Language Testing Profession." IRAAL **28**(4): 67-78.

Davies, A. and A. Brown (1990). Designing Instruments to Measure Language Proficiency. IWG/ Assessment of Proficiency Levels in JFL, Canberra.

Davies, A., A. Brown, et al. (1999). Dictionary of Language Testing. Cambridge, Cambridge University Press.

Davies, A. and C. Elder (1991). The Language Confidence of Non-Native Speaking Teachers. RELC Seminar, Singapore.

Davies, A., Glendinning, E. McLean, A. (1984). Survey of English Language Teaching in Nepal. Kathmandu, British Council/ODA and HMG/N Ministry of Education and Culture, HMG/N.

Davis, E. (1980). Teachers as Curriculum Evaluators. Sydney, George Allen and Unwin.

De Francis, J. (1977). Colonialism and Language Policy in Viet Nam. New York, Mouton Publishers.

Delamont, S. (2002). Fieldwork in Educational Settings: Methods, Pitfalls and Perspectives. London, Routledge.

Denhem, P. (1985a). Language Proficiency and Two Major Thoeries about its Measurement. Assessment of English Language Proficiency of Indonesian and the Effect of Linguistic background and Attitude on Performance: 10-15.

Denhem, P. (1985b). Assessment of English Language Proficiency of Indonesians and the Effect of Linguistic Background and Attitude on Performance.

Devkota, B. (2002). The Problems Created by School Closure. Kantipur, 21 July 2001. Available: http://www.kantipuronline.com/. Kathmandu.

Dhakal, S. (2001a). A Hornet's Nest. Spotlight. **20:** 7-9.

Dhakal, S. (2001b). Poor Examination Results: Who is to Blame for Poor Results? Spotlight. **20:** 17-23.

Dhakal, S. (2003). SLC Results and Public Schools. The Kathmandu Post, 21 September 2003. Available: http://www.kantipuronline.com/. Kathmandu.

Dhakal, T. P. (2000). Effectiveness of Discrete Point Tests and Integrative Tests as Measures of English Language Proficiency: A Comparative Study. Unpublished Master Thesis. Kathmandu: Department of English Language Education, Faculty of Education. Tribhuvan University.

Dick, B. (1997). The Naive Philosophy of a Naive Philosopher [On Line]. Melbourne, http://www.scu.edu.au/schools/gem/ar/arp/naive.html. **2002**.

Dick, B. (1999). What is Action Research?, Available online at http://www.scu.edu.au/schools/gem/ar/whatisar.html. **2003**.

Dietel, R. J., Herman, J. L. and Knuth, R. A. (1991). What does Research Say about Assessment?, NCREL. **2000**.

Dixit, K. (2003). The Business of SLC. The Nepali Times. 11 April 2003. Available: http://nepalnews.com. Kathmandu.

Donald, E., M. Schedl, et al. (1999). "Validation and Revised Test of Spoken English against a Criterion of Spoken English." Language Testing 16(4): 399-425.

Douglas, D. (1987). Testing Listening Comprehension in the Context of the ACTFL Proficiency Guidelines. Evaluation of Foreign Language Proficiency. A. Valdman. Bloomington, IN, Indiana University: 97-11.

Dunkel, P. G., G. Henning, et al. (1993). "The Assessment of an L2 Listening Comprehension Construct: A Tentative Model for Test Specification and Development." The Modern Language Journal 77: 180-191.

Dutcher, P. (1990). Authentic Reading Assessment. Practical Assessment, Research and Evaluation. Washington, D.C., ERIC Clearing House on Assessment and Evaluation. **2001**.

Dyson, B. (2002). "Focus on Learnable Form in a Communicative Context; A Framework for Second Language Acquisition in the Classroom." Australian Review of Applied Linguistics 25(1): 52-70.

Eckman, F. R. (1994). The Competence Performance Issue in Second Language Acquisition Theory: A Debate. Methodology in Second Language Acquisition. E. Tarone, S. Gass and A. Cohen. Cambridge, Cambridge University Press.

Eckstein, M. and H. Noah, Eds. (1992). Examinations: Comparative and International Studies. Oxfoxd, Pergamon Press.

Elder, C. (1996). "The Effect of language Background on 'Foreign' Language Test Performance: The Case of Chinese, Modern Greek and Italian." Language Learning 46(2): 233-282.

Elder, C. (1997). Is it Fair to Assess Native and Non-native Speakers in Common on School Foreign Language Examinations? Language Testing Research Colloquium, Orlando, FA.

Elder, C. and B. Lynch (2000). <u>Investing Test Impact: Public Perceptions of the LAP Tests in Victoria</u>. Melbourne, The University of Melbourne (in Progress).

El-Koumy, A. (1997). Effect of Cloze Instruction in EFL Listening Compression, ERIC ED 411702.

Elliott, J. (1996). <u>Action Research for Educational Change</u>. Philadelphia, Open University Press.

Ellis, R. (1995). <u>The Study of Second Language Acquisition</u>. Oxford, Oxford University Press.

Ellis, R. (1997). <u>A Study of National Curriculum Testing Across the Interface at Key Stage 2/ Key Stage 3</u>. Annual Conference, York, University of York.

Ellis, S. W. (1998). <u>Developing Whole School Approaches to Curriculum and Assessment in Secondary Schools</u>. Annual Conference, Belfast, Queen's University of Belfast.

Emel, R. L. and D. A. Frisbie (1991). <u>Essentials of Educational Measurement</u>. New Jersey, Prentice Hall.

Enright, M. K. (2004). "Research Issues in High Stakes Communicative Language Testing: Reflections on TOEFL's New Directions." <u>TESOL Quarterly</u> **39**(1): 148-111.

ERIC (1995). Cost of a National Examination. Washington, D.C., ERIC Clearing House on Assessment and Evaluation. **2000**.

ETS (1984). <u>SLEP Manual</u>. Princeton, Educational Testing Service.

ETS (1996). "Computer-Based TOEFL." <u>Language Testing Update</u>(19): 10-11.

ETS (2001). <u>TOEFL Bulletin</u>. Princeton, Educational Testing Service.

Farhady, H. (1979). "The Disjunctive Fallacy between Discrete-point and Integrative Tests." <u>TESOL Quarterly</u> **13**(3): 347-357.

Farhady, H. (1983). New Directions for ESL Proficiency Testing. <u>Issues in Language Testing Research</u>. J. Oller. Rowley, Newbury House: 34-46.

Farhady, H. (1996). A Text Driven Method for the Deletion Procedure in CLoze Passages, ERIC EJ 529503.

Ferries, D. and J. Hedgecock (2001). <u>Teaching ESL Composition: Purpose, Process and Practice</u>. New Jersey, Erlbaum.

Firth, J. R. (1957). Papers in Linguistics 1934 -1951. London, Oxford University Press.

Fisher, J. F. (1997). Living Martyrs. Delhi, Oxford University Press.

Fishman, J. and R. L. Cooper (1978). The Sociolinguistic Foundation of Language Testing. Advances in Language Testing. B. Spolsky. Washington, DC, Centre for Applied Linguistics.

Flick, U. (1992). "Triangulation Revisited: Strategy for Validation or Alternative?" Journal of the Theory of Social Behaviour 22: 175-198.

Foster, D. (1992). Commentary [on Heyneman and Ransom 1992]. Examination: Comparative and International Studies. M. Eckstein and H. Noah. Oxford, Pergamon Press.

Fotos, S. S. (1991). "The Cloze as an Integrative Measure of EFL Proficiency: A Substitute for Essay on College Entrance Examinations." Language Learning 41(3): 313-336.

Franken, M. (2001). "The Parameters of Speaking for Writing Tasks." Australian Review of Applied Linguistics 24(1): 51-61.

Freedman, S. (1991). Evaluating Writing: Linking Large-scale Testing and Classroom Assessment. Berkeley, University of California, Centre for the Study of Writing.

Fried, L. (1985). On the Validity of Second Language Tests. Modelling and Assessing Second Language Acquisition. K. a. P. Hyltenstam, M. Clevedan, England, Multilingual Matters Ltd. 18.

Fulcher, G. (1997). The Testing of Speaking in a Foreign Language. Encyclopaedia of Language and Education: Language Testing and Assessment. C. Clapham and D. Corson. Dordrecht, The Netherlands, Kluwer Academic Publishers. 7: 75-85.

Fullan, M. (1991). The New Meaning of Educational Change. London, Cassell.

Fullan, M. (1993). Changing Forces: Probing the Depth of Educational Reform. London, The Falmer Press.

Gass, S. M. and J. Schachter, Eds. (1989). Linguistics Perspectives of Second Language Education. Cambridge, Cambridge University Press.

Gautam, B. (2003). [Demands on Private Education should be Scientific and Balanced]. Kantipur, 19 June 2003. Available: http://www.kantipuronline.com/. Kathmandu.

Gautam, R. B. (2001a). SLC Results, More Questions than Answers. The Kathmandu Post, 1 October 2001. Available: http://www.kantipuronline.com/. Kathmandu.

Gautam, R. B. (2001b). [Change Essential in the Form and Function of Education]. Kantipur, 18 September 2001. Available: http://www.kantipuronline.com/. Kathmandu.

Gautam, S. (2001a). Intellectual's Role in Nation Building. The Kathmandu Post, 17 July 2001. Available: http://www.kantipuronline.com/. Kathmandu.

Gautam, S. (2001b). Teaching English at College. The Rising Nepal, 7 February 2001. Available: http://www.gorkhapatra.gov.np/. Kathmandu.

Gautam, S. (2002). Tribhuvan University and Noam Chomsky. The Kathmandu Post, 6 February 2002. Available: http://www.kantipuronline.com/. Kathmandu.

Gibson, B. (1997). Talking the Test: Using Verbal Report Data in Looking at the Processing of Cloze Tasks, ERIC ED 409713.

Giri, C. M. (1995). A Survey into People's Attitudes towards the Existing SLC Examination in Nepal. Department of English Education. Kathmandu, Tribhuvan University.

Giri, R. A. (1996a). "Cloze Tests in the Foreign Language Classroom." Journal of NELTA 1(1): 26-31.

Giri, R. A. (1996b). "The Diagnostic Technique of Error Correction." NELTA Newsletter 2(2): 3-7.

Giri, R. A. (2002). "Approaches to Language Testing." Journal of NELTA 7(1-2): 11-25.

Giri, R. A. (2003). "Language Testing Then and Now." Journal of NELTA 8(1): 49-67.

Giri, R. A. (2003). "The Testing of Reading." Young Voices in ELT 3(1): 5-12.

Gordon, P. (1981). The Study of the Curriculum. London, Batsford Academic and Educational Ltd.

Gouvias, D. (1998). The National Examination's System in Greece and its Function as a Mechanism of Social Selection. The European Conference on Educational Research, Ljubijana, Greece, National Centre of Educational Research.

Gowans, C. J. (1999). Second Language Learning Strategies and Factors Affecting Their Use: A qualitative Study of the Experiences of

Missionaries in Nepal. <u>Faculty of Arts</u>. Melbourne, Victoria University of Technology.

GP (2004). Education. <u>Gorkhapatra, 12 July 2004. Available: http://www.gorkhapatra.gov.np/</u>. Kathmandu.

Grabe, B. (1991). "Current Developments in Second Language Reading Research." <u>TESOL Quarterly</u> **25**: 375-397.

Grant, J. (1991). Individual and Co-operative Completion of Cloze. ERIC ED 341259.

Greenfield, G. R. (1999). Classic Readability Formulas in an EFL Context: Are they Valid for Japanese Speakers? <u>Department of Education (Unpublished)</u>. USA, Temple University.

Greenwood, D. J. and M. Levin (1998). <u>Introduction to Action Research</u>. London, Sage Publications Inc.

Gregson, J. (2002). <u>Blood Against the Snow</u>. London, Fourth Estate.

Griffin, P. (1985). The Use of Latent Trait Models in the Calibration of Tests of Spoken Language in Large Scale Selection- Placement Programs. <u>New Directions in Language Testing</u>. Y. P. Lee, C. Fok, R. Lord and G. Low. Oxford, Pergamon Press: 149-162.

Grove, L. (1996). "Recent Research in Language Testing: NILLIA Language Testing Research Centre." <u>Language Testing Update</u>(19): 34-38.

Grundy, S. (1995). <u>Action Research as Professional Development</u>. Murdoch, School of Education, Murdoch University.

Hada, V. (2002). [Promoting Quality Education: Today's Necessity]. <u>Kantipur, 21 March 2002. Available: http://www.kantipuronline.com/</u>. Kathmandu.

Haines, C. (2004). <u>Assessing Students' Written Work: Marking Essays and Reports</u>. London, Routledge Falmer.

Hamilton, J., M. Lopes, et al. (1993). "Rating Scales and Native Speaker Performance on a Communicatively Oriented EAP Test." <u>Language Testing</u> **10**(3): 337-353.

Hammer, J. (1991). <u>The Practice of English Language Teaching</u>. London, Longman.

Hammersley, M. (2002). <u>Action Research: A Contradiction in Terms</u>. Annual Conference, England, University of Exeter.

Hamp, L. (1992). Holistic Writing Assessment of LEP Students, ERIC ED 020649.

Hamp-Lyons, L. (1986). "Testing Writing Across the Curriculum." PALM 2(1): 16-29.

Hamp-Lyons, L. (1990). Second Language Writing: Assessment Issues. Second language writing: research insights for the classroom. B. Kroll. Cambridge, Cambridge University Press: 69-87.

Hamp-Lyons, L. (1991a). "Combining Holistic Scoring with Diagnostic Feedback in a Large Scales Writing Assessment." Language Testing Update(1): 12-15.

Hamp-Lyons, L., Ed. (1991b). Assessing Second-language Writing in Academic Contexts. Norwood, NJ, Ablex.

Hamp-Lyons, L. (1997). "Language Test Washback, Impact and Validity: Ethical Concerns." Language Testing 14(3).

Hamp-Lyons, L. and S. Prochnow (1991). The Difficulties of Difficulties. Current Development if Language Testing. S. Anivan. Singapore, SEAMEO Regional Language Centre: 59-76.

Hanania, E. and M. Shikhani (1986). "Interrelationships Among Three Tests and Language Proficiency: Standardised ESL, Cloze and Writing." TESOL Quarterly 20(1): 97-109.

Harris, D. and L. Palmer (1986). CELT Manual. New York, McGraw-Hill.

Harrison, A. (1983). Communicative Competence: Jam Tomorrow? Current Developments in Language Testing. A. Hughes and D. Porter. London, Academic Press: 77-86.

Harrison, A. (1991). A Language Testing Handbook. London, Modern English Publications.

Heaton, J. B. (1988). Writing English Language Tests. London, ELBS.

Heaton, J. B. (1990). Classroom Testing. Essex, Longman.

Henning, G. (1987). A Guide to Language Testing: Development, Evaluation, Research. Cambridge, Newbury House Publishers.

Henning, G. (1990). "Priority: Testing." Foreign Language Annals 23(5): 379-384.

Herath, J. P. (1997). Sri Lanka Country Paper on Monitoring Achievement. Re-engineering Education for Change: Educational Innovation for Development. UNESCO. Bangkok, Thailand, UNESCO - ACEID.

Heyneman, S. P. and A. Ransom (1992). Using Examinations and Testing to Improve Educational Quality. Examinations: Comparative and International Studies. M. Eckstein and H. Noah. Oxford, Pergamon Press.

Higuchi, C. and Susan (1991). Critical Issues: Integrating Assessment and Instruction in Ways That Support Learning, ERIC ED 352116.

Hill, C. and K. Parry (1994). From Testing to Assessment: English as an International Language. New York, Longman.

Hinofotis, F. B., K. M. Bailey, and S.L. Stern. (1981). Section II: Empirical Research. The Construct Validation of Tests of Communicative Competence. M. Palmer, P. Groof and G. Trosper. Washington DC: 106-126.

Hisama, K. (1976). Design and Empirical Validation of the Cloze Procedure for Measuring Language Proficiency of Non-native Speakers, ERIC ED 147800.

Holtzman, P. D. (1967). English Language Proficiency and the Individual, ERIC ED 012922.

Hood, S. and N. Solomon (1994). Reading and Writing Assessment Kit. Sydney, NCELTR, Macquarie University.

Hout, B. (1990). "Reliability, Validity and Holistic Scoring: What we Know and What we Need to Know." College Composition and Communication 41(2): 77-84.

Howard, F. (1980). "Testing Communicative Proficiency in French as a Second Language: A Search for Procedures." Canadian Modern Language Review 36(2): 272-280.

Howatt, A. P. (1984). A History of English Language Teaching. Oxford, Oxford University Press.

Hughes, A. (1989). Testing for Language Teachers. Cambridge, Cambridge University Press.

Hughes, A. and D. Poter, Eds. (1983). Current Developments in Language Testing. Applied Language Studies. London, Academic Press.

Hulstijn, J. H. (1985). Testing Second Language Proficiency with Direct Procedures: A Comment on Ingram. Modelling and Assessing Second Language Acquisition. K. a. P. Hyltenstam, M. Clevedon, England, Multilingual Matters Ltd. 18: 277-281.

Hunt, J. and M. Burnet (1996). "The Graduating Students' Language Proficiency Assessment Project." Language Testing Update(19): 13-14.

Hutt, M., Ed. (1994). Nepal in the Nineties. Delhi, Oxford University Press.

Hymes, D. H. (1972). On Communicative Competence. Sociolinguistics. J. B. Pride and J. Holmes. England, penguin: 269-293.

Ingram, D. E. (1984). Report on the Formal Trailling of the Australian Second Language Proficiency Ratings. Canberra, Australian Government Publishing Service.

Ingram, D. E. (1985). Assessing Proficiency: An Overview of Some Aspects of Testing. Modelling and Assessing Second Language Acquisition. K. Hyltenstam and M. Pienemann. Clevedon, England, Multilingual Matters. 18: 215-276.

Ingram, D. E. (1991). The International English Language Testing System IELTS: Its Nature and Development. Current Development in Language Testing. S. Anivan. Singapore, SEAMEO Regional Language Centre: 185-198.

Ingram, E. (1978). The Psycholinguistic Issue. Advances in Language Testing. B. Spolsky. Washington, DC., Centre for Applied Linguistics: 1-14.

Isonio, S. (1993). Implementation and Validation of the Combined English Language Skills Assessment at Golden West College, ERIC ED 353023.

Iwashita, N. (1997). The Validity of Paired Interview Format in Oral Performance Assessment. Language Testing Research Colloquium, Orlando, Florida.

Jacobs, H.L. (1981). Testing ESL Composition: A Practical Approach. Rowley, MA, Newbury House Publishers.

Jakovovits, L. A. (1970). Foreign Language Learning: A Pycholinguistic Analysis of the Issues. Rowley, MA, Newbury House Publishers.

Jenks, F. (1987). Michigan Test of English Language Proficiency. Reviews of English Language Proficiency Tests. J. Alderson, K. Krahnke and C. Stansfield. Washington DC, TESOL.

Jha, S. K. (1992). "Problems of Teaching English in Nepal." Curriculum Development Journal 29: 21-31.

Jitendra, A. K. and E. Rohena-Diaz (1996). "Language Assessment of students Who are Linguistically Diverse: Why a Discrete Approach is not the Answer." School Psychology Review 25(1): 40-57.

Johnson, D. M. (1992). Approaches to Research in Second Language Learning. London, Longman.

Johnson, P. (1987). Testing Skills in Expository Writing through Reading Comprehension, ERIC ED. 281391.

Johnson, R., B. Preston and F. Olive (1999). "Teaching the Second Language Testing Course through Test Development by Teachers in Training." Teacher Education Quarterly 26(3): 71-78.

Jones, L. V. (1999). National Tests and Educational reform: Are they Compatible? Pic@ets.org. 2001.

Jones, N. (1994). "How Language Proficiency Develop: Evidence from an Item Bank." Language Testing 11(2): 102-110.

Jones, R. (1977). Testing: A Vital Connection. The Language Connection: From the Classroom to the World. ERIC ED 161264.

Jones, R. L. (1980). An International Survey of Research in Language Testing, ERIC ED 192579.

Jones, R. L. (1985). Some Basic Considerations in Testing Oral Proficiency. New Directions in Language Testing. Y. P. Lee, C. Fok, R. Lord and G. Low. Oxford, Pergamon Press: 77-84.

Jonz, J. (1990). Another Turn in the Conversation: What Does Cloze Measure?, ERIC EJ 411861.

Joshi, D. (2001). Time to Rethink on "Iron Gate'. The Kathmandu Post. 18 September 2001. Available: www.kantipuronline.com

Joshi, R. (1996). Importance and Significance of SLC Examination in the Secondary Education in Nepal. SLC Examination in Nepal. G. B. Singh. Kathmandu, CERID, Tribhuvan University: 33-39.

K. C., M. (2001a). Challenges Confront Education Sector. The Rising Nepal, 21 May 2001. Available: http://www.gorkhapatra.gov.np/. Kathmandu.

K.C., M. (2001b). Education System Needs Careful Scrutiny. The Rising Nepal, 26 July 2001. Available: http://www.gorkhapatra.gov.np/. Kathmandu.

Kalantzis, M. and A. Harvey (2003). Active Voice. The Age, 13 October 2003. Melbourne.

Kalnberzina, V. (1996). "The Year 12 Project in Latvia." Language Testing Update(19): 23-24.

Kam, H. W. and R. Y. L. Wong, Eds. (2004). English Language Teaching in South East Asia. Singapore, Eastern Universities Press.

Kam, H. W. and R. Y. L. Wong, Eds. (2004). Language Policy and Language Education. Singapore, Eastern Universities Press.

Kane, M. B. and R. Mitchell (1996). Implementing Performance Assessment: Promises and Challenges. New Jersey, Lawrence Erlbaum Associates Publishers.

Kantipur (2002). [Incomplete Courses and its Impact on the SLC Students]. Kantipur, 13 March 2002. Available: http://www.kantipuronline.com/. Kathmandu.

Kantipur (2004). [Recommendation for Making Educational Standard Equal for rural and Urban Schools]. Kantipur, 30 May 2004. Available: http://www.kantipuronline.com/. Kathmandu.

Katchen, J. E. (1994). Learning to Listen to Authentic English from Satellite TV, ERIC ED 366222.

Keeves, J., Ed. (1997). Educational Research Methodology and Measurement: An International Handbook. Oxford, Pergamon Press.

Kellaghan, T. (1992). Examination Systems in Africa: Between Internationalisation and Indigenisation. Examinations: Comparative and International Studies. M. Eckstein and H. Noah. Oxford, Pergamon Press.

Kember, D. (2000). Action Learning and Action Research. London, Kegan Page Limited.

Kemmis, S. (1994). Action Research. The International Encyclopaedia of Education. T. Husen and T. N. Paselethwaite. Oxford, Pergamon Press. 1.

Kemmis, S. (1997). Action Research. Educational Research Methodology and Measurement: An International Handbook. J. P. Keeves. Oxford, Pergamon Press: 173-179.

Kemmis, S. and R. McTaggart (1988). The Action Research Planner. Geelong, Deakin University Press.

Kerr, R. (1999). Planning and Practice: Factors Impacting on the Development of Initial Education in Nepal, with Special Reference to English Language Teaching, 1950-1995. Unpublished PhD Thesis Faculty of Human Development. Melbourne, Victoria University of Technology.

Kerr, R. (2000). Teachers are Unqualified. The Kathmandu Post: 6 May 2000. Kathmandu.

Khadaka, N. (1991). Foreign Aid, Poverty and Stagnation in Nepal. Delhi, Vikas Publishing House Pvt Ltd.

Khadka, P. (2001). Quality Education for Better Tomorrow. The Rising Nepal, 21 October 2001. Available: http://www.gorkhapatra.gov.np/. Kathmandu.

Khanal, K. P. (1997). A Study on the Effectiveness of the Cloze Test over the Conventional Objective Tests in Testing Reading Comprehension in English. Unpublished Master Thesis. Kathmandu: Department of English Language Education. Tribhuvan University.

Khaniya, T. R. (1990). Examinations as Instruments for Educational Change: Investigating the Washback Effect of the Nepalese English Exams. Unpublished PhD Thesis. Edinburgh: Department of Applied Linguistics. University of Edinburgh.

Khatiwada, P. (2001). SLC Tantrums. The Kathmandu Post. Kathmandu.

Kirk, R. E. (1990). Statistics: An Introduction. London, Holt, Rinehart and Winston.

Kirschner, M., E. Spector-Cohen, et al. (1996). "A Teacher Education Workshop on the Construction EFL Tests and Materials." TESOL Quarterly 30(1): 85-111.

Klenowaski, V. (1990). Connecting Assessment and Learning. Annual Conference, Lancaster, University of Lancaster.

Kokkota, V. (1991). "The Standardisation of Two EFL Tests in Universities in the Soviet Union." Language Testing Update(9): 22.

KOL (2004). School-based Assessment. Kantipuronline, 27 March 2004. Available: http://www.kantipuronline.com/.

Koseli (2001). [Community Participation in Education]. Kantipur, 18 September 2001. Available: http://www.kantipuronline.com/. Kathmandu.

Krakowian, P. (1997). "Towards a Reform of the Entrance Examination in English at the Institute of English Studies." Language Testing Update(22): 15-18.

Krashen, S. (1985). Writing: Research, Theory and Applications. Oxford, Pergamon Press.

Kumar, R. (1996). Research Methodology. London, Sage Publications.

Kunnan, A., Ed. (1998). Validation in Language Assessment. Mahwah NJ, Erlbaum.

Kunnan, A. J., Ed. (1995). Test Taker Characteristics and Test Performance. 'Studies in Language Testing. Cambridge, Cambridge University Press.

Lado, R. (1961). Language Testing. London, Longman.

Laesck, K. B. and A. V. Kleeck (1987). "The Cloze Test as an Alternate Measure of Language Proficiency of Children Considered for Exit from Bilingual Education Programmes." Language Learning 37(3): 172-186.

Lange, D. L. (1990). "Priority Issues in the Assessment of Communicative Language Abilities." Foreign Language Annals 23(5): 430-407.

Latheaf, M. (1997). The Maldives. Re-engineering Education for Change: Educational Innovation for Development. UNESCO. Bangkok, Thailand, UNESCO - ACEID.

Lee, J. F. (1990). Issues in Defining and Assessing Reading Comprehension: The ACTFL Guidelines. Language Proficiency: Defining, Teaching and Testing. L. A. Arena. London, Plenum Press: 93-98.

Lee, J. J. (2000). Language Testing as a Technique to Enhance EFL Teaching Effects on Vocabulary Acquisition at the Intermediate Level. Faculty of Education. Philadelphia, The Pensylvania State University.

Lee, S. (1996). The Concurrent Validity of the Cloze Test with Essay Test Among Korean Students, ERIC ED 416713.

Lee, Y. P. (1985). Investigating the Validity of the Cloze Scores. New Directions in Language Testing. Y. P. Lee, C. Fok, R. Lord and G. Low. Oxford, Pergamon Press: 137-148.

Lee, Y. P., C. Fok, et al., Eds. (1985). New Directions in Language Testing. Oxford, Pergamon Press.

Leech, K. B. and J. Svartvik (1994). A Communicative Grammar of English. London, Longman.

Leedy, P. D. (1981). How to Read Research and Understand. New York, MacMillan Publishing Co., Inc.

Li, W. (1996). Research on the Cloze Procedure as Measurement of Reading Comprehension and its Use in the EFL Classroom, ERIC ED 274204.

Lindell, J. (1979). Nepal and the Gospel of God. Kathmandu, United Mission to Nepal.

Liskin-Gasparro, J. (2000). Proficiency Movement. Routledge Encyclopaedia of Language Teaching and Learning. M. Byram. London, Routledge.

Lombard, J. V. (1988). An Empirical Comparison of Direct and Indirect Method of Assessing Writing Proficiency, ERIC Clearinghouse TM012787.

London, N. A. (1997). A National Strategy for Systemwide Curriculum Improvement in Trinidad and Tobago. From Planning to Action: Government Initiative for Improving School level Practice. D. Chapman, L. Mahlck and A. Smulders. London, Pergamon Press.

Long, S. (2002). Action Research. Research-In-Action Symposium, Melbourne, Graduate School, Swinburne University.

Lowe, P. J. (1987). Interagency Language Roundtable Oral Proficiency Interview. Reviews of English Language Proficiency Tests. J. C. Alderson, K. Krahnke and C. Stansfield. Virginia, TESOL.

Lowe, P. J., Ed. (1988). Second Language Proficiency Assessment: Current Issues. New Jersey, Prentice Hall.

LTU (1990). National Language Certification Project. Language Testing Unit, Melbourne University. Melbourne.

Lumley, T. (1994). ESL Development Language and Literacy School Project: Teachers' manual. Canberra, The National Language and Literacy Institute of Australia.

Lumley, T., B. K. Lynch, et al. (1994). A New Approach to Standard Setting in Language Assessment. Annual Conference of the American Association for Applied Linguistics, Baltimore, MD.

Lynch, B. (1997). "In Search of the Ethical Language Test." Language Testing **14**(3).

Lynch, B. and F. Davidson (1994). "Criterion-Referenced Language Tests Development: Linking Curricula, Teachers, and Tests." TESOL Quarterly **28**(4): 727-743.

Lynch, B. K. and F. Davidson (1997). Criterion-referenced Testing. Encyclopaedia of Language and Education: Language Testing and Assessment. C. Clapham and D. Corson. Dordrecht, Netherlands, Kluwer Academic Publishers: 263-273.

MacFarlance, D. (1994). Fatalism and Development in Nepal. Nepal in the Nineties. M. Hutt. Delhi, Oxford University Press: 106-127.

Madsen, H. S. (1983). Techniques in Testing. Oxford, Oxford University Press.

Maheshwori, A. N. (1997). Assessment of School Achievement in India. Re-engineering Education for Change: Educational Innovation for Development. UNESCO. Bangkok, Thailand, UNESCO - ACEID.

Major, R., S. FitzMaurice, et al. (2002). "The Effect of Nonnative Accents on Listening Comprehension: Implications for ESL Assessment." TESOL Quarterly 36(2): 175-190.

Malla, K. P. (1975). English in Nepal: An Investigation. English in the Nepalese Education. K. P. Malla. Kathmandu, Ratna Pustak Bhandar.

Mansheng, Z. and S. Chunmeng (1997). Monitoring Achievement in China. Re-engineering Education for Change: Educational Innovation for Development. UNESCO. Bangkok, Thailand, UNESCO - ACEID.

Marsh, C. (1999). Perspectives: Key Concepts for Understanding Curriculum. London, Falmer Press.

Marsh, D. and J. R. Brown (2000). "Making Most of the High school Exit Examination." Thrust in Educational Leadership 29(5): 16-19.

Mazuoliene, Z. (1996). "The Year 12 Project in Lithuania (1995)." Language Testing Update(19): 25-27.

McGarrell, H. (1981). "Language Testing: A Historical Perspective." Medium 6(4).

McGill, I. and L. Beaty (2001). Action Learning. London, Kegan Page Limited.

McGrath, I. (1996). "Teaching, Learning, Testing and Evaluation." Journal of NELTA 1(1): 59-64.

McKay, P. (1992). ESL Development: Language and Literacy in Schools Project. Melbourne, National Language and Literacy Institute of Australia.

McKay, P. (1995). Developing ESL Proficiency Descriptions for the School Context: the NLLIA ESL Bandscales. Language Assessment in Action. G. Brindley. Sydney, National Centre for English Language Teaching and Research, Macquarie University. 31-63.

McKeller, E. (2002). Change Our Assessment Practices? Why should we? The Theory behind Assessment Practices. Paper presented at Learning Communities and Assessment Cultures Conference, East London, South Africa, University of Northumbria.

McKenna, A. (1996). The Development of the National Reporting System as a Social Literacy Event. Melbourne, University of Melbourne.

Mcleod, J. and J. Anderson (1975). Development of Standard Reading Test Designed to Discriminate Effectively at the Adolescent Level, ERIC ED 155602.

McNamara, T. F. (1991). The Role of Item Response Theory in Language Test Validation. Current Development in Language Testing. S. Anivan. Singapore, SEAMEO Regional Language Centre: 165-181.

McNamara, T. F. (1996). Measuring Second Language Performance. London, Longman.

McNamara, T. F. (1998). Policy and Social Considerations in Language Assessment. Annual Review of Applied Linguistics. B. Grabe. New York, Cambridge University Press. **18:** 304-319.

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

McNamara, T. F. and T. J. Quinn (1987). Australian Second Language Ratings. Reviews of English Language Proficiency Tests. J. C. Alderson, K. Krahnke and C. Stansfield. Virginia, TESOL: 7-9.

McNamara, T. F. and T. J. Quinn (1988). Issues in Second Language Learning: General and Particular. Brisbane, Deakin University Press.

McNeil, J. D. (2003). Curriculum: The Teacher's Initiative. New Jersey, Merril, Prentice Hall.

McNiff, J. (1994). Action Research for Professional Development. England, Hyde Publications.

Mehrens, W. A. (1989). Preparing Students to Take Standardised Achievement Tests. Practical Assessment, Research and Evaluation. Washington, D.C., ERIC Clearing House on Assessment and Evaluation. **2001**.

Messick, S. (1989). Validity. Educational Measurement. R. L. Linn. New York, McMillan.

Messick, S. (1996). "Validity and Washback in Language Testing." Language Testing **13**(3): 241-256.

Michael, G. (2002). The Training of Foreign Language Teachers: Development in Europe. Conference of the European Educational Research Association, Lisbon, Portugal, European Educational Research Association.

Milanovic, M. and N. Saville, Eds. (1996). Performance Testing, Cognition and Assessment. Studies in Language Testing. Cambridge, Cambridge University Press.

Miles, M. B. and A. M. Huberman (1984). Qualitative Data Analysis. London, Sage Publications.

Miller, H. and K. Piper (1986). Priorities in Language Education: A Survey of Community Views. Melbourne, Victoria, Australian Council for Educational Research (ACER).

MMW (1997). Mark My Words Video Series. Melbourne. Language Testing Research Centre. The University of Melbourne

MOE/COE (1998). SLC Examination (At a Glance). Kathmandu, Ministry of Education/Office of the Controller of Examinations.

MOE/OCE (2000). SLC Examination at a Glance. Kathmandu, Ministry of Education/ Office of the Controller of Examinations.

MOE (1992). National Education Commission Report. Kathmandu, Ministry of Education.

MOE (1996). Ministry of Education Nepal. Kathmandu, Ministry of Education/HMG Nepal.

MOE (1997). High Level National Education Commission Report. Kathmandu, Ministry of Education.

MOE (2003). Nepal: The Country Context. Kathmandu, Ministry of Education/HMG Nepal.

MOE/VIC (1990). Change is a Process, Not an Event: The Social Justice Curriculum Collection. Melbourne, Victoria, Ministry of Education, Victoria.

MOEC (1956). Education in Nepal. Kathmandu, Nepal, NNEPC.

MOEC (1961). All Round National Education Committee Report. Kathmandu, Ministry of Education and Culture.

Mohan, B. (1986). Where Language and Content should not be Confused. Language and Content. B. Mohan. Reading, MA, Addison-Wesley: 122-135.

Morrow, K. (1979). Communicative Language Testing: Revolution or Evolution? The Communicative Approach to Language Teaching. C. J. Brumfit and K. Johnson. Oxford, Oxford University Press: 143-157.

Moss, P. A. (1994). "Can there be Validity without Reliability?" Educational Researcher 23(2): 5-12.

Mossenson, L., Hill, P., Masters, G. (1997). <u>TORCH, Tests of Reading Comprehension</u>. Melbourne, Australian Council for Educational Research.

Moulton, K. (2000). Tests fail to Tell the True Story of Many Clever Children's Talents. <u>The AGE, 3 May 2000</u>.

Moya, J. and M. O'Malley (1994). "A Portfolio Assessment Model for ESL." <u>The Journal of Educational Issues of Language Minority Students</u> **13**(Spring): 13-36.

Munby, J. (1978). <u>Communicative Syllabus Design</u>. Cambridge, Cambridge University Press.

Nairn, A. (1980). <u>The Reign of ETS: the Corporation that Makes up Minds</u>.

Naughtin, P. (2003). A Measure of Success. <u>The Age, 13 October 2003</u>. Melbourne.

NEC (1992). Report of the National Education Commission. Kathmandu, national Education Commission, Nepal.

Nepal, T. R. (2001). Task Force Recommends to Categorise Schools. <u>The Rising Nepal: The National English Daily Newspaper</u>. Available: <u>http://www.gorkhapatra.gov.np/</u> Kathmandu.

Nidhi, B. (2004). Systematise Education Sector. <u>Gorphapatra, 12 July 2004.</u> <u>Available: http://www.gorkhapatra.gov.np/</u>. Kathmandu.

Noah, H. and M. Eckstein (1992a). Comparing National Systems of Secondary School Leaving Examinations. <u>Examinations: Comparative and International Studies</u>. M. Eckstein and H. Noah. Oxford, Pergamon Press.

Noah, H. and M. Eckstein (1992b). The Two faces of Examinations: A Comparative and International Perspective. <u>Examinations: Comparative and International Studies</u>. M. Eckstein and H. Noah. Oxford, Pergamon Press.

Noffke, S. E. and R. B. Stevension, Eds. (1995). <u>Action Research</u>. New York, Teachers' College Press, Columbia University.

Norris, J., J. Brown, et al. (1998). <u>Designing Second Language Performance Assessments</u>. Honolulu, University of Hawaii.

North, B. (2000). <u>The Development of Common Framework Scale of Language Proficiency</u>. New York, Peter Lang.

NT (2001). Private Education, Public Diktak. <u>Nepali Times, 16 May 2001</u>. Available: <u>http://nepalnews.com</u>. Kathmandu.

Nunan, D. (1987). Applying Second Language Acquisition Research. Adelaide, NCRC.

Nunan, D. (1999). Language Teaching Methodology. New York, Prentice Hall.

O'Brien, R. (1998). An Overview of the Methodological Approach of Action Research, www.web.net. 2002.

OCE (2000). The Trialling of an English Examination Based on the Grade 9 Curriculum. Kathmandu, Office of the Controller of Examination, MOE.

Ojerinde, A. (1980). Cloze Test as a Measure of English Language Proficiency Among Nigerian Primary School Pupils, ERIC EJ 238908.

Olaofe, I. A. (1992). "A Communicative Model For Assessing Second Language Performance." IRAL 30(3): 207-223.

Oller, J. (1972a). "Assessing Competence in ESL Reading." TESOL Quarterly 6(4): 313-323.

Oller, J. (1972b). "Scoring Methods and the Difficulty Levels in Cloze Tests of Proficiency in ESL." Modern Language Journal 56(3): 151-158.

Oller, J. (1972c). "Cloze Tests in English, Thai and Vietnamese. Native and Non-native Performances." Language Learning 2(1): 1-15, ERIC ED 078787.

Oller, J. (1977). "How Important is Language Proficiency to IQ and Other Educational Tests?" Occasional Papers in Linguistics 1: ERIC ED 144394.

Oller, J. (1978). Pragmatics and Language Testing. Advances in Language Testing. B. Spolsky. Washington, DC, Centre for Applied Linguistics: 39-58.

Oller, J. (1979). "The Psychology of Language and Contrastive Linguistics: The Research and the Debate." ERIC EJ 206643.

Oller, J. (1991). Current Research/Development in Language Testing. Current Development in Language Testing. S. Anivan. Singapore, SEAMEO Regional Language Centre: 2757.

Oller, J. (1992). "Foreign Language Testing; Part 2: Its Depth." ADFL Bulletin 22: 5-13.

Oller, J. and F. Hinofotis (1980). Two Mutually Hypotheses about Second Language Ability, ERIC ED 139267.

Oller, J. W. (1979). Language Tests as Schools. London, Longman.

Oller, J. W., Ed. (1983). Issues in Language Testing Research. Rowley, Mass, Newbury House.

Olmedilla, J. M. (1992). Tradition and Change in Examination Systems: A Comparison of Mediterranean and Anglo-Saxon Countries. Examinations: Comparative and International Studies. M. Eckstein and H. Noah. Oxford, Pergamon Press.

Onta, P., M. D. Chene, et al. (1996). Studies in Nepali History and Society. Kathmandu, Mandala Book Point.

PABSON (2003) PABSON Memoir 9

Papahoan, D. (2002). "Concept Mapping for Rater Training." TESOL Quarterly 36(2): 219-233.

Pennycook, A. (1990). "Towards a Critical Applied Linguistics for the 1990s." Issues in Applied Linguistics 1(1): 9-29.

Pigg, S. L. (1993). Unintended Consequences: The Ideological Impact on Development in Nepal. South Asia Bulletin. 13.

Popham, W. J. (1978). Criterion-referenced Measurement. Eaglewood Cliffs, Prentice Hall.

Popham, W. J. (1992). "A Tale of Two Test-specification Strategies." Educational Measurement: Issues and Practices.

Poudel, H. (2004). [Continuous Student Evaluation: What, Why and How]. Gorkhapatra, 19 October 2004. Available: http://www.gorkhapatra.gov.np/. Kathmandu.

Prabhu, N. S. (1987). Second Language Pedagogy. Oxford, Oxford University Press.

Pratt, D. (1980). Curriculum Design and Development. London, Harcourt, Brace and Jovaovich.

Pye, L. (1972). Aspects of Political Development. Boston, Newbury House.

Quinlan, m. and A. Scharaschkin (1999). National Curriculum Testing: Problems and Practicalities. British Educational Research Association Annual Conference, Brighton, University of Sussex.

Rafsanjani, M. (1995). Public Enterprises Reform: Problems and Prospects: A Comparative Study of Iran and Malaysia. Melbourne, Victoria University of Technology.

Rahman, A. (2002). Education for Development. Singapore, Institute of Southeast Asian Studies.

Rana, D. (2003). Education for All, All for Education. The Nepali Times, 24 January 2003. Available: http://www.nepalnews.com/. Kathmandu.

Raza, M., A. Ahmed, et al. (1990). School Education in India. New Delhi, National Institute of Educational Planning and Administration.

Rea, P. M. (1985). Language Testing and the Communicative Language Teaching Curriculum. New Directions in Language Testing. Y. P. Lee, C. Fok, R. Lord and G. Low. Oxford, Pergamon Press: 15-30.

Read, J. (1981). Directions in Language Testing. Singapore, SEAMEO/RELC.

Read, J. (1991). The Validity of Writing Test Tasks. Current Development in Language Testing. S. Anivan. Singapore, SEAMEO Regional Language Centre: 77-102.

Rea-Dickins, P. (1997). "So, Why do we Need Relationships with Stakeholders in Language Testing? A View from the U.K." Language Testing **14**(3).

Rea-Dickins, P. and K. Germaine (1996). Evaluation. Oxford, Oxford University Press.

Recela, L. B. (1997). Student Assessment from an International Perspective with Particular Reference to Link Between Curriculum and Assessment: The Philippines Experience. Re-engineering Education for Change: Educational Innovation for Development. UNESCO. Bangkok, Thailand, UNESCO - ACEID.

Reid, J. (1993). Teaching ESL Writing. Eaglewood Regents, Prentice Hall.

Reising, B. (1997). "Do We Need a National Language Policy?" Clearing House **70**(5): 231-232.

Ricento, T. (2000). "Historical and Theoretical Perspectives in Language Policy and Planning." Journal of Sociolinguistics **4**(2): 196-214.

Richards, C. and R. Schmidt, Eds. (2002). Longman Dictionary of language Teaching and Applied Linguistics. London, Longman.

Richardson, M. (1991). Evaluation of Whole Language and Traditional Language Arts Instruction Using a Cloze Procedure Test for Reading Comprehension., ERIC ED 339012.

Rijal, M. (2001). Development Planning: Focus on Decentralised Process. The Rising Nepal, 9 October 2001. Available: http://www.gorkhapatra.gov.np/. Kathmandu.

Rijal, M. (2001). The New Scheme of Education: Implementation Matters. <u>The Rising Nepal, 6 November 2001. Available:</u> <u>http://www.gorkhapatra.gov.np/</u>. Kathmandu.

Rivera, C. (1983). <u>An Ethnographic/Sociolinguistic Approach to Language Proficiency Assessment</u>. Clevedon, Multilingual Matters Ltd.

Rivera, C. (1984). <u>Language Proficiency and Academic Achievement</u>. Clevedon, Multilingual Matters Ltd.

Rivera, C. (1994). <u>Communicative Competence Approaches to Language Proficiency Assessment: Research and Application</u>. Clevedon, Multilingual Matters Ltd.

Rizvi, F. and S. Kemmis (1987). <u>Dilemmas of Reform: The Participation and Equity Program in Victorian Schools</u>. Geelong, Australia, Deakin Institute for Studies in Education, Deakin University.

Robinson, D. (1988). "Language Policy and Planning." <u>ERIC DIGEST</u> **12**(1): 1-4.

Robinson, F. (1989). <u>The Cambridge Encyclopaedia of India, Pakistan, Bangladesh, Sri Lanka, Nepal, Bhutan and the Maldives.</u> Melbourne, CUP.

Rudman, H. C. (1989). Integrating Testing with Teaching. <u>Practical Assessment, Research and Evaluation</u>. Washington, D.C., ERIC Clearing House on Assessment and Evaluation. **2001**.

Rudner, L. M. (1994). Questions to Ask When Evaluating Tests, http://ercae.net/pare/getnv.asp/. **2001**.

Rudner, L.M. (1989). A Testing Primer. <u>Understanding Achievement Tests: A Guide for School Administrators</u> L.M. Rudner, J. Conotey, and B. Plake. Washington DC: ERIC Clearing House

Saif, S. (2000). Theoretical and Empirical Considerations in Validating Washback: A study of ESL/EFL Learners. Canada, University of Victoria.

Salter, R. G. (2000). Swaraj and Sweepers: The JP Movement and Future of Transformational Politics. Melbourne, Victoria University of Technology.

Saussure, F. d. (1966). <u>Course in General Linguistics (Translated by Harris, R. in 1983)</u>. London, Duckworth.

Savignon, S. (2000). Communicative Language Teaching. <u>Routledge Encyclopaedia of Language Teaching and Learning</u>. M. Byram. London, Routledge.

Schaeffer, R. K. (2003). Understanding Globalisation: The Social Consequences of Political, Economic and Environmental Change. Oxford, Rowman and Littlefield Publishers, Inc.

Schedl, M. (1996). "The Test of English as a Foreign Language." Language Testing Update(19): 12.

Schedl, M., Gordon, A.,Carey, Patria, A.,Tang, K. L. (1996). "An Analysis of the Dimensionality of TOEFL Reading Comprehension Items." Language Testing Update(19): 31-32.

Scholz, G. and C. M. Scholz (1979). Testing in an EFL/ESP Context, ERIC ED 259548.

Scholz, G. and C. M. Scholz (1981). Multiple-Choice Cloze Tests of EST Discourse: An Explanation., ERIC ED 208656.

Schulz, R. A. (1977). "Discrete-Point versus Simulated Communication Testing in Foreign Language Testing in Foreign Languages." The Modern Language Journal 61(3): 94-101.

Scott, D. (2001). Curriculum Assessment. London, ABtex Publishing.

Seddon, D. (1994). Democracy and Development in Nepal. Nepal in the Nineties. M. Hutt. Delhi, Oxford University Press: 128-164.

Seidlhofer, B. (2000). Pronunciation Teaching. Rouledge Encyclopaedia of Language Teaching and Learning. M. Byram. London, Rouledge.

Selinker, L. (1972). "Interlanguage." International Review of Applied Linguistics 10: 201-231.

Sharma, A. (2001). Bitter Lesson. Spotlight. The National English Weekly Newsmagazine. 20.

Sharma, I. S. (1986). A Study of the Effects Psycho-Socio Factors on the Comprehensibility of Language Use in the Textbooks of Primary. New Delhi, NCRT.

Sharma, K. (1995). A Critical Evaluation of Primary Level Government Textbooks of the Nepali Language. Unpublished PhD Thesis. Kathmandu: Faculty of Humanities and Social Sciences. Tribhuvan University.

Shohamy, E. (1985a). A Practical Handbook in Language Testing for the Second Language Teacher. Ramat Aviv, Tel Aviv University.

Shohamy, E. (1985b). Al Alternative Model for Assessing Language Proficiency at the End of Secondary School. Washington DC, ERIC Clearing House. 2001.

Shohamy, E. (1990). "Language Testing Priorities." Foreign Language Annals 23(5): 385-394.

Shohamy, E. (1994a). The Role of Language Tests in The Construction and Validation of Second Language Acquisition Thoeries. Methodology in Second Language Acquisition. E. Tarone, S. Gass and A. C. Cohen. Cambridge, Cambridge University Press: 133-141.

Shohamy, E. (1994b). The Use of Language Tests for Power and Control. Georgetown University Roundtable on Language and Linguistics. J. Alatis. Washington DC, Georgetown University Press: 57-72.

Shohamy, E. (1994c). "The Validity of Direct versus Semi-direct Oral Tests." Language Testing 11(2): 99-123.

Shohamy, E. (1996). Competence and Performance in Language Testing. Performance and Competence in Second Language Acquisition. G. Brown, K. Malmkjaer and J. Williams. Cambridge, Cambridge University Press: 138-151.

Shohamy, E. (1997). Critical Language Testing and beyond. American Association for Applied Linguistics (AAAL) Annual Conference, Orlando, FA.

Shrestha, B. D. (1993). Nepalese Economy in Retrospect and Prospect. Kathmandu, Mandala Book Point.

Simon, L. (2000). Examination Orientation and the Opportunity Structure in Chinese Education: Case Studies of Kunming High Schools. Unpublished PhD Thesis. Canberra: The Australian National University.

Singh, G. B. (1996). SLC Examination in Nepal. Kathmandu, Research Centre for Educational Innovation and Development, Tribhuvan University.

Siwakoti, P. P. (1996). An Analysis of Reading Proficiency of the Secondary School Students of Jhapa District. Unpublished Master Thesis. Kathmandu: Department of English Language Education. Tribhuvan University.

Skehan, P. (1989a). "Language Testing'. State of the Art Article Part I." Language Testing 1: 211-221.

Skehan, P. (1995). Task-based Approaches to Language Testing. Conference on Testing and Evaluation in Second Language Education, Hong Kong, University of Science and Technology.

Skehan, P. (1998). A Cognitive Approach to Language Learning. Oxford, Oxford University Press.

Skehan, P.-. (1989b). "Language Testing: The State of the Art Article Part II." Language Testing 18(1): 1-9.

Skinner, B. F. (1957). Verbal Behaviour. New York, Appleton Century Croft.

Skinner, J. L. (2001). Environmental Turbulence and its impact on Strategic Change and Organisational Culture: The Case of the Queensland Rugby Union. Melbourne, Faculty of Human Development, Victoria University of Technology.

Spandel, V. (1997). Creating Writers: Linking Writing Assessment and Instruction. New York, Longman.

Speedy, S., Ed. (2003). Women Using Action Learning and Action Research. Australia, Southern Cross University Press.

Spolsky, B. (1969). Introduction: Linguists and Language Testers. Approaches to Language Testing. B. Spolsky. Washington D.C, Centre of Applied Linguistics: 1-19.

Spolsky, B. (1975). Language Testing: Art or Science? Association Internationale de Linguistique Appliquecé World Congress, France.

Spolsky, B., Ed. (1978a). Advances in Language Testing. Washington, DC, Centre for Applied Linguistics.

Spolsky, B. (1978b). Introduction: Linguists and Language Testers. Advances in Language Testing. B. Spolsky. Washington, DC, Centre for Applied Linguistics.

Spolsky, B. (1979). Some Major Tests: Advances in Language Testing Series 1. Washington DC, Centre for Applied Linguistics.

Spolsky, B. (1983a). Fourteen Years on - Later Thoughts on Overall Language Proficiency. Current Developments in Language Testing. A. Hughes and D. Porter. London, Academic Press.

Spolsky, B. (1993b). Testing the English of Foreign Students in 1930, ERIC. 2002.

Spolsky, B. (1995). Measured Words: The Development of Objective Language Testing. Oxford, Oxford University Press.

Spolsky, B. (1997). "The Ethics of Language Gatekeeping Tests: What have we Learned in a Hundred Years?" Language Testing 14(3).

Spost (2001). Education: Who is to Blame? The Kathmandu Post, 8 July 2001. Available: http://www.kantipuronline.com/. Kathmandu.

Spotlight (2001). SLC Results: Low Pass Rate. Spotlight: The National Newsmagazine. **20**.

Stake, R. E. (1998). Critical Issues. Integrating Assessment and Instruction in Ways that Support Learning. Tempe, NCREL Education Policy Analysis Archives. **2000**.

Stansfield, C. (1990). "Some Foreign Language Test Development Priorities for the Last Decade of the Twentieth Century." Foreign Language Annals **23**(5): 395-401.

Stenhouse, L. (1984). An Introduction to Curriculum Research and Development. London, Heinemann.

Stibbs, A. (1983). Assessing Children's Language: Guidelines for Teachers. London, Ward Lock Educational.

Storey, P. (1997). Examining the Test-Taking Process: A Cognitive Perspective on the Discourse Cloze Test., ERIC EJ 549224.

Stow, H. (1996). Mark my Words: Assessing Second and Foreign Language Skills. Parkville, Australia, NLLIA Language Testing Research Centre, University of Melbourne.

Stubbs, J. B. and G. R. Tucker (1974). The Cloze Test as a Measure of English Proficiency, ERIC EJ 107480.

Stump, T. (1977). Cloze and Dictation Tasks as Predictors of Intelligence and Achievement, ERIC ED 144402.

Sunuwar, L. R. (2000). ELT in the 21st Century in Nepalese Context. The Kathmandu Post, 16 April 2000. Available: http://www.kantipuronline.com/. Kathmandu.

Swaffer, J. A. and H. Byrnes (1991). Reading for Meaning: An Integrated Approach to Language Learning. New Jersey, Prentice Hall, Inc.

Swain, M. (1976). The Cloze Test as a Measure of Second Language Proficiency for Young Children, ERIC ED 129076.

Swain, M. (1985). Large-scale Communicative Language Testing: A Case Study. New Directions in Language Testing. Y. P. Lee, A. C. Fok, C. Lod and G. Low. Oxford, Pergamon: 35-46.

Syed, Z. (2003). "TESOL in the GULF: The Sociocultural Context of English Language Teaching in the Gulf." TESOL Quarterly **37**(2): 337-341.

Taylor, L. (2004). "Current Issues in English Language Testing." TESOL Quarterly **39**(1): 141-146.

Taylor, W. (1953). "Cloze Procedure: A New Tool for Measuring Readability." Journalism Quarterly **30**: 36-47.

Taylor, W. (1956). "Recent Development in the Use of the Cloze Procedure." Journalism Quarterly **33**(44-49).

Taylor, W. (1957). "Cloze Readability Scores as Indices of Individual Differences in Comprehension and Aptitude." Journal of Applied Psychology **41**(1): 19-26.

Thakur, A. (2002). On a New Course. Spotlight. **21:** 8-10.

Timsina, N. N. (2004). New Measures to Avoid Dismal SLC Results. The Kathmandu Post, 19 February 2004. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2000). SLC Papers not to be Checked at Home. The Kathmandu Post, 14 November 2000. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2001a). Non-elite Students Fare Well. The Kathmandu Post. 17 July 2001. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2001b). New Curriculum Blamed for Dismal Results. The Kathmandu Post, 17 July 2001. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2001c). Budget for Education Sector Inadequate. The Kathmandu Post, 11 July 2001. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2001d). SLC: A Test of Cheating Skills. The Kathmandu Post, 22 April 2001.Avaliable: http://www.kantipuronline.com/. Kathmandu.

TKP (2001e). Our Teacher has Become Lazy. The Kathmandu Post, 8 February 2001. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2001f). Education: Who is to Blame? The Kathmandu Post, 8 July 2001. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2001g). Uncertainty Shrouds SLC Results. The Kathmandu Post, 26 June 2001. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2001h). Parsa Bags Better Results in SLC Examinations. The Kathmandu Post, 23 July 2001. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2001i). SLC Signals. The Kathmandu Post, 17 July 2001. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2002a). There is a Dire Need of Administrative Reform. The Kathmandu Post, 22 April 2002. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2002b). Changing Curriculum. The Kathmandu Post, 15 march 2002. Available: http://www.kantipuronline.com/. Kathmandu.

TKP (2004). Nepal's Primary School Drop Out Rate 40 Percent. The Kathmandu Post, 10 February 2004. Available: http://www.kantipuronline.com/. Kathmandu.

Tomal, D. R. (2003). Action Research for Educators. Oxford, The Scarecrow Press Inc.

TRN (2001a). Private Sector's Role Vital in Upgrading Education. The Rising Nepal, 10 July 2001. Available: http://www.gorkhapatra.gov.np/. Kathmandu.

TRN (2001b). Revamping Education. The Rising Nepal, 6 May 2001. Available: http://www.gorkhapatra.gov.np/. Kathmandu.

TRN (2001c). Quality Education. The Rising Nepal, 17 July 2001.Available: http://www.gorkhapatra.gov.np/. Kathmandu.

Truus, K. (1996). "The Year 12 Project in Estonia (1995)." Language Testing Update(19): 19-22.

Turner, C. E. and J. A. Upsher (2002). "Rating Scales Derived from Student Samples: Effects of the Scale Maker and the Student Sample on Scale Content and Student Scores." TESOL Quarterly 36(1): 49-67.

Tyabane, M. (2003). Encouraging Reflective Teaching Using Action Research. Women Using Action Learning and Action Research. S. Speedy. Australia, Southern Cross University Press.

UCLES (1990). The IELTS Handbook. London, The British Council, IDP Australia, University of Cambridge Local Examination Syndicate.

UCLES (1995). The IELTS Handbook. London. The British Council and University of Cambridge Local Examination Syndicate

UCLES (2000). The IELTS Handbook, UCLES, The British Council and IDP.

Umar, J. (1997). Assessment and Examinations in Indonesia. Re-engineering Education for Change: Educational Innovation for Development. UNESCO. Bangkok, Thailand, UNESCO - ACEID.

Underhill, N. (1987). Testing Spoken Language. Cambridge, CUP.

UNESCO (1984). Diagnostic Studies of Educational Management: Nepal. Bangkok, UNESCO, Regional Office.

UNESCO, Ed. (1997). Re-engineering Education for Educational Innovation for Development. Bangkok, Thailand, UNESCO-Thailand.

UNESCO (2002). UNESCO Statistics Yearbook. Paris, UNESCO.

Upadhyay, N. P. (2001). Development Administration in Nepal: Multidimentional Approach Essential. The Rising Nepal, 15 August 2001. Available: http://www.gorkhapatra.gov.np/. Kathmandu.

Ushiro, U. (2000). Reading. Routledge Encyclopaedia of Language Teaching and Learning. M. Byram. London, Routledge.

Vallette, R. M. (1977). Modern Language Testing. New York, Harcourt and Brace, Inc.

Van Ek, J. A. (1982). The Threshold Level for Modern Language Learning in Schools. Strasbourg, France, Council of Europe.

Verma, Y. P. and B. D. Pande (1989). Causes of Failure in English in SLC Examination. Kathmandu, Research Centre for Educational Innovation and Development.

Vir, D. (1988). Education and Polity in Nepal: An Asian Experiment. New Delhi, Northern Book Centre.

Viswat, L. and L. A. Jackson (1994). The Effect of Strategies Training on Student Error on a Listening Cloze, ERIC ED 411669.

Vollmer, H. J. (1983). The Structure of Foreign Language Competence. Current Developments in Language Testing. A. Hughes and D. Porter. London, Academic Press: 3-30.

Wadsworth, Y. (1997). Do it Yourself Social Research. Melbourne, Australia, Allen and Unwin.

Walker, B. C. and T. Haslett (2002). Action Research and Systematic Thinking. Working Paper Series. Melbourne, Department of Management, Monash University.

Wall, D. (1996a). "Introducing New Tests into Traditional Systems: Insights from General Education from Innovation Theory." Language Testing 13(3): 334-354.

Wall, D. (1996b). "The Baktic States Year 12 Examination Project." Language Testing Update(19): 15-18.

Wall, D. and J. C. Alderson (1996). Examining Washback: The Sri Lankan Impact Study. Validation in language testing. A. Cumming and R. Berwick. Clevendon, Avon, Multilingual Matters: 194-221.

Waters, G. A. (1998). "Critical Evaluation for Education Reform." Education Policy Analysis Archives 6(20): 1-36.

Watkins, D. and M. Regmi (1990). "An Investigation of the Approach to Learning of Nepali Tertiary Students." Higher Education 20(1): 459-469.

Weaver, G. C. (1979). Using the CLoze Procedure as a Teaching Technique, ERIC EJ 201036.

Weiler, H.N. (ed.), (1992). National Educational Policies and Implications. Educational Planning and Social Change: Report on IIEP Seminar. Weiler. H.N. Paris: UNESCO

Weir, C. J. (1988). The Specification, Realization and Validation of an English Language Proficiency Test. Testing English for University Study. ELT Documents 127. A. Hughes. London, Modern English Publications/The British Council: 45-110.

Weir, C. J. (1990). Communicative Language Testing. New York, Prentice Hall.

Weir, C. J. (1993). Understanding and Developing Language Tests. New York, Prentice Hall.

Weir, C. J. (1994). Reading as Multi-divisible or Unitary: between Scylla and Charybdis. Annual Seminar, Regional Language Centre (RELC), Singapore.

Weir, C. J. (1997). The Testing of Reading in a Second Language. Encyclopaedia of Language and Education. C. Clapham and D. Corson. Netherlands, Kluwer Academic Press: 39-49.

Weir, C. J. (1999). The Formative and Summative Uses of Language Tests Data: Present Concerns and Future Directions. Studies in Language Testing. A. Davies. Cambridge, Cambridge University Press. 11.

Wesche, M. B. (1985). Introduction. Language Performance Testing. C. Philip, L. Raymond and M. B. Wesche. Ottawa, University of Ottawa Press.

Wesche, M. B. (1987). "Second Language Performance Testing: The Ontario Test of ESL as an Example." Language Testing 4(2): 28-47.

West, A., A. Edge, et al. (1999). Secondary Education Across Europe: Curricula and School Examination Systems. London, Centre for Educational Research.

White, E. M. (1994). Teaching and Assessing Writing. San Francisco, Jossey-Bass Publishers.

White, R. C. (1991). Process Writing. London, Longman.

Widdowson, H. G. (1978). Teaching Language as Communication. Oxford,. Oxford University Press.

Widdowson, H. G. (1999). Communicative Language Testing: The Art of the Possible. Studies in Language Testing. A. Davies, A. Brown, C. Edler, N. Iwashita and T. F. McNamara. Cambridge, Cambridge University Press. 11.

Wiersma, W. (1975). Research Methods in Education. Illinois, F.E. Peakock Publishers.

Wilde, J. and S. Sockey (1995). Evaluation Handbook. Albuquerque, Evaluation Assistance Centre, New Mexico High Land University.

Wilkins, D. A. (1976). Notional Syllabuses: A Taxonomy and its Relevance to Foreign Language Curriculum Development. London, Oxford University Press.

Williams, P. L. (1989). Using Customised Standard Tests. Practical Assessment, Research and Evaluation. Washington, D.C., ERIC Clearing House on Assessment and Evaluation. 2001.

Witt, E. A. (1995). Issues in Constructing an Analytical Scoring Scale for a Writing Assessment, ERIC TM 023627.

Wylie, W. (1997). International Second Language Rating Scales: Available at www.griffith.edu.au. Western Australia, Griffith University. 2003.

Yalden, J. (1987). Principles of Course Design for Language Teachers. Oxford, Oxford University Press.

Zeichner, K. M. and S. E. Noffke (2001). Practitioner Research. Handbook of Research on Teaching. v. Richardson. Washington, D.C., American Educational Research Association.

Zhong, Y. H. and G. S. Chun (1985). The English Proficiency Test Used in China. New Directions in Language Testing. Y. P. Lee, C. Fok, R. Lord and G. Low. Oxford, Pergamon Press: 59-66.

Zuber-Skerritt, O. (1996). Emancipatory Action Research for Organisational Change and Management Development. New Directions in Action Research. Zuber-Skerritt, O. London. The Falmer Press

Appendix A
Test Specifications

The Format

Target Language Use	General Description	Prompts	Sample Tasks	Response Characteristics	Scoring Scheme	Supplementary Information

Figure: The test specification format used in the present research

The seven components of the test specification are described below:

Test Specifications

1. LISTENING   Full Mark –10

1.1 Target Language Use

1.1.1 Sub-skills

- ❖ To listen and do
- ❖ To listen and arrange
- ❖ To listen and match
- ❖ To listen and label
- ❖ To listen and complete
- ❖ To listen and answer
- ❖ To listen and make notes
- ❖ To listen and summarise

1.1.2 Aspects

- ❖ Informal Listening
- ❖ Academic listening
- ❖ Formal (non-academic) listening
- ❖ Comprehension of implicit meaning
- ❖ Comprehension of explicit meaning
- ❖ Low (Local) level meaning
- ❖ High (global) level meaning (see General Description below)

1.1.3 Domains

- ❖ Home and family
- ❖ Community
- ❖ School and classroom
- ❖ Education
- ❖ Public places – restaurant, shopping centres, parks
- ❖ Entertainment
- ❖ Movies
- ❖ Games and sports
- ❖ Exercises
- ❖ Transportation
- ❖ Communications and technology
- ❖ Everyday life

1.2 General Description

Listening in the target language is an important source of exposure and crucial for the learners of a foreign language because this is from where they pick up the models of the spoken forms. Listening is also important because nearly half of any conversational act is listening. A successful conversant is one who demonstrates the ability to comprehend what others say during a conversation and to respond to it. The lack of such ability results in communication breakdowns and misunderstanding. Ability to listen well is also extremely important in academic settings where learners must listen to lectures, presentations, taped materials or live newscasts, and understand them and record information or ideas in the form of notes.

The candidates will demonstrate their ability to (i) understand and retrieve information from listening to a wide variety of academic and non-academic texts including conversations; (ii) convert textual information into non-text forms such as notes, tables, charts, and diagrams; and follow and respond to spoken prompts in completing a task or a process; or labelling parts of a whole. The candidates will demonstrate their listening ability in four major areas of cognitive operations – orientations, knowledge, comprehension and application which are described below.

The specific listening skills on which the candidates will demonstrate their ability are:

a. Orientation meaning
- establishing relationship between persons;
- establishing references between elements of an utterance;
- recognising explicit (stated) and/or implicit implied) elements in the text, for example, time, date, event, names of people, places and objects, etc.

b. Detail meaning
- recognising simple lexical meaning;
- recognising propositional meaning(s) of the text;

c. Main ideas
- identifying topic of the text;
- identifying main headings of the text;
- paraphrasing the text

d. Implications
- deriving meanings from the candidates' application of background knowledge and logic to the text.

Abilities stated in 'a' and 'b' above are bottom-up elements requiring lower-level of listening proficiency whereas elements 'c' and 'd' are top-down elements requiring higher level of proficiency.

1.2.1 Materials

The candidates will listen to two authentic texts of short length (100- 150 words) and appropriate difficulty level.

Text One will be a one-to-two minute long text taken from one of the following:
- academic lectures
- advertisements
- announcements
- instructions
- directions
- narratives
- descriptions
- newscasts (local radio)

Text Two will be a one-to-two minute long conversation taken from one of the following situations:

- informal/casual exchanges
- family/friendly conversations
- interviews
- group discussions
- telephonic conversations

1.3 Prompts

Each of the two listening texts will have 4-8 test items requiring about 5 minutes to complete. Two test tasks will be selected from the following types of tasks.

A. Listen and complete a table, a chart or a diagram with eight blanks;

   [Example – Task Type A]

B. Listen and label objects or parts of an object with four to six blanks;

   [Example – Task Type B]

C. Listen and match eight ideas listed in two columns;

   [Example – Task Type C]

D. Listen and arrange six ideas presented in a mixed order;

   [Example – Task Type D]

E. Listen and recognise eight true or false statements based on paraphrase of the text;

   [Example – Task Type E]

F. Listen and answer eight questions based on the explicit (stated) information;

   [Example – Task Type F]

G. Listen and answer four questions based on the implicit (implied) information requiring answers resulting from the application of background knowledge and logic;

   [Example – Task Type G]; and

H. Listen and complete a summary or note of text presented in 25-35 words with four blanks.

   [Example – Task Type H]

Candidates will be required to record their answers only in the answer sheets provided to them.

The blanks in Task Type A should be numbered 1-8 requiring the candidates to record only answers against their numbers.

In Task type B, the candidates should be asked to identify objects or write names of the objects or parts of an object.

In task type C, items in the left column and items in the right column should be named Column A and Column B respectively. Items given in Column A should be numbered 1-8, and items given in Column B should be numbered a-i. (Note – Column B has an additional item, which is an incorrect distracter.) Candidates will be required to record the numbers of Column A with their matching alphabets in Column B.

For task D, the statements should be numbered a-f, Candidates will be required to put the alphabet a, b, c, d, e or f, according to their matching number 1, 2, 3, 4, 5 or 6.

Statements in task type E will be numbered 1-8. Candidates will be required to write T for true statements or F for false statements, whichever they choose, against the number of the statements.

Candidates will be required to write answers for questions in task types F and G.

For task type H, the blanks will be numbered 1-6, the candidates will be asked to record answers only against the number.

1.4 Sample Tasks

See Appendix A1.1

1.5 Response Characteristics

The candidates will listen to the tape played once and (in task type A) fill-in the numbered blanks with appropriate words from the text or supply their own words. They are required to write only the word or phrase of their choice.

In some task types, they are required to indicate the answers by simply writing alphabets or numbers of their choice or just by ticking the right box or item.

When answering questions, they are required to answer questions in their own words.

1.6. Scoring Scheme

1.6.1 Criteria

- Accuracy. Accuracy refers to the choice of item required for the intended meaning of an utterance. In a more general sense, it may also mean, in relation to grammatical accuracy, the candidates ability to understand and produce well-formed structures. The estimates of such abilities generally relate to a proficiency scale; for example the one given in Appendix 3.
- Contextual appropriateness. This refers to the acceptability of the selected item (which is not the restoration of the exact or original item) in the overall context of situation in which it is to be used.

1.6.2 Mark Distribution

The allocation of marks to the two selected tasks depends on the nature of the tasks and the discretion of the test developers. The total mark to be allocated is 10.

In the case of sample tasks given here, each correct answer in sections A and B (in Task Type A) should be awarded ½ a mark. Each appropriate answer in C should be awarded 1 mark. Spelling errors should be ignored. Grammar errors may be considered. The full mark for this question is 6.

Similarly, in Task Type B, the candidates put a tick against the things mentioned in the tape. There are eight items to be ticked each one for ½ a mark. The full mark for this task is 4.

1.7 Supplementary Information

1.7.1 Resources

A tape player and a taped material for each of the examination rooms

Two invigilators in each of the exam rooms, one of whom should be adequately trained to handle the equipment (cassette player).

1.7.2 Timeframe

Time for listening will be allocated as follows:

	To listen to text	To read instructions and tasks	To record answers
Text 1	11/2-2 mns	11/2-2 mins	5 mins
Text 2	11/2-2 mins	11/2-2 mins	5 mins
Total	**3-4 mins**	**3-4 mins**	**10 mins = 16-18 mins.**

Figure: Time allotted to listening activities

For listening texts and sample tasks, see appendix A1.1.

The invigilators should make sure that (i) the candidates are all seated before they play the tape, (ii) they have answer-sheets, (iii) they have a pen or a pencil, (iv) the cassette player is placed at a higher level so that every one can see/hear it, (v) the volume of the cassette player is loud enough for the whole room, (vi) they write their registration or symbol number and other required information on the answer-sheets before listening, and (vii) the answer-sheets has the chart, etc. clearly printed.

The invigilators should allow 2 minutes to read and fill in the required information (name of the centre, registration number, etc.)

After the tape has been played, the invigilators should allow 11/2-2 minutes before moving on to the next listening text. The tape should be played only once.

1.7.3 Answers

For answers of the Listening Sample Tasks, see Appendix A2.1

2. SPEAKING   Full Mark 10

2.1 Target Language Use

2.1.1 Sub-skills

- Using correct pronunciation
- Speaking coherently
- Describing a person, a place or a process
- Participating in social conversations
- Questioning
- Supporting and clarifying or giving opinions
- Paraphrasing
- Summarising
- Comparing, analysing or interpreting non-text materials
- Narrating experiences or events

2.1.2 Aspects

2.1.2.1 Phonological skills
- Pronunciation
- Stress
- Intonation

2.1.2.2 Interaction skills
A. Expressing Ideational Meaning
- Factual information
- Argument
- Likelihood

B. Expressing Modal Meaning
- Mood
- Emotion
- Attitude

C. Expressing Phatic Meaning
- Greetings
- Acknowledgement
- Empathy

2.1.2.3 Discourse skills
- Cohesion
- Enumeration
- Addition
- Logical sequence
- Contrast
- Apposition

293

2.1.2.4 Academic/Study Skills

- Asking for/giving meaning of words, phrases, or sentences;
- Asking for/giving instructions or directions as to how to do certain things;
- Asking for/giving/reporting messages;
- Asking for/giving information
- Describing/interpreting information presented in a non-text form;
- Narrating personal experiences
- Asking for/supporting/giving opinions or arguments
- Asking for/giving descriptions of a person, thing or process;
- Using social skills such as participating in a friendly/ family conversation
- Paraphrasing/summarising an utterance/text for clarification

2. 1.3 Domains

- ♦ Family/Friendship
- ♦ School/Education
- ♦ Future Ambitions
- ♦ Hobbies
- ♦ Town life/ country life
- ♦ Community
- ♦ Work/Employment

2.2 General Description

Speaking in a language involves linguistic ability (knowledge of sounds, vocabulary, grammar and skills) as well as communicative ability (knowledge of what to say when, why, how and with whom). In other words, knowledge of various components and skills of the language alone is not sufficient for an individual to enable him/her to communicate competently and effectively. A competent communicator should know, along with the language elements and skills, how to make choices of linguistic elements specific to a given situation.

A speech event consists of a speaker, an interlocutor, a purpose of speaking, a topic and a context of situation.

Developing speaking ability is the primary goal of most language programmes. In Nepal, however, given the fact that English is learned as a foreign language primarily

for academic purposes, the skill of speaking receives less importance, only 10% of the total weightage, than other skills.

The candidates will demonstrate their ability to (a) engage effectively in conversations in informal and formal communicative situations such as friendly chats, shopping conversations, telephonic conversations, family conversations, expressing opinion, arguing and so on; (b) describe a person, an object, a process, and narrate an experience or describe events; and (c) ask for/ give/report information or messages in formal and informal situations.

The nature of speaking is transitory, rudimentary and often ill-formed. It is also the most complex, labour-intensive and time-consuming activity (Berry, 1997). A measurement or assessment scheme for this skill must maintain a proportional balance among accent, fluency and accuracy.

The pressure on the test developers will be develop an oral test that reveals the actual ability present in the candidate or reflect on his or her current communicative language ability. A test of speaking, therefore, should incorporate (a) the type of ability involved, (b) authenticity of the material(s), (c) authenticity of the task (s) on which the candidates are required to perform, and (d) a context in which the performance is to take place.

The following test formats are often used to test oral or speaking ability:
- **Verbal essays**, which require the candidates to speak for a given length of time on one (or more) specified general topic;
- **Oral descriptions**, which require the candidates to speak about or perform tasks on a picture, a picture sequence, event or situation provided to them by engaging them in a conversation;
- **Oral presentations**, these may take the forms of lecturettes or prepared short-talks on the topic(s) supplied in advance;
- **Controlled interviews**, which follow pre-determined procedures or tasks to elicit responses from the candidates;

- **Free interviews**, for which no pre-determined procedures or tasks are set/used, but the conversation unfolds in an unstructured fashion;

- **Pair or group interactions**, in which participants are required to participate in a conversation on an assigned topic.

- **Role-plays**, in which candidates are required to engage actively in oral interaction in artificially enacted or contrived situations.

Of the test formats mentioned above, only the first two formats, **verbal essays** and **oral descriptions** are recommended in the current situation because -

(a) they are easy to construct and administer,
(b) they allow a wide range of criteria such as fluency, accuracy, accent, vocabulary, etc. to be applied,
(c) they are comparatively less costly, and finally,
(d) they involve, lesser degree of administrator/ tester's training as the testers or test administrators do not need to effectively engage in interactions with the candidates to elicit the required data.

For other formats mentioned, expertise on the part of the testing personnel and infrastructure development will be needed before they can be practised.

2.2.1 Materials

Speaking tasks may be developed on the materials from the following fields:

- family and relationships
- school/time table/studies/study habits
- future ambitions
- work/employment/profession
- preferences on city/town/village life
- games and sports
- dresses/clothes
- entertainment/TV/movie/hobby
- vacation/holidays/
- festivals/celebrations
- earning and spending
- environment/ climate/health/sanitation/diseases
- people/places/events/house and housing
- food
- transportation/travelling/tourism

2.3 Prompt

Two types of tasks should be developed on the selected materials/fields, each one of 9 minutes.

2.3.1 Task One – Verbal Essay (on a prepared topic)

This is an open-ended task in which the candidates are not restricted in what they say and how they complete the task. The candidates will be required to choose on the spot one of the three topics. After they have a topic, they will have at least 2 minutes to prepare, and 5 minutes to speak on it. The allotted time is distributed as follows:

Interlocutor/assessor to set a context	-	2 minutes
Time to prepare for the candidates	-	2 minutes
Time to speak on the selected topic	-	5 minutes
Total	-	9 minutes

Figure: Time allotted to speaking activities

2.3.2 Requirement for the selection of a topic/tasks

Topics or tasks developed on the materials selected from the aforesaid fields should be:
- ◆ Meaningful;
- ◆ Relevant;
- ◆ Motivating;
- ◆ Authentic;
- ◆ Specific;
- ◆ Requiring strategies common in academic speaking.

2.3.3 Task Two: Oral Descriptions (impromptu)

This is an open-ended task. Candidates answer the question or respond to the task in their own way. The candidates are given a subject matter (a picture, a series of pictures, a topic or a context) and a task to respond to by engaging them in a conversation. The interlocutors/assessors should maintain informality of the situation so that the candidates feel comfortable and relaxed. They should be given opportunity to ask questions and respond to the interlocutors/assessors' questions which should set a context for an impromptu response to the context.

The conversation should lead the candidates to speaking on the topic specified or respond to the task arising from the conversation. The subject matter for this task should be selected from the list given under 'materials' (see General Description).

The time for this task is allotted as follows:

Conversation between the candidate and interlocutor/assessor
to set a context and specify a task or a topic   -   4 minutes

Time to speak on the selected topic   -   5 minutes
_____
                                                     Total   -   9 minutes

Figure: Time allotted to speaking activities

2.4 Sample Tasks

See Appendix A1.2

2.5 Response Characteristics

The candidates in **Task Type One** will choose one of the three topics, written and folded in separate slips of paper. The candidates select the slip of their choice and read out the topic.

Topics will not be disclosed prior to selection.

After they find out about the topic to speak on, they are given two minutes to prepare after which they can speak for five minutes.

The interlocutor/assessors do not interrupt during this period. The responses may be tape-recorded for future responses.

In **Task Type Two**, the candidates participate in a conversation initiated by the interlocutors/assessors. They will lead the conversation to provide a topic or a situation for an impromptu response or speech. The candidates are not given any time for

preparation. They speak on the task or respond to the situation provided for five minutes. The interlocutor and/or assessor may ask question while the candidate is speaking.

Task Type One will reveal the candidate's ability in general or academic speaking whereas Task Type Two will show their communicative ability.

This session may be recorded for future reference.

2.6 Scoring Scheme

The following criteria and their breakdowns should be used to score performance on speaking. The candidate's performance on the criteria is assessed on the basis of overall or global communicative effectiveness.

The criteria for scoring on speaking are as follows:
1. Accent
2. Appropriateness
3. Language range
4. Accuracy
5. Fluency
6. Formality
7. Content
8. Interpersonal skills
9. Ease at conversation
10. Confidence

The scale given in Appendix 3 will be used for scoring on this skill.

2.7 Supplement Information

2.7.1 Resources

An audiocassette recorder for every examination centre will be needed. The sessions with students should be recorded for future references, for rechecking marking, for monitoring marking or in case there is complaints or disagreement between the assessors.

Two adequately trained assessors who play the role of both, interlocutor and assessor will be needed.

2.7.2 Time Allotment

The time allocated to the recommended two tasks are as follows:

Task Type One   9 minutes
Task Type Two   9 minutes

2.7.3 Patterns of Conducting Speaking Tests

The following two patterns of conducting a speaking test are recommended:

Pattern 1

The candidate responds to the task or speaks for the period of allotted time. Neither the interlocutor nor the assessor engages in a conversation with the candidate. Task Type One falls in this category on which there is a minimum interaction with the candidate. Both the interlocutor and assessor do the scoring. A joint score is submitted.

Candidate	Interlocutor/assessor
	Interlocutor/assessor
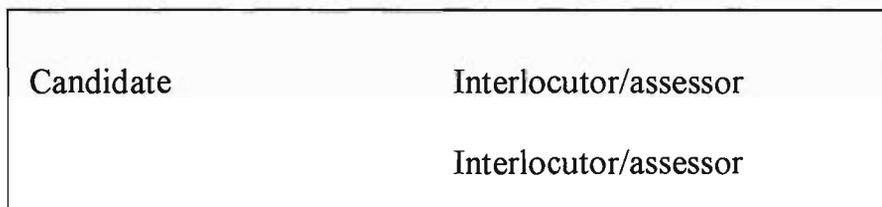
Figure: Format for conducting speaking test

Pattern 2

The candidate interacts with both the interlocutor and assessor and both of them do the scoring. A joint scoring is submitted.

Candidate	Interlocutor/assessor
	Interlocutor/assessor
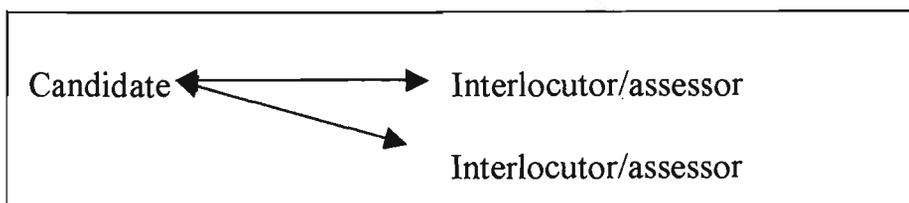
Figure: Format for conducting speaking test

The assessor(s) should record all the details of the candidates such as name, registration number, name of the school, name of the examination centre, etc. before any testing begins. They should make it sure that the candidate settles down before starting the process of testing.

3. READING   Full Mark – 45

3.1 Target Language Use

3.1.1 Sub-skills

- Reading a text quickly and identifying the main idea of the text
- Reading a text quickly and locating specific information in the text
- Reading a text quickly and establishing references and co-references of language items
- Read a text and draw inferences from it
- Getting the gist of a text
- Identifying the main ideas of a text and distinguishing them from supporting details
- Deducing implicit meaning from a text
- Understanding intra- and inter-sentence/paragraph relationship
- Understanding the communicative function of a sentence/paragraph
- Comparing, analysing and interpreting non-text materials
- Describing a person, a place, an event or a process.
- Synthesising information, making notes from a text and summarising

3.1.2 Aspects

- Scanning to locate specifically required information
- Skimming to obtain the gist of the text or general impression of the text
- Intensive reading of short/medium length texts for comprehension and language skills
- Extensive reading of short/medium length texts for information and pleasure
- Transcoding information presented in diagrammatic display involving completing a diagram/table/chart, constructing one or more diagrams/tables/charts and interpreting a diagram/table/chart.
- Basic reference skills, i.e. understanding and use of graphic presentation, dictionaries, and so on
- Note-making skills
- Reading literary pieces for enjoyment

3.1.3 Domains

- Family/neighbourhood/community/friendship/relationship
- Study/academy/school/education
- Work/employment/job
- Religions
- Health/diseases/fitness/exercise/games/sports
- Transportation/travel/tourism
- Entertainment/TV/movie
- Vacation/holidays/festivals

- Food/cooking
- Hobbies/ambitions
- Town/city/country life/places
- Earnings/spending/shopping

3.2 General Description

Reading in the secondary English curriculum of Nepal has the most important place as it carries the most weightage, 45 per cent of the total. It is only natural for this skill, then, to dominate most English lessons. It is, however, important to understand that reading, from teaching as well as testing point of view, is no longer a mere process of decoding information or message from a text. Rather it is an integration of the bottom-up process (which involves decoding meaning from the printed text), top-down or data driven process (which requires sophisticated or higher level linguistic skills) and schema which utilises the prior or background knowledge of the reader. A good testing scheme of reading should incorporate all these three aspects of reading.

The candidates should be able to demonstrate their ability to –

(a) read and understand a variety of texts of short and medium length, and retrieve specific and general meaning through skimming, scanning and synthesising (note-making and summarising);
(b) read and recognise the structure and organisation of text and the use of cohesive devices;
(c) read and deduce meaning through contextual, syntactic and other linguistic clues;
(d) read and interpret non-text materials; and
(e) read and describe cultural significance of a text.

3.2.1 Materials

The materials will be selected from the following sources:

- readers (prescribed texts)
- fictional stories
- poems
- information leaflets/brochure
- dramas
- memoirs
- letters
- essays on health/food/eating or dietary habits
- conversation/anecdotes

- ◆ newspaper articles
- ◆ interviews
- ◆ autobiographies
- ◆ advertisements
- ◆ announcements
- ◆ reports

3.2.2 Nature of texts

Materials should not be selected from a specific discipline, which may disadvantage certain candidates due to unfamiliarity of the topics. They should be selected from the areas or fields of general interest. They should not be over simplified so that it looses its authenticity. Minor simplifications may be done to maintain anonymity.

3.2.3 Length of the text

Both medium length texts and short texts may be selected. By medium length it is meant that the text should be 300 words long. By short length it is meant that the text should be 150 words long.

3.3 Prompts

Four texts should be selected from the list of sources given in General Description.

3.3.1 Text One

The first text should be selected from the readers, which consists of the prescribed texts. The length of the text should be about 300 words. The tasks on this text should follow the format of the tasks given in the textbooks. There should be four tasks on this text. They are:

Task A: a vocabulary task, there will be 10 items based on the matching format, In this task, the candidates will demonstrate their ability to guess the meaning of words with the help of the contextual as well as linguistic clues.

The words selected from the first reading text of the question paper will be of general nature rather than of technical nature. The words to be given in the task should be

examples of lexical as well as syntactic importance and should fall within the range recommended by the secondary curriculum.

The tasks for vocabulary can be matching, multiple choice or fill-in. There may be five to ten items on each. If multiple-choice items are given, each item should consist of four alternatives which in nature, should be as follows:

- the correct answer,
- a synonym but inappropriate to the context,
- an antonym,
- a randomly selected word not fitting the context.

For the matching task, words and their meanings should be arranged in two separate columns named as column A and Column B. There should be an additional item in the column B to avoid automatic selection of the last item. The meanings in column B will be arranged in a mixed order.

An alternative to this task is to give meanings in an orderly fashion in the task, asking candidates to find the word for each of the meanings in the text as they read it.

For the fill-in task, a list of words will be provided. Candidates will be required to fill in the numbered blanks of the summary of the text, selecting appropriate words from the list.

Task B: a comprehension task of reordering 7 statements. The statements should be based on the simple meaning of the text highlighting its main ideas or events.

Task C: a comprehension task on the Question-and-Answer format. There will be five comprehension questions based on the implicit meaning of the text. These questions may be based on the main ideas or themes or events of the text.

Task D: a grammar task with five items on completion format. In this task, the candidates show their ability to frame sentences on the basis of their awareness of one of the following prescribed grammatical points:

♦ expression of probability and conditionals,
♦ uses of modal auxiliaries
♦ tenses and inflections
♦ clauses

- ♦ causatives
- ♦ prepositions
- ♦ articles and determiners
- ♦ reported speech

Framing sentences on a given sentence pattern, completing incomplete sentences, fill-ins and multiple-choice items are some of the tasks types that may be given. For the present specification, the task on grammar has been developed on the completion format.

3.3.2 Test Two

The second text will be a short tone, about 150 words, and will be selected from a reader consisting of reading materials comparable to the prescribed texts. The task on this text will be a cloze test (Task E). Every fifth word of the text will be deleted to infer the candidate's ability in comprehension, vocabulary, grammar, and in the understanding of contextual and linguistic devices.

3.3.3 Text Three

The third text should be an unseen one, taken from a local newspaper. This newspaper article should be about 300 words long. The text selected from local newspapers should not be over-simplified. Simplification, if needed should be restricted to names, places, etc to maintain anonymity. There should be three tasks on this text all designed to test faster reading.

Task F will be to read the text and find information in the text quickly. There will be ten items on the fill-in format. Task G will be to establish co-references between the elements within a sentence or paragraph. There will be five items on the question-and-answer format. Task H will be to write the gist of the text in about 50 words.

3.3.4 Text Four

The fourth text should be selected either from a bill, a menu, a brochure or a chart from a report. The length of this should text should be about 150 words. There should be two types of tasks on this text both requiring interpretation of the text.

Task I will be question and answer. There will be five items requiring an interpretation of the text. Task J will consist of five items on the true-or-false format.

3.3.5 The Time Allotment

The total weightage given to Reading is 45 per cent. So 45 per cent of the total time available, i.e., three hours, is 81 minutes. The suggested time allotment for the texts and the tasks is as follows

Text One	**31 minutes**
Reading	6 minutes
Task A	10 mins
Task B	5 mins
Task C	5 mins
Task D	5 mins
Text Two	**15 minutes**
Task E. Reading and filling the blanks (cloze)	
Text Three	**20 minutes**
Reading	5 (2+2+1) minutes
Task F	5 mins
Task G	5 mins
Task H	5 mins
Text Four	**15 minutes**
Reading	5 minutes
Task I	5 mins
Task J	5 mins
Total	81 minutes

Figure: Time allotted to reading comprehension tasks

3.4 Sample Tasks

See appendix A1.3

3.5 Response Characteristics

In Task A, the candidates are required to find out the exact word or phrase in the text (text 1) for each of the meanings given, and to write the chosen word or phrase against their number in the answer book. In Task B, they will rearrange the statements which are given in a mixed order. In this task, they are required to write the alphabet of the chosen statement against the number 1, 2, 3 and so on. They are not required to copy the statements. However, they will not be penalised if they do so. Task C is a comprehension question. The candidates are expected to answer the questions on the basis of their understanding of the text. They are required to answer the questions in their own words. Task D is a grammar exercise, which requires the candidates to complete the given incomplete sentences using 'because'.

Task E (Text 2) is a cloze exercise. The candidates are expected to complete the text by filling in with the exact word or words that are contextually appropriate.

Task F (Text 3) is a rapid reading or scanning exercise in which the candidates are required to read the text quickly and fill in the blanks with suitable information from the text. Task G is to establish co-references between elements of the text. The candidates should read the text quickly and find the word(s) the given term refers to. They are required to write only answers in their answer sheet. Task H is a comprehension task. The candidates will write a 50-word summary of the text in their own language.

Text 4 is an information leaflet, which mostly consists of non-text materials or information. The candidates will study the leaflet and respond to the two types of tasks on it. Task I is a comprehension question. The candidates are required to write their answers in their own language. Task J is a true-or-false task. The candidates will write T for true statements and F for false statements.

3.6 Scoring Scheme

Some tasks on reading are bottom-up tasks requiring quick reading and responding to items that are based on language or surface level information. Such tasks have been allotted ½ to 1 point each depending on the level of comprehension and effort required. Items in tasks A, E, and F, thus, have been allotted ½ point each and items in tasks B and J have been given 1 point each.

Tasks types C, D, F, G and H are top-down tasks requiring higher level of reading and comprehension skills and so the items in these tasks have been allotted 2 points each. The total point, thus, allotted to all items in tasks from A to J is 85. The aggregates of the obtained marks should be converted to 45 per cent.

A summary of Tasks and the mark allotment are given below:

Text 1: Seen			
Tasks	No. of test items	Time allotted (mins)	Mark allotted
Reading		6	
Task A Vocabulary	10	5	5
Task B Reordering	7	5	7
Task C Q/A	5	10	10
Ask D Grammar	5	5	5
Text 2: Unseen			
Task E Reading Cloze	20	15	20
Text 3: Unseen			
Reading		5	
Task F Fill-in	10	5	5
Task G Co-references	5	5	10
Task H Gist	50 words	5	10
Text 4: Unseen			
Reading		5	
Task I Q/A	5	5	5
Task J True-False	8	5	8
Total	70+gist	81 mins	85

Figure: Summary of tasks, time allotment and mark distribution on reading comprehension tasks

As it is mentioned earlier, the aggregate of the score obtained will be converted to 45 per cent.

All correct answers in A, B, E, F and I will be awarded full mark. For the Task D, i.e. the cloze, the answer should be scored on the basis of what is known as the 'Exact Restoration Method' (Oller 1979), 'Contextually Appropriate method' (Oller 1979), and 'Acceptable Method' (Brown 1980). Each blank is worth one point and will be scored as follows:

A. One point will be awarded for   restoration of the exact word, or for contextually as well as grammatically appropriate word;

B. Half a point will be given for contextually appropriate but grammatically unacceptable responses.

For tasks requiring writing, an adapted analytical scale will be used (See Appendix 3).

3.7 Supplement Information

The raters should be familiar with the analytical scale and its uses, scoring methods on a cloze, and other criteria to be used for marking written answers.

The difficulty level of unseen texts to be selected should fall within the prescribed vocabulary and structure range. They should also meet the requirement of length.

Each line of the text 1 and 3 will be numbered.

3.7.1 Answers

For Answers of Reading Sample Tasks, see Appendix A2.3.

4 WRITING   Full Mark 35

4.1 Target Language Use

4.1.1 Sub-skills

♦ Expression of information/opinion/argument/emotions
♦ Writing letters/postcards/memos/ads/announcements/brochure/information leaflets)
♦ Describing a picture or a picture sequence, a person, a place, or an object or event or a process
♦ Narrating experiences
♦ Answering questions
♦ Writing on a given topic or a situation
♦ Writing reports and newspaper articles
♦ Writing summary of a text
♦ Writing a running text into a dialogue form or vice -versa
♦ Interpreting non-text materials
♦ Paraphrasing writing on a character of a story

4.1.2 Aspects

♦ Content
♦ Form (organisation, cohesion)
♦ Vocabulary
♦ Grammar
♦ Mechanics
♦ Fluency (registers/discourse/styles/consistency in the use of language)

4.1.3 Domains

♦ Official/semi-official contexts
♦ Print/audio/video media
♦ Immediate surroundings
♦ Academy and academic matters
♦ Health and related matters including food/diet
♦ Travel/tourism/ vacation/holidays
♦ Hobbies/interest/entertainment/TV/movie/leisure
♦ Work/employment
♦ Games/sports/exercises/
♦ Everyday life

4.2 General Description

If the ability to read and understand and handle information in the target language is the primary purpose of an EFL programme, the ability to express one's knowledge, ideas and opinion in a written form is a major achievement. The development of writing skills needs constant practice during the instruction phase and a careful consideration of all major aspects involved in it during the testing phase.

The classroom instruction must facilitate the process the students go through to produce a piece of product. The testing of this skill, on the other hand, must create conditions to enable candidates to process their thinking before they produce it. In other words, candidates must be allowed to think and plan before they actually start writing it. Evidences of processing and planning should accompany the product as a proof that what the candidates generate is the production of their thinking and planning, and not the reproduction the memorised materials.

The candidates will demonstrate their ability to

(a) plan, organise and write descriptions, experiences, opinion, argument, etc. in clear and grammatically correct English using a variety of registers and styles as demanded by the context of situation.
(b) write official and personal including letter to the editor, a complaint letter, letter to the principal, etc.
(c) transfer information from non-text writing text form and produce a variety of text types such as postcards, questionnaires, phone messages, CV, bio-data, job application, notices, and advertisements,
(d) write creatively on personal/academic matters using context, style of registers and mechanics appropriately.

4.2.1 Materials

Topics or situations for writing will be derived from the materials/topics/situations contained in the textbooks. Such topics or situations should present novelty on which candidates should write on the basis of what they have learned during the course. The topics/situations should not be reproduced as they are given in the textbooks. In general, the topics or situations for writing should be –

♦ a replication of the situations they practise during the course,

- an extension of what they learn,
- a portrayals of the skills the develop during the practice phase.

The topics or tasks, which should come from the domains specified in the TLU section above should reflect the sub-skills and should -

- be meaningful to the candidates
- be relevant to their need,
- be motivating,
- be authentic
- be specific, and
- require strategies common in academic writing.

4.3 Prompts

Three tasks of writing will be given to the candidates, each one requiring different sets of writing skills.

Task One is to write a well-organised essay of 150 words after reading a text. The writing will demonstrate the candidates' ability to link reading, personal experience and writing. This task is for 25 minutes. The candidates will be required to spend about 5 minutes on thinking and planning, and 20 minutes on writing.

Task Two is to write a letter as a response to a given situation. The situation may be presented in the form of a letter or a running text. This task requires the candidates to write a letter of complaints to a book retailer concerning an order of books, which has not been delivered in the specified date. The situation will be given in the form of a text. The candidates will write an official/business letter complaining about the delay in the delivery.

The total time allotted to this task is 20 minutes of which about 5 minutes should be spent on thinking and planning and 15 minutes on writing.

Task Three is to prepare an information leaflet or brochure about one's village, town or city. The candidates will be required to write on the suggested points such as -

- location/geography,

- population/demography/culture
- places of historical importance
- places of interest for tourism
- current development projects
- travelling facilities
- contact places.

This task is for 18 minutes. The candidates should spend about 5 minutes for planning and rest of the time on writing. Information in the brochure may be presented in the non-text form.

4.3.1 Time allotment

This skill has been allotted 35 per cent of the total weightage. Thus, 35 per cent of 3 hours is 63 minutes. The suggested time allotment is as follows:

	Time (mins.)	Expected length of writing
Task One	**25**	**1½ Pages**
Planning	5	
Writing	20	
Task Two	**20**	**1 Page**
Planning	5	
Writing	15	
Task Three	**18**	**1 Page**
Planning	5	
Writing	13	
Total	**63 minutes**	

Figure: Writing tasks, time allotment and expected length

4.4 Sample Task

See Appendix A1.4

4.5 Response Characteristics

In Task One, the candidates are free to use vocabulary and structure of their choice. They can also plan and organise the structure of the essay in their own way. The candidates are expected to write about 150 words in three to five paragraphs.

Task Two is letter writing. They will write a letter of complaint to the director of the said company. This is a formal letter and should include the following -

- name, title and full address of the addressee
- date of writing and address of the writer
- subject of writing
- address/ salutation
- body of the letter
- closing
- signature/ name of the writer

In Task Three, the candidates are not expected to provide information on the given topics in full sentences, rather, they can write in fragments or phrases.

They are expected to plan or design as to what information they are to put in what part of the leaflet. The brochure should contain all relevant information.

Planning is an important phase of writing. Whatever they write during the planning phases of tasks One, Two and Three should be done on a separate sheet, and must be attached with the answer-booklet.

4.6 Scoring Scheme

The criteria for scoring on writing are -

- content
- organisation
- cohesion
- grammatical accuracy
- vocabulary
- fluency (style, register, discourse, consistency, flow in the use of language)
- mechanics

Points will be awarded '0' to 3, according to the level exhibited in each criterion. The definition of the number/categories is as follows:

0 = Unattempted, unintelligible, poor, failure
1 = Satisfactory, pass
2 = Good, Second class (division)
3 = Very good, First Class (division)
4 = excellent, Distinction

For the analytical scale to be used for scoring, its bench marking on the categories and their descriptors, please see Appendix 3. The aggregates of the marks secured under each task should be converted to the allotted percentage. The writing tasks and the percentage of marks allotted are as follows:

Task	Task Type	Percentage of marks allotted
Task One	Essay	15
Task Two	Letter writing	10
Task Three	Brochure Preparations	10
	Total	35

Figure: Writing test tasks and marks allotment

4.7 Supplement Information

The raters should have training to familiarise them with the analytical scale, categories of the scale and their bench marking.

The candidates' scripts of the planning phase should be reviewed during the scoring/marking stage to match it with the writing.

For answers of the Writing Sample Tasks, see Appendix A2.4

Appendix A 1: Sample Tasks

A1.1 Sample Tasks: Listening

Text 1:   **Task Type A**

You will need some oil, 3 large potatoes, and 4 eggs. First peel the potatoes and chop them into small cubes. Next fry the potatoes in a frying pan until light brown. Set aside on a plate for a few minutes. Break the eggs into a bowl and beat them with a fork for one or two minutes. Mix the eggs and the potatoes together. Heat a teaspoon of oil in a frying pan before pouring the mixture in it. Add some salt. With the help of a plate, turn the omelette over and cook the other side until it is golden brown.

(103 words) (adapted from Learning to Cook. WWW.rhlschool.com. Accessed on 14-09-2001.)

Instruction: *Listen to the tape and complete the table given in your answer sheet by supplying the missing parts:*

Table:

A. The ingredients needed for the recipe are: 1. *some* oil 2. ...... large potatoes 3. ...... eggs
B. The four utensils mentioned in the text are: *1.   A frying pan* 2.   ...... 3.   ...... 4.   ......
C. The steps to cook a potato omelette are: 1.   ...... 2.   Fry the potatoes in a frying pan. 3.   ...... 4.   Mix the eggs and potatoes together. 5.   ......

Text 2:   **Task Type B**

Receptionist:	Good evening, Ma'am! How can I help you?
Visitor:	I'd like to book a room, please. Do you have a single with attached bath?
Receptionist:	We certainly do, Ma'am.
Visitor:	And do you have TV in the room?
Receptionist:	Yes, all our rooms have a TV.
Visitor:	Good! And a telephone?
Receptionist:	Yes, Ma'am
Visitor:	How much do I have to pay for a night?
Receptionist:	300.00 rupees for a night or Rs. 525 for two nights. How long would you stay?
Visitor:	Two nights starting tonight. I'll pay 525 rupees on arrival, if that's OK with you.
Receptionist:	That's certainly fine, Ma'am. Your room number is 105 on the second floor.
Visitor:	Thank you.
Receptionist:	You're welcome.

Instruction – *Tick (√) the things mentioned in the tape.*

A double room	
A single room	
Attached bath	
Common bath	
TV	
Fax	
Phone	
Computer	
525 rupees	
600 rupees	
Room no. 105	
Room No. 501	

Text 3:   **Task Type C**

Here's a little advice for writing fiction. First, make your characters believable. Make sure that they behave and talk as individuals. In real life everyone is unique. If all your characters speak in the same way, you will loose your readers from the start.

Once your readers believe in your characters, you must get them to care. Each reader must be able to identify with at least one character, to almost become that character in his or her mind. You can do this by developing characters with genuine human traits, both good and bad. The individuals who populate your story should have human strengths and weaknesses.

Now it's time to weave your tale, to create a plot. Your readers are part of the story now; they are involved.

One last thing. Your story must touch the readers' emotions. If you can make them laugh and cry along with your characters, you will be successful writer.

(155 words) (adapted from A Little Advice. WWW.rhlschool.com. Accessed on 14-09-2001.)

Instruction – *Match the words in Column A with their meanings in Column B. Column B has an additional item which does not have its match in Column A. Write in your answer sheet the alphabet of the item in Column B against their matching number in Column A.*

Column A
1. believable
2. unique
3. identify with
4. genuine
5. populate
6. characters
7. be involved
8. emotions

Column B
a. not fake
b. people in the story
c. feelings
d. be a part of
e. to put oneself in another's place
f. inhabit
g. appear to be true
h. different from all others
i. permit

Text 4:   Task Type D

KATHMANDU, Sept 28:Classical guitarist, Kishor Gurung, is to leave for the US on Monday, to participate and represent Nepal at the "Music in Mountain Culture" festival, to be held in Denver, Colorado.

The two-day music festival is being organised by the Society for Ethno-Musicology. Before his US programmes, Gurung is to represent Nepal in the seven-day "International Guitar Festival" in Thailand and participate in the National Guitar Competition as a judge.

"I am proud to represent Nepali music throughout the world with the help of these programmes," said Gurung to The Kathmandu Post.

Before the programme, Gurung is scheduled to give a lecture on Nepali music and present guitar music at the St. Lawrence University.

Gurung, a graduate in Music from the US, has already presented his solo and group guitar music in various cities of the US and Europe.

(140 words)(adapted from The Kathmandu Post, 29-09-2002)

Instruction - *Listen to the tape and tick (√) the best answer:*

1. Mr. Kishore Gurung is a ....

a. singer.   b. musician.   c. guitarist.   d. drummer

2. Mr. Gurung is Leaving Nepal on ....

a. Sunday   b. Monday.   c. Tuesday.   d. Friday

3. Mr. Gurung will participate in the national Guitar Competition as ....
a. a competitor.   b. a judge.   c. an organiser.   d. a visitor

4. Mr. Gururng was speaking to the Post reporter in ....

a. Denver   b. Thailand   d. Europe   d. Kathmandu

5. The International Guitar Festival is organised for.......... days.

a. 28.   b.2.   c. 7.   d. 1

6. 'Music in Mountain Culture' is being held in ....

a. Nepal.   b. USA.   c. Thailand.   d. Europe

7. Mr. Gurung will give a lecture at St. Lawrence University on .... Music.

a. classical   .   b. Pop.   c. .Nepali   .   d. American

8. Mr. Gurung has received his education in ....

a. Nepal   b. Thailand.   c. Europe.   d. USA

Text 5:   Task Type E

I, at 55, look back at my childhood and compare things to the way life is for today's kids. Some things have certainly changed. One area of change is television. Some changes have been improvements. Others have been setbacks.

When I started school, most people didn't have a television. TV was just beginning to get popular. My father decided to buy a 16-inch black and white TV set. I still remember watching Fulbari in which there was a different story everyday. That was exciting.

The variety and quantity of programming has increased greatly. There are dozens of channels and more shows than one can possibly watch. There are fine entertainment and educational shows. There is also a lot of garbage, stuff that most parents don't want their kids exposed to. Overall we have more choices, and that is good.

I wonder what TV will be like when today's kids are my age.

(152 words) (adapted from How Television Has Changed. WWW.rhlschool.com. Accessed on 14-09-2001.)

Instruction - *Listen to the tape and say whether the statements below are true or false. If a statement is true, write 'T' against its number in the answer sheet, if it is false, write 'F'.*

1. The writer of this text is a young adult.
2. Some changes in the TV have been for the better.
3. The writer had a coloured TV when he was young.
4. The writer didn't like Fulbari.
5. Today's TV shows a lot of garbage.
6. His father bought an 18" TV.
7. The writer is happy about the variety today's TV offers.
8. Some stuff shown on TV is not suitable for children.

Text 6:   **Task Type F**

The new painting was hanging in the National Art Gallery, Kathmandu. Throughout the day, visitors stood before the new exhibit and speculated as to what it represented.
"I think it's a terrifying monster," said Maya. "There's no name on it, but I'm sure it was drawn by a famous artist."

"I would say that it's a picture of a park as seen from an aeroplane," mused Janak. "I see many trees and roads."

Ms. Karki, the local art critic for the Himal Weekly made her pronouncement. "This is clearly a masterpiece created by an artist who wishes to be anonymous at this stage of his or her career. The brilliant strokes of colour symbolise a person who is looking back at an exciting lifetime of accomplishments."

At the moment, Mr Lama, the gallery curator entered the room. "Great, isn't it?" he asked with a big smile on his face. "My three year old did that this morning. He says it's a maze. I think it's amazing!"

(165 words) (adapted from The Painting. WWW.rhlschool.com. Accessed on 14-09-2001.)

Instruction – *Listen to the tape and answer the following questions. Write the answers against their numbers in the answer sheet.*

1. Where was the new art exhibit being held?
2. What did Maya think the paining was of?
3. What did Janak see in the painting?
4. Who does Ms. Karki work for?
5. What is the name of the gallery curator?
6. Who, according to the curator, did the painting?
7. What, according to the painter, is the painting of?
8. What does the curator think of the painting?

Text 7:   Task Type G

The country has a new Prime Minister. The people are hopeful that the Prime Minister will succeed in dealing with some of the major problems that our nation is facing. Some feel that the task is extremely difficult as the nation is facing acute shortage of resources and the people are divided as to how the political situation can be resolved. Others are confident that our new leader is a man who will bring people together to get things done.

The Prime Minister's top priority is to improve education in our country. He is determined to see that no child ever fails. He believes that local school districts should decide how to meet national standards and insists that parents should have greater choices about which schools their kids attend. He will explore areas of agreement between the opposing parties so that important education reforms will become law. (147 Words)

Instruction – *Listen to the tape and fill in the blanks. Write the answers against the number of the question in your answer sheet:*

The country has a new Prime Minister. The people are .... (1) that the Prime Minister will succeed .... (2) dealing with some of the major .... (3) that our nation is facing. Some .... (4) that the task is extremely difficult .... (5) the nation is facing acute shortage .... (6) resources and the people are divided .... (7)to how the political situation can .... (8) resolved. Others are confident that our .... (9) leader is a man who will .... (10) people together to get things done.

.... (11)Prime Minister's top priority is to .... (12) education in our country. He is .... (13) to see that no child ever .... (14). He believes that local school districts .... (15) decide how to meet national standards .... (16) insists that parents should have greater choices about which schools their kids attend. He will explore areas of agreement between the opposing parties so that important education reforms will become law.

Text 8:   **Task Type H**

Instruction: *Listen to the tape and complete the summary by filling in the appropriate words or phrases. You may fill in the blanks by supplying your own words.*

Madan always had a great idea for celebrating Holi. He didn't even have any ancestors from the Terai, but that didn't matter. This festival invites everyone to be colourful for one day each year.

Last year, Madan had got his friends and classmates together to form a Holi celebration committee. They went up and down the main street playing Holi and had a really good time. Madan decided that this year he would have a small party and invited his closest friends. They prepared good food and played Holi by putting coloured water and coloured dust on each other.

Madan wasn't sure what sort of music they should play during the party. So, he asked Binod who brought some music tapes appropriate for the occasion. After meal, they played music and danced together. (133 words)

Summary

Holi is a (1) ...... of colours. Madan and closest friends celebrated this year's (2) ...... by staying in, (3) ...... food, (4) ...... music, (5) ......together, and putting (6) ...... on each other.

Task Type One

In English, speak for about five minutes on what you would do if you found a million rupees.

OR

In English, speak for about five minutes on whether the life of the present generation is better or worse than the life of the previous generation.

OR

The chart below shows how Mr. Pradhan spends his income. Study the chart and speak in English for about five minutes on how Mr. Pradhan spends his money, whether or not he spends it wisely, how can he spend more wisely, if he should spend more on some items and less on others, whether he should spend less on some items to save more and why.
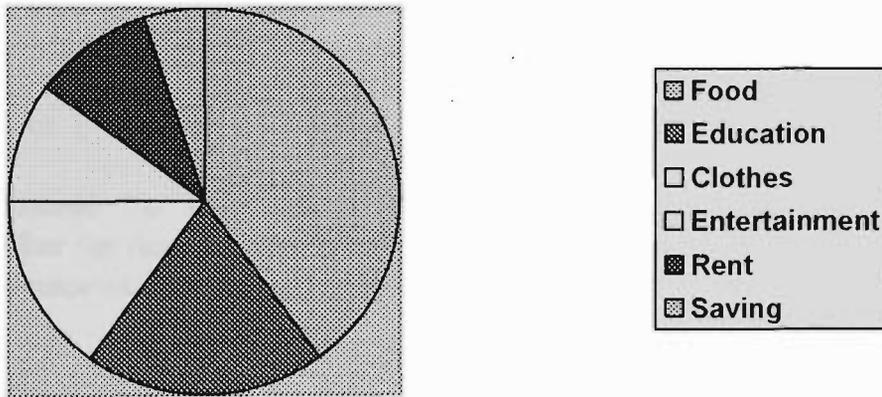
Task Type B

[The format of the conversation may take several forms, and it may differ from one situation to another and from one topic to another. The following conversation may be taken as an example.]

Interlocutor (I): [Asks the candidate to take a seat.] Hi! My name is Kishore Gurung. This is (if there is a second interlocutor/assessor) my colleague Mr. Nurul Ansari. We will conduct the session together.

Candidate (C): Good morning/afternoon. Thanks

I: Would you like to tell us about yourself?

C: Yes Sir. My name is Indu Sharma. I am from Junta Secondary School, Chainpur. I live with my parents in a small village in Siraha district called Dhanagadhi. I have two brothers. Both of them go to the same school. My parents are farmers.

I: All right, Indu. Most Nepalese celebrate and enjoy festivals. I, for example, celebrate Dasai and Tihar. Mr. Ansari celebrates Ida and Ramadan. Yet others celebrate Thanks Giving and Christmas. Do you celebrate any festivals at all?

C: Yes sir.

I: What festivals do you and your family celebrate?

C: I come from the Terai. So, I and my family celebrate Holi, Chhatha and Deepawali. We celebrate Dushehara too.

I: Good. Now, suppose that my colleagues and I don't know anything about the festival of Holi. Can you describe the Holi festival for us, what it is, its significance, the way you celebrate it, and so on. You have about five minutes to do so.

C: Yea sir. (And speaks for about five minutes)

[The interlocutor/assessor can interrupt the candidate during his or her speech in order to ask questions. After the time allotted is over or if the candidate is finished within five minutes, the interlocutor/assessor thanks him/her and takes leave]

1. *Read the text carefully and do the question given below:* (6 mins.)

My parents did not believe in boys wearing trousers, so I wore shorts to school. I wanted to wear trousers but every time I opened my mouth, my parents told me to shut up. One day, the class teacher told us that we were going to have a group photograph taken and he asked us to wear trousers for the occasion. I was in a dilemma, as I did not possess a pair of trousers.

After school I asked some of my friends to lend me a pair of trousers but no one had a pair which would fit me. I even tried on a pair of my father's trousers. They were too short for me and much too large at the waist. I gave up in despair. The next day arrived. I thought of not going to school but my parents told me to get ready. They told me to stop grumbling because there was nothing wrong in wearing shorts. So, I turned up as usual at school. The class teacher was aghast when he saw me. He and some other students wanted me to be excluded from the photograph but the head teacher intervened. She asked me to join the group. The class teacher told me to sit in the front with two students on either side of me.

The photograph was duly taken A few days later, a copy of the picture appeared on the notice-board. Crowds of students gathered to see it. It was a well-taken photograph. All the students were there in their well-ironed, spotlessly white shirts and trousers - except me in the front row, with my knobbly knees conspicuously exposed. Needless to say, the picture gave rise to a lot of laughter and unkind remarks. You can imagine how I felt. (298 words)

Task A. *Find the words in the text which means the similar to the words or phrases given below. Write the answers in the space provided:* (5 mins.)

a. a time when an event takes place ....

b. a situation of uncertainty ....

c. a sense of hopelessness ....

d. complaining ....

e. arrived ....

f. kept out ....

g. interfered to prevent something from happening ....

h. noticeably ....

i. left uncovered ....

j. comments ....

Task B. *The following sentences are not in proper order. Rearrange then in a sensible order. Put the letters next to the numbers in the space provided following the example of number one (5 mins.)*

a. The writer brought home a copy of the photograph.

b. The class teacher told the writer to wear trousers for a group photo.
c. The group photo was taken and it created a great deal of laughter when it appeared on the notice board.
d. The class teacher refused to include him in the group.
e. His father told him that it was all right to wear shorts.
f. He tried to find a pair of trousers that fits him without a success.
g. He felt terribly embarrassed.
h. He was included in the group because the head teacher intervened.

Example -

1. .... *h* .....
2. ....
3. ....
4. ....
5. ....
6. ....
7. ....
8. ....

Task C. *Write the answer of the following questions in your own words. The answers should be brief and to the point. You need not copy the questions. (10 mins.)*

i.   What was the dilemma the writer was in?

....
....
....

ii.   Why was the class teacher aghast?

....
....
....

iii.   How would a seat in the back row have helped the writer?

....
....
....

iv.   Why were students amused when they looked at the photograph?

....
....
....

v.   Write in brief how the writer felt after the photograph appeared on the notice board.

....
....

....
....
....
....
....

Task D. Complete the following sentences. Use the example as a guide (5 mins).

Example –

The writer's parents told him to stop grumbling because ....
.. .

Answer -
The writer's parents told him to stop grumbling because they thought there was nothing
wrong in wearing shorts.

1. The writer always wore shorts because ....
   ....
2. One day, the class teacher asked them to wear trousers because
   ....
3. He couldn't borrow a pair of trousers from his friends because
   ....
4. .... because he did not
   wear trousers.
5. .... because the head
   teacher intervened.

2. Task *E. Read the following story. Some of the words have been left out. Read the whole story quickly before you try to fill in the missing words. Then fill in the blanks with an appropriate word. Try every blank. Blank number one has been done for you. (15 mins.)*

Some people say that you should not run away if you meet a tiger. *They* (1) say that you must .... (2) still and face the .... (3). They say that it .... (4) not attack a human .... (5) unless the person panics .... (6) runs away. When these .... (7) hear of a person .... (8) killed by a tiger, ... (9) say, "He should not .... (10) run away. He ought .... (11) have stood still. If .... (12) had, he might have .... (13) alive today.

In the .... (14) days in Nepal, they .... (15) to call this 'talking .... (16) the tiger'. 'When you .... (17) a tiger,' they used .... (18) say, 'you should speak .... (19) it. You should tell .... (20) that there is plenty .... (21) food in the jungle and that it need not have to eat you.

Unfortunately, most people who were attacked by a tiger did not have courage to speak to it.

3. Read the following passage and answer the questions given below (5 mins):

The temple of Tripurasundari, one of the most famous goddess temples of Far Western Region, is situated in the border district of Baitadi. Baitadi is 83 km from nearby Dadeldhura, 200 km from Dhangadi and about 800 km from the capital city, Kathmandu. This famous temple is 10 minutes drive and 40 minutes walk from the district headquarters.

Fairs (Jaat in local language) occur twice a year i.e. in Kartik and Asar. Visitors from neighbouring districts of Dadeldhura, Darchula, Doti, Kailali and Kanchanpur and even from many bordering parts of India throng here to observe the fairs and to worship the goddess. The fair is celebrated for two days. The first day fair is called Nani Jaat (small fair) and the second Thuli Jaat (big fair). In the remote villages where women hardly get time to visit their relatives due to domestic works, these fairs provide them an opportunity to share their feelings with their relatives, parents and friends. Villagers look dazzling and glowing in neat and clean dresses, ornaments and a beautiful tika on their forehead.

There are seven temples of seven Bhagwatis in different parts of Bajhang, Doti and Baitadi. Of these, five are situated in Baitadi. According to the ancient mythology, these seven Bhagwatis are sisters. In these different temples, fairs occur at different times.

Accommodation is cheap and easily available. Taxi and day/night bus services are available from most neighbouring cities and towns to Tripurasundari and Nepal India border of Jhulaghat, where people can enjoy horse riding, trekking and swimming in the Mahakali River. Air services from Kathmandu to Nepalgunj or Dhangadi are also available. One can enjoy popular sight seeing in the Royal Bardia National Park when travelling by a car.

Task F. *Complete the following sentences by filling in blanks with words or phrases from the text (5 mins).*

a. Baitadi is about .... km away from Kathmandu.
b. One can reach the temple of Tripurasundari in .... minutes when driving from the district headquarters.
c. Fairs take place in the month of Kartik and ....
d. Naani Jaat refers to the .... fair
e. Of the two days of celebrations, Thuli Jaat takes place on the .... day.
f. Of seven temples of Bhagawatis, ... are in Baitadi.
g. One can go swimming in ....
h. .... services are available from Kathmandu to nearby towns of Nepalgunj and Dhanagadhi.
i. Royal Bardia National Park is popular for ....
j. Jaat is a local term for ....

Task G. *Answer the following questions in one word or phrase. Write your answer in the space provided against their number (5 mins).*

a. Name the district the term 'district' (para 1, line 4) refers to. ....

b. What does the word 'here' (para 2, line 3) refer to? ....

c. What does the word 'them' (para 2, line 6) refer to? ....

d. What does the term 'these' (para 3, line 2) refer to? ....

e. What does the word 'one' (para 4, line 5) refer to? ....

Task H. *In the space below, write a gist of the text 4 above in about 50 words (5 mins)*

....

...

....

....

....

....

....

4. Read the following information leaflet carefully and then answer the questions given below (5 mins)

Annapurna Bus Service (A Private Company) Estd 1999	Mahendranagar		Nepalgunj		Bharatpur		
	Arrival	Departure	Arrival	Departure	Arrival	Departure	
Bus Time Table	AM					PM	Passengers, please note that buses depart every two hours, AM to PM from Mahendranagar and Nepalgunj, and PM to AM from Bharatpur
Route – 25	5.30	6.00	9.30	10.00	2.00	3.00	
	7.30	8.00	11.30	12.00	4.00	5.00	This is an express service. So passengers will not be picked up from any other places except the designated ones.
Mahendra Nagar – Bharatpur *Via Nepalgunj*	9.30	10.00	1.30PM	2.00	6.00	7.00	
	11.30	12.00	3.30	4.00	8.00	9.00	
Effective January 2003	PM						Passengers are encouraged to do the reservations prior to their travelling date. Tickets at windows are subject to availability.
For more Information	1.30	2.00	5.30	6.00	10.00	11.00	
	3.30	4.00	7.30	8.00	12.00	1.00	The buses run as scheduled everyday of the week including Saturdays, Sundays and Public Holidays.
Call – 593 333	5.30	6.00	9.30	10.00	2.00	AM	
Visit our website www.annapurnabus.com	7.30	8.00	11.30	12.00	4.00	3.00	
Head Office **15 Shaheed Chowk, Mahendranagar 13** Kunchanpur						5.00	Tickets may be collected from our site offices in Mahendranagar, Nepalgunj and Bharatpur and Head Office in Mahendranagar.

Task I. In the space provided against each question, write the answer of the following questions in your own words. Your answers should be brief (5 mins).

    a. When was the Annapurna Bus Service Company established?................

    b. How many main terminal stops are there in Route 25?   ....

    c. How many buses depart from Nepalgunj everyday?   ....

    d. In the afternoon, what is the fist bus one can take to go to   ....
        Nepalgunj from Mahendranagar?   ....

    e. Which of the three bus terminals is busy all night long?   ....

Task J. Read the text once again and say whether the statements below are true or false. If a statement is true, write 'T' against the number of the statement, if it is false, write 'F'. (5 mins)

    a. The timetable for route 25 is effective in 2003.   .....

    b. Passengers can take buses after 6 PM at the Mahendranagar terminal.   .....

    c. Buses run in every three hours from each of the three terminals.   .....

    d. No passengers will be taken from any other places than the three terminals. ......

    e. Passengers can book tickets in advance.   .....

    f. The main office of this bus company is in Bharatpur.   .....

    g. Passengers can collect their tickets from the site offices.   .....

    h. Tickets cannot be collected from the company's head office.   .....

Task One (25 Minutes) (Full Mark - 15)

Describe the character of the writer of the reading text one (see question 1). Write in about 150 words what he was like and how he felt when he saw crowds of students looking at the photographs and laughing.

Spend about five minutes on planning, i.e. thinking, putting down ideas and organising them. Spend the rest of the time on writing the essay.

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

(Please use back of the sheet if needed)

Task Two (20 minutes) (Full Mark - 10)

Imagine that you are a regular passenger of Annapurna Bus Services. You had made your reservation two weeks prior to your travel date with a provision that you would collect your ticket from its Nepalgunj office on the date of your travel. However, when you contacted the office half an hour before your departure, the counter informed you that there was no reservation made in your name. Obviously, you were very angry about the carelessness of the bus company.

Write a letter of complaint in about 150 words to the Director of the bus company in its head office at Mahendranagar asking him to take an immediate action regarding the matter. In your letter, say what might happen if the company allows continuing such a problem.

Include in your letter, who you are writing to and why, what about, the date and place of reservation, and the name of the booking clerk. You must write the letter in the standard, official format and formal language.

Spend about 5 minutes in planning and rest of the time on writing the letter.

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

(Please use back of the sheet if needed)

Task Three (18 minutes) (Full Mark - 10)

Suppose you are a tourism development officer of your village/town/city. In about 150 words, prepare a brochure/ information leaflet of your village/town/city for a tourist who may not have heard about it before. Following the brochure of Annapurna Bus Services (see question 4) as a model, present information to attract tourists. Include in your brochure information about - location (geography/climate), population (demography/culture), places of historical importance, places of interest, current development projects, travelling and accommodation facilities, contact places, etc. Spend about 5 minutes to plan and rest of the time on writing the brochure. You need not write the information in complete sentences.

(Please use back of the sheet if needed)

Appendix A2: Answers of the Sample Tasks

Task Type A

A.
1. Some (already provided)
2. *3*
3. *4*

B.
1. A frying pan
2. *A plate*
3. *A bowl*
4. *A fork*

C.
1. *Peel and chop the potatoes into cubes;*
2. Fry the potatoes in a frying pan;
3. *Break the eggs into a bowl and beat them with a fork;*
4. Mix the eggs and potatoes together;
5. Heat the oil in a frying pan, pour the mixture in it and cook both sides until brown.

The tasks are open-ended. The answers, therefore, are going to be different from different candidates.

Answer of Question A. Find the words in the text which means the same as the following:

a. occasion
b. dilemma
c. despair
d. grumbling
e. turned up
f. excluded
g. intervened
h. conspicuously
i. exposed
j. remarks

Answer of Question B. Rearrange the following statements into a sensible order:

a. The class teacher told the writer to wear trousers for a group photo.
b. His father told him that it was all right to wear shorts.
c. He tried to find a pair of trousers that fits him without a success.
d. The class teacher refused to include him in the group.
e. He was included in the group because the head teacher intervened.
f. The group photo was taken and it created a great deal of laughter when it appeared on the notice board.
g. The writer brought home a copy of the photograph.
h. His father brought him a pair of trousers.

A2.4 Answer of Sample Tasks: Writing

The tasks are open-ended. So, the answers are going to be different from different candidates.

Appendix B
The Pilot Test

School of Education, Victoria University Melbourne, Australia

Question/Answer Booklet

Identification Code ....

....

Identification Code ....

Name of the student: ....
Name of the school: ....
Section ....   **Roll no.** ....

Note to the student –

♦ *Your participation in the research is entirely voluntary and you can withdraw from the research at any stage.*

♦ *Your withdrawal from the research will not jeopardise you in any way.*

♦ *Your performance in the test will not jeopardise your assessment position in the forthcoming examinations.*

♦ *If you have any difficulty or you fell that you have not been treated well, you may talk to your teacher.*

Task A: Listening Text 1 (For the testers' reference only, not to be supplied to the candidates)

Laxmi, a friend of Arun finds him in a bad mood. She wants to know why. Listen to the text and complete the table given:

Laxmi:	Hello Arun, you look very up set. What's the matter?
Arun:	Hello Laxmi. I haven't got any trousers to wear for the school photo tomorrow.
Laxmi:	Did you ask anyone whether you could borrow some?
Arun:	Yes, I did. I spoke to Raju and your brother Hari too. Unfortunately, they don't have any extra pair of trousers. I even tried on my father's trousers.
Laxmi:	They didn't fit, did they?
Arun:	No, they didn't. I don't know what to do now.
Laxmi:	Just wear your shorts for the photo.
Arun:	Do you think that'd be OK? Won't the teacher be angry?
Laxmi:	I'm sure it'll be fine. I don't think Mr. Pradhan will mind. I'll see you at school tomorrow.
Arun:	Yes. See you tomorrow. Bye.
Laxni:	Bye.

Task A: *Listen to the tape and complete the table given. Write your answers in the 'Answers' Column in table 1*

Table 1

Questions	Answers
When Laxmi saw Arun, how did he look?	
What was Arun looking for?	
Whose brother is Hari?	
Why couldn't Arun wear his father's trousers?	
What is their teacher's name?	
Why did Arun have to wear trousers?	
What would Arun be wearing the next day?	
How many persons' names are mentioned in the text?	

Listening Text 2:

Task B: (For the testers' reference only, not to be supplied to the candidates)

Receptionist:	Good evening, Ma'am! How can I help you?
Visitor:	I'd like to book a room, please. Do you have a single with attached bath?
Receptionist:	We certainly do, Ma'am.
Visitor:	And do you have TV in the room?
Receptionist:	Yes, all our rooms have a TV.
Visitor:	Good! And a telephone?
Receptionist:	Yes, Ma'am
Visitor:	How much do I have to pay for a night?
Receptionist:	300.00 rupees for a night or Rs. 525 for two nights. How long would you stay?
Visitor:	Two nights starting tonight. I'll pay 525 rupees on arrival, if that's OK with you.
Receptionist:	That's certainly fine, Ma'am. Your room number is 105 on the second floor.
Visitor:	Thank you.
Receptionist:	You're welcome.

Task B: *Listen to the tape and tick (√) against the things mentioned in the tape in the empty space in table 2.*

Table 2

A double room	
A single room	
Attached bath	
Common bath	
TV	
Fax	
Phone	
Computer	
525 rupees	
600 rupees	
Room no. 105	
Room No. 501	

Task C: Note to the assessors /raters -Candidates will be asked to choose one of the following three tasks through lottery.

'

Have you (or any of your friends) been in an embarrassing situation? How did you (or they) when you (or your friends) were embarrassed? Speak in English for five minutes describing it.

OR

In English, speak for about five minutes on what you would do if you found a million rupees.

OR

The chart below shows how Mr. Pradhan spends his income. Study the chart and speak in English for about five minutes on how Mr. Pradhan spends his money, whether or not he spends it wisely, how can he spend more wisely, if he should spend more on some items and less on others, whether he should spend less on some items to save more and why.

▨	Food
▩	Education
☐	Clothes
☐	Entertainment
▧	Rent
▨	Saving

Task D

[The format of the conversation may take several forms, and it may differ from one situation to another and from one topic to another. The following conversation may be taken as an example.]

Interlocutor (I): [Asks the candidate to take a seat.] Hi! My name is Kishore Gurung. This is (if there is a second interlocutor/assessor) my colleague Mr. Nurul Ansari. We will conduct the session together.

Candidate (C):   *Good morning/afternoon. Thanks*

I: *Would you like to tell us about yourself?*

C: *Yes Sir. My name is Indu Sharma. I am from Junta Secondary School, Chainpur. I live with my parents in a small village in Siraha district called Dhanagadhi. I have two brothers. Both of them go to the same school. My parents are farmers.*

I: *All right, Indu. Most Nepalese celebrate and enjoy festivals. I, for example, celebrate Dasai and Tihar. Mr. Ansari celebrates Ida and Ramadan. Yet others celebrate Thanks Giving and Christmas. Do you celebrate any festivals at all?*

C: *Yes sir.*

I: *What festivals do you and your family celebrate?*

C: *I come from the Terai. So, my family and I celebrate Holi, Chhatha and Deepawali. We celebrate Dushehara too.*

I: *Good. Now, suppose that my colleagues and I don't know anything about the festival of Holi. Can you describe the Holi festival for us, what it is, its significance, the way you celebrate it, and so on. You have about five minutes to do so.*

C: *Yea sir. (And speaks for about five minutes)*

[The interlocutor/assessor can interrupt the candidate during his or her speech in order to ask questions. After the time allotted is over or if the candidate is finished within five minutes, the interlocutor/assessor thanks him/her and takes leave]

Read the text carefully and do the question given below: (6 mins.)

My parents did not believe in boys wearing trousers, so I wore shorts to school. I wanted to wear trousers but every time I opened my mouth, my parents told me to shut up. One day, the class teacher told us that we were going to have a group photograph taken and he asked us to wear trousers for the occasion. I was in a dilemma, as I did not possess a pair of trousers.

After school I asked some of my friends to lend me a pair of trousers but no one had a pair which would fit me. I even tried on a pair of my father's trousers. They were too short for me and much too large at the waist. I gave up in despair. The next day arrived. I thought of not going to school but my parents told me to get ready. They told me to stop grumbling because there was nothing wrong in wearing shorts. So, I turned up as usual at school. The class teacher was aghast when he saw me. He and some other students wanted me to be excluded from the photograph but the head teacher intervened. She asked me to join the group. The class teacher told me to sit in the front with two students on either side of me.

The photograph was duly taken A few days later, a copy of the picture appeared on the notice-board. Crowds of students gathered to see it. It was a well-taken photograph. All the students were there in their well-ironed, spotlessly white shirts and trousers - except me in the front row, with my knobbly knees conspicuously exposed. Needless to say, the picture gave rise to a lot of laughter and unkind remarks. You can imagine how I felt. (298 words)

Task E. *Find the words in the text which means the similar to the words or phrases given below. Write the answers in the space provided:* (5 mins.)

a. a time when an event takes place   . . . .
b. a situation of uncertainty   . . . .
c. a sense of hopelessness   . . . .
d. complaining   . . . .
e. arrived   . . . .
f. kept out   . . . .
g. interfered to prevent something from happening   . . . .
h. noticeably   . . . .
i. left uncovered   . . . .
j. comments   . . . .

Task F. The following sentences are not in proper order. Rearrange then in a sensible order. Put the letters next to the numbers in the space provided following the example of number one (5 mins.)

i. The writer brought home a copy of the photograph.
j. The class teacher told the writer to wear trousers for a group photo.
k. The group photo was taken and it created a great deal of laughter when it appeared on the notice board.
l. The class teacher refused to include him in the group.
m. His father told him that it was all right to wear shorts.
n. He tried to find a pair of trousers that fits him without a success.
o. He felt terribly embarrassed.
p. He was included in the group because the head teacher intervened.

Example -

1. .... *h* .....
2. ....
3. ....
4. ....
5. ....
6. ....
7. ....
8. ....

Task G. *Write the answer of the following questions in your own words. The answers should be brief and to the point. You need not copy the questions. (10 mins.)*

q. What was the dilemma the writer was in?

....
....
....

ii. Why was the class teacher aghast?

....
....
....

iii. How would a seat in the back row have helped the writer?

....
....
....

iv. Why were students amused when they looked at the photograph?

....

....

....

v. Write in brief how the writer felt after the photograph appeared on the notice board.

....

....

....

....

....

....

Task H. Complete the following sentences. Use the example as a guide (5 mins).

Example –

The writer's parents told him to stop grumbling because ... ...
... ...

Answer -
The writer's parents told him to stop grumbling because they thought there was nothing wrong in wearing shorts.

1. The writer always wore shorts because ... ...
... ...

2. One day, the class teacher asked them to wear trousers because ... ...
... ...

3. He couldn't borrow a pair of trousers from his friends because ... ...
... ...

4. .... because he did not wear
trousers.

1. .... because the head
teacher intervened.

Task I. *Read the following story. Some of the words have been left out. Read the whole story quickly before you try to fill in the missing words. Then fill in the blanks with an appropriate word. Try every blank. Blank number one has been done for you. (15 mins.)*

Some people say that you should not run away if you meet a tiger. *They* (1) say that you must .... (2) still and face the .... (3). They say that it .... (4) not attack a human .... (5) unless the person panics .... (6) runs away. When these .... (7) hear of a person .... (8) killed by a tiger, ... (9) say, "He should not .... (10) run away. He ought .... (11) have stood still. If .... (12) had, he might have .... (13) alive today.

In the .... (14) days in Nepal, they .... (15) to call this 'talking .... (16) the tiger'. 'When you .... (17) a tiger,' they used .... (18) say, 'you should speak .... (19) it. You should tell .... (20) that there is plenty .... (21) food in the jungle and that it need not have to eat you.

Unfortunately, most people who were attacked by a tiger did not have courage to speak to it.

Read the following passage and answer the questions given below (5 mins):

The temple of Tripurasundari, one of the most famous goddess temples of Far Western Region, is situated in the border district of Baitadi. Baitadi is 83 km from nearby Dadeldhura, 200 km from Dhangadi and about 800 km from the capital city, Kathmandu. This famous temple is 10 minutes drive and 40 minutes walk from the district headquarters.

Fairs (Jaat in local language) occur twice a year i.e. in Kartik and Asar. Visitors from neighbouring districts of Dadeldhura, Darchula, Doti, Kailali and Kanchanpur and even from many bordering parts of India throng here to observe the fairs and to worship the goddess. The fair is celebrated for two days. The first day fair is called Nani Jaat (small fair) and the second Thuli Jaat (big fair). In the remote villages where women hardly get time to visit their relatives due to domestic works, these fairs provide them an opportunity to share their feelings with their relatives, parents and friends. Villagers look dazzling and glowing in neat and clean dresses, ornaments and a beautiful tika on their forehead.

There are seven temples of seven Bhagwatis in different parts of Bajhang, Doti and Baitadi. Of these, five are situated in Baitadi. According to the ancient mythology, these seven Bhagwatis are sisters. In these different temples, fairs occur at different times.

Accommodation is cheap and easily available. Taxi and day/night bus services are available from most neighbouring cities and towns to Tripurasundari and Nepal India border of Jhulaghat, where people can enjoy horse riding, trekking and swimming in the Mahakali River. Air services from Kathmandu to Nepalgunj or Dhangadi are also available. One can enjoy popular sight seeing in the Royal Bardia National Park when travelling by a car.

Task J. *Complete the following sentences by filling in blanks with words or phrases from the text (5 mins).*

a. Baitadi is about .... km away from Kathmandu.
b. One can reach the temple of Tripurasundari in .... minutes when driving from the district headquarters.
c. Fairs take place in the month of Kartik and ....
d. Naani Jaat refers to the .... fair
e. Of the two days of celebrations, Thuli Jaat takes place on the .... day.
f. Of seven temples of Bhagawatis, ... are in Baitadi.
g. One can go swimming in ....
h. .... services are available from Kathmandu to nearby towns of Nepalgunj and Dhanagadhi.
i. Royal Bardia National Park is popular for ....
j. Jaat is a local term for ....

Task K. *Answer the following questions in one word or phrase. Write your answer in the space provided against their number (5 mins).*

a. Name the district the term 'district' (para 1, line 4) refers to. ....

b. What does the word 'here' (para 2, line 3) refer to?   ....

c. What does the word 'them' (para 2, line 6) refer to?   ....

d. What does the term 'these' (para 3, line 2) refer to?   ....

e. What does the word 'one' (para 4, line 5) refer to?   ....

Task L. *In the space below, write a gist of the text 4 above in about 50 words (5 mins)*

....

....

....

....

....

....

Read the following information leaflet carefully and then answer the questions given below (5 mins)

Annapurna Bus Service (A Private Company) Estd 1999	Mahendranagar		Nepalgunj		Bharatpur		
	Arrival	Departure	Arrival	Departure	Arrival	Departure	
Bus Time Table	AM					PM	
Route – 25	5.30	6.00	9.30	10.00	2.00	3.00	
	7.30	8.00	11.30	12.00	4.00	5.00	
Mahendra Nagar – Bharatpur *Via Nepalgunj*	9.30	10.00	1.30PM	2.00	6.00	7.00	
	11.30	12.00	3.30	4.00	8.00	9.00	
Effective January 2003	PM						
For more Information	1.30	2.00	5.30	6.00	10.00	11.00	
	3.30	4.00	7.30	8.00	12.00	1.00	
Call – 593 333	5.30	6.00	9.30	10.00	2.00	AM 3.00	
Visit our website www.annapurnabus.com	7.30	8.00	11.30	12.00	4.00	5.00	
Head Office **15 Shaheed Chowk, Mahendranagar 13** Kunchanpur							

Passengers, please note that buses depart every two hours, AM to PM from Mahendranagar and Nepalgunj, and PM to AM from Bharatpur

This is an express service. So passengers will not be picked up from any other places except the designated ones.

Passengers are encouraged to do the reservations prior to their travelling date. Tickets at windows are subject to availability.

The buses run as scheduled everyday of the week including Saturdays, Sundays and Public Holidays.

Tickets may be collected from our site offices in Mahendranagar, Nepalgunj and Bharatpur and Head Office in Mahendranagar.

Task M. *In the space provided against each question, write the answer of the following questions in your own words. Your answers should be brief (5 mins).*

a. When was the Annapurna Bus Service Company established?.................. .
b. How many main terminal stops are there in Route 25?   ....
c. How many buses depart from Nepalgunj everyday?   ....
d. In the afternoon, what is the fist bus one can take to go to   ....
   Nepalgunj from Mahendranagar?   ....
e. Which of the three bus terminals is busy all night long?   ....

Task N. *Read the text once again and say whether the statements below are true or false. If a statement is true, write 'T' against the number of the statement, if it is false, write 'F'. (5 mins)*

a. The timetable for route 25 is effective in 2003.   .....
b. Passengers can take buses after 6 PM at the Mahendranagar terminal.   .....
c. Buses run in every three hours from each of the three terminals.   .....
d. No passengers will be taken from any other places than the three terminals. ......
e. Passengers can book tickets in advance.   .....
f. The main office of this bus company is in Bharatpur.   .....
g. Passengers can collect their tickets from the site offices.   .....
h. Tickets cannot be collected from the company's head office.   .....

Task O (25 Minutes) (Full Mark – 10)

Describe the character of the writer of the reading text one (see question 1). Write in about 150 words what he was like and how he felt when he saw crowds of students looking at the photographs and laughing.

Spend about five minutes on planning, i.e. thinking, putting down ideas and organising them. Spend the rest of the time on writing the essay.

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

(Please use back of the sheet if needed)

Task P (20 minutes) (Full Mark - 10)

Imagine that you are one of the regular passengers of Annapurna Bus Services. You had made your reservation two weeks prior to your travel date with a provision that you would collect your ticket from its Nepalgunj office on the date of your travel. However, when you contacted the office half an hour before your departure, the counter informed you that there was no reservation made in your name. Obviously, you were very angry about the carelessness of the bus company.

Write a letter of complaint in about 150 words to the Director of the bus company in its head office at Mahendranagar asking him to take an immediate action regarding the matter. In your letter, say what might happen if the company allows continuing such a problem.

Include in your letter, who you are writing to and why, what about, the date and place of reservation, and the name of the booking clerk. You must write the letter in the standard, official format and formal language.

Spend about 5 minutes in planning and rest of the time on writing the letter.

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

(Please use back of the sheet if needed)

Task Q (18 minutes) (Full Mark - 15)

Suppose you are a tourism development officer of your village/town/city. In about 150 words, prepare a brochure/ information leaflet of your village/town/city for a tourist who may not have heard about it before. Following the brochure of Annapurna Bus Services (see question 4) as a model, present information to attract tourists. Include in your brochure information about - location (geography/climate), population (demography/culture), places of historical importance, places of interest, current development projects, travelling and accommodation facilities, contact places, etc. Spend about 5 minutes to plan and rest of the time on writing the brochure. You need not write the information in complete sentences.

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

(Please use back of the sheet if needed)

Appendix C

Workshop Questions

Workshop1

Group 1: Administrators' Group

1. Do you think the test specifications and structure are workable?

2. How do you think the test specifications could be improved?

3. What in your mind should be done differently?

5. Are there additional issues that could be addressed?

6. What sort of professional development will be most useful for test developers to use the proposed test structure/test specifications?

Group 2: Test Developers' Group

1. Do you think the test specifications and structure are workable?

2. How do you think the test specifications could be improved?

3. Does the document provide sufficient information for the test developers on test objectives, making of the tests, time and mark allocations, scoring and reporting schemes, resources required and constraints or limitations?

5. What in your mind should be done differently?

6. What sort of professional development will be most useful for test developers to use the proposed test structure/test specifications for developing tests?

Group 3: Examiners' Group

1. Are the proposed scoring and reporting schemes more appropriate compared to the existing ones?

2. Do the schemes appropriately cover all four micro-skills of language?

3. Would the proposed scoring and reporting schemes improve the quality and accuracy of scoring and reporting of scores?

4. How do you think the scoring and reporting schemes could be improved?

5. Do you think the test specifications and structure are workable?

6. What in your mind should be done differently?

7. What sort of professional development will be most useful for examiners/ raters to implement the test structure in general and scoring and reporting frameworks in particular?

Group 4: Teacher Trainers' Group

1. Do the proposed test structure/test specifications provide a workable way to help assessors assess more accurately against the curriculum of the SLC English?

2. Do the proposed testing framework and test specifications adequately address the foreign language testing principles?

3. How do you think the test specifications could be improved?

4. What in your mind should be done differently?

5. What sort of professional development will be most useful for teachers and examiners to implement the proposed test framework and test specifications?

Workshop Questions

Workshop 2

Group 1. Teachers' Group: Listening

1. Do you think the test specification and structure for listening is workable?

2. How do you think the test specification for listening could be improved?

3. What in your mind should be done differently?

4. Are the proposed allocations of time, marks and resources for listening adequate?

5. Does the document provide enough information on resources, equipment and constraints for effective implementation of the proposed test specification for listening?

Group 2. Teachers' Group: Speaking

1. Do you think the test specification and structure for speaking is workable?

2. How do you think the test specification for speaking could be improved?

3. What in your mind should be done differently?

4. Are the proposed allocations of time, marks and resources for speaking adequate?

5. Does the document provide enough information on resources, equipment and constraints for effective implementation of the proposed test specification for speaking?

Group 3. Teachers' Group: Reading

1. Do you think the test specification and structure for reading is workable?

2. How do you think the test specification for reading could be improved?

3. What in your mind should be done differently?

4. Are the proposed allocations of time, marks and resources for reading adequate?

5. Does the document provide enough information on resources, equipment and constraints for effective implementation of the proposed test specification for reading?

Group 4. Teachers' Group: Writing

1. Do you think the test specification and structure for writing is workable?

2. How do you think the test specification for writing could be improved?

3. What in your mind should be done differently?

4. Are the proposed allocations of time, marks and resources for writing adequate?

5. Does the document provide enough information on resources, equipment and constraints for effective implementation of the proposed test specification for writing?

Group 5. Scoring and Reporting Group

1. Do you think the scoring and reporting frameworks are workable?

2. How do you think the proposed scoring and reporting schemes could be improved?

3. What in your mind should be done differently?

4. Are the proposed scoring and reporting schemes more appropriate compared to the existing ones?

5. Do the schemes appropriately cover all four macro-skills of language?

6. Would the proposed scoring and reporting schemes improve the quality and accuracy of scoring and reporting of scores?

Appendix D

The Proposed Rating and Reporting Schemes

An Analytical Rating and Reporting Scale

A five-point analytical scale for scoring on the four skills of language has been adapted. The four points of the scale roughly correspond to the labels used by the Office of the Controller of Examinations (OCE) as the four levels of performances. The four levels of performances are as follows:

- Failure
- Pass (Third Division)
- Second Division
- First Division
- Distinction

Many scales use more than four points or levels of performance. Some scales use as many as 9 performance levels. The present research proposes only four levels as it is expected to be difficult for the Nepali users of the scales to realise the subtle differences between any two levels of a 7 or 9-point scale.

The terms used in the five point rating scale are described below:

The Five points of the rating Scale and their Definitions

0 (Zero)	=	No evidence of knowledge Unattempted Poor Failure Incoherent Lack of basic knowledge of the target language
1 (One)	=	Limited but satisfactory Accuracy confusing Serious grammatical errors but reasonable correct Inconsistency in the use of the target language Pass
2 (Two)	=	Moderate or good Medium target language use range Some inconsistencies Second class (division) Simple but accurate realisation of tasks
3 (Three)	=	High, very good or excellent Large target language use range Minimum, non-basic grammatical errors Complete realisation of the task First class (division)
4 (Four)	=	Excellent Effortless and smooth language Near-native proficiency Almost flawless language Distinction

Analytical Scoring Scale with Skill-wise Descriptions for Proficiency Assessment

Level	Listening	Speaking	Reading	Writing
'0'	- Unattempted, - No comprehension of spoken language	- Unattempted, - Oral production limited, - Unintelligible production - No communicative ability - Native language like pronunciation, - Poor stress and intonation,	- Unattempted - No comprehension of simple isolated words or phrases - Unable to comprehend written text, -	- Unattempted Gr= No accuracy, unintelligible Vo= Poor range, inadequate for the most basic intended communication Org= No evidence Cn= No evidence R= No evidence Ct= Answer not related to the task, inadequate answer P= unfamiliar with the conventions of punctuation Sp= most spellings inaccurate Hw= Unintelligible Fl= Use of half-learned words/structures, misused words or structures, broken language, no communication
1	- can comprehend short utterances on familiar materials/ tasks - can ask and respond to simple questions - have limited ability to cope with longer texts	- Some spontaneity, - Fluency uneven, - Limited vocabulary - Articulation comprehensible, - Problems with stress and intonation, - No understanding of social conventions, - Cannot initiate or sustain	- Comprehension is limited and dependent on the explicitly stated information - Can comprehend complex sentences if simplified - Can answer questions on surface level information	Gr= Errors, inconsistencies, frequent grammatical inaccuracies Vo= limited range, frequent lexical inappropriateness or repetition, 5000 words range Org= Poor organisation Cn= few markers of cohesion R= Limited knowledge of formal language, formulaic use of language Ct= Answers of limited relevance,

371

#	Listening	Speaking	Reading	Writing
		- conversations, - Can cope with routine situations, - Can express themselves in simple and two-clause sentences, - Errors in tense/inflections		major gaps in the treatment of the topic, pointless repetition P= Some familiarity of punctuation conventions Sp= Frequent spelling errors Hw= Reasonable readable handwriting but frequent confusions in handwriting Fl= Little sense of ease at communication, frequent evidence of hesitations and repetitions, inconsistent use of language
2	- can comprehend most simple conversations in social situations - have limited comprehension in work/ employment/academic settings - have difficulty in following complex instructions	- can describe simple topics, - can comprehend discourse with simple linguistic features, - can follow/give instructions - can interpret non-text materials in simple language, - can cope with most social situations, - describe events, processes, works, experiences - accent intelligible, - fluency reasonable	- Can read and comprehend most texts written in simple language and are familiar in context - Can read and enjoy simply structured prose and poetry - Can deduce limited implicitly stated meaning from the text	Gr= Some grammatical inconsistencies/ inaccuracies Vo= Medium range (7000 words), some lexical inappropriacy Org= clear organisation but some flaws Cn= medium range use of cohesive devices, sometimes confusing R= Good use of formal language but some inconsistencies Ct= answers relevant, major areas covered, some gaps or redundancies P= Some inaccuracies in the use of punctuation Sp= Some inaccuracies Hw= Good handwriting, not legible at times Fl= Reasonable ease of communication, some inconsistencies in the use of

	Listening/Speaking	Speaking	Reading	Writing
				language items, some gaps and hesitations
3	- Can comprehend and speak to the speakers of English - Comprehension not affected by the use of complex words/structures - Can understand utterances without paraphrase or repetition	- can handle most social/work situations confidently' - can discuss interest, hobbies, preferences with reasonable 'ease' - accent reasonably well, - can use a wide range of vocabulary and language devices, - fluency is rarely disrupted by hesitations, - can speak confidently on topics or situations presented, - control of grammar good	- Can read, understand and respond to information texts. - Can enjoy reading prose and poetry - Can comprehend intended meaning of texts - Can read and comprehend texts at a reasonable speed.	Gr= Good control on grammar, no grammatical inaccuracies Vo= Large range (10000 words), Few lexical inappropriateness Org= Good organisation Cn= proper use of a wide range of cohesive devices R= Good control in the use of formal/ informal language as demanded by the context of situation Ct= answers relevant and adequate to the task P= Few inaccuracies in the use of punctuation Sp= Few inaccuracies in spellings Hw= clear, beautiful handwriting Fl= Consistent and appropriate use of language, no hesitations, ease at communication
4	- can understand both formal and colloquial speech unrestricted by lexical and grammatical deficiencies	- can initiate, sustain and conclude a wide variety of communicative tasks - can use language appropriate to context, function and intention	- can read a variety of language forms and respond to social, academic and vocational needs, - can read and interpret various forms texts independently	-can convey meaning precisely and accurately unrestricted by lexical, morphological, syntactic and spelling deficiencies, - can transform language accurately from one form to another without loosing the intended meaning, - can perform a wide variety of written tasks confidently and

373

		flexibly with language features including formal accuracy, structural variation, word choice, idioms, register, appropriateness, discourse structure, and stylistic convention.

[adapted from ASLPR Ingram and Wylie (1982), Bachman and Palmer (1996), Alderson et al (1995) , North (2002)]

Gr. = Grammar, Vo. = Vocabulary, Org. = Organisation, Cn. = Cohesion, R = Register, Ct. = Content, P = Punctuation, Sp. = Spelling, Hw = Handwriting, Fl = Fluency

Figure: The proposed Rating Scale

374

The Proposed Test Result Reporting Scheme

Alphabetical Grade	Candidate's score (75)	Score Range	Division	Level and their Descriptors
A+ A	⋄	86-100	Distinction	**4** - native-like proficiency - can understand both formal and colloquial speech - can initiate, sustain and conclude a wide variety of communicative tasks - can use language appropriate to context, function and orientation - can read at a reasonable speed and comprehend implicit meaning of texts - can convey meaning precisely and accurately - can perform a wide variety of written tasks confidently
A- B+		71-85	First Div	**3** - can understand utterance without paraphrase or repetition - can comprehend and speak to English speakers - can hold conversations on most topics with reasonable fluency - can read, understand and respond to questions on most texts within the structure and vocabulary range, - can perform on most written tasks confidently and without hesitation - occasional grammatical flaws,
B B- C+		56-70	Second Div	**2** - can comprehend most simple conversations in social situations - limited comprehension in academic settings - can describe familiar things, personal feelings, events and processes - can read and comprehend most texts written in simple language, - can read and enjoy simple, graded prose and poetry - can answer questions on surface level information - medium range of use of cohesive devices - can perform on familiar written tasks - fluency and accuracy inconsistencies
C C-	▬	40-55	Pass Div	**1** - can comprehend short conversations on familiar materials and tasks - can ask and respond to simple questions - can hold simple conversations on familiar topics - limited comprehension of explicitly stated information - can answer questions on surface level information - limited range and frequent lexical inappropriateness - hesitation, repetition and inconsistencies in writing - lexical and grammatical errors
D		39 or less	Fail	**0** - unattempted - no comprehension of spoken language - limited oral production of memorised stock phrase - no communicative ability - unable to comprehend running texts - no accuracy - inadequate vocabulary and structure to most tasks - poor hand writing

⋄   = A candidate's achievement in the skill
▬   = Average performance
▭   = National score range

Figure: Proposed result reporting format

Appendix E
Statistics and Statistical Procedures

Results of the Pilot Test: Rater 1

Item	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	Total
	I	I	s	s	r	r	r	r	r	r	r	r	r	r	w	w	w	
Full Mark	**7**	**3**	**5**	**5**	**5**	**3.5**	**5**	**5**	**10**	**2.5**	**2.5**	**5**	**2.5**	**4**	**10**	**10**	**15**	**100**
									Private Schools									
1	5	3	3	4	4	2.5	4	5	9	2	2.5	1.5	2	3	4.5	4	9	
2	5	3	4	4	4.5	3.5	3.5	5	10	2.5	1.5	3	1.5	4	7	6	10	
3	4	3	4	3	4.5	3	3	5	6.5	2.5	2.5	3.5	2	3.5	7	6	6	
4	4.5	3	2	3	4.5	3.5	1.5	4	5	2.5	2	1	2.5	3	7	6	8	
5	2.5	3	2	3	4	0	3	4	5	2.5	1.5	3	1.5	2.5	6	5	10	
6	5	3	3	4	4.5	2	3	4	7	2.5	2	2	1.5	3	7	7	11	
7	4.5	3	2.5	2.5	4.5	3.5	3.5	4	8	2.5	1.5	2	2	4	7	7	9	
8	2.5	3	3.5	4	5	3.5	3	4	8	2.5	2.5	3.5	2.5	3	8	7	11	
9	.5	1	1	1	2.5	.5	1.5	3	3.5	1.5	1	1.5	1.5	2	3	3	5	
Public Schools																		
10	1.5	2	1	1	2	2	1.5	3	7	2.5	1.5	1.5	1.5	3	4	3	6	
11	1.5	2	1	2	3	2	1.5	3	5.5	1.5	1.5	1.5	1.5	2	4	3	6	
12	1.5	1	1	2	4	2	1.5	4	6.5	2	1.5	1	1.5	2	4	4	7	
13	1.5	1	0	1	2	1.5	1.5	2	1.5	1.5	1	1	1	1.5	2	2	4	
14	1.5	1	1	1	4	2	1.5	4	6	1.5	1.5	1.5	2	1.5	3	3	5	

Results of the Pilot Test: Rater 2

Item	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	Total
	1	1	s	s	r	r	r	r	r	r	r	r	r	r	w	w	w	
Full Mark	7	3	5	5	5	3.5	5	5	10	2.5	2.5	5	2.5	4	10	10	15	100
Private Schools																		
1	4	3	4	4	4	2.5	4.5	4.5	9	1.5	2.5	1.5	2.5	3	4.5	5	9	
2	5	3	4	4	4.5	3.5	4	4.5	10	2.5	1.5	3	1.5	3.5	6	6	8	
3	4.5	3	4	3	4	3.5	3.5	5	6	2.5	2.5	3	2.5	3	8	7	6	
4	4	3	3	3	4.5	3	2	4	5	2.5	2	1	2.5	3	6	6	8	
5	2.5	3	3	3	4.5	1	3	4	6	2.5	1.5	3	1.5	2.5	7	5	9	
6	4.5	3	4	3	4	2.5	3	4	8	2.5	2	2	1.5	3	6	7	10	
7	4	3	2.5	2	4	3.5	3.5	4	8	2.5	1.5	2	2	4	7	6	9	
8	3	3	3.5	2	4.5	3.5	3.5	4	9	2	2.5	3	2.5	3	6	7	11	
9	1.5	2	1	1	1.5	.5	1	3	3.5	1.5	1	1.5	1.5	2	4	3	4	
Public Schools																		
10	1.5	2	1.5	1	2	2	1.5	3	7	2.5	2	1.5	2	3	3	4	6	
11	2	1.5	1.5	2	3	2	1.5	3	5.5	1.5	2	1.5	1.5	2	4	3	6	
12	2	1	1.5	2	3	2	1.5	3.5	6.5	2	1.5	1	2	2	5	4	8	
13	1.5	1	0	1	2	1.5	1	2	1.5	1.5	1	1	1	1.5	3	3	4	
14	2	1	1	1	4	2	1.5	4	6	1.5	1.5	1.5	2	1.5	4	4	6	

377

Results of the Test Trial: Combined Score (R1 & R2)

Item	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	Total
	1	1	s	s	r	r	r	r	r	r	r	r	r	r	w	w	w	
Full Mark	7	3	5	5	5	3.5	5	5	10	2.5	2.5	5	2.5	4	10	10	15	100
Private Schools																		
1	4.5	3	3.5	4	4	2.5	4	5	9	2	2.5	1.5	2	3	4.5	4.5	9	
2	5	3	4	4	4.5	3.5	4	5	10	2.5	1.5	3	1.5	3.5	6.5	6	9	
3	4	3	4	3	4	3	3	5	6.5	2.5	2.5	3	2	3	7.5	6.5	6	
4	4.5	3	2.5	3	4.5	3	1.5	4	5	2.5	2	1	2.5	3	6.5	6	8	
5	2.5	3	2.5	3	4	.5	3	4	5.5	2.5	1.5	3	1.5	2.5	6.5	5	9.5	
6	5	3	3.5	3.5	4	1.5	3	4	7.5	2.5	2	2	1.5	3	6.5	7	10.5	
7	4	3	2.5	2	4	3	3.5	4	8	2.5	1.5	2	2	4	7	6.5	9	
8	3	3	3.5	3	5	3.5	3	4	8.5	2.5	2.5	3	2.5	3	7	7	11	
9	1	1.5	1	1	2	1	1	3	4.5	1.5	1	1.5	1.5	2	3.5	3	4.5	
Public Schools																		
10	1.5	1.5	1.5	1	2	2	1.5	3	7	2.5	1.5	1.5	1.5	3	3.5	3.5	6	
11	1.5	2	1.5	2	3	2	1.5	3	5.5	1.5	1.5	1.5	1.5	2	4	3	6	
12	1.5	1	1.5	2	3.5	2	1.5	3.5	6.5	2	1.5	1	1.5	2	4.5	4	7.5	
13	1.5	1	0	1	2	1.5	1	2	1.5	1.5	1	1	1	1.5	2.5	2.5	4	
14	1.5	1	1	1	4	2	1.5	4	6	1.5	1.5	1.5	2	1.5	3.5	3	5.5	
Total	41	33	32.5	33.5	50.5	31	33	53.5	91	30	23.5	26.5	24.5	37	73.5	67.5	105.5	
Mean	2.9	2.3	2.3	2.4	3.6	2.2	2.4	3.8	6.5	2.1	1.7	1.9	1.8	2.6	5.2	4.8	7.5	
P-Value	.41	.78	.46	.48	.72	.62	.48	.76	.65	.84	.68	.38	.72	.65	.52	.48	.50	

378

Item Difficulty Level

Item	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	Total
Full Mark	7	3	5	5	5	3.5	5	5	10	2.5	2.5	5	2.5	4	10	10	15	100
Total Score	41	33	32.5	33.5	50.5	31	33	53.5	91	30	23.5	26.5	24.5	37	73.5	67.5	105.5	
Mean	2.9	2.3	2.3	2.4	3.6	2.2	2.4	3.8	6.5	2.1	1.7	1.9	1.8	2.6	5.2	4.8	7.5	
P-Value	.41	.78	.46	.48	.72	.62	.48	.76	.65	.84	.68	.38	.72	.65	.52	.48	.50	

Total no. of examinees – 14

379

Discrimination Index: Examinees vs. Test Tasks

Item/Examinee	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	.64	1	.7	.8	.8	.71	.8	1	.9	.8	1	.3	.8	.8	.5	.5	.6
2	.71	1	.8	.8	.9	1	.8	1	1	1	.6	.6	.6	.9	.7	.6	.6
3	.57	1	.8	.6	.8	.86	.6	1	.7	1	1	.6	.8	.8	.8	.7	.4
4	.64	1	.5	.6	.9	.86	.3	.8	.5	1	.8	.2	1	.8	.7	.6	.5
5	.35	1	.5	.6	.8	.14	.6	.8	.6	1	.6	.6	.6	.6	.7	.5	.6
6	.71	1	.7	.7	.8	.42	.6	.8	.8	1	.8	.4	.6	.8	.7	.7	.7
7	.57	1	.5	.4	.8	.86	.7	.8	.8	1	.6	.4	.8	1	.7	.7	.6
8	.42	1	.7	.6	1	1	.6	.8	.9	1	1	.6	1	.8	.7	.7	.7
9	.14	.5	.2	.2	.4	.29	.2	.6	.5	.6	.4	.3	.6	.5	.4	.3	.3
10	.21	.5	.3	.2	.4	.57	.3	.6	.5	1	.6	.3	.6	.8	.4	.4	.4
11	.21	.67	.3	.4	.6	.57	.3	.6	.7	.6	.6	.3	.6	.5	.4	.3	.4
12	.21	.33	.3	.4	.7	.57	.3	.7	.7	.8	.6	.2	.6	.5	.5	.4	.5
13	.21	.33	.0	.2	.4	.42	.2	.4	.2	.6	.4	.2	.4	.4	.3	.3	.3
14	.21	.33	.2	.2	.8	.57	.3	.8	.6	.6	.6	.3	.8	.4	.4	.3	.4

Correlation among Tests

X Reading	Y Cloze	XY	X2	Y2
76	90	6840	5776	8100
83	100	8300	6889	10000
80	65	5200	6400	4225
67	50	3350	4489	2500
64	55	3520	4096	3025
67	75	5025	4489	5625
76	80	6080	6840	6400
83	85	7055	6889	7225
41	45	1845	1681	2025
53	70	3710	2809	4900
50	55	2750	2500	3025
53	65	3445	2809	4225
36	15	540	1296	225
56	60	3360	3136	3600
885	910	61020	60099	65100

X Reading	Y Writing	XY	X2	Y2
76	51	3876	5776	2601
83	61	5246	6889	3721
80	57	4560	6400	3249
67	59	3953	4489	3481
64	60	3840	4096	3600
67	69	4623	4489	4761
76	64	4864	6840	4096
83	71	5893	6889	5041
41	31	1271	1681	961
53	37	1961	2809	1369
50	37	1850	2500	1369
53	46	2438	2809	2116
36	26	936	1296	676
56	34	1904	3136	1156
885	703	47215	60099	38197

X Cloze	Y Writing	XY	X2	Y2
90	51	4590	8100	2601
100	61	6100	10000	3721
65	57	3705	4225	3249
50	59	2950	2500	3481
55	60	3300	3025	3600
75	69	5175	5625	4761
80	64	5120	6400	4096
85	71	6035	7225	5041
45	31	1395	2025	961
70	37	2590	4900	1369
55	37	2035	3025	1369
65	46	2990	4225	2116
15	26	390	225	676
60	34	2040	3600	1156
910	703	48415	65100	38197

X Listening	Y Reading	XY	X2	Y2
75	76	5700	5625	5776
80	83	6640	6400	6889
70	80	5600	4900	6400
75	67	5025	5625	4489
55	64	3520	3025	4096
80	67	5360	6400	4489
70	76	5320	4900	6840
60	83	4980	3600	6889
25	41	1025	625	1681
30	53	1590	900	2809
35	50	1750	1225	2500
25	53	1325	625	2809
25	36	900	625	1296
25	56	1400	625	3136
730	885	50135	45100	60099

X Speaking	Y Reading	XY	X2	Y2
75	76	5700	5625	5776
80	83	6640	6400	6889
70	80	5600	4900	6400
55	67	3685	3025	4489
55	64	3520	3025	4096
70	67	4690	4900	4489
45	76	3420	2025	6840
65	83	5395	4225	6889
20	41	820	400	1681
25	53	1325	625	2809
35	50	1750	1225	2500
35	53	1855	1225	2809
10	36	360	100	1296
20	56	1120	400	3136
660	885	45880	38100	60099

X Listening	Y Cloze	XY	X2	Y2
75	90	6750	5625	8100
80	100	8000	6400	10000
70	65	4550	4900	4225
75	50	3750	5625	2500
55	55	3925	3025	3025
80	75	6000	6400	5625
70	80	5600	4900	6400
60	85	5100	3600	7225
25	45	1125	625	2025
30	70	2100	900	4900
35	55	1925	1225	3025
25	65	1625	625	4225
25	15	375	625	225
25	60	1500	625	3600
730	910	52325	45100	65100

X Listening	Y Writing	XY	X2	Y2
75	51	3825	5625	2601
80	61	4880	6400	3721
70	57	3990	4900	3249
75	59	4425	5625	3481
55	60	3300	3025	3600
80	69	5520	6400	4761
70	64	4480	4900	4096
60	71	4260	3600	5041
25	31	775	625	961
30	37	1110	900	1369
35	37	1295	1225	1369
25	46	1150	625	2116
25	26	650	625	676
25	34	850	625	1156
730	703	40510	45100	38197

X Speaking	Y Cloze	XY	X2	Y2
75	90	6950	5625	8100
80	100	8000	6400	10000
70	65	4550	4900	4225
55	50	2750	3025	2500
55	55	3025	3025	3025
70	75	5250	4900	5625
45	80	3600	2025	6400
65	85	5525	4225	7225
20	45	900	400	2025
25	70	1750	625	4900
35	55	2275	1225	3025
35	65	2275	1225	4225
10	15	150	100	225
20	60	1200	400	3600
660	910	48200	38100	65100

384

X Speaking	Y Writing		X2	Y2
75	51	3825	5625	2601
80	61	4880	6400	3721
70	57	3990	4900	3249
55	59	3245	3025	3481
55	60	3300	3025	3600
70	69	4830	4900	4761
45	64	2880	2025	4096
65	71	4615	4225	5041
20	31	620	400	961
25	37	925	625	1369
35	37	1295	1225	1369
35	46	1610	1225	2116
10	26	260	100	676
20	34	680	400	1156
660	703	36955	38100	38197

X Listening	Y Speaking	XY	X2	Y2
75	75	5625	5625	5625
80	80	6400	6400	6400
70	70	4900	4900	4900
75	55	4125	5625	3025
55	55	3025	3025	3025
80	70	5600	6400	4900
70	45	3150	4900	2025
60	65	3900	3600	4225
25	20	500	625	400
30	25	750	900	625
35	35	1225	1225	1225
25	35	875	625	1225
25	10	250	625	100
25	20	500	625	400
730	660	40825	45100	38100

Correlation with the Pilot Test

X	Y	XY	X2	Y2
	Pilot Listening Test			
75	69	5175	5625	4761
80	77	6160	6400	5929
70	69	4830	4900	4761
75	63	4725	5625	3969
55	60	3300	3025	3600
80	70	5600	6400	4900
70	67	4690	4900	4489
60	75	4500	3600	5625
25	35	875	625	1225
30	44	1320	900	1936
35	43	1505	1225	1849
25	47	1175	625	2209
25	27	675	625	729
25	<u>42</u>	1050	625	<u>1764</u>
730	790	45580	45100	47746

X	Y	XY	X2	Y2
Pilot Test	Speaking			
69	75	5175	4761	5625
77	80	6160	5929	6400
69	70	4830	4761	4900
63	55	3465	3969	3025
60	55	3300	3600	3025
70	70	4900	4900	4900
67	45	3015	4489	2025
75	65	4875	5625	4225
35	20	700	1225	400
44	25	1100	1936	625
43	35	1505	1849	1225
47	35	1645	2209	1225
27	10	270	729	100
<u>42</u>	20	840	<u>1764</u>	400
790	660	41780	47746	38100

X	Y	XxY	X2	Y2
	Pilot			
Reading Test				
76	69	5244	5776	4761
83	77	6391	6889	5929
80	69	5520	6400	4761
67	63	4221	4489	3969
64	60	3840	4096	3600
67	70	4690	4489	4900
76	67	5092	6840	4489
83	75	6225	6889	5625
41	35	1435	1681	1225
53	44	2332	2809	1936
50	43	2150	2500	1849
53	47	2491	2809	2209
36	27	972	1296	729
56	42	2352	3136	1764
885	790	52955	60099	47746

X	Y	XY	X2	Y2
Pilot				
Test	Writing			
69	51	3519	4761	2601
77	61	4697	5929	3721
69	57	3933	4761	3249
63	59	3717	3969	3481
60	60	3600	3600	3600
70	69	4830	4900	4761
67	64	4288	4489	4096
75	71	5325	5625	5041
35	31	1085	1225	961
44	37	1628	1936	1369
43	37	1591	1849	1369
47	46	2162	2209	2116
27	26	702	729	676
42	34	1428	1764	1156
790	703	42505	47746	38197

387

X Pilot Test	Y Cloze	XY	X2	Y2
69	90	6210	4761	8100
77	100	7700	5929	10000
69	65	4485	4761	4225
63	50	3150	3969	2500
60	55	3300	3600	3025
70	75	5250	4900	5625
67	80	5360	4489	6400
75	85	6375	5625	7225
35	45	1575	1225	2025
44	70	3080	1936	4900
43	55	2365	1849	3025
47	65	3055	2209	4225
27	15	405	729	225
42	60	2520	1764	3600
790	910	54830	47746	65100

Results from the SLC English

Private School

SN	Score from the SLC English	Trial Test
1	76	69
2	80	77
3	70	69
4	60	63
5	60	60
6	75	70
7	72	67
8	80	75
9	30	35

Public Schools

SN	Score from the SLC English	Trial Test
10	46	44
11	34	43
12	52	47
13	30	27
14	44	42

Discrimination Index (D-Values)

$D = P\mu - P\iota$

Where   D = Discrimination value

$P\mu$ = Mean P-value of 27.5% of upper percentage [high achievers]

$P\iota$ = Mean P-value of 27.5% of lower percentage [low achievers]

Item	Pμ	-	Pι	=	D value
A	0.62	-	0.19	=	0.43
B	1	-	0.45	=	0.55
C	0.73	-	0.17	=	0.56
D	0.73	-	0.25	=	0.48
E	0.88	-	0.55	=	0.33
F	0.78	-	0.46	=	0.32
G	0.7	-	0.25	=	0.45
H	0.9	-	0.6	=	0.3
I	0.9	-	0.5	=	0.4
J	0.95	-	0.6	=	0.35
K	0.85	-	0.5	=	0.35
L	0.48	-	0.27	=	0.21
M	0.75	-	0.6	=	0.15
N	0.83	-	0.45	=	0.38
O	0.65	-	0.47	=	0.18
P	0.63	-	0.3	=	0.33
Q	0.65	-	0.35	=	0.3

Correlation Between The SLC and the Pilot Test

Examinee	Trial Test (X)	SLC (Y)	XY	X^2	Y^2
1	69	65	4485	4761	4225
2	77	80	6160	5929	6400
3	69	70	4830	4761	4900
4	63	60	3780	3969	3600
5	60	55	3300	3600	3025
6	70	75	5250	4900	5625
7	67	72	4824	4489	5184
8	75	70	5250	5625	4900
9	35	30	1050	1225	900
10	44	46	2024	1936	2116
11	43	34	1402	1849	1156
12	47	44	2068	2209	1936
13	27	30	810	729	900
14	42	40	1680	1764	1600
Total	790	771	46973	47746	46467

Correlation between the scores of Rater 1 and Rater 2

Examinee	Rater 1 (X)	Rater 2 (Y)	XY	X^2	Y^2
1	68	69	4692	4624	4761
2	78	75	5850	6084	5625
3	69	71	4899	4761	5041
4	63	63	3969	3969	3969
5	59	62	3658	3481	3844
6	72	70	5040	5184	4900
7	71	71	5041	5041	5041
8	77	73	5621	5929	5329
9	33	34	1122	1089	1156
10	44	46	2024	1936	2116
11	43	44	1892	1849	1936
12	47	49	2303	2209	2401
13	26	28	728	676	784
14	41	45	1846	1681	2025
	791	800	48691	48513	48928

Correlation Coefficient

Examinees	Score	-	Mean			(Score - Mean)^2
1	69	-	56	=	13	169
2	77	-	56	=	21	441
3	69	-	56	=	13	169
4	63	-	56	=	7	49
5	60	-	56	=	4	16
6	70	-	56	=	14	196
7	69	-	56	=	13	169
8	75	-	56	=	19	361
9	35	-	56	=	-21	441
10	44	-	56	=	-12	144
11	43	-	56	=	-13	169
12	47	-	56	=	-9	81
13	27	-	56	=	-29	841
14	42	-	56	=	-14	196
	790					3442

Statistical Procedures

Pearson Brown Correlation Coefficient Formula

$$N\Sigma XY - (\Sigma X)(\Sigma Y)$$

$$\overline{\sqrt{(\Sigma X2)} - (\Sigma X)2 \ X \ \sqrt{(\Sigma Y2)} - (\Sigma Y02}$$

Where - ΣXY is the sum of\ XY products

ΣX is the sum of X scores

ΣY is the sum of Y scores

$\Sigma X2$ is the sum of squared X scores

$\Sigma Y2$ is the sum of squared Y scores

N is the number of paired scores (examinees)

STANDARD DEVIATION

$$SD = \sqrt{\Sigma \frac{(X-M)^2}{N}}$$
$$= \sqrt{\frac{3442}{14}}$$
$$= \sqrt{246}$$
$$= 16$$

MEAN $\quad = \quad \dfrac{\Sigma X}{N} \qquad$ where $\dfrac{\Sigma X}{N} \quad$ = Total Score

= Number of Participants

Standard Deviation (SD) σ

$$SD \ = \sigma \ = \sqrt{\Sigma \frac{(X-M)^2}{N}}$$

Where $\qquad$ X = Total Score

M = Mean

N = Number of Examinees

Appendix F

Contact Summary Sheet

1. Contact Details

Contact type		Site -
- group note (#)		Contact Date -
- individual note (#)		Today's date -
- taped interview		Written by -
- taped response		
- interview		

2. Main issues and themes

3. Summary of questions and information received

Questions	Information
1. Do you think the test specification/structure of the said skill is workable?	
2. How do you think that the test specification for this skill could be improved?	
3. What in your mind should be done differently?	
4. Are the proposed allocations of time, marks and resources for this skill adequate?	
5. Does the document provide enough information on resources, equipment and constraints for effective implementation of the proposed test specification for this skill?	

4. Salient features (including quotes)

5. Reflective remarks

Appendix G

Codes and Coding of field notes, observations and archival materials

1. Description of Codes

A code is an abbreviated form of a word or a symbol applied to a segment of transcribed field notes, observation, and archival materials in order to analyse and classify the materials.

Codes, which are categories, derived from research questions hypotheses, key concepts or themes, are retrieval and organising divides that allow a researcher to spot pull out and then reorganise them into groups relating to a particular question hypotheses or theme being investigated (Miles and Huberman 1991 Miles, M.B and Huberman, AM, 1991, Qualitative Data Analysis, London, Sage publications)

2. Types of Codes

- Descriptive codes
    - Summarising notations which are expandable as needed.

- Interpretive codes
    - Illustrative issues/concepts/themes coming out of the intended meaning.

- Inferential Explanatory codes
    - Illustrate emergent issues or patterns in the course of analysis and interpretation

3. Creating Codes

- Start with a master codes
- Transform them into sub-codes as required

4. Revising Codes

Whether the codes are created or adapted they change and develop over the course research and analysis. The Researchers encounter new sometimes-unexpected ideas/issues then they originally expect. The original set of codes therefore, gets expanded, elaborated and revised.

5. Structuring Codes

The codes, adapted or created should have a governing structure, which is done by giving them a conceptual order in the importance of the research being carried of.

6. Defining Codes

A clear and operational definition of codes is essential in order for them to be read and interpreted in the same way. A code or a term used as a code may have multiple meanings interpreted differently in different contexts. For the purpose of using them precisely, a clear and simple definition of a code helps to share the same thing meaning among the readers.

7. Naming codes

A code should be given a name that is closes to its concept, pattern, question issue or theme. This is done to make it easy for the analyst or researcher/reader to get back quickly as possible without naming to translate what it represents.

Codes and their Definitions

		The Educational context
		The context, in which, education is provided in Nepal. It includes factors like general educational goals, policy, provisions or practices and organisational structure. It also includes the constraints of education and discrepancies that there are in the system. In addition, Nepali education as a culture in which examinations dominate all practices is also included in this theme.

1. Educational Context	EdCON
1.1 Attitude/Expectations to Education	EdCON/Expect
• Attitude toward education	EdCON/Expect/Attitude
• Expectations	EdCON/Expect/Expect
• Social Issues	EdCON/Expect/Social
• Political Issue	EdCON/Expect/Political
• Economic Issues	EdCON/Expect/Economic
1.2 Constraints	EdCON/Constraints
• No or poor training	EdCON/Constraints/training
• No support materials	EdCON/Constraints/materials
• Lack of sufficient resource	EdCON/Constraints/resource
• Poor management	EdCON/Constraints/Management
1.3 Discrepancies	EdCON/Discrepancies
• Theory and practice	EdCON/Discrepancies/TheoryPractice
• Public and private schools	EdCON/Discrepancies/PublicPrivate
• Regional variations	EdCON/Discrepancies/regional
• Teaching and testing	EdCON/Discrepancies/TeachTest
1.4 Exam Disease	EdCON/Exam
• Examination as an end	EdCON/Exam/end
• The SLC as a norm	EdCON/Exam/norm
• Backwash effects	EdCON/Exam/backwash
• Reform	EdCON/Exam/reform

			The ELT Context
2. ELT Context			
- Demographics	ELTCON		The context in which the English taught and learned as a foreign language. It includes demographic factors concerned with ELT as well as environmental factors. Attributes such as teachers, learners, trainers, teaching and assessment designs, teaching and learning materials etc. are included.
	ELTCON/DEMO		
- Schools	ELTCON/DEMO/Schools		
- Teachers	ELTCON/DEMO/Teachers		
- Trainers	ELTCON/DEMO/Trainers		
- Personnel	ELTCON/DEMO/Personnel		
- Environment	ELTCON/ENV		
- Physical	ELTCON/ENV/Physical		
- Equipment	ELTCON/ENV/Equipment		
- Materials	ELTCON/ENV/Materials		
3. Existing ELT Practices	EELT PRA		**The Existing ELT practices**
- Endorsement	EELT PRA/Endorse		The approaches schemes and activities relating to English Language Teaching. Inclusive are factors such as English language teachers, ELT classroom practices endorsement or non-endorsement of the existing ELT practices, constraints or limitations of ELT.
- Non-endorsement	EELT PRA/Non-Endorse		
- Teaching Learning	EELT PRA/Teach-Learn		
- Teachers	EELT PRA/Teachers		
- Learners	EELT PRA/Learners		
- Test Development process	EELT PRA/Test Develop		

399

4. The SLC Examination

- Administration
- Capability
- Personnel
- Resource

The School leaving Certificate (SLC) examination in Nepal, its administrative structure, expertise required, personnel, resources problems associated with the SLC in general.

SLC	
SLC/Admin	
SLC/Capability	
SLC/Personnel	
SLC/Resource	

5. Existing English Language Testing of SLC

- Endorsement
- Non-endorsement
- Parameters
  - o Expertise
  - o Resources
  - o Equipment
  - o Teaching and learning
  - o Environment
  - o ELT curriculum
  - o EL Testing purposes
  - o Stakeholders

The Existing EL Testing of the SLC

The current English language testing practices of the School Leaving Certificate (SLC) examination, endorsement or non-endorsement, parameters (constraints or problems influencing the examination schemes.)

EELAN Test	EELAN Test
EELAN Test/Endorse	
EELAN Test/Non-Endorse	
EELAN Test/Para	
EELAN Test/Para/Expertise	
EELAN Test/Para/Resource	
EELAN Test/Para/Equipment	
EELAN Test/Para/Teach-Learn	
EELAN Test/Para/Env.	
EELAN Test/Para/Curriculum	
EELAN Test/Para/Test-Purpose	
EELAN Test/Para/Stakeholder	

6. The Proposed Testing Package
- Endorsement   PTP
- Non – endorsement   PTP/Endorse
- Requirements   PTP/Non-Endorse
                                       PTP/Require
  - Expertise   PTP/Require/Expertise
  - Resources   PTP/Require/Resource
  - Equipment   PTP/Require/Equipment
  - Training   PTP/Require/Training
- Improvement   PTP/Improve
- Test Structure   PTP/Structure
- Test development process   PTP/TestDevPro
- Problems   PTP/Problem
- Issues   PTP/Issue
  - Specifications   PTP/Issue/Specification
  - Listening   PTP/Issue/Listening
  - Speaker   PTP/Issue/Speaking
  - Reading   PTP/Issue/Reading
  - Writing   PTP/Issue/Writing
  - Marking/Reporting   PTP/Issue/Mark-Report

The Proposed EL Testing Package

The package developed for the purpose of collecting data and feedback for the present research. Inclusive factors are – endorsement, non-endorsement, requirements for its implementation (expertise, training, resources, provisions, equipment, etc) test development process, test structure, problems and issues. Specifications of listening, speaking, and reading, writing, marking and reporting of performance outcomes are also included under this.

401

Appendix H

Target Language Skills at Tribhuvan University and Affiliated Colleges

Proficiency Certificate Level (PCL)
Higher Secondary Level (10+2)
Undergraduate Level

Listening Skill

Objectives

- Develop students' ability to receive spoken messages;
- Understanding and responding to sender of spoken messages;
- Listen as a prelude to using other skills in authentic communication tasks;
- Listen to lectures and other sources for academic and study purposes;

Sub-skills

- Recognising language elements through knowledge of phonological system and grammatical system,
- Understanding how a particular sentence relates to what else has been said and its function in the communication,
- Selecting what is relevant to the purpose of the communication,
- Listening to achieve a purpose
- Identifying the purpose
- Listening for specific purposes,
- Listening for main points
- Listening for the sequence of main events,
- Listening for the organisation of ideas
- Listening for lexical, syntactical, functional items,
- Listening for cultural interest,

Materials for Listening

- Descriptions
- Dialogues and monologues
- Stories
- Advertisements
- Instructions
- Directions

- Telephone conversations
- Interviews
- Lectures
- News broadcasts
- Talks
- Songs

Listening Activities

- Listen and discriminate
- Listen and draw
- Listen and follow directions
- Listen and place in order
- Listen and compare
- Listen and identify
- Listen and select
- Listen and take notes
- Listen and write
- Listen and complete a table
- Listen and complete a flow chart
- Listen and continue a dialogue
- Listen and discuss
- Listen and solve a problem

Speaking Skill

Objectives

- Communicate verbally in social and academic settings
- Using appropriate language form to express intended meaning
- Verbally interpreting formulated information, i.e. non-text materials
- Take part in a pair or group discussions

Activities

- Pair work
- Group work
- Interviews
- Information gap activities
- Prepared talks
- Report
- Description
- Explanation
- Conversation

- Role-plays and dramatisations
- Games and problem solving

Reading Skill

Objectives

- Reading for intellectual content to develop or add to intellectual skills nto manipulate ideas
- Reading for inferential content to get factual information with which to operate on environment
- Reading for pleasure or self improvement
- Reading for 'local' level comprehension
- Reading for 'higher' level comprehension
- Purposive reading
- Scanning to locate a known item
- Search reading for information relevant to the reader's purpose
- Skimming to get the general idea within the text
- Receptive reading to discover accurately what the writer wants to say
- Responsive writing to reflect upon what the writer has written
- Deducing the meaning and use of unknown words
- Understanding the meaning and implications of grammatical structures
- Recognising discourse markers
- Recognising and identifying the function of a sentence
- Understanding cohesive and coherence devices
- Understanding inter-, intra- and super-sentence relationships
- Reading and interpreting non-text materials charts, graphs, menus, bills, etc.)

Activities

- Mastery of the mechanics
- Silent reading
- Intensive reading
- Extensive reading
- Reading for facts
- Reading for meaning
- Reading jumbled sentences
- Recognising the link words
- Reading jumbled key points
- Doing cloze exercises
- Giving titles to paragraphs
- Seeing how a text is organised
- Speeded reading
- Reading for implied details
- Reading and changing text into non-test forms

Writing Skill

Objectives

Write cohesive paragraphs dialogues, letters and essays
Use English for functional, academic and communicative purposes
Use Library and reference materials

Activities

- Write a paragraph on a given topic
- Writing a conversation on a given situation
- Taking notes and summarising
- Writing formal and informal letters
- Describing people, place, objects, events, processes
- Using discourse markers, cohesion and cohesive devices
- Narrating stories and experiences
- Writing longer texts
- Writing essays on given situations/topics
- Write several paragraphs on imaginary situations
- Writing reviews and reports
- Writing resume and bibliography
- Using non-textual forms; transferring information from one form to another

Appendix I:

The Consulted Rating Scales

1. ESL Composition File

<table>
<tr><td colspan="3" align="center">**ESL COMPOSITION PROFILE**</td></tr>
<tr><td>**Aspect**</td><td>**Score**</td><td>**Criteria**</td></tr>
<tr>
<td rowspan="4">**Content**</td>
<td>30-27</td>
<td>**Excellent to Very Good**: knowledgeable, substantive, thorough development of thesis, relevant to assigned topic.</td>
</tr>
<tr>
<td>26-22</td>
<td>**Good to Average**: some knowledge of subject, adequate range, limited development of thesis, mostly relevant to topic, but lacks details.</td>
</tr>
<tr>
<td>21-17</td>
<td>**Fair to Poor**: Limited knowledge of subject, little substance, inadequate development of thesis</td>
</tr>
<tr>
<td>16-13</td>
<td>**Very Poor**: does not show knowledge of subject, non-substantive, not pertinent, OR not enough to evaluate</td>
</tr>
<tr>
<td rowspan="4">**Organisation**</td>
<td>20-18</td>
<td>**Excellent to Very Good:** fluent expression, ideas clearly stated/ supported, succinct, well organised, logical sequencing, cohesive</td>
</tr>
<tr>
<td>17-14</td>
<td rowspan="2">**Good to Average**: somewhat choppy, loosely organised but main ideas stand out, limited support, logical but incomplete sequencing</td>
</tr>
<tr>
<td>13-10</td>
</tr>
<tr>
<td>9-7</td>
<td>**Fair to Poor:** non-fluent, ideas confused or disconnected, lacks logical sequencing and development
Fair to Poor: does not communicate, no organisation. OR not enough to evaluate</td>
</tr>
<tr>
<td rowspan="4">**Vocabulary**</td>
<td>20-18</td>
<td>**Excellent to Very Good:** sophisticated range, effective word/ idiom choice and usage, word form mastery, appropriate register</td>
</tr>
<tr>
<td>17-14</td>
<td>**Good to Average**: adequate range, occasional errors of word/idiom form, choice, usage but meaning not obscured</td>
</tr>
<tr>
<td>13-10</td>
<td>**Fair to Poor:** limited range, frequent errors of word/idiom form, choice, usage but meaning confused or obscured</td>
</tr>
<tr>
<td>9-7</td>
<td>**Fair to Poor:** essentially translation, little knowledge of English vocabulary, idioms, word form OR not enough to evaluate</td>
</tr>
<tr>
<td rowspan="3">**Language Use**</td>
<td>25-22</td>
<td>**Excellent to Very Good:** effective, complex constructions, few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions</td>
</tr>
<tr>
<td>21-18</td>
<td>**Good to Average**: effective but simple constructions, minor problems in complex constructions, several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions but meaning seldom obscured</td>
</tr>
<tr>
<td>17-11</td>
<td>**Fair to Poor:** major problems in simple/complex constructions, frequent errors of negation, agreement, tense, number, word order/ function, articles, pronouns, prepositions and/or fragments, run-ons,</td>
</tr>
</table>

	10-5	deletions, meaning confused or obscured **Fair to Poor:** virtually no mastery of sentence construction rules, dominated by errors, does not communicate, OR not enough to evaluate
Mechanics	5 4 3 2	**Excellent to Very Good:** demonstrate mastery of conventions, few errors of spelling, pronunciation, punctuation, capitalisation, paragraphing **Good to Average:** occasional errors of spelling, punctuation, capitalisation, paragraphing but meaning not obscure **Fair to Poor:** frequent errors of spelling, punctuation, capitalisation, paragraphing, poor hand writing, meaning confused but not obscured **Fair to Poor:** no mastery of conventions, dominated by errors of spelling, punctuation, capitalisation, paragraphing, handwriting illegible, OR not enough to evaluate

(From Jacobs, H.L. et al (1991). Testing ESL Composition: A Practical Approach. Rowley, MA. Newbury Hose Publishing House)

2. Rating Scale for Evaluating Fluency and Accuracy in Written Language (Analytical Scales)

	Fluency	Accuracy
0	Incomprehensible	Entirely Inaccurate
1	Vague statement of authors main intent (not necessarily explicit) Can deduce author's intent No supporting information No expansion of ideas No awareness of organisational features (opening; body; wind-up) Lacking Cohesion	Poor grammar and vocabulary Strongly interfering with comprehensibility Elementary
2	Main intent expressed Supporting ideas expressed No expansion of ideas Ideas not always clearly expressed Lacking some organisational features Mostly lacking cohesion	Frequent consistent errors yet comprehensible Basic structure and simple vocabulary
3	Main intent and supporting ideas expressed clearly Ideas generally expressed clearly Occasional lack of cohesion Exhibits proper organisational elements Limited expression of ideas	Consistent errors Accurate use of varied/richer vocabulary Longer sentence structure
4	Main intent and supporting ideas expressed clearly Expansion of ideas Proper organisational elements Sporadic awkwardness of cohesion	Few sporadic mistakes More sophisticated, complex sentence structures Idiomatic expressions Accurate use of sophisticated vocabulary
5	Native-like Main intent and supporting ideas expressed clearly Expansion of ideas Clarity of expression Proper organisational elements Entirely cohesive	Near-native accuracy

Developed as a part of Bagrut Alternative Assessment Project 1984 (Shahomy 1985)

3. ACTFL /ETS Scale: Sample Page

		Language Functions	Accuracy/ Description of Learner Language	Text/ Content
5	Superior			
4+ 4				
3+ 3				
2+	Advanced Plus			
2	Advanced			
1+	Intermediate High			
1	Intermediate Middle			
	Intermediate Low			
0+	Novice Middle			
	Novice Low			
	0			

Details Available at WWW. ACTFL.org

4. ASLPR: Sample Page

Proficiency levels	Skill-wise General Description	Example of Specific Tasks	Comments
0 Zero or Initial proficiency			
1 – Elementary Proficiency			
1+ Minimum Survival Proficiency			
Survival Proficiency			
2 Minimum Social Proficiency			
3 Minimum Vocational Proficiency			
4 Vocational Proficiency			
5 Native-like Proficiency			

Details Available at WWW. Griffith.edu au

APPENDIX J

PROPOSAL
FOR
A NEW APPROACH TO IMPLEMENTING A REFORM
OF THE School Leaving Certificate
IN NEPAL

The proposed implementation plan, as the diagram below demonstrates, is a progressive one in that the preceding phases are continually revised which then become a basis for the implementation of succeeding phases:

Phased implementation plan

Phase 1: Initiation of the Implementation Process

⇩

Phase 2: Implementation of Part I of the New Testing Package Tests Reading and Writing

⇩

Phase 3: Implementation of Part II of the New Test Package Tests of Listening and Speaking

⇩

Phase 4: Continuation of Implementation of Parts I & II of the New Package

⇩

| Phase 5: Review of the Implementation |
|---|
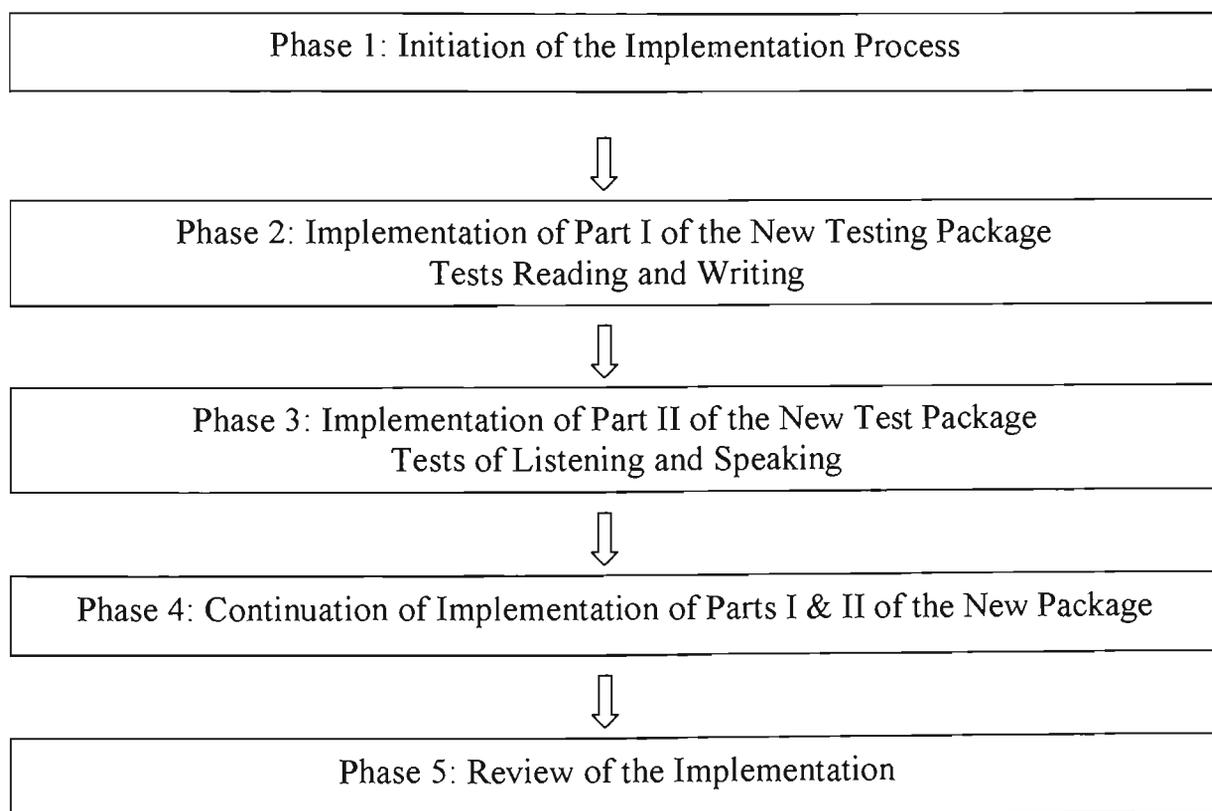
Figure: Phased implementation plan

A brief description of the phases follows:

1 Phase One

Phase one is basically making preparation and creating necessary conditions for the next phases. Five areas of provisions are addressed in this phase. They are coordination, relevance, readiness, resource and organisational structure.

1.1 Consensus/ Coordination

Before any changes are introduced and institutionalised, it is an important task of the government to receive public support to legitimise the reform. This may be difficult in a country like Nepal where education, as it was argued in the initial chapters, is highly politicised and is often used as a political weapon (as explained in Chapter 1). Nonetheless, I would like to suggest the following four strategies to gain public confidence in the favour of the reform. Consensus or legitimacy of the reform may be established by – procedure, expertise, action, and participation

Consensus on the reform is obtained when the government authorises the procedures used to bring about changes under which the policy for reform is formulated through a careful planning process. The task of legitimising becomes relatively convenient for the government when it is backed by expertise or research. Research findings give confidence to donors as well as the government to decide on a reform and reform processes. Governmental actions in the form of commitments, better resources, qualified practitioners and other provisions can also bring support for the reform. However, the most popular strategy to bring support is the participatory strategy in which all stakeholders are involved in the decision making process (Fullan 1997). For Nepal, I recommend a combination of all these four strategies.

Coordinating the concerned individuals and agencies is going to be one of the most important factors to ensure the success of the reform project. This should, therefore, be an important task of the examination administration body (i.e., OCE).

1.2 Relevance

In order for the general public and the stakeholders to support the reform, it is necessary that they understand the need for such a reform, what it offers to the ELT in general, and to individuals and institutions in particular. For example, they need to know how the reform is compatible with the changing demands of the ELT. An awareness campaign involving all stakeholders and all available means, such as media, conferences, workshops, seminars, consultative meetings and so forth, needs to be organised at all levels but more importantly at the grass-root levels.

1.3 Readiness

This is mainly the capability building of the people and institutions concerned for the development, implementation, monitoring and evaluation of the reform project. It requires training and support mechanism and preparing resource materials for the reform in phases. This can be achieved by creating a training and support cell within the Test Development Division with the responsibility of training and supporting the people involved.

1.4 Resources

Developing necessary infrastructure and generating required funds is the most important task for the entire project. Funds will be required for paying the staff of the newly created administrative division and research division, for training people involved, for purchasing equipment such as computer software, cassette players, and for producing support materials. As the reform package is implemented in phases, training of people, accumulation of equipment and developing support materials can also take place in phases. This would then mean that funds will be needed in instalments, rather than all at once. Funds needed for implementing different phases should be worked out, and Secondary Education Development Project (SEDP) and different donors can be mobilised to support the different stages of the project.

1.5 Organisational Restructuring of the OCE

In order to administer and manage the SLC more effectively and efficiently and create a mechanism for communication among its stakeholders, I would like to recommend that an organisational provision be made in the current structure of the Department of Education and bring the OCE, as is in the case of the Philippines, for example, under its administration with a new name of 'National Education Testing and Research Centre' (NETRC). The NETRC will have three divisions, Test Development Division, Test Administration Division and Research Division. The Division of Test Development coordinates people and agencies involved in the teaching and testing of the SLC English and undertakes all activities to help the concerned agencies with conceptualisation and development of the tests. The Test Administration Division coordinates all local level administrative units such as the DEOs and SEDUs with a purpose of empowering them for the reform. The Research Division of the centre conducts research to evaluate the effectiveness of the tests and other programmes of the centre. In this way, communication from the grass roots level practitioners to the policy level workers can be maintained.

I believe that the creation of this provision may not require a lot of monetary resources because it only requires reorganising the existing provisions. The new structure of OCE will look like the one given below:

MOEC

Department of Education

NETRC

Test Development   Test Administration   Research
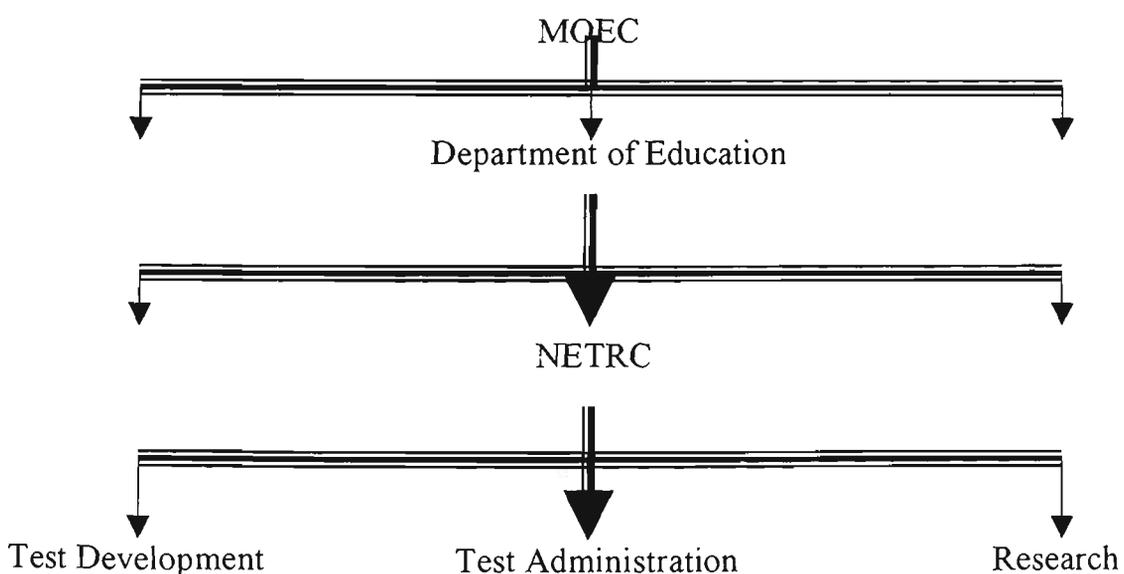
Figure: Proposed Structure of the OCE

414

Under the proposed provisions, development of a test should undergo the following steps (Umar 1997):

- test items are written and reviewed by specially trained people;

- test items pass through two levels of review – theoretical review done at the centre, and practical done in the test conditions;

- items go through the process of empirical validation to examine the extent to which each item provides valid information about what it is supposed to do;

- items found valid are calibrated into national, common scale of difficulty level. Validation and calibration of items are established;

- items calibrated are stored in an item bank from where they are selected for the national examinations.

2. Phase Two: Implementation of the Reform Part I

Part I of the package is reading and writing tests. Revising reading and writing tests is recommended in the first part of the implementation because –

- it would not upset drastically the current provisions as all involved would be familiar with some form of testing reading and writing;

- a short training package for reading and writing can be formulated and implemented;

- It would not require any specific equipment to administer the tests;

- it would allow time for the preparation for the second part of the project, i.e. listening and speaking;

- it would allow accumulation and management of resources more efficiently as all resources will not be needed at once. Resource will only be needed for developing and piloting the new tests and for training the concerned personnel.

3. Phase Three: Implementation of the Reform Part II

The Implementation will be in its third year by the time this phase of the package is implemented. The implementation of this part of the reform takes place, as the figure above shows, in a similar fashion to that of part I except that this part of the reform will require longer training programmes and greater degree of resources as the tests are new. As these tests will require playing and recording devices, audio cassette recorders will be required for all test centres.

4. Phase Four: The Continuation

Phase four of the new test package is the continuation of the project in its fourth year. In this phase, both parts – Part I (reading and writing) and Part II (listening and speaking) are implemented together. Necessary equipment and materials for this phase will have been acquired through second and third phases. The process should make sure that there is required level of coordination and necessary support provisions. It should also see that there is enough interest, expertise and resources for the reform.

5 Phase Five: The Review

While the reform package continues in its fifth year, an independent review of the entire project should be launched. The independent evaluation team should consist of experts, teachers, institutional heads and representatives from the governmental agencies. The review should focus on the outcome of the new test package, people's interest and concerns, the preparation, monitoring and support mechanism, expertise and equipment, training staff development, and resources. The outcome of the review should form the basis for further modification in the reform package. It should be noted that reform is a journey, not an end in itself. Total reform cannot be achieved in one attempt. The reform, in its revised and modified form should continue. The review should also address the areas of concerns, the problems arising from the implementation and how they should be addressed. The review should also suggest ways of sustaining and institutionalising the reform.

Samples of Recent SLC English Tests

2058 (2002)
Compulsory English
Reading and Writing

English (XR-401A1)

Read the passage and answer the following questions.   10

Parsa Wildlife reserve was established in 1984 with an area of 499 sq. km. It occupies parts of Chitwan, Makwanpur, Parsa and Bara districts in central Nepal. The reserve headquarter is at Adibar on the Hetauda-Birgunj highway (22 km. South of Hetauda and 20 km. north of Birgunj) the reserve supports a good population of resident wild elephant, tiger, leopards, sloth bears, blue bulls and wild dogs other animals are the sambar chital, hog deer, barking deer, langur, stripped hyena, palm civet and jungle cat.

There are nearly 300 species of birds in the reserve. The giant Hornbill, one of the endangered species to be found in certain forest patches. Peafowl, red jungle fowl, flycatchers, and woodpeckers are a few of the other common birds found in the reserve.

A. Find the best answer and write them in your answer book.
   i) The word, 'reserve' in the first line mean
      a) an animal park   b) a zoo   c) military force
   ii) The word, 'nearly' in the tenth line means
      a) closely   b) maximally   c) about
   iii) The word, 'species' in the tenth line can be replaced with the
        word : a) common   b) kinds   c) specific
   iv) The word, 'endangered' in the eleventh line means
      a) may soon no longer exist   b) in danger   c) dangerous

B. Decide whether the statements are true or false
   a) Parsa Wildlife Reserve was established in late 80's
   b) The Reserve's headquarter is at Hetauda
   c) The Reserve extends in more than one district.
   d) The Giant Hornbill is not a common bird in the reserve.

C. Answer the following questions
   a) Where is Adibar? What's its importance?
   b) Mention 5 common and 3 uncommon animals in the reserve.
   c) How many kinds of birds are found in the reserve? Which one of them is/are rare?

Read the news story and answer the following questions.

Striking Protest at TIA

Kathmandu, July 26
The departure lounge at the Tribhuvan International Airport which is the only international airport in the country, was a sight to see today.

417

Travellers standing in the check-in queue were covering their noses with hankies and tissues. Their relatives and friends who came to see them were no longer there. The departure lounge was littered all over with rubbish thrown by the staff as a protest against losing their jobs.

Fill in the blanks:

On ...... Tribhuvan International airport was covered with .... The travellers were covering their faces with their hankies. The .... thrown by the ..... Staff. They did it as a ....

Read the poem and answer the following questions:

What are they thinking ?
The circus elephants,
As they tramp round the ring each night ?
Do they wish that
They were far away
In the forest's leafy light ?
Where they'd roam at will
And could eat their fill
Far from human sight,
As the gentle giants
Perform their tricks
The children stare with delight.
But is it fair ?
Should they be there ?
Is keeping them captive right ?

A. From the poem find the words which mean :
   a) walk with heavy steps
   b) want something impossible
   c) walk aimlessly
   d) look with unmoving eyes
   e) Skillful acts
   f) pleasure
   g) just
   h) prisoner

B. Choose the best answer. Then write them in your answer book.
   i) In this poem, the poet talks about
      a) the happy elephants   b) the wild elephants
      c) the circus elephants
   ii) The elephants probably wish to
      a) stay in the ring   b) go to the jungle
      c) play with children
   iii) The poet is suggesting that keeping elephants in a circus is
      a) good b) not good c) neither good nor bad
   iv) The poet is asking the question in the last line of the poem to
      a) the elephants b) himself c) man
   v) The best title for the poem is
      a) Elephants and Children b) The Circus Elephant
      c) The Elephants think

C. Answer the following questions :
   a) Who are happy and why ?
   b) What might the circus elephants be thinking ?

34

c) Why is it not fair to keep elephants in a circus ?

Read the story and answer the following questions : 15

John Keats, the eldest child of his parents, was born in 1795 in London. When he was only eight, his father died, and before he was fifteen, his mother died of tuberculosis, leaving him in charge of the orphan family. In 1818, his brother, George left for America. The same year another of his brother, Tom died of tuberculosis foreshadowing Keats's own death from tuberculosis after three years later. He was alone, bewildered, financially pressed, and deserted by his ladylove. In addition to all these tragedies, his poems were unfairly criticized by the critics. Yet he continued to hold before himself the ideals of sincerity, open mindedness, and generosity.

His poetic craftsmanship and artistic talent are best reflected in his odes, On Grecian Urn, To Autumn, To a Nightingale, and his long classical poem Endymion and the unfinished Hyperion. His poetry was bitterly and unjustly condemned by the critics of his time. Nevertheless, Keats as a genuine artist went on writing, and the fame denied in his life was lavishly granted after his death.

Keats is the classic example of the gifted poet who died young. Before his death at the age of twenty-five, he succeeded in writing poems that, in talent and imaginative power and in mastery of phrase, ranks among the highest achievements in English poetry. It is clear that his talent had only started to unfold which his untimely death closed forever. Without any relatives, deserted by his love, unjustly criticized by the critics, and seized in the fatal grip of tuberculosis, disheartened and lonely, the best talented poet, Keats died on 23rd February 1821, and buried in Rome. Had he lived a bit longer he would have excelled even Shakespeare.

A. Match the words in column 'I' with their meanings in column 'II': 3

I	II
	(i) left and unsupported
(a) orphan	(ii) in great extent
(b) bewildered	(iii) surprised
(c) deserted	(iv) resulting in death
(d) condemned	(v) a child whose parents are both dead
(e) lavishly	(vi) totally confused
(f) fatal	(vii) serious
	(viii) disapproved

B. Find whether the statements are true or false : 3
   (a) Keats died of tuberculosis.
   (b) He was very much praised for his poetry in his life and after his death.
   (c) Keats lived a happy life.

C. Make a table in your answer book and complet it with correct information about Keats' life : 3

35

Time	Place	Persons
He was born ?	Where he was born	
His mother died ?	Where did his brother go ?	
He died ?	Where was he buried ?	

D. Answer the following questions :
   (a) Who was 'in charge of the orphan family', and why ?
   (b) What troubles and sufferings Keats faced in his life ?
   (c) Why did the writer say, 'his talent had only started to unfold'
       which his untimely death closed forever'?

5. Where are you from ? - a village or a town ? Do you like to live in a village or a town ? Pour your own opinions and feelings in about 125 words. Consider the following clues given in the box ?

   Introduction - why people tend to settle in villages / towns - advantages and disadvantages of village / town life - conclusion

6. You and your friend, Gautam Yadava, were working in Gyan Kunja Secondary school in Sanothimi, Bhaktapur. Unfortunately, he was killed in a motorcycle accident in Bagbazar, Kathmandu. Write a formal notice of condolence to be published in a newspaper, using the clues given in box

   Gautam Yadav - 30 years old - science teacher - died on 1st April, 2002 helpful - hard working and sociable

7. Complete the following dialogue with the appropriate sentences in the box given below :

   Tourist : Good morning ....
   Receptionist : Yes, sir. Double or single ?
   Tourist : I want one double on the ground floor.
   Receptionist : I'm soory, sir ....
   Tourist : Hum ! ....
       What about on the first floor ?
   Receptionist : Yeas, sir, I've a double room on the first floor. ....
   Tourist : Can I know something about hte facilities you provide here ?
   Receptionist : ....
   Tourist : Well, I want to book a room for three days
   Receptionist : ....

   a) There's no double room vacant at the moment on the ground floor
   b) Of course ! Here's a brochure giving all information about the hotel.
   c) Good morning. Can I help you, sir?
   d) O.K. Thank you, sir.
   e) Have you any room vacant ?
   f) But I do need a double one
   g) Do I have to pay only in advance ?
   h) Could I book one of the rooms for you in advance, sir?

36

420

8. Rewrite the following sentences selecting the correct words from the brackets. 11
  i. Did Sunitra .... the exam ? (take, took, taken)
  ii. The cost of apples .... risen at present. (has, have, had)
  iii. Each of the girls .... here on time yesterday. (was, is, were)
  iv. Niru never gets angry ......? (does she, doesn't she, did she)
  v. Bhim and Nahakul will finish the work, ...? (won't he, won't they, don't they)
  vi. "What's wrong ..... you?" She asked me. (in, for, with)
  vii. Who is .... best player of your team ? (a, an, the)
  viii. Sushila met him .... the way to work. (in, on, at)
  ix. New Zealand is .... island. (a, an, the)
  x. Man Bahadur drives a car but he .... a bus.
                (doesn't drive, doesn't drives, don't drive)
  xi. If Amita .... the exam, she will not come to school
                 (fail, fails, failed)
  xii. If she invited, I .... her wedding
          (will attend, will have attended, would attend)
  xiii. His shoes are dirty, he will get the cobbler .... them.
                (to polish, polish, polished)
  xiv. Samjhana is very cruel. She always makes her children .... the hard work. (to do, do, did)
  xv. The school was closed .... strike. (because, so that, because of)
  xvi. You will be punished for .... you have done.
                (what, when, while)
  xvii. Priyanka said to her friend, " ....?" She replied that she was from Dadeldhura.
      (Where she is from, Where she was from, Where are you from)
  xviii. Shibani said that she .... Pokhara.
            (has visited, had visited, was visited)
  xix. Meal is being prepared by my sister. That is to say ....
    (My sister is preparing meal, My sister has been preparing meal, My sister prepares meal)
  xx. Someone told the students a story. It means ....
    (The students are told a story, The students have been told a story, The students were told a story)
  xxi. Rupak .... yet. (hasn't arrived, didn't arrive, won't arrive)
  xxii. When I saw them, they ....
      (are dancing, have been dancing, were dancing)

★ ★ ★ ★ ★

English (XR-401A2)

1. Read the passage and answers the following question. 10

A panda looks like a bear, but it has its own distinctive black-and-white colouring, short limbs, small ears and peculiarly shaped head and face. It is a herbivorous animal and eat only one kind of food-bamboo. This

37

421

2057 (2001)

Compulsory English
Reading and Writing
(Written Test)

Full Marks : 80
Pass Marks : 26

Time : 2½ hours

1. Read the following passage carefully and do the exercises given below : 10

Computers have become a part of modern life. Today's school children carry around calculating power which would have filled a large room 40 years ago. Computer scientists are now working on the next generation of computers, ones which will have true intelligence.

What are the parts of a computer ? The word 'hardware' refers to the physical parts of a computer. The main parts are the monitor, the key board, the mouse and the CPU or Central Processing Unit, which contains the hard disk.

The monitor is also known as the screen or VDU (Visual Display Unit). It looks like a television screen and displays what we do on the computer. It comes in different sizes.

The actual computer, i.e. the part that performs tasks we ask it to do is in a unit holding the CPU, memory, hard and floppy disks. It comes in a rectangular box.

CPU stands for Central Processing Unit. This could be called the brain of a computer. It performs calculation and manipulates data. The keyboard functions like a typewriter. The computer is given commands by typing them on the keyboard. Information is then displayed on the monitor.

The mouse is a hand-held pointing device. By moving the mouse around on a flat surface, a pointer or arrow on the monitor will make corresponding movements. By clicking, holding and dragging with the mouse various things can be done on the monitor. The hard disk is a fixed device for permanent storage which stays within the computer. A hard disk is capable of holding tremendous amounts of information, yet it is only a few inches in width and length. A computer also has a floppy disk drive. It is a slot for the insertion of portable disks also known as floppy disks.

A. Find the words in the passage which have the opposite meaning to the words given below : 3

lost, hides, ancient, taken, temporary, disable

B. Answer the following questions : 5

i. What does the monitor look like ?
ii. What do you mean by 'hardware' ?

---

xi. When I saw them, they .... (are dancing, have been dancing, were dancing)

xii. Krishna .... yet. (hasn't arrived, didn't arrive, won't arrive)

xiii. Someone told the students a story, the students have been told a story. It means.........
(the students are told a story, the students have been told a story, the students were told a story)

xiv. Food is being prepared by my sister. That is to say ....
(my sister is preparing food, my sister has been preparing food, my sister prepares food)

xv. Dipa said that she .... Pokhara. (has visited, had visited, was visited)

xvi. Rita said to her friend, "................. ?" She replied that she was from Doti.
(Where she is from, Where are you from, Where she was from,)

xvii. You will be punished for .... you have done. (what, when, while)

xviii. The school was closed .... strike. (because, so that, because of)

xix. She is very cruel. She always makes her children .... the hard work. (to do, do, does)

xx. My shoes are dirty. I will get the cobbler .... them (to polish, polish, to be polished, polished)

xxi. If she invited me, I .... her wedding. (will attend, attend, will have attended, would attend)

xxii. If Mira .... the exam, she will not come to school. (fail, fails, failed, has failed)

— 0 —

422

iii. Why is the CPU called the brain of a computer ?

iv. What is the use of a floppy disk drive ?

v. How does the keyboard function ?

C. Read the following statement and write T for true and F for false statements :   2

   a) CPU is a hard-held pointing device.

   b) Information is displayed on the monitor.

   c) 'Hard Disk' is contained in a keyboard.

   d) The mouse is moved around on a flat surface.

2. Read the following poem and do the exercises that follow :   5

What is this life if, full of care
We have no time to stand and stare.
No time to stand beneath the boughs
And stare as long as sheep or cows.

No time to see, when woods we pass
Where squirrels hide their nuts in grass.

No time to see, in broad daylight,
Streams full of stars, like skies at night.

A. From the poem find the words that are similar in meaning to the following words :   2

   under, a small river, forest, large

B. Answer the following questions :   3

   i. What is the main subject of the poem ?

   ii. Why do people have no time to enjoy nature ?

   iii. What are the streams compared in the poem with ?

3. Read the following text and do as instructed below it :   15

The total number of cars in Britain now exceeds 11,000,000. Traffic goes on increasing all the time and the roads of most big cities are almost permanently blocked by a slow moving procession of metal cages. Complicated systems of one-way streets and the universal use of traffic lights have not provided a real solution of the problem. As far as the motorist is concerned, driving in crowded towns is far from being a pleasure.

Some time ago, a friend of mine who works in a part of the city I do not know very well, invited me to call on him. It took me hours to get there and I drove round and round looking for a convenient spot to park my car. At last I found a small space in a back street. As I was already three quarters of an hour late, I parked my car quickly and hurried off on foot. Making my way rapidly along the street, I could not help reflecting that, nowadays, it is much easier to walk than to drive.

At noon, just as I was leaving my friend's office, it suddenly struck me that I had no idea where I had parked my car. I could

hardly go up to a policeman and tell him that I had lost a small green car somewhere! I would simply have to look for it myself. Walking down street after street, I examined each car closely and was greatly relieved to see a small green car just behind an old cart. But how disappointed I was to discover that though the car was exactly like my own, it belonged to someone else! Feeling quite tired now, I gave up the search and went off for lunch. Some time later, I left the restaurant and walked idly down the street. Turning the corner, I nearly jumped for joy: my car was right in front of me and there was no mistaking it this time. I could not help smiling as I drew near. Pasted on the wind-screen was a little ticket which informed me that the car had been visited by a policeman in my absence. To add it all, I was now guilty of a traffic offence.

A. Answer the following questions: 8
(i) Why is driving in crowded towns not a pleasure?
(ii) What did the writer do after he arrived near his friend's office?
(iii) Why was he feeling tired?
(iv) What did the little ticket pasted on the windscreen say?

B. Fill in the appropriate words: 3
i. The writer drove round and round to find a .... place to park his car.
ii. He was late by .... minutes to get to his friend.
iii. He was guilty because he offended the .... rules.

C. Match the words in column A with the words in column B: 4

A	B
universal	difficult
rapidly	international
drew near	quickly
complicated	came close

4. Read the following letter and answer the questions given below: 15

3/95 Fareway Street
WA 6539 Perth
Australia
Feb. 25, 2001

The Manager,
Brown's Sport Store,
164 Victoria Parade,
Sydney, Australia.
Dear Sir/Madam,
Recently, I turned fifteen and I am currently seeking part-time work. This year I am completing class 10 at St. Catherine's college. The course includes English, Japanese, science, mathematics, computer studies, physical education, history and geography.

Eventually, I wish to take up a career related to physical education, and I am therefore interested in part-time work involving equipment in this field. Should there be a vacancy for part-time work in your store, I would appreciate the opportunity of an interview at a time convenient to you? I can be contacted at the above address or by telephone on 3569872.
Yours faithfully,
Peter Smith

A. Write the words from the letter which are closest in meaning to the following: 3
i. profession ii. place to be filled iii. looking for

B. Answer the following questions: 6
i. Who is this letter addressed to?
ii. What is the qualification of the candidate?
iii. How old is he?
iv. How can he be contacted?
v. What languages are included in the course?
vi. Where is this letter sent from?

C. Fill in the gaps with appropriate words from the above letter: 6
i. The letter is a request asking for a .... job.
ii. The candidate wants to attend the interview at a .... time the manager.
iii. He is going to complete class 10 from .... college.
iv. He has special interest in .... education.
v. The letter is written on ....
vi. The manager is a responsible person in ....

5. Suppose you are Mina Wanta from Dadeldhura. Read the given advertisement and write an application for the post: 12

Yeti Parcel Service
3/60 Mahendrapool, Pokhara
Wanted Immediately

YPS, a leading parcel service in the region, is looking for an Office Secretary.
Requirements:
- B.Com. or equivalent
- Strong computer skills
- Excellent written & spoken English
- 25-35 years
- Nepali citizen, preferably a woman

Application along with PP sized photograph, photocopies of certificates and a complete bio-data should reach the above address not later than March 25, 2001.

---

6. Write a readable story with the help of the guidelines given below :

A king disappointed - his people lazy - to teach them a lesson he had a big stone put in the middle of the road one night - next day a big stone put in the middle of the road one night - next day the merchants pass and go round it - an officer driving his car did the same - a young soldier came cycling, did the same - all cursed the government for not removing it - then the king had the stone removed - many people watched it - under it was an iron box, marked, "For the man who moves away the stone" - inside was a purse full of money - the people were ashamed - moral.

7. Write out this conversation in full using the appropriate sentences in the box given below it :

Sharan : This is a nice house-warming party. There are a lot of people and they are all dancing. ....

Shanti : I don't want to dance, ...

Sharan : Bimala will give you some tablets for your headache.

Shanti : No, thanks

Sharan : ....

Shanti : Oh, I don't know.

---

4 I went to a party yesterday evening

2 I have a terrible headache.

5 What's the matter with your, Shanti ?

3 Then you will be better.

1 Do you want to dance, Shanti ?

4 I don't want to be better.

---

8. Rewrite the following sentences selecting the correct word from the brackets :

i. It is .... one rupee note. (a, an, the)

ii. That is .... book I like most. (a, an, the)

iii. She is .... the sari today. (at, in, by)

iv. Madan married .... an American lady. (to, with, for)

v. No one solved the problem .... ? (did they, didn't they, did he)

vi. I'm a nurse .... ? (amn't I, aren't I, Don't I)

vii. Look the deer (is, are, were) grazing.

viii. Neither Mohan nor they (has, have, had) understood my question till now.

ix. I said I need a shirt when she asked me, "...................?" (What you need, What do you need, What do I need)

x. Mohan drives a car but he .... a bus. (doesn't drive, doesn't drives, don't drive)

(m) What do poachers kill tigers for ?

(The Disappearing Tiger)

(n) Why is it important for development that people are responsive to change ?   *(An Interview with Prof. Sharma)*

2. Read the following passages and answer the questions given below :

(a) It is still winter, but on Sunday the weather was fine. The boys decided to go on a trip to the seaside. The sun was bright and warm, but we did not go into the water. It was too cold. Instead, we walked along the shore and looked at the ships. In the evening we returned home. Then a strong wind began to blow and it began to rain. So we sat round a warm fire. We did not mind about the weather at all.

   (i) Where did the boys decide to go ?
   (ii) Why didn't they go to the water ?
   (iii) What did they do instead ?
   (iv) What did it begin when they returned home ?

8

(b) Two thieves had stopped a train and stolen mail-bags containing a lot of money. This news was broadcast on the radio. A motorist had heard it and when he saw two men along the road carrying heavy bags, he suspected them and informed the police at once. The police soon arrived on the scene. They questioned both men but neither could speak English. They kept shouting at the police and simply pointed at their bags. The police opened them at once and then realised that they had made a mistake. The men were French onion-sellers and their bags were full of onions.

   (i) What had the thieves done ?
   (ii) Why did the motorist inform the police ?
   (iii) Why couldn't both men speak English ?
   (iv) What proved that the police had made a mistake ?

12

3. Pretend that you are an angry farmer. Write a newspaper article of about 150 words explaining why the tiger must be killed off.

10

ENGLISH

As far as practicable candidates are required to give their answers in their own words indicative of individual style and personal expression. Credit will be given to such answers, not to memory work, rote-learning or copying from the text book.

Time—3 hours

Full Marks—100
Pass Marks— 32

1. Attempt any *ten* questions :

$3 \times 10 = 30$

(a) Why and how was Aesop killed ?   *(The Fables of Aesop)*

(b) Why are paper notes more convenient than metal coins ?   *(The History of Money)*

(c) Why did the beggar refuse to get off the horse ?   *(The Prince and the Judge—I)*

(d) How does a steam engine work ?   *(Transport—I)*

(e) What national anthems were sung on the final day of the Olympic Games in Munich ?   *(The Olympic Games in Munich)*

(f) What will Gambhir Man do if he does not see Jim at the airport ?   *(A Letter to a Pen Friend)*

(g) How did the doctor treat Miss Rai's eyes ?   *(A Visit to the Doctor)*

(h) How does Gopal Bahadur avail of his every visit to Kathmandu ?   *(The Storyteller)*

(i) What did the enquiry find out about the "Mary Celeste" ?   *(An Unsolved Mystery)*

(j) What evidence is there in the passage that Kiran's grandfather is a useful member of his community ?   *(A Visit to My Grandfather)*

(k) How is R. N. A. C. doing the nation a great service ?   *(R. N. A. C.)*

(l) How did Bal Bahadur succeed in protecting his wife from the tiger ?   *(An Encounter with a Tiger)*

Contd.

426

Or

The teacher always makes............

3. Write a paragraph on "Tribhuvan Highway" with the help of the following notes: **3**

Birgung—Kathmandu, —179 Km., —Terai, Sewalik Hills, —Mahabharat range—highest point 2,500 m—1953—1956—blacktopped—King Tribhuvan—India.

4. Choose the correct alternative: (*any four*) **4**

(a) The word "go" has the same vowel sound as the word............
(*do, sew, how*)

(b) The word "they" has the same initial consonant sound as the word............
(*day, that, thief*)

(c) In the word "mathematics" the stress falls on the............ syllable.
(*first, second, third*)

(d) "Are you having orange squash or tea?" has normally a............tone. (*falling and falling, rising and rising, rising and falling, falling and rising*)

(e) In the word "increase" the stress falls on the............ syllable when it is used as noun. (*first, second*)

(f) The pair "............" has the same pronunciation.
(*fool, full; hut, hot; sit, seat; sow, sun*)

Group "A"

Attempt *any four*:

6 × 4 = 24

5. (i) Put the verbs in brackets into the correct tense: (*any four*) **4**

(a) He (walk) in the park when I met him.

(b) I wish I (know) the answer to that question.

(c) He (learn) English since he was a child.

(d) He (not go) there unless you follow him.

(e) My brother (come) to visit me next week.

(f) Medical scientists (eradicate) malaria by 1994.

(ii) Complete the following sentences in such a way that it conveys the same meaning as contained in the first statement: **2**

(a) The boys did not want to play the match. But the teacher told them they had to play.

The teacher made............

6. (i) Rewrite the following sentences, filling the blanks with appropriate prepositions, choosing from those given within the brackets: (*any three*) **3**

(a) He has been drinking............half past six.

(b) It took them an hour to put............the fire.

(c) The pencil rolled............the table.

(d) You must work hard............success.

(*for, since, down, over, off, of, out, to*)

(ii) Add "a", "an" or "the", where necessary: (*any three*) **3**

(a) The clouds over............sea are lovely today.

(b) A horse is............animal.

(c) Copper is............useful metal.

(d) ....page of............book is torn.

7. Rewrite the following passage into Indirect Speech: **6**

"It's too bad," said Narayan, "I was going to play hockey this evening. But it is raining; so we cannot play. I can think of nothing to do."

"Then I will give you some funny sums," said his father.

8. Transform *any six* of the following sentences as indicated: **6**

(a) More rice has to be grown by farmers. (*into active*)

(b) We must go into this matter. (*into passive*)

(c) The child is being fed now. (*into active*)

(d) Nobody heard a sound. (*into passive*)

(e) The man has to solve some problems. (*into yes/no question*)

(f) He started several projects in the village. (*into negative*)

(g) Shyam always walks slowly. (*into who-question*)

(h) Don't tell anybody. (*into affirmative*)

9. (i) Rewrite the following sentences using the appropriate question-tags: (*any three*) **3** (4)

(a) I'm very late,............?

(b) Let's go back,............?

Contd.

(c) You never used to drink, ...?
(d) He wasn't feeling well, ...?

(ii) Join each of the following pairs of sentences into one sentence with the help of the given clues : (any three)   3

Clues : *Whose, Who, Where, Whom*

(a) This is the place. We visited it last year.
(b) What is the name of the man ? His wife has run away and left him.
(c) Hari lives in the next village. Hari is a very close friend of mine.
(d) I met a man yesterday. This is the man.

10. (i) Join the following sentences into one sentence with clues given within the brackets : (any three)   3

(a) They cancelled the football match. It was raining. (*because of*)
(b) He was wealthy. He never bought a watch. (*in spite of*)
(c) Hari can't read. He buys a newspaper everyday. (*although*)
(d) He is very popular among boys. He is sociable, good-humoured and helpful. (*because*)

(ii) Complete any three of the following sentences with infinitives or -ing forms of verbs given within the brackets :   3

(a) He succeeded in...........(defeat) his enemy.
(b) Don't you enjoy...........(watch) television ?
(c) I want...........(go) out.
(d) The teacher wanted them...........(study) hard.

- Group 'B'

Attempt *any three* :   4 × 3 = 12

11. Give one word for : (any four)   4

(a) Someone who does not like sudden change.
(b) A person who travels through an unknown place to find out about it.
(c) All persons born about the same time.
(d) Place of residence.

(e) Rub hard, specially with something rough.
(f) Quality of fair play.

12. Use *any four* of the following phrases in sentences of your own :   3

(a) out of control
(b) to look forward to
(c) in front of
(d) in search of
(e) for a while
(f) to take advantage of

13. Make nouns or verbs of the following words : (*any four*)

(a) explode (b) discover (c) ignorance (d) move (e) life (f) beauty

14. (i) Give the opposites of *any two* of the following words :

(a) confuse. (b) lure (c) remote

(ii) Give the opposites of *any two* of the following words :

(a) dull (b) truth (c) dangerous

15. Choose the correct word from within the brackets to fill in the blank : (*any four*)

(a) She...........me an extra suitcase because I had only one. (*borrowed, lent*)
(b) I can...........your difficulties but I can't help you. (*realize, understand*)
(c) He...........everything about motor cycles. (*informs, knows*)
(d) He works in a...........which makes matches. (*industry, factory*)
(e) A tiger...........him a year ago. (*died, killed*)
(f) The milkman...........the milk to my house. (*took, brought*)

428

Appendix L
The Existing SLC Examination Specification Grid

English (Secondary Level)
Examination Specification Grid based on New Curriculum

Competence Level (Knowledge & Understanding)	Performance Level (100 Marks) (Practical Abilities)				Remarks
	Language/Linguistics				
	Listening (8 Marks)	Speaking (12 Marks)	Reading (45 Marks)	Writing (35 Marks)	
This will be integrated to PA. No separate examination will be given in this area.	**Time: 15 minutes** Question type: 1. Multiple Choice 2. Fill in ..... 3. True/False 4. Tick the correct pictures/answers 5. Ordering	**Time: 15 minutes** Question type: 1. Interview/Conversation 2. Cued situation 3. Story telling	**Time: 1 : 30 hrs.** Question type: 1. Seen texts (5+10=15 marks) Any two texts can be asked a. Yes/No questions b. Wh questions c. Vocabulary d. Ordering e. True/False 2. Unseen texts (15+15=30 marks) Any two texts can be asked. a. Multiple choice b. Close c. Fill in the table/ordering d. Wh questions e. True/False f. Fill in	**Time: 1 hr.** Question type: 1. Controlled writing (11 marks) Any one of the following - Substitution table - Scrambled sentences - Gap filling - Close - Broken dialogue 2. Guided writing (2x6=12 marks) Any two of the following - Skeleton - Parallel writing - Set of questions - Beginning & End of the story - Describing pictures 3. Free writing (1x12=12 marks) (120 words) Any one of the following - Describing - Replying - Explaining - Narrating	
	Materials: 1. Cassette/Teacher's own voice 2. Pictures, maps diagrams etc.	**Materials:** 1. Teacher as an interviewer 2. Pictures, maps, diagrams etc.	**Materials:** 1. Seen texts taken from set 9/10 textbooks. 2. Different kinds of unseen authentic texts. (advertisements, stories, notices, etc.)	**Materials:** 1. Pictures, maps, clues, diagrams 2. Authentic reading texts etc. should be measured	
	Note: 1. Ask any 3 questions 2. Objectives L1-L3 of Grade 9/10 Curriculum should be measured.	**Note:** Functions 1-20 of Grade 9 and 1-16 of Grade 10 Curriculum should be measured.	**Note:** 1. Objectives R1-R11 of Grade 9/10 Curriculum Grade 10 Curriculum 2. The unseen texts must not be extracted from the set textbooks.	**Note:** Objectives W1-W10 of Grade 9/10 Curriculum should be measured	

Instruction & Marking Scheme for the Examiner
1. Clear instruction should be given to the learners for measuring listening and speaking skills.
2. One or more question types can be asked. But it is not necessary to use all at once.
3. The oral exam should be recorded and sent back to OCE.