



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Examinations as Instruments
for Educational Change: Investigating
the Washback Effect of the Nepalese English Exams

By

Tirth Raj Khaniya



Ph. D.

University of Edinburgh

1990

ABSTRACT

This study examines the washback effect of a final examination. Despite the general criticisms of a final examination for its negative influence on education, no empirical evidence was noted in the existing literature; rather some evidence for positive washback was found.

The study was based on the assumption that the detrimental effect of a final exam is not inherent; whether the washback is negative or positive is dependent on the design of the exam. Furthermore, the power that the exam has to influence teaching and learning, if exploited properly, can make it work as an instrument for educational change.

The context of the study was English language teaching and testing at school in Nepal. The washback effect of the School Leaving Certificate English exam, an exit exam based on prescribed textbooks in terms of its content and discrete-point approach in terms of its format, was examined. A new exam of reading, grammar, note-taking and writing based on the course objectives of the SLC English course in terms of its content, and integrative-communicative approach in terms of its format was designed to use as the criterion measure of the English proficiency of the students.

The new exam was administered to school leavers and students of the previous year. The performance of the school leavers on the new exam was compared with their performance on the SLC English exam, and with the performance of the students of the previous year on the new exam.

The results indicate that SLC English exam had a negative washback on the teaching and learning of the SLC English course because it failed to allow the students and the teacher to work for the course objectives of the SLC English. It was concluded that washback is an inherent quality of a final exam; people whose future is affected by the exam-results work for the exam regardless of the quality of the exam. Whether the washback is negative or positive is dependent on what the exam measures; if it is congruent with the sentiment and the purposes of the course objectives, it can achieve beneficial washback; if not it is bound to produce harmful washback. Innovations through the former type of exam would lead the teaching for the exam to be in accordance with it.

The implications for language testing in general, and the Nepalese ELT situation in particular are presented. It is suggested that the SLC English exam should be replaced by an exam similar to the one used, in order to bring about change in the teaching of the SLC English course. Recommendations for further research are made.

ACKNOWLEDGEMENTS

This work could not have been completed without the cooperation of many people in Nepal and Britain.

I am most grateful to Dr Alan Davies for his constant guidance and encouragement. This study would not have been completed without him leading me to work on something very useful for Nepal and finding funds for the third year of my studies. I was very lucky to have the opportunity to work with him. I am deeply indebted to Dr Robert Hill, who has been more than an excellent supervisor, for his professional supervision and friendly encouragement throughout the whole study. It was only because I could discuss my work with him anytime I needed that this study has been completed within two years.

I would like to express my gratitude to Mr Keith Mitchell and Mr Alastair Pollitt for their invaluable suggestions. I have greatly benefited from comments by them.

I am indebted to my Gurus Prof. Dr Shishir Kumar Sthapit, Prof. Sharada Bhadra, Mr Chandra Bir Singh, Dr Indra Vilash Adhikary, Mr Vinod Shrestha, and Mr Mahendra Rana, and my friends Birendra Pokhrel, Kapil Shrestha, Bal Krishna Khan and Tulasi Neupane for their help to accomplish this study. Dr Sthapit's help has been invaluable.

I am immensely indebted to the students who took the exams, and the teachers and the experts who answered the questionnaires without which the study would not have been completed. I would like to thank Dr Nurshing Narayan Singh, the

then Education Secretary, and Mr Aryal, the Controller of the SLC Examinations for permitting me to visit SLC exam centres and sending to me the results of the SLC English exam.

I am grateful to David Hill, Cathy Benson, Catherine Crossman, Shelagh Dougall, Lindsay Friedman, Bishnu Thapaliya and many others who were kind enough to offer me support and advice at various moments in marking answer sheets and the writing of the thesis.

The financial support for the first two years of the study was provided by the Overseas Development Administration. I would like to record my thanks to the ODA. However, the study would have remained incomplete if special assistance from the British Council in London and the IALS 10th Anniversary Scholarship Fund had not been given in the third year of my studies. I am grateful to the British Council and Institute for Applied Language Studies for their support. I am grateful to Mr Bissett, the then British Council Representative in Edinburgh, and Mrs Ann Hay, my Programme Officer, and Mr Eric Glendinning, the Director of the Institute for Applied Language Studies, for the arrangements of the funds. At the same time, I would like to express my gratitude to the Vice Chancellor and the Rector of the Tribhuvan University, and my colleagues in the Faculty of Language Arts for granting me study leave for three years to pursue this research.

Lastly, but by no means least, I would like to thank my wife, Sabitri Khaniya, who not only looked after the children in my absence, but also worked hard in collecting relevant data for the study in Nepal and sending them to me in Edinburgh, and Kishor, Kalpana, Kopila, and Kabita Khaniya, who were deprived of my warm lap for three years.

ABBREVIATIONS

- APA - American Psychological Association
- CDC - Curriculum Development Centre
- CERID - Centre for Educational Research, Innovation and
Development
- CRT - Criterion-Referenced Test
- CTSDC - Curriculum, Textbook, Supervision Development Centre
- DEO - District Education Officer
- DES - Department of Education and Science
- ELT - English Language Teaching
- ELTS - English Language Testing Service
- ESP - English for Specific Purpose
- FLS - Foreign Language School
- GCSE - General Certificate of Secondary Education
- HMG - His Majesty's Government of Nepal
- HMI - Her Majesty's Inspectorate of Schools
- HMSO - Her Majesty's Stationery Office
- JEMC - Janak Education Materials Centre
- MET - Matriculation English Test
- NCE - National Certificate of English
- NRT - Norm-Referenced Test
- NUEE - National Unified Enrolment Examination
- PCA - Principal Component Analysis
- SLC - School Leaving Examination
- TEFL- Teaching English as a Foreign Language

CONTENTS

Abstract

Acknowledgements

Abbreviations

CHAPTER 1: EXAMINATIONS AS INSTRUMENTS FOR EDUCATIONAL CHANGE

	page
1.0. Introduction	1
1.1. The Structure of the Thesis	2
1.2. What an Examination is	4
1.3. Functions and Purposes of Examinations	7
1.4. Examinations: Desirable or Inevitable?	11
1.4.0. Introduction	11
1.4.1. Criticisms of Examinations	11
1.4.1.1. The Nature of Exams	13
1.4.1.2. The Use Made of Exam-results	13
1.4.1.3. Effects on Education	14
1.4.2. Search for an Alternative	15
1.4.2.1. Sought-after School-based Assessment	15
1.4.2.2. Attack on School-based Assessment	15
1.4.3. Will Examinations Survive?	17
1.5. The Role of Examinations in Education: Washback Effect	19
1.5.0. Introduction	19
1.5.1. Washback Effect: What it is	20
1.5.2. Washback Effect: Development of the Concept	20
1.5.3. Washback Effect of Examinations on Teachers	23
1.5.4. Washback Effect of Examinations on Students	24
1.6. Examinations as Instruments for Educational Benefit	26
1.6.0. Introduction	26
1.6.1. Undesirable Effects of Examinations: A	

Contingent Factor or an Inherent One?	27
1.6.2. Examinations for Educational Benefit:	
Positive Washback	30
1.6.3. Need for a Good Combination of the Examination and the School-based Assessment	33
1.7. The Role of Examinations in Educational Change	35
1.7.0. Introduction	35
1.7.1. The Interdependence of Curriculum Elements	36
1.7.2. Strategy for Change	38
1.7.3. An Examination as the Element to Begin With	40
1.7.4. The Role of an Examination Within the Teaching Syllabus: Four Case Studies	41
1.7.5. Conclusions	46
1.8. Summary and Comments	47
CHAPTER 2: ENGLISH LANGUAGE TESTING IN NEPAL	
2.0. Introduction	49
2.1. English Language Teaching in Nepal	50
2.1.0. Introduction	50
2.1.1. Linguistic Situation	50
2.1.2. The need for English in Nepal: A Rationale	52
2.1.2.1. Educational Purposes	53
2.1.2.2. Occupational Purposes	54
2.1.3. English as a School Subject: Compulsory or Optional?	54
2.1.4. Summary	59
2.2. The Existing ELT Situation at School Level	59
2.2.0. Introduction	59

2.2.1. The General Objectives of Teaching English at School Level in Nepal	60
2.2.2. English as a Compulsory School Subject	60
2.2.3. Types of School	61
2.2.3.1. Type A	61
2.2.3.2. Type B	62
2.2.3.3. Type C	63
2.2.3.4. Type D	63
2.2.4. Methodology	63
2.2.5. Course Materials	64
2.2.5.1. Example 1: Language	65
2.2.5.2. Example 2: Content	65
2.2.5.3. Example 3: Exercise	66
2.2.6. Examinations	67
2.2.6.0. Introduction	67
2.2.6.1. Functions of SLC: the Crux of the Matter	67
2.2.6.1.1. SLC: A Sense of Achievement	67
2.2.6.1.2. SLC: A Licence for Prestigious Work	68
2.2.6.1.3. SLC: A Permit for Higher Education	68
2.2.6.2. The SLC English Exam Content	69
2.2.6.3. Criterial Level of Performance	70
2.2.6.4. Problems in Determining Proficiency Level	71
2.2.6.5. Cutoff Points	72
2.2.6.6. Analysis of the SLC English Exam, 1989	72
2.2.7. Summary	77
2.3. Criticisms of SLC English: Need for Change?	78
2.3.0. Introduction	78
2.3.1. Statements of the Problems	79
2.3.1.1. Results in the SLC English Exam	80
2.3.1.2. Performance of College Entrants	82

2.3.2. Criticisms of the SLC English Exam	84
2.3.2.0. Introduction	84
2.3.2.1. Theoretical Considerations: The Efficiency of an Examination	86
2.3.2.1.0. Introduction	86
2.3.2.1.1. Validity	87
2.3.2.1.2. Content Validity	89
2.3.2.1.3. Predictive Validity	93
2.3.2.1.4. Inclusiveness of Construct Validity	97
2.3.2.1.5. Face Validity	98
2.3.2.1.6. Summary and Comments	99
2.3.2.1.7. Reliability	100
2.3.2.1.8. Validity and Reliability: Tension or Compromise?	103
2.3.2.1.9. Practicality	104
2.3.2.1.10. Conclusions	106
2.3.2.2. Negative Washback Effect of the SLC English Exam	107
2.3.2.3. SLC English Exam as a Lever for Change	110
2.3.3. Summary	112

CHAPTER 3: CONTENT VALIDATION OF NEW EXAM

3.0. Introduction	113
3.1. Theoretical Considerations: Approaches to Language Testing	114
3.1.0. Introduction	114
3.1.1. What is to be Tested?	114
3.1.2. What is to be Tested: A Changing Concept	116
3.1.3. Pre-discrete Point Testing	117
3.1.4. Discrete Point Testing	117

3.1.5. Integrative Testing	119
3.1.6. Functional Testing	122
3.1.7. Communicative Testing	123
3.1.8. Summary and Comments	125
3.2. Guidelines for the Design of the New Exam	127
3.2.1. Needs-orientated Approach: A Practical Consideration	127
3.2.2. Assessment of Listening and Speaking: A Practical Problem	130
3.2.3. General Proficiency and Language Skills Construct	131
3.2.4. Washback Effect of the Exam	132
3.3. Assessment of Reading and Writing	133
3.3.0. Introduction	133
3.3.1. The Nature of Reading	133
3.3.1.0. Introduction	133
3.3.1.1 Reading Comprehension: Unitary Competence or Combination of a Set of Subskills	134
3.3.2. The Nature of Writing	140
3.3.3. Writing from the Testing Point of View	143
3.4. Specifications of the New Exam	145
3.4.0. Introduction	145
3.4.1. Reading	146
3.4.2. Linguistic Competence	146
3.4.3. Note-Taking	148
3.4.4. Writing	149
3.5. Formats of the the New Exam	150
3.6. Summary and Comments	151

CHAPTER 4: DESIGN OF EMPIRICAL INVESTIGATIONS

4.0. Introduction	152
4.1. Formulation of Predictions	152
4.1.1. Background to Predictions	152
4.1.2. The Main Research Questions and the Procedures to Address them	154
4.1.3. Formulation of Hypotheses	155
4.2. Methods of Investigation	157
4.2.1. Construction of Two Sets of Tests	157
4.2.2. General Principles of the Test Construction	157
4.2.2.1. Contextualization	158
4.2.2.2. Authentic-like Materials	158
4.2.2.3. Unseen Tasks	159
4.2.2.4. Criteria for Text Selection	159
4.2.2.5. Criterion-Norm Referenced Tests	161
4.2.3. Detailed Specifications of the New Exam	163
4.2.3.1. Objectives of the Exam	163
4.2.3.2. Contents and Formats of the Exam	165
4.2.4. Design of Questionnaires	167
4.2.4.1. Questionnaire for People Involved in SLC English	168
4.2.4.2. Questionnaire for College Students	169
4.2.4.3. Questionnaire on Needs Analysis	170
4.2.5. Scoring Procedures	175
4.2.5.0. Introduction	175
4.2.5.1. Scoring Reading, Cloze, and Note-taking Tasks	176
4.2.5.2. Marking Compositions	177
4.2.5.3. What is a Composition?	177

4.2.5.4. What We Mean When We Say 'Scoring Compositions'	177
4.2.5.5. Approaches to Marking Compositions	178
4.2.5.6. General Impression Method	180
4.2.5.7. Analytic Method	181
4.2.5.8. Reliability in Holistic Evaluation: Problems and Suggestions	184
4.2.6. Design of a Marking Scheme	186
4.2.6.0. Introduction	186
4.2.6.1. Theoretical Considerations	187
4.2.6.2. Purpose and Nature of the Test Tasks	188
4.2.6.3. Practicality	189
4.2.7. Focused Impression Method	190
4.2.7.1. Criteria and Weighting	190
4.2.7.2. Rationale for Inclusion of Criteria	191
4.2.8. Summary and Comments	192
4.3. Description of Field Study	192
4.3.0. Introduction	192
4.3.1. Piloting Tests in Edinburgh	193
4.3.2. Piloting Tests in Nepal	193
4.3.2.1. Subjects	194
4.3.2.2. Procedures of the Pilot Study	194
4.3.2.3. Visiting SLC Exam Centres	197
4.3.3. Insights Gained from Piloting Tests	197
4.3.4. Final Version of the Exam	199
4.3.5. Administration of the Exam	199
4.3.5.1. Selection of School	199
4.3.5.2. Selection of College	201
4.3.5.3. Selection of Subject	201
4.3.5.4. Type of Subject	202

4.3.6. Administration of Questionnaires	204
4.3.6.1. Questionnaire for People Involved in SLC English	204
4.3.6.2. Questionnaire for College Students	205
4.3.6.3. Questionnaire on Needs Analysis	205
4.3.7. Problems Encountered While Administering the Exam	207
4.3.8. Arrangements for Collecting SLC Results	209
4.3.9. Summary and Comments	209

CHAPTER 5: RESULTS

5.0. Introduction	211
5.1. Reliability	212
5.1.0. Introduction	212
5.1.1. Internal Consistency	212
5.1.2. Intermarker Reliability	213
5.1.3. Single-Rater Reliability	214
5.1.4. Test Retest Reliability	214
5.2. Descriptive Data	216
5.2.0. Introduction	216
5.2.1. Results of New Exam: Gradewise (Grade 10 and SLC)	217
5.2.1.1. Grade 10: First Administration in February 1989	217
5.2.1.2. Grade 10: Second Administration in December 1989	218
5.2.1.3. Results Obtained by SLC Students	218
5.2.2. Results of New Exam: Schoolwise	221
5.2.3. Results of SLC English Exam	224
5.2.4. Comparison of Performance Obtained by SLC students in the New Exam and the SLC English Exam	225

5.2.5. Results of New Exam Obtained by College Students	234
5.2.5.1. Administered to Students at the Institute of Science	234
5.2.5.2. Administered to Students at the Institute of Humanities and Social Science and the Institute of Management	235
5.2.6. Questionnaire Results	236
5.2.6.1. Questionnaire for People Involved in SLC English	236
5.2.6.2. Questionnaire for College Students	239
5.2.6.3. Questionnaire on Needs Analysis	240
5.2.7. Summary	245
5.3. Hypothesis Testing	245
5.3.0. Introduction	245
5.3.1. Main Hypothesis 1 and Subsidiary Hypotheses (1a-1c) Tested	246
5.3.2. Main Hypothesis 2 and Subsidiary Hypotheses (2a-2c) Tested	250
5.3.3. Summary of the Hypotheses Tested	254
5.4. Exam Validation	255
5.4.1. Test Validation: A Rationale	255
5.4.2. Problems in Test Validation: A Case of of the New Exam	259
5.4.3. Validation of the New exam	261
5.4.3.0. Introduction	261
5.4.3.1. Traits Tested	262
5.4.3.2. Practicality	263
5.4.3.3. Item Statistics	264

5.4.3.4. Item Analysis: Facility Values: Reading and Cloze	264
5.4.3.5. Comments on the Reading Tests	265
5.4.3.6. Response Frequency Distributions	268
5.4.3.7. Comments on the Cloze Test	269
5.4.3.8. Test Analysis: Difficult/Easy: Note-taking, Letter Writing, and Essay Writing	270
5.4.4. Validity	271
5.4.4.0. Introduction	271
5.4.4.1. Face and Content Validity	271
5.4.4.2. Construct and Concurrent Validity	273
5.4.4.2.0. Introduction	273
5.4.4.2.1. Relationship between New Exam and SLC English Exam	274
5.4.4.2.2. Unique Variation between New Exam and SLC English Exam	277
5.4.4.2.3. Relationships between New Exam, SLC English Exam, and Cambridge O' Level Language Test	27
5.4.4.2.4. Relationships between New Exam, SLC English Exam, and College Teachers' Judgements of their Students' English	28
5.4.4.2.5. Relationships between Subtests	28
5.4.4.2.6. Principal Component Analysis	28
5.4.5. Is the Existing ELT at School in Nepal Adequate?	28
5.4.5.0. Introduction	28
5.4.5.1. Meaning of Passmark in the New Exam: An Index of Language Proficiency	28

5.4.5.2. Meaning of Passmark in SLC English Exam: The Minimum Level of Language Proficiency Required for SLC English Exam	29
5.4.5.3. Summary and Comments	29
5.4.5.4. Logical Evidence of the Construct Validity of the New Exam	29
5.5. Summary and Comments	30

CHAPTER 6: DISCUSSION AND IMPLICATIONS

6.0. Introduction	30
6.1. Meaning of the Findings	30
6.1.1. Washback Effect of the SLC English Exam on the Teaching and Learning of the SLC English Course	30
6.1.2. Relationships between Teaching in Different Types of Schools and Learning	30
6.1.3. Adequacy of the SLC English Course	30
6.2. Efficiency of the New Exam	31
6.2.1. Reliability	31
6.2.2. Validity	31
6.2.2.1. Statistical Validity	31
6.2.2.2. Logical Validity: Face, Content, and Washback	31
6.3. Implications of the Study for Language Teaching and Testing	31
6.3.1. Exams as Powerful Instruments for Gearing up Classroom Activities in a Desired Direction	31
6.3.2. Relationship between Formal Instruction and Learning	31
6.3.3. Relationship between Course Objectives and Exam Content	31

6.3.4. Direct and Indirect Testing	31
6.4. Implications of the Study for the Nepalese ELT Situation	32
6.5. Methodological Limitations	32
6.6. Summary	32

CHAPTER 7: CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

7.1. Conclusions	32
7.1.1. Washback: An Inherent Quality of a Final Exam	32
7.1.2. Ingredients of the Exam Determine whether the Washback is Negative or Positive	32
7.1.3. Teaching for a Final Exam: Inevitable or Desirable?	32
7.2. Suggestions for Further Research: Washback Validity: Need to Innovate	32

Bibliography	33
---------------------	----

FIGURES

Figure 1: Elements of curriculum	37
Figure 2: Types of subject who took the new exam	203
Figure 3: Comparison of frequency distributions of score of SLC and Grade 10 students in the new exam	220
Figure 4: Comparison of scores obtained by SLC and Grade 10 Grade 10 students in the new exam according to school type	222
Figure 5. Mean scores for the SLC English exam and the new exam according to school type	228
Figure 6: Comparison of frequency distributions of scores in the two exams	229
Figure 7: Scattergram of the scores in the new exam and the SLC	

English exam	275
Figure 8: Scattergram of the scores in the SLC English exam and the Cambridge O' level language test	280
Figure 9: Scattergram of the scores in the SLC English exam and the Cambridge O' level language test	281

TABLES

Table 1: The result in the SLC English exam, 1986	80
Table 2: Failure rate at proficiency certificate level	83
Table 3: Reading skills included in the questionnaire	172
Table 4: Writing activities included in the questionnaire	174
Table 5: Traits to be emphasised in teaching and testing writing	175
Table 6: Students who took part in the pilot	194
Table 7: Students who participated in the questionnaire	205
Table 8: Tertiary tutors who participated in the questionnaire	206
Table 9: Item Analysis: Discrimination Index	213
Table 10: Frequency of scores obtained by Grade 10 students	217
Table 11: Frequency of scores obtained by SLC students	219
Table 12: Descriptive statistics of the scores obtained by Grade 10 and SLC students	219
Table 13: Description of the scores according to school type	223
Table 14: Results in the SLC English exam according to type of school	224
Table 15: Percentage of the students according to cutoff-points	227
Table 16: Percentage of the students who exhibited equal performance on both exams	231

Table 17: Percentage of the students who reached 0-31 cutoff in the new exam and 32 and above in the SLC English exam	232
Table 18: Percentage of the students who obtained below 32 cutoff in the new exam but 32 and above in the SLC English exam	233
Table 19: Correlationships between the scores in the SLC English exam, the new exam, I Sc first and second year English exam obtained by the students from the Institute of Science, Bharatpur	235
Table 20: Summary of the responses to questionnaire for the people involved in SLC English	236
Table 21: Raw scores obtained by each item	241
Table 22: T-Test for the performance of Grade 10 and SLC	248
Table 23: T-Tests for the comparison of group means obtained by Grade 10 and SLC groups from different schools	249
Table 24: Matched-T-Test for the comparison of group means obtained by Grade 10 students from Type D schools in the two administrations of the new exam	249
Table 25: ANOVA for the difference between school types at SLC level	251
Table 26: Post-hoc Scheffe comparison of the performance of SLC students according to type of school	252
Table 27: Oneway ANOVA for performance of Grade 10 students according to type of school	253
Table 28: Post-hoc Scheffe comparison of the performance of Grade 10 students according to type of school	253
Table 29: Summary of the Hypotheses Tested	254
Table 30: Item Analysis: Facility Value	264

Table 31: The percentage of responses attracted by each distractor	268
Table 32: Percentage of Notes, Letters, and Essays categorised as good, moderate, and poor	271
Table 33: Correlations between the SLC English exam and the new exam according to school type	276
Table 34: Correlation between new exam, SLC English exam, and Cambridge O' level language test	279
Table 35: Correlations between new exam, SLC English exam, and College Teachers' Judgement on their students' English	283
Table 36: Correlations between subtests	284
Table 37: Principal components and eigenvalues for four subtests by extracting four components	286
Table 38: Principal components and eigenvalues for four subtests by extracting three components	287
Table 39: Item Analysis: Facility Value	290
Table 40: Item analysis: Facility Value: Reading and Cloze	295
Table 41: Average score on the four subtests obtained by the two group of students: cutoff points 32-44 and 0-31	301

APPENDICES

Appendix I: A sample of the SLC English Exam, 1989	
Appendix II: A sample of the questionnaire for the people involved in the teaching and testing of SLC English	
Appendix III: A sample of the questionnaire for college students	
Appendix IV: A sample of the questionnaire on needs analysis	
Appendix V: A sample of the ESL composition profile	
Appendix VI: A sample of the new exam	
Appendix VII: Scores in the new exam obtained by Grade 10 students	

Appendix VIII: Scores in the new exam and the SLC English exam
obtained by SLC students

Appendix IX: A sample of the questionnaire for college teachers

Appendix X: Results obtained by performing Principal Component
Analysis

Appendix XI: A sample of the Cambridge O' level language test

CHAPTER 1: EXAMINATIONS AS INSTRUMENTS FOR EDUCATIONAL CHANGE

1.0. Introduction

The system of examinations, generally, has been the focal point for expressions of dissatisfaction. The sources of discontent are, partly, the way exams (note: examination and exam are used interchangeably) are carried out - the selection of content, the organisation of exams, and the marking procedures, and partly, the system itself - the obstacles it places which students are required to surmount to go on to the next level of their education. The requirement to pass their exams to advance further leads students to work for their examinations rather than to study for the acquisition of useful knowledge and skills. In other words, examinations by virtue of their association with screening have been criticised for having a detrimental effect on teaching and learning.

It will be worth investigating how far the association of an examination with the detrimental effect on teaching and learning is an inherent attribute. One would also ask: Does an examination have any potential to contribute to the educational process?

These are the issues that the present study is aimed to address.

This chapter considers the theoretical concepts under which the present study was conceived.

First, for the purpose of the present study, a distinction between a test and an exam will be made, and then the purposes and functions of an examination will be discussed.

Then the arguments presented for and against the system of examinations will be considered. It will be argued that, despite severe criticisms, examinations do not seem to be disappearing. Moreover, they are likely to play important roles in education for the foreseeable future.

Next the criticisms that an exam can have a negative influence on teaching and learning will be discussed. It will be argued that the negative effect of an exam is not an inherent but a contingent factor. This study considers that the educational effect of an examination is unavoidable. It will be argued that if an examination is designed in relation to the purpose for which we want it, and in congruence with the course objectives, it can be educationally beneficial.

Finally, the role of the examination in curriculum development will be discussed. It will be argued that the exam can have the potential to speed up the movement of curriculum innovation if it has given a lead. It will also be argued that ignoring the potential role of examinations in educational change may lead the whole process of change to stagnate. To make this point more explicit, four case studies will be presented.

1.1. The Structure of the Thesis

The thesis contains seven chapters. This chapter considers the theoretical aspect of the educational effect of a final examination. The remaining chapters will present the context, and the methods adopted, and the empirical evidence of washback found in the investigation.

In **Chapter 2**, first English Language Testing at school in Nepal, the context of the study, will be discussed. The existing

SLC English exam will be examined against the efficiency criteria of an exam. It will be argued that the exam is not efficient enough for the purpose it has been used for. It will be proposed that a new exam which can serve the purpose should be designed in order to replace the SLC English exam.

To obtain the desired purposes, considerable attention was paid to the a priori validation of the exam. This constitutes the content for **Chapter 3**. Different approaches to language testing, and the nature of reading and writing will be discussed in order to determine the specifications and the formats of the new exam.

The details of the investigations will be presented in **Chapter 4**. First the hypotheses of the study will be presented. The scoring procedures used in the study will also be considered in this chapter. Then the description of the instruments, and the field study will be given.

Chapter 5 presents all the relevant data about the empirical investigation obtained by the administration of the exam and the questionnaires. The statistical analyses carried out on the data will also be discussed. The hypotheses will be tested. The results obtained by the validation of the exam will be presented.

The results of the study will be discussed in **Chapter 6**. The implications of the study for language testing in general, and the English language testing in the schools in Nepal in particular will be presented.

Chapter 7 will conclude the study. Suggestions for further research will also be made.

1.2. What an Examination is

In the testing literature, the terms 'test' and 'examination' are often used synonymously (Davies 1977:49): the Cambridge Proficiency Exams, the Michigan Test, the ELTS Test, etc. For the purpose of the present study, however, it seems extremely important to make a distinction between the two terms in order to avoid confusion between an assessment involving judgement about the adequacy of a pupil's achievement (e.g. pass/fail decision), and an assessment involving remedial instruction (e.g. testing for teaching). To make a distinction between the two, however, is difficult. On the one hand, sometimes, an exam seems to be included in a test in the sense that a test can have different realisations - a class progress test, a proficiency test, a summative test - and an examination is one of them. On the other hand, a test appears also to be included in an exam in the sense that an exam can have different forms - a test, assessment of course work, interview - and a test is one of them.

The test and the examination are also not different from the point of view of the way they require testees to perform: both contain a list of tasks or test items requiring students to act upon them. However, they are different from the point of view of the way their test items are selected, and the purposes behind them. In other words, they differ in terms of exam-syllabus, and the interpretation of the results. In most cases, the examination is concerned with achievement. So it would be useful to discuss these terms, first, within the framework of achievement testing.

Briere (1971) (quoted in Abdrabou 1984:2) defines an achievement test as a measure of "the extent to which an individual student has mastered the specific skills or body of information which have been presented in a formal classroom situation". This definition seems too narrow because it confines the achievement test's syllabus to 'the information which has been presented in a classroom situation'. There is another way of looking at the achievement test's syllabus if we follow Ingram (1968:71), Davies (1977:45), and Heaton (1975:163), who consider the achievement test's syllabus to be what the examinees are supposed to have learned within a total educational system, which may not necessarily be taught in a class-room situation. Taking this view into consideration, it would appear that the achievement test's syllabus covers a wide range in which what has been taught in a class-room falls on the one extreme, and the course's objectives falls on the other. So far as the types of test which fall along the continuum are concerned, Heaton (1975:163) and Hughes (1988a:39) discuss two types of attainment test: i.e. 'class progress test' and 'final achievement test'. They further argue that a class progress test is designed to measure the extent to which students have achieved what has been taught in the classroom. This type of testing is solely based on, and intended for, classroom teaching as it is diagnostic and formative, whereas a final achievement test is designed to measure the extent to which students have achieved what they are supposed to have achieved. This type of testing is based on a course's objectives, not necessarily on what has been taught. A final achievement test sometimes takes

the form of a school leaving certificate examination. This type of assessment, particularly in school leaving exams, is generally associated with judgement about grades and the pass/fail decision. It is this type of assessment which is the main concern of the present study. A final achievement test appears to be different from a proficiency test in the sense that the latter is not based on any course objectives.

Another point that helps differentiate exams from tests is the way students perceive them as instruments for measurement. The issue will be clear from the following definition. Boisson (1887) (quoted in Agazzi 1967:65) describes an examination in the following words:

In every country, examinations represent the stamp of official approval set on the completion of the pupil's school career. Besides marking the end of that career, they are indispensable for inciting adolescents to a serious and sustained effort. When there is an examination ahead of them, pupils and teachers can no longer behave as though they were at home, in an easy, casual way, with allowances made for good intentions, all must march in step and try to keep in line. Pupils pay more attention to their work and teachers make their lessons more accurate and concise. The examination must be passed at all costs.

Though the definition is very old, it is still relevant to the present situation in the sense that students consider exams something to be passed at all costs. It leaves room for doubt whether they perceive their performance in tests in the same way, as Black and Devine (1986:21) find students less anxious about diagnostic tests than summative tests and examinations. That means an exam can be a stronger motivating force for students than a test.

Brereton (quoted in Tibble, 1969:347) distinguishes between a test and an examination in a similar vein. He defines a test as:

designed to provide an accurate standardized measurement of certain abilities or skills without influencing teacher or student and without creating any tension in the student.

His definition of an examination (1969:347) is:

a dynamic part of the whole educational process, involved with motivation and with defining the character of the education in which the student is involved - a type of happening which incorporates many of the characteristics of human life itself - competition and rivalry, purposeful activity directed towards a goal, the planning of a course of work and maintenance of standards.

On the whole, it would seem that examinations and tests are not different physically because they require examinees to perform. However, they differ in influencing the psychological set of students and teachers.

For the purpose of the present study, to make a distinction between the two terms, 'exam' will be used to refer to an assessment of students' achievement in a course where the syllabus is based on course objectives, where external control is included and where the result contributes to the pass/fail decision and where this assessment takes place at the end of a course of teaching. 'Test' will refer to an assessment of students' achievement in a course which is associated with testing for teaching.

1.3. Functions and Purposes of Examinations

The question 'why bother with examinations' is very closely related to other questions: 'who will be examined?' and 'who will use the results?' Regarding the present study, it is the school leavers whose performance will be assessed, and there can be several agents including the students themselves who will make use of the results of the examinations. Fabian (1982:24) seems to agree with this opinion when he writes "an examination is ... to all intents and purposes a dialogue between supplier

and consumer". His interpretation of the supplier is those who create examination structures and strategies: i.e. teachers, schools, course planners, examining boards, and students. By consumer, he means those who need the information: e.g. the educational authority, teachers, students, employers.

The primary purpose of a final examination is to assess the current knowledge, achievement, and skills of candidates (French et al 1988:16). The assessment is carried out by asking the candidates to perform certain tasks/activities. A candidate's achievement level is expressed by assigning a mark or grade or number to his performance. So it might be supposed that the final grade/mark/number is the quantified measure of a candidate's overall achievement. The quantified measure of a candidate's achievement expressed by such an examination can and will be used by different agencies, such as employers, schools, colleges, and universities for different purposes. The use of examination results can also be considered part of the functions and the purposes of examinations.

Before proceeding to the discussion of the different uses of examination results, another important function of a final examination should first be discussed: i.e. the representation of a course's objectives. If we accept the argument that a final achievement test is designed to measure the extent to which students have achieved a course's objectives, the items which students are required to perform, whether they are assessed in one or two papers or by continual assessment, will be the representation of a course's objectives. To put it another way, an important function of a final examination is to discover the

extent to which a course's objectives are being achieved (Thyne 1974:33). For students those course objectives stand for the target to be reached.

Ingenkamp (1977:130) mentions the following as the main functions of an examination:

- a. information for students concerning their progress, to enable them to improve their subsequent learning process,
- b. motivation for students,
- c. selection, in order to single out students with superior ability and achievement for further studies or specific professions;
- d. information on the effectiveness of certain teaching methods, curricula, or form of organisation.

For Oppenheim et al (1967:341), and Pilliner (1973:3), the purposes and functions of an examination are to provide students with a sense of achievement, feedback, permission to go on to further education, teachers with feedback about the effectiveness of their teaching, and employers with a guarantee of competence in those examined to perform the tasks demanded of them by the jobs or professions they take up.

Tyler (1969:346) and Findley (1963:3) state the functions of examinations as "to guide or select students for further education, monitor the educational program of school system, and aid the work of teaching and learning".

Agazzi (1967) also thinks in the same direction. For her, the functions of examinations are to provide teachers and parents with information about the pupils' progress and performance, useful guidance and information for selection, and with information about the overall results of the education system.

However, all the uses of final examination results cannot be generalised, because they vary from place to place. What can be done is to present some generalisable purposes and functions

which should be relevant to the purpose of the present study.

Looking at the nature of the purposes and functions of an exam discussed above, they can be seen as being used, in a broad sense, for two purposes: 'backward looking' and 'forward looking' (terms from Wiseman 1961:151). An examination serves a backward looking purpose in the sense that the information which can be obtained from it can be used for bringing about improvement in students' achievement. The explanation is that an exam:

- i. stands for a stimulus, guidance and a goal for students, and provides them with evidence of achievement to date,
- ii. provides teachers with feedback about the effectiveness of their teaching,
- iii. provides administrators with feedback about "the appropriateness of the curriculum they devise and the extent to which curricula objectives are being achieved" (Pilliner 1973:33).

A final examination also serves a forward looking purpose in the sense that the quantified measure of a student's knowledge, skills, and achievement is approved as some guarantee of competence in him to perform the tasks demanded of him by the further education he goes on to, or the job and profession he takes up. In other words, an exam serves as a selection procedure.

The assumption underlying this purpose is that the students who perform best in current examinations benefit from further educational investment. Their best performance in the examinations is a proof that they have greater ability than

others.

1.4. Examinations: Desirable or Inevitable?

1.4.0. Introduction

This section deals with the arguments presented for and against the examination. In the literature, the exam has been criticised for providing imperfect information about examinees and having a detrimental effect on teaching and learning. It has been suggested that the exam should be replaced by a school-based assessment system. However, despite the criticisms, the exam is very likely to survive in future. In that case, it will be suggested that the exam should be used for educational benefit.

1.4.1. Criticisms of Examinations

Examinations have been considered to have the potential to serve several purposes as mentioned in the preceding section; however, it is not new to argue whether they should continue or not. For a long time, exams have been accused of being responsible for several ills. Some have gone even to an extreme point like O'Meara (1944:10) who criticises examinations for being associated with:

carelessness, hatred, favouritism, labor unrest,
unprogressiveness, dishonesty, discontent, poverty,
fraudulency, laziness, a generator of mental
defectiveness and physical degeneration, serfdom,
radicalism, suffering, death, strikes, and war.

It can be said that several improvements have taken place regarding examination procedures in the last few of decades, especially during the 1960's and 1970's (Morris 1972:74). Nonetheless, the criticisms of exams have not stopped, though they are expressed in contemporary terminology. As Cooper and Leiter (1980:35) point out:

misapplication and misinterpretation of test results can injure individual students and erode curriculum and instruction... create social and intellectual segregation, foster elitism, fashion a punishment/reward syndrome, reduce learning to rote and regurgitative modes, deprecate, stigmatize, exclude.

Agazzi (1967:56) also holds the view against an external examination when she argues:

much of the blame for the present illogical unsatisfactory system can be traced to the mistaken idea that the main purpose of exams is to provide a guaranteed 'control' ... based on lack of confidence in ... the work of teachers, and the degree of conscientiousness... Given this lack of confidence in the teachers, the retention - as also the distortion - of the examination system is inevitable.

Agazzi (1967:42) talks about the modern approach in education that emphasises an individual as the pivot. In this case, she argues, no one examination can be accepted as the core of the whole system. In the reformed education system under this approach, assessment of pupils' capacities tends to be based on reports from all the teachers who have contact with them. As long as this situation prevails, an examination, since it has intrinsically a negative effect, should not be allowed to have an influence on their education, and their future career.

If we accept the above arguments, an exam can be seen as an obstacle to an effective teaching/learning program, because it can distract attention from the school's basic intents, and arouse unnecessary excitement and fear. It is pointless to use the exam for any purpose if it is full of only these ills. In that case, it is worth investigating how far the accusations that an exam as an instrument has a harmful effect on teaching and learning are true and fair. To make any comment on this aspect of an exam, it would be necessary to look at the major points made against the exam.

If we look at the fundamental aspects of examinations that have been criticised, they can be discussed under three headings: the nature of the exam, the use made of exam results, and the effects of the exam on education.

1.4.1.1. The Nature of the Exam

Examinations have been criticised on the ground that they are based on a core exam-syllabus which assumes that all students have had the same learning experiences because they have followed the same curriculum. Kellaghan (1982:4), however, argues that it is not true that all schools share the same learning experiences. The point is that if what predominates in most schools and in the exam-syllabus do not go together, the measurement may be imperfect.

They are also criticised for poor sampling of the activities students are asked to perform on the exams. The point is that the limited content of an exam makes it trivial to assess, and provides no true information about the range of pupils' abilities. Other criticisms of examinations are: norm-referenced interpretation, fluctuation in marking, and the time and costs involved in administering them. The whole criticism is based on the design level of the exam papers. In other words, they are concerned with the validity and reliability of the exam used.

1.4.1.2. The Use Made of Exam-results

The application of exam results has been another aspect of the criticism. Examinations have been criticised also on the grounds that exam results are used for selection and classification of students, which are obviously not the

immediate purpose of a final achievement test. The argument is that the exams should not be for ranking students. The result of using exam-results for pass/fail decisions can have a very bad effect on students because, as Rowntree (1977:16) points out, "for the majority of candidates, many such tests function as rejection tests". Failing an examination can cause a stigma for life, but Wong (1969:362) defies the pass-fail classification arguing, "it is not clear whether many among those who have not succeeded at an exam have not been wasted because of the process". The point is that, on the one hand, the decision (i.e. pass/fail) made on the basis of exam results is very important, and can have serious consequences for students future, on the other hand, the processes which are usually followed in examining are not satisfactory.

1.4.1.3. Effects on Education

Some critics have described examinations as having negative effects on education. This criticism is largely based on the criticisms discussed above: the nature of exams, and the uses made of exam results.

The claim is that the consequences of exam-results are so important for students' future that examinations pressurize the teachers and students into teaching and learning for the exams, and as a result, the whole effort "leads to a narrowing of the curriculum into score producing areas" (Kellaghan 1982:5). In this connection, Wong's (1969:363) summary of the criticism seems to be applicable to many situations. He has written:

examination dictates the activities in schools. Syllabuses... are issued by examination syndicates and central authorities. Interpretation of the syllabus is carried out chiefly by reference to past examination papers which ... tend to carry questions similar in type

and content year after year. In certain countries where school-leaving and pre-university examinations are still external and conducted in English, the high sale annually of the 'Ten Years Series' booklets points to the untiring efforts of pupils and teachers to devine the mind of the examiner. Teaching and learning on such a basis sets a premium on a happy combination of guesswork and memorization.

This description clearly reflects at least the Nepalese situation, which will be discussed in detail in Chapter 2. In this situation, exam-syllabuses take the place of teaching syllabuses and, as a result, teachers and students do not become interested in anything apart from what they think will appear in the examination. In situations like this an examination can no doubt have a detrimental effect on teaching and learning.

1.4.2. Search for an Alternative

1.4.2.1. Advocacy of School-based Assessment

People who take the above criticisms seriously hold the view that it is the teacher who occupies 'a key position' in keeping a close look to the student's performance and identify the student's difficulty in learning (Summer 1987:5). Perhaps Curriculum and Examinations in Secondary Schools (1943) (quoted in Nuttall (1984:164) had this view in mind a long time ago, and, therefore, it (p.164) even suggested that:

in the interest(s) of the individual child and of the increased freedom and responsibility of the teaching profession, change... should be in the direction of making the examination entirely internal; that is to say, conducted by the teachers at the school on syllabuses and papers framed by themselves

The point to be emphasised here is that examinations have been criticised because of the nature of external control. Therefore, an argument in favour of school-based-assessment is presented.

1.4.2.2. Attack on School-based Assessment

School based assessment has been no less immune from

attack for being responsible for harmful effects on education, and the lack of confidence in the certification. Bowe and Whitty (1984:180) seek to explain why the freedom given to teachers to develop their courses and methods of assessment was being progressively eroded. One of the possible explanations was, they argue (p. 180), that, "many such schemes were dangerously subversive and allowed teachers to abuse their professional mandate in the interests of extremist political goals". Though the context of the expression in itself was related to British politics - Conservative versus Labour - there is some truth in it which is applicable to several situations.

Agazzi (1967:67) is of the opinion that some sort of 'control' is indispensable as a means of regulating the activities of students, and teachers in every education system. The study carried out by Bowe and Whitty (1984:185) strongly supports this view in which they find that the public are concerned that "assessment should be independently conducted by an outside person or agency". They further say that people associate external assessment with impartiality, and to many people "that the teacher is in fact responsible for examining his own candidates is abhorrent".

Nuttall (1984:168) holds the view that the result of inexorably growing freedom for schools in England led to variations in exam-syllabus, which, in turn, led to variations in the curriculum from school to school. Nuttall (1984:169) also quotes Burgess and Adams (1980) summarising the charges levelled at examinations as:

their lack of comprehensiveness in covering all the aims of education, their exclusion of a large minority, their

cost and obtrusiveness into the organisation and administration of schools.

Some people even consider that examinations at schools are a preparation for life because "life is full of tests and trails" (Evans 1942:59) and that students who have never experienced a formal examination before are most likely to have enormous tensions when they do face one (Broadfoot 1984:210).

It can, therefore, be argued that general agreement on the desirability of complete dependence on school-based assessment for certification has not been reached. In this connection, England can be taken as an example which intends not to allow teachers "even a limited role in the certification of pupils at 16+" (Broadfoot 1984:99).

1.4.3. Will Examinations Survive?

Examinations appear sure to survive. The explanation is, in the first place, as has already been argued in the preceding section, that school based assessment cannot be accepted as a sole substitute for the examination. Secondly, despite severe criticisms, there is no indication that the importance of examinations will soon decline in the existing situation. Moreover, what seems likely is that examinations will play a maximum role, and their uses will increase (Atiyeh 1969:376; Tyler 1969:346). The argument is that no country is known to have abolished examinations or to have thought of abolishing examinations (Ranson 1984:211). Rather, Britain, for example, which introduced an examination system (GCSE) that has sufficient room for school-based assessment (which is also based on course work) has recently decided to abolish "examinations based entirely on course work, with no final test" (Independent,

9 Aug.1989, p.1). Though nothing has come out in detail, it can be understood that the traditional type of examination in the form of a final achievement test is most likely to come back.

The argument appears to agree with Morris (1972:75) who considers exams necessary instruments to ensure that the curriculum is put into effect, and Wiseman (1961:164) who believes that "good examinations are useful and desirable: without them education would be poorer and much less effective". This idea can be supported also by Scarth (1984:91) who discusses exams from the point of view of the way teachers perceive them. His study reveals that one of the reasons why the exam seems likely to continue in the future may be the teachers' own attitude towards it. In this study, he (1984:98) found teachers viewing exams as having an important enabling capacity:

providing an 'essential service' for pupils, giving teachers a basis for self-evaluation and setting up a framework for teaching, in terms of the knowledge to be presented; its ordering and timing.

If we accept the argument, then the exam appears to be not only inevitable, but also desirable.

On the basis of what has been discussed so far, the possibility of abolishing examinations can be ruled out, at least in the present situation. It can also be argued that, for practical reasons, they have been doing a good job in the sense that there must be some sort of measuring instrument which can be used to ascertain how far the proposed objectives have been achieved.

If an assessment procedure is properly devised, it can be argued, people would not disagree about having a means of checking the standard of what teachers are doing as well as

providing a benchmark "to make sure that pupils are being given the education they are entitled to, to everyone concerned the information on which to make changes and compensate if things are going wrong" (Gow 1989:23).

Here in the present study, it is not proposed to discuss whether the examination ought to continue or not. We consider the examination an inescapable procedure (at least in the Nepalese context) which should be seen in an overall context: i.e. not neglecting its psychological, social, and educational benefit. What is proposed, is to study the various aspects, defects and problems of the examination from an educational point of view. In other words, the study is devoted to the educational role of the examination so as to help it contribute to education. The idea is that if examinations cannot be avoided, efforts should be made to reduce their harmful effects on education, and at the same time, increase their potential role for educational benefit.

1.5. Role of Examinations in Education: Washback Effect

1.5.0. Introduction

In this section, the educational effect of the exam from the point of view of its washback effect will be considered. Attempts will be made first to work out what washback effect is. Then it will be argued that, in contrast with its negative connotation in the past, it should not necessarily be negatively interpreted; it is just an educational effect of the exam. In conclusion, it will be said that the exam has a strong influence on teaching and learning.

1.5.1. Washback Effect: What it is

The 'washback' (or 'backwash') effect of a test is not a new concept in the testing literature. The term has frequently been used to refer to the effect of a test on the teaching and learning of a foreign language and syllabus design (Wilkinson 1968:125; Heaton 1975:161). In other words, it is the way in which a test may influence in a backward direction. Sinclair et al (1987:93) define 'backwash' as "the backwash of an event or situation is the situation, usually unpleasant, that exists after it and as a result of it".

Although the effect of a test on education has been discussed in the language testing literature (see Wiseman 1961; Findlay 1963; Davies 1968, and 1985; Pilliner 1973; Kellaghan 1982; Alderson 1986; Morrow 1986; Pearson 1988; Hughes 1989), there appear to be no research studies which specifically focus on the washback effect of a test. Since this aspect of language testing has recently received more attention, there should be more research evidence about it in a few years' time. For example, studies going on in the ELT situations in Sri Lanka (Alderson 1986) and China (Li Xiaojun 1989) should produce more information about the effect of a test in terms of its educational benefit. The term 'washback' is still so young in applied linguistics that it has not yet entered even the specialist Dictionary (Richards et al 1985).

1.5.2. Washback Effect: Development of the Concept

The use of the term washback has passed through different stages over the years. Following Wiseman (1961:159), the term was used to "describe the deleterious effects of examinations".

Another example of its negative associations is found in the same paper. He describes paid coaching classes which were intended to prepare students for exams, and argues that those coaching classes were not worth the time, because the students were practising exam techniques rather than language learning activities. Referring to that situation, he (p.158) has written:

Educationally, it can be a disaster to some children and to some schools. Much less effort has gone into the examination of how the backwash effects might be mitigated however, than in the technical problems of allocation.

Up to that stage, one of the criteria for a good test was not to have a washback effect on teaching (i.e. not to influence classroom activities). Originally, then, the term washback would seem to have had a negative connotation. There has also been a tendency to use the term 'washback' as a neutral term (i.e. neither negative nor positive) simply to refer to the effect of an examination on education (Wilkinson 1968:125).

For Nisbet (1969) whatever is done for the preparation of an examination is its washback effect. He has written:

young children are being tested constantly in the process of teaching, and the test is real and obvious; it is the test of their mastery of simple skills. Beyond a certain standard, practical achievement is less obvious, and assessment begins to take an artificial form. This artificiality of assessment comes to be reflected in an artificiality in the student's preparation for examination - the 'washback' effect of examinations.

Pearson (1988:101) looks at the washback effect of a test from the point of view of its potential negative and positive influences on teaching. According to him, a test's washback effect will be negative if it fails to reflect the learning principles, and/or course objectives to which it supposedly relates, and it will be positive if the effects are beneficial and "encourage the whole range of desired changes". Pilliner

(1973:4) maintains the view that the most important requirement of a good test is that it should be educationally beneficial. Morrow (1986:6) terms this effect of a test 'washback validity' (i.e. positive influence on teaching) and considers this the most important criterion for a good test, especially if it is to be used as an external examination. He further argues that one of the uses of examinations is to ascertain how much of "the intended washback effect was actually being met in practice".

In the present study, we hold the view that 'washback' is an inherent attribute of an examination. Since an examination is used as an achievement test, asking students to take an exam entails teaching and preparing for it. The explanation is that an exam is administered in the situation in which teachers and students know all about it (i.e. the format, specification of the exam) beforehand, and they are also expected to work for that. Whatever is done all along the way of examination preparation is the 'washback' effect of the examination. This effect can influence the teaching and learning methods employed from beginning to end of a course if examinations require students to cover all what is entailed in the course objectives. But if an exam does not require the students to work for the whole year, the whole preparation will rest on the last couple of weeks/months before the examination.

From the discussion so far, it appears that the concept of washback, which originally had negative connotations, has emerged as "washback validity" which is considered to be one of the most important criteria for a good examination. In other words, to be a good examination, an exam should not only not

exert a negative influence, but it must also have the potential to exert a beneficial influence on teaching, where necessary. If it fails to do that, an exam is not a good exam.

1.5.3. Washback Effect of Examinations on Teachers

Teachers are influenced by what they think their responsibilities to the students are, and what the parents expect of them, in how they will teach (DES 1979:217).

Department of Education and Science (1979:247) finds examination incentives constantly at work on the teacher working for school leavers. They are:

1. the desire to gain fame for his school against the national standard
2. the desire to do the best he can for his pupils
3. the desire to satisfy the demands of those parents who are ambitious for their children.

It is further argued that the pressure from parents is much greater than is generally believed. In this connection, DES (1979:217) finds many schools interpret parents' interest as "a demand solely for good grades in public examinations".

It would appear that one of the responsibilities that a teacher thinks he/she has is to enable students to obtain good marks in the examinations. Harrison (1983:40) also finds teachers anxious about information about marking system and the way they can prepare more students more successfully for the examinations.

Wong (1969:364) and Wiseman (1961:156) hold the view that whatever the teacher thinks right or wrong, it is very difficult for him/her to deviate from what s/he thinks the exam expects of

him/her. The reason is that, they argue, even if the teacher is committed to educational aims, s/he finds him/herself trapped by the pressure of students and consumers (e.g. superior teachers, and parents) to be exam-orientated, because they all know that their students have to compete with students from other schools. That is one of the reasons why coaching for exams has been a social responsibility for the teacher (Wrightstone 1963:56, Wong 1969:365), though the teacher may not accept it wholeheartedly as the best way of doing his/her job.

1.5.4. Washback Effect of Examinations on Students

A common practice of students is to work harder when they know that they are approaching exams than when they do not (Ebel 1979:23). The reason is, Tibble (1969:350) argues, that the intrinsic value of learning is not sufficient to persuade young people to go on to the further education that a society needs. The implication is that, he holds an extreme view as he says, "nobody in his senses applies himself to learning unless there is some extrinsic reward attached to it". Though it seems an extreme statement, Wong's (1969:364) presentation of a Malaysian case supports this view. He describes how when the Malaysian Secondary Selection Entrance Examination was abolished in 1964, the students knew that their way to a place in a secondary school to pursue beyond the primary school was open. As a result, primary teachers found their students not working as they expected them to. Almost immediately the general complaint from primary teachers was that students would not work because of the removal of examinations. The idea that students had about automatic promotion reduced the need to work. Teachers' opinion

was that they could not proceed without the examination incentive. The result was that the public examination at primary level was restored (Wong 1969:364).

It is not surprising that students are vitally concerned with examinations because it has serious consequences for their future lives (Alderson 1986:99). Students know what Himmelweit (1967:364) assumes, "whether we like it or not, society will continue to use the degree as a predictive device". It has frequently been noted that a test has a strong influence on the teaching /learning of a foreign language. Alderson (1986:104) asserts that tests have a strong influence on teaching and learning, but whether the effect is negative or positive depends upon the nature and use of the test. It has even been said that teachers "teach to a test" (Kellaghan 1982:124).

Tyler (1969:342) is of the opinion that examinations have profound educational effects on students. Those students who are confident in their ability to perform in examinations find the information they are given stimulating to study. The marks given to their performance reinforces learning. Wrightstone (1963:54) finds students strongly influenced in preparation for examinations of what ability they would need to exhibit. His suggestion is that an exam should assess the ability of "understanding, generalisation, inferences, and application of principles to their ability to think critically". The effect of external examination is so strong that he (1963:54), in an intensive survey, finds:

the achievement of the students paralleled more closely the objectives tested by the Regents' examinations than the objectives given major emphasis in the local curriculum

Cliff and Imrie (1979:34) agree with Himmelweit (1967) that examinations act as a signal or exert a trigger effect; and suggests that the quickest way to change students' learning is to change the assessment system.

Agazzi (1967:42) argues that at the very least an exam obliges students to make an effort, to rethink and make a plan to put the knowledge they have into order. The students' learning is certainly influenced by examination requirements. The common practice of using past examination papers leads them to understand what is important in the course. To put it another way, an exam defines for the students the content and performance objectives of the course.

It is also true that an examination is not just a measuring instrument, it is also, as Wiseman (1961:138) argues, for the students, a very concentrated period of practice in the "skills involved and an opportunity for learning".

On the whole, it must be accepted that the idea of how to get through the exam, and how to help students to get through exams, considerably influences what goes on in the classroom. It also influences students' learning activities outside the classroom in the sense that the way teachers prepare students for the examination by giving them homework, and the way students do exam preparation at home, are not unaffected by examinations. In this way, examinations have a very strong influence in inducing teachers and students to work.

1.6. Examinations as Instruments for Educational Benefit

1.6.0. Introduction

This section considers whether the attribute associated

with the exam is a contingent or an inherent factor, regardless of its negative or positive effect on education. It will be argued that whether the washback effect is negative or positive depends upon whether the particular exam has been designed according to the purpose for which it is intended. If the exam is designed in consonance with the course objectives and the purpose for which it is designed, it will be argued that the washback effect will be educationally beneficial. For that purpose, it will be suggested that there should be a good combination of school-based assessment and final exam.

1.6.1. Undesirable Effects of Examinations: A Contingent Factor or an Inherent one?

To recapitulate a few of the undesirable effects of the examination, they are: narrowing down the curriculum into score producing areas, study of previous exam papers, leading to guesswork and memorization. To put them together, we would call it 'teaching for the exam'. Following Ebel (1979:8), this can be interpreted as "attempting to fix in pupils' minds the answers to particular test questions". In other words, teaching for the exam may mean preparing students according to what we think students need to do in order to get through the exam. We have already discussed, in section 1.5., that students and teachers cannot avoid working for the examination because it has serious consequences for them. It is also considered a dominant criterion for evaluating schools, whether they are doing what they are supposed to be doing. This is what leads the school to adopt a strategy: e.g. teaching for the examination.

Now the question is: Is teaching for the exam desirable

or not? To answer this question, it is necessary to study whether a harmful effect of an exam on education is a contingent factor or an inherent one. It further leads us to study what it is that makes an exam responsible for having an undesirable effect on education. The point is that if the undesirable effect of the teaching for the exam is inherent, then, apparently, it is undesirable. If the undesirable effect of the exam is a contingent factor, it can be hoped that under certain condition the harmful effect can be reduced, and efforts can be made to exploit the energies that the teacher and the student invest in preparing for the exam for educational benefit.

If we look at the criticisms (section 1.4.1.), it would appear that all they are rooted in the same problem: the efficiency of examinations as measuring instruments. The main issue to be discussed is twofold: challenge to the efficiency, and effectiveness of the exam. The challenge to the efficiency of exams may mean that there have been doubts about what they measure. In other words, they are wrongly designed in relation to the purposes for which they are intended.

The challenge to the effectiveness of exam is the issue of its educational effects. Since the effectiveness is the product of an exam, the problem rests on its efficiency. Then it would appear that the accusation that an exam produces a negative influence on education is mainly based on the design procedures.

In that case, it can be argued that one possible explanation for why an exam is said to have a harmful effect on education is, as Nisbet (1969) suggests (quoted in Heywood

1974:17), that:

We set examination questions which require the student to have at his fingertips a large number of details, accurately remembered dates, processes, articles and so on, then of course the student will spend much of his time preparing for that examination by memorising these facts and figures.

Morris (1972:81) also criticises existing exams on the ground that they emphasise too greatly the acquisition of reproducible knowledge in the form of facts. Issac (1964:136) follows the same pattern when he has written:

We must ... point out that our current forms of examination fail to assess those qualities which the educational system should be able to develop in its students; the capacity to understand the relevance of knowledge to the problems which will confront them in their professional lives; the capacity to analyse and synthesize

If this type of exam prevails, it is true that teaching for the exam is undesirable. But it does not necessarily mean that the exam as an instrument should have a harmful effect when there is something wrong with its design procedure. If we look at Kingdon and Stobart's (1988) study in which they discuss the GCSE's implications for schools, some evidence can be found to defy the challenge made to examinations. Kingdon and Stobart (p.21) say:

From the start the new Government demonstrated a lack of confidence in the teaching profession and set out to use the school exam system as a method of influencing what went on in schools.

In assessing the implications of the GCSE for the schools during the secondary year of GCSE preparation, they (p.87) find:

emphasis is already moving away from a demand for text-book to on more resources in the form of well-prepared source materials, worksheets, and audio visual aids.

That means the exam as an instrument does not necessarily have to have a harmful effect on education. If any exam has a negative effect on education, it may be because of the nature of

the particular examination.

1.6.2. Examinations for Educational Benefit: Positive Washback

Kingdon and Stobart's (1988) study, as was mentioned in the preceding section, gives us an impetus to search for the possibility of using an examination for educational benefit. In that case, can we not use the exam for educational change?

Nisbet (1969) (quoted in Heywood 1974:17) finds it possible to use the exam for educational change. The condition for that, he argues, is:

If the examination is one which requires the student to argue around a question, to put forward the pros and cons of a theory, then he will spend a great deal of time studying theoretical issues from this point of view. Therefore, by choosing very carefully the kind of question we ask, we can control the behaviour of the student and force him to study in the way we think desirable.

Issac (1964:136) considers the exam as a potential instrument to bring about changes in the student's role and the teacher's teaching methods. He has written:

If the questions in the exam are designed to exercise the students' creative imagination to implicate the students' personality in his work, to develop the abstraction of his thinking and to emphasize the social and economic contexts in which he will eventually work, then the students will be put under pressure to accept greater responsibility and a more active role, the teacher's relationship with the student will take a different form and teaching methods will change.

To provide a positive washback effect, the exam should play a role as a leader. Since an exam is a check on achievement of course objectives (Davies 1985:8), a valid exam assesses whether the students have achieved the proficiency as conceived in the course objectives. This implies that it is necessary for the exam to reflect what ought to have been formally presented in the classroom. In this connection, Davies (1977:42) points out:

an exam gives students a real sense of purpose by describing exactly where they are meant to be going.

The implication is that it is the content and format of the exam that determines what should be going on in the classroom or how the syllabus should be used to achieve the course objectives. The student is generally concerned with what is being taught and what he thinks will be asked in the examination. The idea of passing the exam takes precedence over learning unless the teaching syllabus and the testing syllabus become the same. Then it seems necessary to match the curriculum and examination targets to pupils' needs. While the demands of examinations may be taken as a guiding factor in work at the school leaving stages, it should not be incompatible with the wider educational aims of the course. To require that schools follow the educational objectives, it is the exam that has to take the lead.

Failing to do that, Carroll (1973:17) argues;

It is only natural for students to shape their learning efforts so as to be maximally successful on tests, and if the tests measure objectives that are in some ways different from those of the instruction, students will work towards those objectives and pay less attention to achieving other objectives.

It has already been argued that one of the functions of the examination is to reflect the course objectives. Hughes (1988a:42) emphasizes this argument when he argues:

One function of testing is to provide the kind of information that will help keep its partner on the right track. It can best do this when achievement test content is based not on the syllabus and textbooks but on course objectives.

It is, therefore, argued that the examination should work as a leader to attract classroom teaching and syllabus to itself by requiring the students to practise the activities which are congruent with the course objectives, as well as necessary to pass the examination. In this way, an examination of this type,

which reflects the objectives, requires classroom teaching to follow its leadership.

Hughes (1986:33) discusses the washback effect of a criterion-referenced test. He argues that the main advantage of a criterion-referenced test would be its "potential influence on teaching and learning - its washback effect". He goes on to argue that the advantage of a criterion-referenced test is that:

the requirement of direct interpretation entails the performance on the test of tasks which are directly related to course objectives.

Then it can be hoped that the teaching and learning will be accelerated to the attainment of authentic objectives. As long as this situation prevails, teaching for the exam develops into teaching for genuine objectives.

By the same token, if the examination fails to require the examinees to exhibit the abilities as envisaged in the course objectives, it can also fail to ensure that the teaching encapsulates the spirit of the course objectives. Then the washback of the exam will be negative.

Given that the objectives of a course are made explicit to a reasonable extent, it can be argued that an exam should represent the course objectives in such a way that whatever is done to help students pass the exam (i.e. teaching for the exam) will contribute to achievement of the course objectives. In such a situation, it is obvious that those students who pass the exam cannot avoid having achieved the abilities envisaged in the course objectives.

We can take the Sri Lankan ELT situation as an example. In Sri Lanka, since the introduction of the National Certificate of English (NCE), sample test materials and a teacher's guide to

the NCE have been produced, including the rationale and specifications of the exams. The guide also includes a range of recommended learning activities which would not necessarily appear in the examination. Here the guide has two functions: examples of how teachers should prepare students for the exam, and how exam preparation can be used to promote language learning. The underlying principle of this type of examination is that the texts used will not appear in the exam, so there is no value in memorising them; but they are important from the point of view of developing the necessary language abilities in the students in order to enable them to do well in the examination (Alderson 1986:105).

Nothing more can be said before the effectiveness of this type of instruction is measured. However, on the basis of the arguments presented in the study, there are grounds for optimism.

1.6.3. Need for a Good Combination of the Examination and the School Based Assessment

Most criticisms made against an examination are based on the assumption that an exam is a 3-hour test representing a given syllabus. The examination is intended to be a 'sampling' to elicit information from the pupil in order to become familiar with him, determine his level of development, assess qualitatively and quantitatively how far he has prepared himself. At the same time, it is also a test of the effectiveness of the teaching he has received which would also, for administrators, or the society as a whole, work as an assessment of the knowledge and ability of his teachers: i.e.

efficiency of the schools (Taylor and Richards 1985:146).

If we go back to section 1.5., we will see that, as long as the exam is properly devised, and the information collected from it is properly used, the exam can serve very useful purposes. Nevertheless, there is room for doubt whether a 3-hour exam can provide a basis for all the fine purposes it is expected to serve.

It is possible for a 3-hour exam to gauge some functional abilities that the student is expected to have learnt from a course. These abilities are only samples of all that has been included in his/her education. The point is that there are other aspects - physical, social, intellectual, emotional - which have also contributed to his/her educational development. These aspects cannot be appraised in a 3-hour examination.

As was discussed earlier in section 1.4.1., a 3-hour exam has been criticised for poor sampling of the behaviour that a course purports to develop in the student. It has also been argued that the teacher and the student tend to ignore the behaviours which are not appraised. In that case, a 3-hour exam will not do any good from the educational point of view, mainly because it is nearly impossible for it to assess all the objectives proposed by the course.

Needless to say, testing plays an important role in teaching (see Heaton 1975; Barruta 1967; Abdrabou 1984). From the educational point of view, testing is an integral part of teaching. It has played a dominant role in the development of Applied Linguistics, and has spawned a vast literature. We do not propose to review that literature in detail. What we would

like to stress is that, since it has been evident that an external testing program has a strong influence on school-based assessment (see Kellaghan 1982:246), the exam's other potential role in education should be to facilitate the teacher to use a test as his useful ally.

This again leads us back to the discussion of the merits and demerits of a school-based assessment system. Whatever the criticisms made against school-based assessment, on the basis of the discussion presented above, it can be said that the whole education process can be handicapped if the school-based assessment is not accelerated. It is, therefore, argued that there should be a 'good combination' of the exam and the school-based assessment. In other words, provision should be made for internal assessment with external control so as to ensure that the expected standard has been maintained, and that people's confidence in the certificate has been preserved. Another important thing which must be done is to give the student the belief that every piece of his/her coursework contributes to the exam: i.e. provision of continuous assessment. Taylor and Richards (1985:146) maintain that an assessment system which has a wide spectrum is less likely to distort the curriculum than the one which has a narrow spectrum. On the whole, the intention behind the exam should be to facilitate teaching and testing rather than determine them.

1.7. The Role of Examinations in Educational Change

1.7.0. Introduction

In the preceding sections, the role of the examination in terms of its washback effect has been discussed. It has also

been argued that the exam can have a beneficial effect on the way teaching and learning take place in the schools.

Nevertheless, it must not be forgotten that improvement in the exam is not a panacea, it is only a tactic. The point is that a desired improvement in education is possible only after bringing about changes also in other elements of a curriculum: content, methods, and objectives, in order to make them all go together. What is advantageous in adopting an innovation through an exam is that it can play a significant role in speeding up the movement of its sister elements towards that direction. It is this issue that is taken up in this section.

1.7.1. The Interdependence of Curriculum Elements

It is necessary to understand the nature of the interdependence of the elements of a curriculum in order to understand what effect a decision made about one of these elements can have on the others. A decision made without understanding the interrelationship of the elements is bound to be faulty. Tharu (1981:1) has pointed out that several educational reforms have been unsuccessful in the past in obtaining the desired effects. One of the reasons, he argues, is that the reforms 'remained isolated and piecemeal operations', so the effect of the reforms was only on the surface level of the system. Keeping this view in mind, he suggests that persons involved in curriculum development should keep 'the whole system in perspective and appreciate the functions of the various components' (p.2) in order to obtain a desired change.

In the literature (see Audrey and Nicholls 1985; Taba 1962; Taylor and Richards 1985; Yalden 1987; Clark 1987), the

following are often mentioned as the curriculum elements:

- 1.determination of objectives,
- 2.selection of learning experiences,
- 3.organisation of learning experiences,
- 4.evaluation.

Nevertheless, it has often been argued that curriculum innovation, in practice, does not follow the step-by-step procedures listed here. The argument is that whenever a decision about any one element has to be taken, this cannot, and should not, be done solely based on that element. It is necessary to make reference to the other elements as well, because the different components of the curriculum are interwoven with each other. They should, therefore, be treated as four interdependent elements of the same body. In this connection, Taba (1962:425) presents these curriculum elements in a diagram as follows:

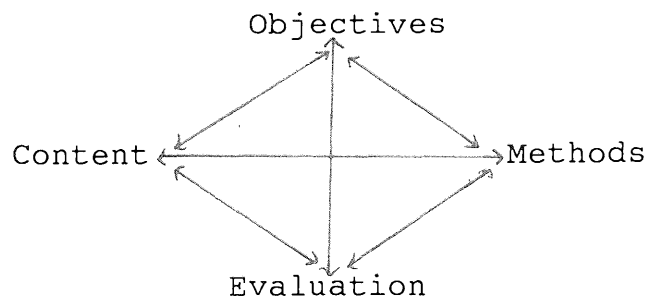


Figure 1: Elements of curriculum

Here the four elements - objectives, content, methods, and evaluation - are described in an interacting scheme. The arrows indicate the ways one element interacts with the others.

Following this model, curriculum development is not merely a sequential process in which the elements can be treated individually at different stages. Rather these elements form a constellation in which each one is in a relationship of

interdependence with the others. Accepting this view means denying the discreteness of the four components as well as denying their sequential stages of development.

The argument for the denial of the sequential process of curriculum development is that, as Audrey and Nicholls (1983:96) state, 'there is constant moving backwards and forwards'. They go on to argue:

in considering content, there is constant reference back to objectives and forwards to methods. In considering methods constant reference is made to content and objectives. Evaluation cannot be considered without reference to objectives and content.

To sum up what has been discussed so far, it would appear that there is general agreement that the curriculum elements are usually discussed under separate headings for the purpose of analysing and clarifying the major points about which decisions are to be made in the curriculum innovation. However, they are not discrete elements in terms of the way they contribute to the curriculum innovation. In other words, they are interdependent. The implication is that whenever a decision about any point needs to be made, whatever the element, it should not be forgotten that the affected element has to work in combination with the others. It will not work if it is a piecemeal operation.

1.7.2. Strategy for Change

A strategy for curriculum change is largely based on the reason why change is felt necessary. An innovation in a curriculum presupposes that there is already a curriculum being followed. So the reason why a change in the curriculum is being sought leads those involved to look at why the existing educational program is felt to be inadequate. The reasons could

be that the current curriculum has not been working as expected for unknown reasons, or as Taylor and Richards (1985:44) point out:

changing conditions, such as those brought by socio-economic pressures, or by the 'knowledge explosion', necessitate revision of what ought to be taught, the way it ought to be taught, and the way it ought to be assessed.

And, sometimes, both reasons can be prominent. Whatever the reason, it is necessary to examine the educational program being followed in order to identify the areas of concern. The first step in constructing a strategy for an innovation in education is, therefore, a critical analysis of the elements of the present curriculum.

It is also necessary to look at the existing educational program from the point of view of current practices in that discipline, because education is always changing, and the responsibility of an institution is to provide its pupils with an up-to-date education as far as possible. Davies (1985:5) maintains that an educational program must make a provision for accommodating the new ideas and demands.

The next step in that direction is "a systematic sequence of work which deals with all aspects of the curriculum ranging from goals to means" (Taba 1962:455). In that case, a decision has to be taken about where one should begin. The issue is: what element should be considered first to give the best contribution to accomplishing the task of keeping the whole system (e.g. all the elements of a curriculum) in perspective while making decisions about any of the elements. This issue is crucial in curriculum innovation in the sense that the whole business of innovation can be accelerated if it begins with an element that

can work as a potential catalyst.

1.7.3. An Examination as the Element to Begin with

Davies (1985:7) argues that change in education could be effectively implemented through the 'syllabus and the examination and the teacher'. In case of making a choice between them 'in order to move quickly', then it is obvious that 'the exam is the most sensitive; it is the most controllable,... it has most certainty in terms of its goals'.

He (1985:7) goes on to state:

The test/examination is a major and a creative influence for change and development in language teaching, and if there is a need to choose, then that is what should always change first.

Morris (1972:82) supports Davies' view. According to him, reform in school work must begin by reforming examinations. His suggestion is that the desired objectives in teaching should be identified first, then they should be embodied in the content and format of the examinations.

Change in education is not once and for ever. It is a continuous process in the sense that a progressive education system is expected to assimilate new ideas and approaches recently innovated. The education system, therefore, must have room to effectively incorporate new ideas so as to circulate them through all the elements. Following Davies (1985:8), an exam appears to have potential to serve this purpose:

creative and innovative testing, starting often in a proficiency guise,can, successfully, attract to itself a syllabus change or a new syllabus which effectively makes it into an achievement test. Indeed, it could be said that the proper accolade for a good proficiency test is that it allows itself to be outdated- it becomes the achievement test for a teaching syllabus and thereby permits a new proficiency test to be constructed, more appropriate for developing ideas of language teaching and learning.

Morris (1972:83) considers the exam in a similar way. He has

written:

To attempt to improve our teaching at the present time of education scarcity, 'through the use of examinations as instruments of reform' is therefore not merely to undertake a task made necessary by existing evils but is to contribute directly to the fundamental and long-term improvement of education.

Another advantage of beginning a change in education from an exam is as Tharu (1981:2) argues:

the insights gained while making and using tests may lead to improvements in the quality of the learning experiences and the clarity and practicality of objectives.

1.7.4. The Role of an Examination Within the Teaching

Syllabus: Four Case Studies

To make the aforementioned points more explicit, four case studies, specifically related to English Language Teaching situations, are presented below. Two of them (i.e. the Chinese, and the Turkish situations) deal with positive changes brought into teaching syllabuses through examinations. The other two cases (i.e. the Malaysian, and the African situations) are examples which deal with the other side of the coin: the negative influence of an exam in restricting changes in the teaching syllabuses.

Case Study 1

Li Xiaoju (1989) assesses the washback effect of the Matriculation English Test (MET) after four years of its implementation in China. The MET replaced the National Unified Enrolment Examination (NUEE), which was founded in the early 1950s. One of the reasons why the NUEE had to be replaced by the MET, Li Xiaoju (1989:5) writes, was that:

complaints about its validity kept coming in, citing cases of candidates gaining fairly high in NUEE English scores but subsequently doing rather poorly after being enrolled.

In 1984, the State Education Commission decided that the NUEE be reformed on the assumption that the change in the exam would lead the school teaching to compensate for what caused the students to perform differently in the NUEE and after being enrolled (as was quoted above). Gaining insights from this, the MET was designed to require the students to exhibit 'the knowledge of all the structures found in the unified English textbooks' as well as 'the ability to use those structures in relevant sociolinguistic contexts to serve communicative purposes'. In a word, the intention behind the MET was to emphasise the use of the language rather than the knowledge of it.

In 1987, a survey was administered to investigate "how much has been achieved in the way of making people realize the need for change and in bringing about changes itself" (p.11).

Li Xiaojun (1989) describes the following as the positive washback effect of the MET:

1. Teaching materials have been expanded to include a greater use of imported and self-compiled materials.
2. There have been changes in teachers' approaches to what is to be emphasised in teaching in the classroom. For example, reading skills attract the most attention because they are most heavily weighted in the MET. Translation, which used to be dominant in the UNEE, on the contrary, has been excluded from the new format.
3. Students have been found to be conscious of using their time and resources for learning English. The explanation she gives is that in one province it was revealed that all simplified

English readers were sold out after the province participated in the MET.

4. She reports that there has been change in the teacher's attitude about what to teach and how to teach.

On the basis of the survey, she has argued that the MET has proved to be powerful enough to bring about some positive changes in a short period in a country like China.

Case Study 2

Another example to show that an examination, provided that it is designed properly, can bring about changes in education is taken from Hughes (1986). Bogazaci University in Turkey accepts students on the basis of academic ability irrespective of proficiency in English. What the university does for those whose English is considered inadequate is to require them to enter the University's Foreign Language School (FLS) for one or two semester/s.

Despite the English medium nature of the university, most of the subjects were taught in Turkish because the students' English, even after completing the FLS, was so inadequate that it was impossible to lecture successfully in English. Their English was also inadequate for reading and writing purposes at the undergraduate level. Hughes offers the following explanations for why even those who attended the FLS did not have adequate English to cope with the University education:

1. the courses, largely grammar based, were not wholly directed to the development of the English skills that the students would need as undergraduates.
2. the FLS's testing system was such that, despite the inadequacy of many of its students' English, 99 percent of those completed their time there were allowed to pass on to their university studies proper.

In this situation, the university decided to introduce an English proficiency test at the end of the FLS courses with the 1982-3 academic year. It was also decided that those who passed the test could proceed to their academic studies, and those who could not pass had to leave the university. The idea of introducing the test was to require the FLS to equip the students with English language skills so that the university could maintain the English medium standard. For that purpose, a test was to be constructed, and Hughes was responsible for that. A criterion-referenced test with the three components - listening, reading, and writing - was constructed to create "the greatest possible backwash" (p.35).

After the introduction of the test, its washback effect was assessed. Hughes (1986:36) describes the washback effect as follows:

There was an almost immediate change in syllabus and materials to ones more obviously related to the development of the language skills needed by university undergraduates.

The effect of the changes made in the syllabus and materials on the standard of English attained in the school was also determined by using the written part of Michigan Test of English Language Proficiency. He describes the results:

2. Before 1982-3, the percentage of students reaching the recommended Michigan minimum for study in their field had always been below fifty, some years far below. At the end of the second semester of the 1982-3 session, however, 72 percent of students had reached the minimum for arts subjects... Of all students, completing their FLS course, 86 percent passed the new test and were allowed to proceed to their academic studies. (p.36)

The university teachers' reactions to the students' standard of English, after the implementation of the test, was collected by means of a questionnaire, which confirmed that the first year students in the 1983-4 session "markedly more proficient than those of the previous year, sufficiently so for teaching to be conducted in English once again" (p.36).

It is evident that a test, if properly designed, can act as a catalyst.

Case study 3

Davies (1985) discusses the ELT situation in Malaysia.

Some years ago, the Curriculum Development Centre (CDC) introduced a new communicative syllabus into the secondary schools, intending to integrate the two language-medium systems, English and Malay, as quickly as possible. Another purpose behind that was to develop in school graduates the ability to communicate in English. The Examinations Syndicate, a separate agency from the CDC, did not show its willingness to change the examination so as to make it congruent with the syllabus. The result was, as Davies (1985:7) describes, a disaster. He has written:

the examination did not in any case test the syllabus. Failure was severe except among those from the more elite English-medium schools; many average English-medium students and most of the Malay-medium schools failed. Remember that the purpose of the new syllabus was to help those very Malay-medium students who had had no access to English medium, and that English medium was then on its way out.

There could have been several factors responsible for the disaster. One of them was certainly the mismatch between the syllabus and the examination. As Davies (1985:7) says:

it was right to be concerned about promoting appropriate use of English, but wrong not to check with its sister ... examination institution. ... Change is essential but it needs a fabian lead.

Case study 4

Madsen (1976) discusses the school leaving examination in a large African state (which he does not name).

There was severe criticism of the essay and precis in the exam for leading the teachers to spend most of the time on practising exam techniques rather than on the English fundamentals which the students needed. The people concerned felt that something had to be done in that direction. As a

result, the exam was objectivised, the relevant courses changed, and in-service training was initiated. For a short time, the exam reflected the necessary changes needed.

However, since the idea that an exam has always to take a lead so as to incorporate the new ideas and demand in the curriculum had not been institutionalised, the ELT situation again stagnated. The lack of the idea led to the similar type of problem again. Madsen's (p.138) further explanation is:

when strict behaviourism began to give way worldwide to cognition and more integrative evaluation, the tightly controlled and largely discrete-point objective test remained unchanged in this African state. Teachers in the upper grades were inclined to model instruction on the now sacrosanct objective examination... the backwash effect on the schools became just as devastating as that produced by the earlier precis-essay examination.

In this situation, the negative washback effect of the exam was not inherent, but contingent: "the exam became 'petrified' and therefore sacrosanct" (Pearson 1988).

1.7.5. Conclusions

What seems to be emerging, on the basis of the discussion presented so far, is that an exam can play a potential role in accommodating the new ideas in it, first, and then to convey them to its sister elements effectively. It may also mean that the exam has the potential to contribute to keeping the whole system in perspective when a change has to be integrated in it. When looking at the other side of the coin, on the basis of the case studies, it appears that a change in a teaching syllabus, if the change is not incorporated in the exam, or if an exam does not take a lead, may not take place effectively. To conclude, it can be said that change in education can be effectively implemented through examination strategy.

1.8. Summary and Comments

In this chapter, the theoretical basis of the present study has been considered. Reviewing the testing literature, it was revealed that most of the criticisms made against the examination are based on, first, the design procedures, and then its negative washback effect on education. However, despite severe criticisms, no symptom was found to indicate that the examination was going to be abolished. Rather, it was most likely that the examination will continue to play an important role in education. It was, therefore, concluded that the examination will survive for the foreseeable future.

The washback effect of the examination has been discussed. It was found that the concept of the washback effect, which originally had negative connotations, has acquired positive connotations. Over the last few years, it was found to be, and is regarded as, one of the essential qualities of a good examination. On the basis of the discussion, it has been argued that an undesirable washback effect is not an inherent characteristic of an examination; rather it is conditional on its design. On these grounds, it has been concluded that if an examination is carefully designed in congruence with the course objectives, and in relation to the purpose for which we want it, it can work for educational benefit. It has also been argued that the positive washback effect of the exam can be increased by a good combination of exam and school-based assessment.

Finally, the role of the examination in educational reform was considered. Looking at the elements of a curriculum, it was revealed that although they look discrete on the surface,

they work interactively in educational reform. It was, therefore, suggested that whenever a change in education is desired, all the elements must be kept in perspective.

Based on theoretical considerations, and the insights gained from the case studies presented, it has been concluded that the exam has the potential to take a lead in the process of educational change. By the same token, we have also warned that ignoring the strong influence of the examination may result in stagnation in the process of reform.

CHAPTER 2: ENGLISH LANGUAGE TESTING IN NEPAL

2.0. Introduction

This chapter is devoted to a discussion of English language teaching and testing at school in Nepal which is the context of the empirical investigation of the study. The chapter has four sections.

Section 1 considers the necessity of English language teaching in the Nepalese educational setting. We discuss whether English should be taught as a compulsory or an optional subject in the schools. It will be argued that English as a school subject should remain compulsory as it is now.

In section 2, we concentrate on the teaching of English at school in Nepal. We present the existing ELT situation in terms of the types of schools, the methods of teaching, the textbook materials, and the school leaving examinations used at school leaving stage.

In section 3, we discuss the efficiency of an exam in terms of validity, reliability, and practicality criteria. Comments on the teaching of English, and the School Leaving Certificate (SLC) English examination will be presented. It will be argued that the existing teaching and testing of the SLC English is not adequate for the purpose for which English is taught. The problems with the existing ELT situations will be stated. It will be argued that the system needs change, and the pace of change can be effective if the innovation is led by the SLC English exam.

2.1. English Language Teaching In Nepal

2.1.0. Introduction

In this section, we present a general description of teaching English in Nepal. We argue that Nepal needs English for educational and occupational purposes if we want the development of different aspects of the nation to take place. Since the existing ELT situation faces a lot of problems in teaching English as a school subject, some people involved in teaching argue that the problem can be solved by changing the status of English from a compulsory to an optional subject. But we will argue that English should remain as a compulsory school subject. We accept that there are problems, but that they should be solved by changing the way it is taught rather than by changing the status of English.

2.1.1. Linguistic Situation

Nepal is a mountainous country surrounded by the Tibetan Region of the Peoples' Republic of China on the North, and by India on the other sides. Geographically, the country can be divided into three regions: the mountain region, the hilly region, and the terai region.

The population of Nepal is approximately seventeen millions. The three regions have approximately 10 percent, 52 percent, and 38 percent respectively of the country's total population.

Nepal, the only declared Hindu Kingdom in the World, presents "a semblance of cultural homogeneity" (Hutt 1988:1), and linguistic diversity. Nepali - a member of the Indo-Aryan language family, and the mother tongue for more than half of the total population - is spoken by more than 80% of the population (Hutt 1988:5). This is the official national language. There are

several other languages spoken in Nepal. In that sense, Nepal presents an example of linguistic diversity which Turner (1928:63) puts in the following words:

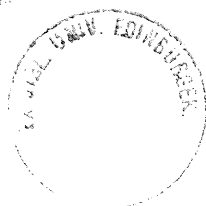
In a population of under six millions in all there are spoken at least scores ... of languages, all mutually unintelligible, some broken up again into numerous and often very different dialects. Even within the limits of a single valley there may be a village the inhabitants of which speak a language completely unintelligible to their neighbours a mile or two away.

Though it was said some 70 years ago, the situation has not changed very much. The language policy of the Nepalese Government is claimed to be liberal and democratic (Hutt 1988). In other words, the Government has not adopted a policy against the development of any languages being spoken in the country. However, its policy about Nepali is clear one way or another, as Hutt (1988:5) observes:

this same language appears in the present circumstances to be a major factor favouring the promotion of national unity and the Government believes that a working knowledge of the language is utterly essential for every Nepalese citizen... The intention of the Government is that, through the establishment of a Nepali medium up to the middle standard in Upper High Schools, the capabilities of Nepalese citizens and officials will be enhanced in the age of progress and democracy as a result of their increasing proficiency.

As a result of the policy, Nepali is taught as a compulsory subject right from the beginning of formal education (age six) to tertiary education. The medium of education is Nepali in most of the schools. However, some private schools use English as the medium of instruction. In any case, all the children who go to school use Nepali extensively.

Putting the importance placed upon Nepali aside, the major concern of formal education has been English because of the extrinsic values associated with it. The growing concern with English constitutes the subject matter for the following sections.



2.1.2. The Need for English in Nepal: A Rationale

In Nepal, there is a saying about English: viz. English is the language across seven seas, meaning - that it is a very different language from Nepali, a language from very far, therefore very difficult to learn. Despite the linguistic and geographical remoteness of English, Nepalese people are desperate for it, and have a close affinity with it. In other words, Nepalese think they need English. The need for English in Nepal can be well understood if we look at the Nepalese situation in the light of what the International Meeting of Specialists (held in London in December 1960) on 'Second Language Learning as a Factor in National Development in Asia, Africa and Latin America' agreed upon:

The language problem in development stems from at least three communication needs which are increasingly being recognised both in the developing countries themselves and in other countries aiding their development: internal communication, transmission of science and technology, and international communication.

(Quoted in Malla 1975:11)

The statement appears to be most relevant in connection with the need for English in Nepal. Since Nepali serves only the internal communication needs, Nepalese desperately need English for two reasons: transmission of science and technology, and international communication. If we put 'the need for English for transmission of science and technology' in a wider perspective, it would appear that we need English for educational purposes. The need appears to arise when we find the Nepali language inadequate to provide the Nepalese students with an access to the knowledge of science and technology of the modern world. As the nation needs an enormous number of people to be involved in development, English can be seen to have an important role to play in the whole education system. To put it in Malla's

(1975:18) words, "English is one of the preconditions for promoting technical education".

Looking at 'the need for English for International communication' from an educational point of view, it would emerge that we need English for occupational purposes. Since the Nepali language is not adequate for International communication, all the manpower involved in the modernisation process of the nation need English. In Nepal, it is only English which is used for international communication.

On the whole, it seems that the purpose for which we need English in Nepal can thus be well put under the following headings: educational purposes and occupational purposes.

2.1.2.1. Educational Purposes

English is necessary in school education for educational purposes. Since one of the purposes of secondary education is to prepare students for higher education, the students who complete their school education are expected to have achieved an adequate knowledge of English in order to cope with the higher education. In the higher education, almost all Faculties expect their entrants to have a working knowledge of English one way or another. Faculties like Engineering, Agriculture, Medicine, Forestry, and Science and Technology use English extensively. Most books are in English. Most subjects are taught in English medium. The schools introduce some English courses; e.g. English for specific purposes. To pass any level, the students have to pass all the subjects they study. Faculties like Humanities, Management, Education, and Law use English a little. They introduce compulsory English courses. Good books for some of the content courses offered in these Faculties are available only in English: e.g. Accountancy, in the Faculty of Management. In this

academic environment, those college entrants who do not have the command over English expected of a school leaver can have enormous difficulty in coping with higher education.

Books for college courses are available also in Hindi, a language widely used in India. Those Nepalese students who have difficulty in reading books in English tend to depend upon books in Hindi though the Hindi books are considered to be of lower quality. However, the Government's language policy does not encourage teaching Hindi as a school subject because of political and social reasons.

2.1.2.2. Occupational Purposes

Though it has been mentioned that Nepali serves as an official national language (section 2.1.1.), in fact, the need for English for occupational purposes is growing. No mention has been made of the fact that a working knowledge of English is necessary for any Government jobs. However, in practice, it was revealed, in Verma and Pande's (1988) study, that the Ministry of Foreign Affairs, the Ministry of Trade and Commerce, and the Ministry of Tourism need English. The Public Service Commission also expects school leavers to have some command over English. Presumably, travel agencies, hotels, foreign missions, air lines also need English.

It is, therefore, true that every educated person in Nepal wants, and is expected, to have learnt some English. Davies et al (1984:47) find the same situation; they were told that, in Nepal, to be educated means to know English. This is how English is associated with social prestige.

2.1.3. English as a School Subject: Compulsory or Optional?

As has been discussed in the preceding section, the need for English in Nepal appears to be associated with national

development in general and individual career development in particular. In that sense, every educated Nepalese needs English one way or another. Since it is generally difficult at the beginning of education (e.g. school stage) to decide who has the potential to contribute to the development of the nation, and who does not, and the way an individual wants to develop his/her career, every educated Nepalese needs English. Malla (1975:12) perhaps accepts this situation, since he argues:

English is undoubtedly of vital importance for accelerating the modernisation process in Nepal, and in so far as education is an agent of such a process the place of English in Nepalese education as a foreign language is secure and unassailable.

It would seem that English is necessary for those who want, and are expected, to contribute to the modernisation process of the nation. It can be argued that a nation should provide its citizens with an equal opportunity for education. However, on the basis of the English Proficiency of school leavers, and the number of SLC candidates that fail the exam, people involved in the field of ELT in Nepal put a question mark on whether English should remain a compulsory school subject. They tend to argue that the situation can be improved if the status of English as a school subject is changed to an optional subject. This issue has been a dispute in Nepal for a long time. It seems important to make our stand clear on the issue.

In respect to this issue, Malla (1975:17) takes the view that English should be an optional subject at secondary education. For example, he suggests:

English is not relevant for the primary and lower secondary stages. It is relevant at the secondary stage only for those who hope to go on to the university stage.

Malla (1978:6) puts this idea more strongly when he argues that:

there is every reason to give English to as few as possible and as late as possible; but give it as

effectively as possible.

What Malla has said about the irrelevance of English at pre-secondary education can be agreed with provided that English is taught effectively at the secondary stage. But the point he makes about making English optional by making it available only for those "who hope to go on to the university stage" is arguable on practical grounds. How is it possible to decide in advance who hope to go on to the university stage and who do not? What are the criteria to be adopted in distinguishing them? The danger with this type of looseness in offering English is that it would lead the children from under-privileged classes to avoid studying English mainly because, at the school stage, they may not have long term aims. They consider education an instrument for getting a job either immediately after they complete their school, or after they complete their tertiary education. They could, of course, have a desire to go on to university, but their financial situation does not permit them to think of doing that when they are 14-16. It is also true that very many students are afraid of English as a school subject because the poor results in School Leaving Examination are ascribed to English (see section 2.3.1.1.). Considering this situation, it can be argued that many students from low income groups may think that studying English is a risk for the exam, so they may avoid it if there is any choice. For immediate benefit, they would look for an easy way to get an academic licence in order to earn their living.

Rogers (1982:146-150) also argues that English as a second or foreign language should only be taught to those who are going to use it. In other words, if English is taught indiscriminately to all those who go to school or college, the result will be a

high percent of drop outs. He also finds the teaching of English in many countries - Pakistan, Bangladesh, and some African countries - associated with a high income job. He criticises this situation for developing a false expectation in students. On this ground, he suggests that English should be taught to those who need it.

From the logical point of view, we agree with Rogers. But as has been shown about, we can not find ways to determine which of our students will need English and which will not. Neither we nor the student himself, can say what he is going to do after his school education because what he will be doing depends upon many factors. In this situation, what we consider important is that we should not make anyone regret not doing some English at school in his adult life.

In reality, every student can have a desire, and potential to do some English so that s/he can look forward to a bright future. It is only the students' socio-economic situation that leads them to avoid studying English. Letting them choose whether they want English as a school subject or not may have a harmful effect in the long term on the development of the nation because the result of not studying English at school may restrict their potential in their future work. It should also be accepted that students of school age cannot give a proper direction to their lives. In other words, they may not be mature enough to envisage the harmful effect of not studying English on their future careers. By the time they realise that they should have studied English, it will be too late to do anything about it.

Another problem in respect of this issue is that, if English remains an optional subject, many schools in the country side will drop it because of the lack of trained teachers, and

the poor result in the SLC exam. The explanation is that, unless the Ministry of Education pays special attention to English teaching, schools in the remote areas would offer only those subjects which students do not find difficult to pass, as results in SLC have an effect on the financial support of the schools. In that situation, those potential students who cannot afford to pay for education away from home will have no opportunity to study English. If this situation prevails, the privileged class will have control over education at the cost of the under-privileged. As a result the existing poor participation of under privileged class in higher education in general, and technical education in particular, will increase, which will cause stagnation in the all round development of the nation.

In this situation, what Malla has said seems to be ideal. Considering the Nepalese socio-economic situation, however, it can be seen that English should remain a compulsory subject in secondary education. What should be admitted, in fact, is the necessity for bringing about changes in the whole ELT situation so that the existing problems can be reduced.

Verma and Pande (1988) carried out an extensive survey on whether English should remain a compulsory or an optional subject. It was revealed that:

Ninety percent teachers, hundred percent headmasters, 82% students, 100%DEOs, 88.3% parents disagreed with the idea of changing the status of English from the present compulsory to an optional subject for SLC students... These responses were in consonance with the response of other respondents- educationists, ... university officials, the SLC Board Officials, the Public Service Commission Officials, the Office of the National education committee and Examinations.
Verma and Pande (1988:22)

In this situation, we do not see any reason why school English should change its status from compulsory to optional. If the students, teachers, educationists, and the parents think that English should be taught as a compulsory school subject, then it

should continue as it is now. What is necessary, of course, is that we should improve the teaching/learning situation so that those who study English are able to use it for the purpose we teach it.

2.1.4. Summary

We have considered the need for English in Nepal, and found that we need English for educational and occupational purposes. On the basis of what has been discussed so far, we hold the view that English should remain a compulsory subject in the school for the all round development of the nation. We agree that the present situation should not continue, but the problems should be solved by bringing about changes in those areas where change is necessary rather than changing the status of English from compulsory to optional.

2.2. The Existing ELT Situation at School Level

2.2.0. Introduction

The existing ELT situation at school in Nepal will be presented in this section. We present the general objectives of teaching English at Grades 9-10 as they appear in the curriculum. Then we describe the types of schools by categorising them into four in terms of the types of teaching they provide. Finally, we describe the teaching methods, textbook materials and assessment system (i.e. the SLC English exam) used for school leavers.

Since the present study is mainly based on the SLC English exam, the present form of it will be reviewed in detail.

The English exam will be seen in the light of the importance placed on SLC. It will be demonstrated that the present SLC English exam is largely based on the prescribed textbooks and previous exam papers.

2.2.1. The General Objectives of Teaching English at School Level in Nepal

The ultimate aims of teaching English at Lower Secondary and Secondary schools in Nepal are stated by the Ministry of Education as being to enable the students to:

- (i) communicate in speech and writing about matters related to everyday activities with other people who speak English, within the range of the language elements provided by the curriculum,
- (ii) understand simple conversation, directions and instructions in everyday activities,
- (iii) develop the essential reading skills and techniques required for the early as well as later stages of language learning,
- (iv) develop an interest in reading for both information and pleasure

Source: CTSDC (1982:25)

A quick glance at the objectives will give us an impression that the intention of the teaching program is to enable students to communicate in the spoken and written forms of English. It is admitted that the way the objectives are presented is not very clear. Rather they are vague in a sense that no further explanation has been given of what are those "everyday activities" that the students should communicate in speech and writing about, what are those "simple conversations" and with whom, and what are those "reading skills" to be developed in the students. Nevertheless, so far as the spirit of the course objectives is concerned, it would seem that the aim of developing language skills to enable the students to communicate is the main concern.

2.2.2. English as a Compulsory School Subject

To achieve the aforementioned objectives, English as a compulsory subject is taught from Grade 4 to 10, that is from age 10 to 16 years in the schools as prescribed by the Ministry of Education. However, the schools are free to introduce some extra course/s in English on top of the compulsory English if they can

afford the salary for the teachers. The medium of instruction, generally, is Nepali. Nevertheless, there are many private schools which use English as the medium of instruction.

2.2.3. Types of Schools

Schools in Nepal can be classified into four types from the point of view of the medium of instruction used, the types of courses offered, and the way English is taught. For the purpose of the present study, the schools in Nepal were divided into the following categories:

- i. English medium schools which emphasise teaching language skills more than exam techniques (hereafter Type A).
- ii. English medium schools which emphasise teaching exam techniques more than language skills (hereafter Type B).
- iii. Nepali medium schools which have also introduced an extra English course (hereafter Type C), and
- iv. Nepali medium schools which teach only the compulsory English (hereafter Type D).

What they have in common is that every school has to teach the compulsory English course prescribed by the Ministry of Education irrespective of the medium of instruction, and the extra English courses they offer. This is the only compulsory English paper which is assessed in the SLC exam. Furthermore, schools are free to introduce any additional courses if they have the necessary resources. Of all the types of schools mentioned above, Type D represents more than 85% of the total number of schools.

2.2.3.1. Type A.

Under this category come those schools like Budhanilkantha and St. Xaviers, which focus on teaching language skills more than on preparing their students for the SLC examination

irrespective of their Grades. (In other types of school, students at Grade 10 (i.e. one year before the SLC exam) spend a considerable amount of time on exam techniques.) These schools are run by English speaking Headmasters. The total environment in the schools, within or outside the classroom, is English speaking. Some of the teachers in the Department of English in these schools are native speakers of English. Budhanilkantha, the British-aided school, has even started teaching British A Level courses. For that purpose, the school is concerned with the Cambridge O' Level Language Test more than the SLC English exam. Generally, the students from the schools of this type have been considered to have a high level of proficiency in English.

2.2.3.2. Type B

This type includes schools like Siddhartha Banasthali school, Balaju and Bhanunhakta Memorial school, Panipokhari. These schools use English as the medium of instruction but, unlike Type A, the school environment outside the classroom in these schools is not English speaking. There were no teachers in these schools who use English as L1. However, these schools have been successful in obtaining excellent results in the SLC exam for several years. For instance, one of the aforementioned schools, Siddhartha Banasthali, has been awarded Gold Medals for coming top of all the schools for several years. Discussing the matter with the students and teachers involved in teaching/ learning English in these schools, it was revealed that they have been spending a considerable amount of time, especially with the students who are at Grade 10, preparing their students for the SLC exam. To differentiate these schools from Type A, we would like to call them 'schools which focus on exam techniques more than teaching language skills especially to the students who are

at Grade 10'.

2.2.3.3. Type C.

The schools like Kanya High School and Janasewa Madhyamik Vidhyalaya were categorised under this type of school which used Nepali as the medium of instruction. In addition to the compulsory English, they have also introduced an extra English course assuming that this would help their students improve their English. These schools have been doing reasonably well in the SLC Exam.

2.2.3.4. Type D

Under Type D, the schools which teach only the compulsory English and use Nepali as the medium of instruction were categorised. This type of school represents the largest number of schools in Nepal. The SLC candidates from this type of school represent approximately 85% of the total number of candidates. They teach only the prescribed course and nothing more in English.

The urgency of bringing about changes in ELT in Nepal rests mainly on this type of school. Taking this point into consideration, in the following section concentration will be given mainly to the text materials prepared for the compulsory English, the methods used in teaching the compulsory English, and the assessment based on the compulsory English.

2.2.4. Methodology

Davies et al (1984:19) remark that "ELT in Nepal is teaching English in difficult circumstances". The number of students in a class is high. In several schools, the classroom situation is not favourable for conducting group work because of the furniture.

Teachers are trained, theoretically, to use a structural

approach to language teaching. However, they are sometimes found to be using the grammar-translation method. The whole teaching/learning situation is text-book based. The schools which have not introduced an extra English course use no other reading and writing materials except the prescribed books.

Mainali et al (1988:36) find lecturing, rote-memorisation, and group drill the most frequent methods used in the Nepalese secondary schools. They ascribe this state of affairs to the lack of training among teachers.

The whole teaching/learning situation is exam-orientated especially at Grades 9-10. Teachers tend to spend their time on practising those exercises which they think will appear in the exam. Evaluating the situation, it seems that most teachers can make a good guess about what will appear in the exam mainly because of the predictability of the SLC English Exam (see section 2.2.6.6.).

Speaking and listening skills are not taught because they are not tested in the SLC English exam, and also because the course materials do not encourage them to do that. Observing this situation, Davies et al (1984:19-20) point out:

the effects of the exam are very marked at grades 9-10 where too much time is spent explaining the content of the reading passage and testing the students ability to recall the content.

2.2.5. Course Materials

At school level, there is the series of 'My English Books' for Grade 4 to 10. The syllabus is traditionally a structural one. It emphasises reading texts: simplified versions, specially written to illustrate specific grammatical points. They are also linked with open ended questions to practise the key points introduced in the reading passage. The whole English syllabus for

school level gives the impression that very few ideas have been drawn from contemporary views of syllabus design or textbook writing techniques.

Students have to buy their own textbooks and supply their own writing materials. There is no tradition of reading around the subject at school. The focus of teaching is on reading and writing only what is required to pass the exam. Only the blackboard can be used as a teaching aid. Every one working on this field realises that there is a desperate need for teachers' guides to enable all teachers to use the textbooks to the greatest advantage.

To give an impression of the language and content of the reading passages that appear in the course books, and exercises which follow, some examples will be presented.

2.2.5.1. Example 1: Language

One of the reading passages in English Reader Part II (a prescribed textbook for SLC) is entitled 'An Interview with Professor Sharma'. To have an impression of the language used, the passage was analysed in the light of discourse principles. The text was compared with a genuine text entitled 'The Royal Interview' published in a British Journal 'The Illustrated London Views'. It was revealed that:

the language used in the course book text was unrealistic. It lacks several salient features of spoken discourse normally to be found in an authentic interview such as contextualisation, framing, the use of cohesive devices, maintenance of discourse theme, closing.

Khaniya (1988:12)

2.2.5.2. Example 2: Content

One of the reading passages in English Reader Part II is 'An Unsolved Mystery' which deals with the sea and a sailing ship in terms of its content. The story in itself seems to be difficult to understand even for those people who have command

over English to a reasonable extent. Furthermore, it can be very difficult to conceptualise the story for those who have no idea about the sea. Even the teachers would find it difficult to teach the text. The words - flying saucer, pirates, crew, - would be difficult for them.

2.2.5.3.Example 3: Exercise

English Reader Part II, Lesson 13, has the following exercise (see p.149). The exercise is intended to be used to teach the past form of can.

Instruction: 'Could' is the past form of 'can' when we are describing past skills and past sensations.

Fill in the spaces with 'Could'.

01. The boy-----see there was a man inside the cab.
02. The doctor----feel a faint pulse.
03. He-----smell petrol.
04. All the pupils in my class.... speak excellent English.

It can easily be seen that to do this exercise, students do not have to think at all. They can do this exercise without having any knowledge of 'Could'. In all the gaps provided in which students are asked to fill in with 'Could', 'Can' also fits in perfectly. Then the question arises: What is the use of this type of exercise? Does learning take place when students spend hours and hours on this type of task? Following the given instruction and practising the given exercises might cause unlearning if the students had some idea about the distinction between 'Could' and 'Can' before. This exercise, therefore, does not serve any purpose.

It can be argued that the books used at Grade 10 do not reflect authenticity in language, familiarity in content, and interaction in tasks. This would also show the least use of

recent ideas about designing a text book.

2.2.6. Examinations

2.2.6.0. Introduction

In Nepal, a child receives 10 years' education at school. After he completes his Grade 10 studies, he sits a SLC qualifying test controlled locally. Those who pass this test sit the SLC examination some three months later. The SLC examination which is controlled by an external body called 'the Board of SLC Examinations' is administered every year in January/February. To pass the exam, a candidate must attain at least 32% in each paper. English is one of the compulsory papers in the SLC examination. The average age of SLC candidates is 17. The SLC examination has serious consequences for the students for two reasons: (i) the results serve a selection purpose for tertiary education, and (ii) the results are taken into account for selecting candidates for any job, so it has a life long effect.

2.2.6.1. Functions of SLC Exam: the Crux of the Matter

We have suggested in section 2.1.2.1., that English serves an educational purposes from School to University. The School Leaving Certificate exam serves other purposes as well. Altogether the exam appears to be the most important event in the whole educational setting in Nepal. The reasons for the importance of the SLC can be seen if we look at the different uses made of the SLC exam for making decisions about individuals. In short, SLC serves the following purposes:

1. It provides a sense of achievement.
2. It is a licence for work.
3. It is an entrance for higher education.

2.2.6.1.1. SLC: A Sense of Achievement

The School Leaving Certificate is the first recognised

qualification that a Nepalese child can have after spending at least 10 years in the school. Any one who holds a SLC is socially recognised as an educated person. All those who go to school have a strong desire to have it whether they go on to higher education or not. Having SLC would mean that at least they have something concrete as evidence for achievement of their ten years' hard work. Failing to have it is disappointing.

Schools are cautious of the SLC exam. The schools are judged to be good or bad mainly on the basis of their students' pass rate in the SLC examination. It also plays an important role in financial matters of the schools. The SLC results provide, therefore, a sense of achievement also for the schools.

2.2.6.1.2. SLC: A Licence for Prestigious Work

SLC serves as a licence for jobs. To apply for most jobs, the minimum qualification required is SLC. The overall performance in SLC is taken into account in selection or promotion of an individual for a job even though he may have passed several other examinations after SLC. Thus, passing the SLC exam qualifies a candidate for relatively high income jobs.

2.2.6.1.3. SLC: A Permit for Higher Education

SLC serves as an entrance for tertiary education as all the higher education institutes accept the face value of SLC for selection purpose. Until 1978, there was an entrance examination for higher education but it was abolished because of political pressure (Gurung 1987:55). Since then, the overall achievement in SLC in general, and the individual marks in English, Maths, and Science in particular provide the basis for selection of university entrants.

It is true, therefore, that SLC stands as a very important educational event in the Nepalese educational setting: it

provides evidence of achievement, serves as a licence for prestigious jobs, and marks an important watershed for an individuals' chance of higher education.

2.2.6.2. The SLC English Exam Content

The structure of the SLC English exam which was used in 1989 was as follows:

Short-answer questions on content of	
Grades 9-10 text passages	30% of the questions
Comprehension of unseen passages	20%
Essay and paragraph writing based on	
Grades 9-10 text exercises	10%
Questions on speech component	4%
Grammar questions	36%

Total	100

The allocation of marks in the SLC English exam shows that it is text-book based. Only 20% of the total marks are allocated to texts that are not taken directly from the text books. Questions on speech components and grammar are taken from previous exam papers based on the exercises practised in the classrooms answers to which are available in commercial cribs. The effect of the textbook-based nature of the SLC English exam is that the students and teachers guess what items are likely to appear in the exam on the basis of the previous exam papers. Then students study the "ready-made answers" prepared by their teachers. Another source of "ready-made answers" for the test items which are likely to appear in the SLC English exam are "commercial cribs" which are easily available in the market. To make the points more explicit, we present some examples based on the analysis of the SLC English exam, 1989.

2.2.6.3. Criterial Level of Performance

Scores in the SLC exam are reported two ways: i.e. raw scores as the total scores in the SLC English exam obtained by a candidate, and an average score based on all the SLC subjects obtained by him. The average score is reported in terms of the scale determined by cutoff points- 60 and above, 45-59, 32-44, and 31-0, as 1st, 2nd, and 3rd division, and failed respectively. Scores in individual subjects are accepted as face values of their performance by the admissions agencies and employers. So far as the description of the performance in terms of language ability is concerned, nothing has been mentioned. There has simply been a tradition of accepting a score in the SLC English exam as an indicator of a student's ability in using English: e.g. a student with a score of 46 in the SLC English exam is considered to be more capable of using English than a student with a score of 40, though nothing has been said about what he can do and what he cannot do with his/her English. In other words, there is no further description of what a score represents in the SLC English exam.

This situation will continue for some years to come. In this situation nothing can be done about changing the reporting of scores in the SLC English exam because the exam does not have any independent role in reporting. Since English is one of the school subjects, it is mixed up with other subjects assessed in the SLC examinations while reporting.

It is for these reasons that we decided to concentrate on the washback part of the test rather than the reporting part. Description of the score is necessary. Alderson (1989) has shown the advantage of describing scores in terms of ability to use language even to increase content validity, hence washback

validity, but it seems difficult, in the present study, to do anything about it for practical reasons.

2.2.6.4. Problems in Determining Proficiency Level

Expressing language proficiency in terms of scores has been a tradition in language testing. There are some tests that describe what a score in that test represents. This is not always the case with other examinations. It is possible to do that only if different skills are measured by different tests so that what a score on a test means can be described. There are exams which have to include several skills in a single test and report accordingly. We do not know what skills in particular compose a given score. We accept the importance of description of a test score. However, we also see the problem of reporting a score in a test especially when we have to deal with an examination like our SLC English exam which is one of several school subjects.

In this study, therefore, we do not propose to change the way of reporting the existing SLC English exam. Nevertheless, we do feel that something must be done about it. For example, if a student can pass the exam attempting only one or two subtests like cloze and writing, or note-taking and reading, then we cannot make sure that the students who pass the exam have achieved the necessary skills as reflected in the exam. But there is nothing that we can do about it at this stage. The researcher would have liked to set the criterial level of performance on the new exam so as to require the SLC students to obtain at least 25% marks on all the subtests so that students who pass the exam will have achieved a minimum level of English proficiency (whatever the proficiency means) required for the tertiary courses. This way of setting criterial levels of performance would also increase the washback validity of the exam. We would also like to

set the passmark in the SLC English exam much higher than it is now. It is difficult to accept an achievement of 32% of the course objectives as a satisfactory levels of achievement. Here again, we are bound to consider the practical problem of setting a pass mark in the SLC English exam different from other school subjects. Taking the real situation into account, we, therefore, agree with Davies (1988b:3) who holds the view that "pass marks properly should be determined by the test users themselves and they should always be related to local circumstances".

2.2.6.5. Cutoff Points

Having accepted the existing way of reporting the exam scores as an indicator of language proficiency, we proposed to investigate the language proficiency of the SLC students within this framework. We were interested in finding out the level of proficiency of the students who passed the SLC English exam and the level of proficiency of the students who received cutoff point 32-44 in the new exam. It was hoped that this sort of analysis would provide a basis for describing the meaning of 'passmark' in the SLC English exam, and the meaning of 'passmark' in the new exam in terms of language ability. This would also provide some evidence of the construct validity of the new exam.

2.2.6.6. Analysis of the SLC English Examination, 1989

The SLC English Exam seems to be a mixture of discrete point and integrative testing on the surface. However, analysis of the test items reveals that the exam is composed of discrete point, memorisation of factual knowledge, and integrative testing.

Here in this section, we present an analysis of the SLC English exam, 1989 (the full text of which is given in appendix I) to exemplify the specifications and the formats of the exam.

Since the present study is largely based on the SLC English exam, a separate section will be devoted to making comments on it (section 2.3.2.).

Question No. 1.

It contains (a-n) 12 short answer questions out of which students are required to attempt only 10. Each item is worth 3 (3x10=30) marks. All these items are based on the reading passages which appear in English Reader, parts I and II prescribed for Grades 9-10. On the surface, looking at the instruction given on the top of the exam paper, it seems that students are expected to answer these items in their own English. But, since all the items were found directly copied from the prescribed text books which students would have practised several times in the classroom, it is hard to believe that the students answer these items in their own words. The test items are so predictable, as some items appear every year, that every year experienced teachers make a good guess, and publish several cribs (e.g. Bazaar notes). In either case, students do not have to write answers in their own English at all, and they do not do that unless they are confident about their English. To exemplify the point, we present some items and the sources to their answers.

1. Surya Prakashan (1988), Gautam Super Lucky SLC Guess Paper

- | | | | |
|-------|----|--------|--------------------------|
| Items | a. | page 2 | (previously SLC 1985) |
| | b. | page 2 | (previously SLC 1984/86) |
| | c. | page 3 | |

2. Sharma (1988) Guess Paper with High Surety for SLC 2045 (1988)

- | | | | |
|------|----|---------|--|
| Item | f. | page 27 | |
|------|----|---------|--|

3. JEMC (1986) English Reader Part II

- | | | | |
|-------|----|----------|-----------------------|
| Items | i. | page 182 | (previously SLC 1982) |
|-------|----|----------|-----------------------|

m. page 244 (previously SLC 1979)

The point is that the answers to all the questions in the exam are available in the sources available in the market.

Question No. 2.

It contains 2 reading passages composed for the exam, followed by some comprehension questions. These passages are intended to assess the reading ability of the students. The language and the content of the passage give the impression that they were contrived only for the exam. In other words, they do not seem to be authentic. Looking at the comprehension questions which follow the passages, it becomes clear that they were not carefully constructed. Further it is not clear which item is meant to test for which reading subskill: e.g. scanning, inference, guessing meaning from context. For example, one of the items in the first passage is this:

Passage (a) i.

Where did the boys decide to go?

The expected answer is: to go on a trip to the seaside which is the exact wording of the passage. On the whole the two passages call for only one reading skill: i.e. locating information from the passage. They do not require the student to process the text in a true sense; e.g. they do not call for inference, guessing meaning from context.

Question No.3.

It has two items out of which the students must answer one. This question is intended to assess the writing ability. The first item requires the students to write an essay on the basis of the given clues, and the second item requires them to write a paragraph on the basis of the given outlines. On the surface, the passages appear to be acceptable to some extent. However, in

fact, these items do not require the students to write their own English because both items were directly copied from the text books prescribed for them. Presumably, the students might have practised those tasks several times in the classroom. If not, their answers are available in the cribs. The first task appears in English Reader part II P.246, and the second task appears in part I p.50. The students can get ready made answers to these questions from different sources. In such a case, how can it be argued that they assess the students writing ability?

Question No.4

This item is intended for assessing pronunciation though it looks a tricky question. The items were borrowed from the previous exam papers answers to which are available in Gautam, B. (1988) Grand Guide to English Reader (I-II) on pages 468-477.

Question Nos. 5 to 10

They are usage based. The examinees are asked to supply the correct form of verbs to the given sentences (No. 5.i.), completion (5.ii.); filling in the gaps with the given preposition (6.i), articles (6.ii); direct/indirect speech (7); transformation from active to passive of the given sentences (8); tag-questions (9.i), joining sentences with whose, who, where, whom (9.ii); joining sentences with connectors (10.i); and verb ending (10.ii).

The general characteristics of these tasks (except Q.N. 7) are that every sentence is presented in isolation. In other words, there is no connection between one sentence and another at all. Most of them are directly borrowed from textbook exercises and previous exam papers, answers to which are available in Gautam (1988) and some other Guess papers. We exemplify some of them below.

Question No.5.i.

- a. He (walk) in the park when I met him. (SLC 1985)
see Gautam (1988) p.520
- b. I wish I (know) the answer to that question. (SLC 1987)
see Gautam (1988) p.546
- f. Medical Scientists (eradicate) malaria by 1994. (SLC 1986)
see Gautam (1988) p.537

Question No.6.i.

- a. He has been drinking.....half past six. see Gautam p.335
- b. It took them an hour to put.... the fire. see Gautam p.335
- c. The pencil rolled.... the table. see Gautam p.336

Question No.6.ii.

- a. The clouds over....sea are lovely today (SLC 1982)
- b. A horse is.... animal see Gautam p.329
- c. Copper isuseful metal. see Gautam p.329

Question No.7.

This task is directly copied from SLC 1984. The answer is available in Gautam p.376.

Question No.8.

- a. More rice has to be grown by farmers. SLC 1982, 1986.
see Gautam p.361
- b. We must go into this matter. see Gautam p.362
- c. The child is being fed now. see Gautam p.362

Question No.9.i.

- a. I'm very late,.....? SLC 1980
see Gautam p.389
- b. You never used to drink.....? see Gautam p.391

Question No.11.

In this question, four sentences are given asking the students to supply a single word (to each) which represents the

meaning of the sentence. The sentences are textbook based which are difficult for any one who is not familiar with the texts from which the sentences are borrowed. For example, we asked a highly educated native speaker of English to do the task, but she could not come up with the answers as given in the marking scheme of the SLC English exam prepared by the board of English teachers. Only those students and teachers who are familiar or made familiar with this type of examination know the answers to these items. These sentences were borrowed from previous exam papers answers to which are available in Gautam (1988):444-446.

Question Nos 12 to 15

These questions are also usage based: e.g. using phrases in sentences (Q.N.12); identifying parts of speech (Q.N.13); supplying antonyms (Q.N.14); and filling in the gaps with given words (Q.N.15).

To sum up, the various characteristics of the exam can be described as:

- i. textbook and previous exam based,
- ii. discrete point testing,
- iii. contrived language,
- iv. 50% chance for guessing.

It should be clear by now that the specifications of the exam have not been the course objectives but the textbooks and the previous exam papers. They will further be discussed in detail in section 2.3.2..

2.2.7. Summary

In this section, we have presented the existing ELT situation in the schools. We have considered the general objectives of teaching English at school. We found that the main concern of the course objectives is to equip the student with the

four language skills. Looking at the ways English is taught in the schools, we have categorised the schools into four types for the purpose of the present study. The methods, the course materials, and the school leaving examinations have been examined. We have found that the methods are traditional, textbooks are not satisfactory, and the exam is largely based on prescribed textbooks and previous exam papers. The exam, therefore, encourages rote-learning and guess work which may not foster language learning.

2.3. Criticisms of the SLC English: Need for Change?

2.3.0. Introduction

It has already been argued that Nepal needs English for educational and occupational purposes. Accepting the importance of English in the modernisation process of the nation, we have made our stand clear that English should remain as a compulsory school subject. However, the existing ELT situation is not conducive to studying English. We feel strongly, therefore, the necessity of upgrading the teaching of English to make it serve the purpose for which it has been introduced.

We believe that any effort to improve the ELT situation in Nepal is likely to be more successful if it is based on the analysis of the situation concentrating on why the existing system has not been effective; rather than a piecemeal operation implemented as mandatory. For that reasons, it is important to consider the purpose for which we need English, and why the existing system has not been serving the purpose.

To find out the effectiveness of teaching English at school, it is, therefore, necessary to investigate the extent to which the existing situation has been serving the purpose for which English has been introduced at school: e.g. educational and

occupation purposes. Considering the practical difficulty, in evaluating the present status of teaching English for occupational purposes, it was decided that the present study should be confined to educational purposes. It was, however, hoped that this would provide a basis to some extent for determining the students' proficiency in English for occupational purposes as well.

Accepting the importance placed on the SLC exam, especially on the SLC English exam (see section 2.2.6), it was presumed that an investigation based on it (i.e. the SLC English exam) would provide the most useful information for making decisions about whether the existing ELT situation has been effective in relation to the purpose for which we want English at school. And if it has not been effective, the study, at the same time, would provide useful information about what is wrong with the existing situation. The whole idea of doing this is to obtain insights which would contribute to making decisions about the changes we need to bring about in ELT in Nepal.

In this section, we describe the problems created by the existing ELT situation which will provide the answer for why the Nepalese ELT situation needs change.

2.3.1. Statements of the Problems

The purpose for which English is taught in schools has been discussed in section 2.1.2.. Whether the existing system has been serving the purpose can be understood if we look at the issue from the point of view of how the achievement of English proficiency at school level is determined (i.e. the SLC English exam), and the work the SLC passers have to do with English for educational purposes (i.e. English for higher education).

It would appear then that, for the purpose of evaluating

the effectiveness of teaching English at school, the SLC English exam and the college entrants' performance in the subjects which require English in tertiary education would be appropriate.

2.3.1.1. Results in the SLC English Exam

The SLC English exam has always been a threat to SLC students. As was discussed earlier, to pass SLC a candidate has to pass all the SLC subjects examined separately. In other words, failing any subject in SLC is failing SLC. In the case of the SLC English exam, it outweighs all other subjects in terms of arousing fear in the students of failing SLC.

Generally, students tend to devote more time to studying the SLC English than other subjects on the assumption that passing it is most likely to lead them to pass SLC. On the contrary, what happens every year is that a high percentage of SLC candidates pass all subjects, but fail the SLC English exam. As a result, despite the hard work, they fail SLC. To make this point clear, we present the following results (cited from Verma and Pande 1988:8):

Table 1: The Result in SLC English Exam, 1986

Year	Total stds	Passed	Failed	Pass %	Fail %
1982	35671	13861	21810	38.86%	64.14%
1983	49327	17202	32125	34.87%	65.13
1984	37028	20282	16746	54.77	45.23
1985	52333	23558	28775	45.02	54.98
1986	49335	26559	22776	53.83	46.17
	44388	202924	24446.4	45.36	54.04

The high percentage of failures appears to be more

serious if we see the results in SLC English in comparison to other subjects. Malla (1977:2) observes that 80% to 90% of the SLC failures are due to failure in English.

Davies et al (1984:30), and CERID (1983:118) find that English and Mathematics are the two main subjects which cause a high percentage of failures in SLC. They point out that the role of English in causing a high percentage of failures is higher than Maths.

The SLC English appears to be, therefore, the single most important factor responsible for a high percentage of failures. This type of result is not pleasant to any education system. The situation, thus, obviously merits a thorough investigation of the reasons why a high percentage of SLC candidates fail the English exam every year.

Verma and Pande (1988) carried out a survey to find out the causes of failures in the SLC English exam. They used questionnaire and interview methods to collect information. The information was collected from the people involved in different sectors: e.g. teachers, administrators, students, educationists.

They found, in order of merit, lack of level-appropriate knowledge of English, lack of teacher training, teachers' irregular attendance in the classroom, students lack of interest, defective textbooks, defective SLC curriculum, and defective SLC English as the causes of failures.

We appreciate the contribution Verma and Pande have made to the ELT situation in Nepal. Nevertheless, we would like to point out that there is a danger of making a faulty decision if we rely entirely on the information collected by using a subjective measure, as they did. The point is that this type of survey presents an impressionistic view of the situation which is

not without suspicion. For example, the researchers have concluded that the SLC English exam is the least responsible factor for a high percentage of failures on the basis of those questionnaires which were answered by only seven students out of 60 (i.e. 11.67%). Other respondents did not answer this question for unknown reasons. Then, this type of information can give a distorted description of the situation. This is how the defectiveness of the SLC English exam has been overlooked in the past. Our argument is that the issue whether an exam is defective should not be judged on the basis of a poll. Rather it should be judged against some objective criteria. An exam is too technical to be evaluated only on the basis of how people feel about it.

Here in the present study, because of logistic constraints, we do not propose to investigate the causes of failure in the SLC English exam. We wish we had access to the situation and enough time and money to investigate the causes of failures because we strongly feel that they need to be analysed.

2.3.1.2. Performance of College Entrants

The discontent with the school English can be more explicit if we look at the problems that those who were lucky enough to pass SLC and get enrolled in tertiary education have in coping with higher education.

Malla (1977:8) observes the inadequacy of college entrants' English for higher education, and suggests that the best thing to do with this is to improve the SLC English. Malla (1977:4) presents the difficulty that college entrants have because of their inadequate English in coping with tertiary education when he says that "of the 22% who pass SLC nearly 90% get into the university, only 30% of them manage to survive"

Shrestha (1979:19) makes a reference to the National

Conference of University teachers, 1968, which recommended that something had to be done about college English "without waiting too long for reforms in Secondary English, particularly in SLC English". This is how the university teachers collectively showed their concern with the unsatisfactory teaching of English at school.

Davies et al (1984:14) report that "since 1981 fewer than 10% nationwide have passed the compulsory English paper" in tertiary education. For this, as Davies et al (1984:34) observe, the college teachers blame the schools for not equipping students with the necessary language ability to cope with higher education. The college teachers consider that it is the schools which should improve their teaching rather than the colleges providing their entrants with some remedial English.

Bhadra and Yadav (1988) carried out a study to investigate the reasons why a high percentage of students fail in tertiary education in Nepal. They report that one of the most responsible factors for that is the college entrants' poor background in English. In the report (p.15), they present the failure rate for different subjects in proficiency certificate level, part II, in the year 1986 as follows:

Table 2: Failure rate at Proficiency Certificate Level

Subjects/	Eng.	/ Nepali/	Nep.Pari/	Economics/	Hist/	Pol.Science
Percent/	76.43/	21.43 /	30.71 /	32.01 /	47.30 /	32.57

The failure rate of 76.43% in English is disappointing. The students who passed other subjects could not complete the level only because they failed the English exam. The researchers

ascribe the higher percentage of failure in tertiary education to SLC English.

CERID (1984:21) finds that different subject committee chairmen (e.g. Science, Engineering, Agriculture) admit that the students with SLC have difficulty in understanding the subject matter at tertiary level.

It can be seen that a high percentage of SLC passers go onto tertiary education without adequate English as needed to cope with the subjects which require English. This situation clearly demands an investigation of the SLC English Exam procedures.

2.3.2. Criticisms of the SLC English Exam

2.3.2.0. Introduction

On the basis of what has been discussed in the above sections, it can be seen that only 40 to 45 percent of SLC candidates pass the SLC English. Looking at this result in relation to the problems the college entrants have in coping with higher education, it becomes clear that even those who pass the SLC English pass it without a knowledge of English sufficient for coping with higher education. In the existing ELT situation, therefore, we face two main problems: i. a mismatch between teaching and testing the SLC English, and ii. a gap between the SLC English and the English expected of college entrants.

It appears to be difficult to locate the causes of failures in the SLC English exam. We have already made two points about the SLC English exam: i.e. that students and teachers tend to devote much more time to the SLC English than any other subject, and the SLC exam is predictable in the sense that the questions are selected either from previous exam papers or from the textbook materials. In either case, students can find answers

to those questions in Bazaar notes. In this situation, the failure rate should not be so high.

Regarding the gap between school English and the English required for tertiary education, we take the view (similar to the university teachers) that school English should be upgraded. In the present ELT situation, we have argued that the SLC textbooks are not satisfactory, and the exam is textbook orientated. Improvement in the ELT situation can only be brought about if change in all these factors is made. To bring about change, one factor has to take a lead so that it leads other factors to change. We consider that the exam is the most effective component to begin with.

The present study is based on the assumptions that if the SLC English exam is improved so that it has a positive washback effect, and if it is designed in relation to the needs of English for college entry, those who pass this exam will be much better equipped to cope with higher education. Changing the methodology, course books in relation to the proposed exam will, it is hoped, bring about a desirable improvement in the ELT situation in Nepal.

In this section, we consider the efficiency of an exam in terms of validity, reliability, and practicality criteria to provide a framework in which what is wrong with the existing SLC English exam can be discovered, and which will also provide a basis for designing a new exam. The SLC English exam will be examined in the light of the content validity and the washback effect of a final exam. It will be argued that an exam like the SLC English exam can have a negative washback if it fails to reflect what is included in the course objectives.

2.3.2.1. Theoretical Considerations: The Efficiency of an Examination

2.3.2.1.0. Introduction

Having said, in Chapter 1, that examinations as instruments can be used for educational change, and that an exam can have an educationally beneficial effect, we have emphasised the efficiency of examinations. The whole argument that an exam can have a positive washback effect, and that it can work as a lever for educational change is based on the assumption that the exam in question is carefully designed. What follows is a discussion of the efficiency of the examination designed according to its desired purpose.

In this section, we consider the three constituents of exam-efficiency: validity, reliability, and practicality. We consider that constructing an exam, at least in the present context, is purpose orientated. In this respect, validity, reliability, and practicality should be seen as relativistic concepts. The whole idea of considering the three constituents of exam-efficiency is to build-up a framework which will provide a basis for examining the existing SLC English exam, and also insights for designing a new exam.

It will be argued that, though all the three constituents are important in exam construction, validity is the primary one. Accepting the importance of validity, a considerable amount of attention has been given to content, and predictive validity, on the ground that they both serve the purpose of the exam, and also provide some evidence for construct validity. Face validity will also be considered. It will be discussed that though we accept

the importance of reliability, we adopted a strategy whereby if the validity of an exam is likely to increase, some reliability can be sacrificed whenever necessary. Finally, it will be argued that the exam should be seen from a practical point of view since it is designed to be able to work in the Nepalese situation.

2.3.2.1.1. Validity

The purpose of a valid final examination is to provide the people concerned with information that gives a basis for making decisions about the extent to which the examinee has achieved skills and ability in relation to the course objectives. Whatever we talk about - the constructs, the traits, the skills, the ability - basically, we are interested in making inferences about the extent to which the examinee can perform the tasks he is expected to, or will be expected to do. We do this by asking the examinee to perform the tasks selected on the basis of the behaviour domain under consideration. The validation of an exam, therefore, would imply whether the exam is sufficient to elicit the genuine information we are looking for from the examinee so that valid decisions can be made. It is, therefore, the ability to perform given tasks that is the main concern in validation procedures. The elicited information is validated against some other independent criteria: e.g. logical validity and statistical validity, in order to determine the extent to which the exam is valid.

It has to be accepted that exam validation is purpose specific (Brown 1976:62), in the sense that a valid exam for one purpose may not necessarily be valid for another purpose. Taking this view, test validity is concerned with eliciting information

for the purpose of testing. By the same token, it can be argued that, as long as an exam is validated so as to justify that the information it elicits, and/or elicited, clearly provides the basis for drawing inferences and making desired decisions, from the practical point of view, it is not necessary to validate the exam against all the types of validities listed above. What is important, of course, is that the validation procedures (i.e. the types of validity) to be adopted must be congruent with the purpose for which the exam is designed.

As the purpose of the study is to use examinations as instruments for educational change, the examination under consideration must have a positive washback effect. Since it is proposed that the type of exam in question should eventually replace the school leaving examination in Nepal, an exam of this type (i.e. exam which is expected to have a positive washback) must have content validity (Anastasi 1982; Brown 1976; Thorndike and Hagen 1977; Ebel 1979). For that purpose, a priori validation is necessary. On this ground, the content validity of the exam will be considered.

From the practical point of view, one source of discontent with the existing examination system was the poor performance of the SLC passers (see section 2.3.2.1.) in their tertiary education. To put it another way, the SLC English exam has not been a valid exam as a selection device. The exam which is intended to replace the SLC, therefore, must have predictive validity. It is, therefore, necessary for the exam to satisfy a posteriori validation. Another important point is that the examination must also be empirically validated to ensure that it

has elicited the information it was supposed to elicit. This can be done by checking the performance of the examinee in the exam against an external criterion. We agree with Davies (1983:141) who pleads for the criterion-related validity in the following words:

The external criterion, however hard to find and however difficult to operationalise and quantify, remains the best evidence of a test's validity. All other evidence, including reliability and the internal validities is eventually circular.

On this ground, the criterion-related (especially predictive) validity will be considered.

In the testing literature face validity is often considered as fake or pseudo-validity. However, we hold the view that if the examinee does not consider an exam a valid one, the information collected from it may not be genuine. Accepting Stevenson (1985:113), who recognises the "seductive appeal of face validity", face validity appears to be important. Face validity of the proposed exam will also be considered.

In the literature, construct validity is discussed as the centre of all methods of validating an exam. Having accepted the inclusiveness of the construct validity, it will be argued, briefly, that if an exam provides evidence that it has content and criterion-related validity, the exam also provides some evidence that it has construct validity as well.

2.3.2.1.2. Content Validity

Content validity is defined as "whether the item composing the test do, in fact, constitute a representative sample of the content domain of concern" (Brown 1976:122-223). In case of a final achievement test, it is said to have content validity if

its test items are considered to be a representative sample of the tasks as can be seen in the course objectives. Despite the "usual problem of selection and sampling for a 2-3 hour test from specifications that could generate a 2-3 year syllabus and which in turn are selections for an even larger whole" (Seaton 1985:116), selection of tasks to be included in the examination is indispensable. The basic question is, therefore, whether the test items that compose the exam constitute an appropriate representative sample of the behaviour domain under consideration (Brown 1976:124). By the same token, the content validity of a final achievement test will ensure its appropriate representative sample of the behaviour domain conceived in the instructional objectives (Moller 1982:37). Anastasi (1982:131) describes the nature of content validity as:

content validity involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured.

The argument is that in an achievement test, the emphasis will be mainly on the coverage of the subject matter. It implies that the test designer must, as clearly as possible, specify what skills the exam is designed to cover. For Brown (1976:123), content validity is "a measure of the adequacy of sampling".

It has already been discussed, in Chapter 1, that the present study proposes to use examinations as levers for change. The main aspect of concern of the exam is, therefore, its washback effect. Following the arguments presented in section 1.5., that an exam cannot avoid influencing teaching and learning, and if an examination requires the examinee to exhibit the ability envisaged in the course objectives, the washback

effect of the examination can be beneficial. An exam of this type can be used for educational change. Looking at the proposed examination from the washback point of view, the exam-content appears to be the most important aspect to deal with.

Regarding the procedures to be followed at the construction stage to make the exam of high content validity, Anastasi (1982:132) (also quoted in Weir 1988), provides the following guidelines:

- 1.the behaviour domain to be tested must be systematically analysed to make certain that all major aspects are covered by the test items, and in the correct proportion.
- 2.the domain under consideration should be fully described in advance, rather than being defined after the test has been prepared.
- 3.content validity depends on the relevance of the individual's test responses to the behaviour area under consideration rather than on the apparent relevance of item content.

What emerges on the basis of these guidelines is the concern of content validity with the a priori stage of exam-construction. This can be agreed with Weir (1983a:148) who argues:

it no longer seems sufficient to rely solely on the more quantitative, post hoc, construct, predictive and concurrent validation studies to establish what it is that we have tested. Unless...the testing system is initially matched against such a framework, it is also difficult to see how we can ever get near to describing accurately the construct that we are attempting to measure.

From the point of view of the benevolent washback effect of an examination, the a priori validation is thus essential. We agree with Thorndike and Hagen (1977:59) who argue that classroom instructions are intimately related to the exam-content, and an exam may be appraised by looking at how truly it represents the instructional objectives. After all, the program of instruction is the main source of exam-content. An appropriate representative sample of the course objectives in an examination can have a beneficial washback effect because it is this type of exam which requires the teaching and learning to be directed towards the course objectives. This type of examination will be a check on the achievement of the course objectives, preparation for which would lead the student to achieve the skills and abilities as can be seen in the instructional objectives. What we should be doing, to make an exam educationally beneficial, as was discussed in Chapter 1, is to make the exam a mirror of the course objectives in order to make the people concerned understand what is expected of them.

Anastasi (1982:132) considers the construction of an educational examination is usually preceded by the thorough examination of the relevant course materials, and instructional objectives. Doing this provides a basis for test specifications on the basis of which test tasks are designed. The discussion of the test specification should show "the instructional objectives or process to be tested, and the relative importance of individual topics, and processes" (p.132).

Following Anastasi's (1982) guidelines presented above, and Brown's (1976:125) suggestions, the following procedures were

adopted in the construction of the exam in question in order to increase its content validity.

1. Attempts were made to delineate, as clearly as possible, the relevant skills to be assessed (i.e. behaviour domain) in relation to the course objectives. The source materials to be used in the exam were also described.

2. Subskills were discussed according to the emphasis they were given in the course objectives. A practical approach was adopted in interpreting the importance the course objectives place on the skills (see section 3.2.).

3. Attempts were made to show the skills and subskills being measured by each item.

2.3.2.1.3. Predictive Validity

Criterion-related validation procedures determine the efficacy of an examination in predicting the examinee's future performance in a prespecified situation (Anastasi 1982:137). In this procedure, the performance of the examinee is examined against a criterion. It is this criterion which represents the behaviour domain the exam is used to predict. A measure of criterion-related validity of an examination is the correlation between the performance in the exam and the performance in the criterion.

In the literature, criterion-related validity is discussed under two heads: concurrent validity and predictive validity. The main difference between the two types of validation procedures is time interval: if the exam scores are validated against the criterion approximately at the same time, it is concurrent validity, and if the exam scores are validated against the

criterion after a stated time interval, it is predictive validity. However, in the present study, looking at the exam from a practical perspective, it is only predictive validity which will be considered. Since the exam is intended to replace the existing SLC English examination so that, apart from its backward-looking nature, it is also being used for forward-looking purpose, the predictive validity of the exam is very important. In other words, since it is used as a selection device, it must provide a basis for accurate decision making. We agree with Weir (1988:30) who argues for empirical validation of a test "whenever the candidate's future may be affected by its results".

It has already been made clear that the SLC result is very important for the examinee since it decides whether, and in which institute, the examinee will be enrolled. For this type of examination, it is necessary to provide a basis for accurate decision making, and a decision is made on the basis of the examinee's exam scores. The proper measure of an exam's predictive validity is thus its contribution to accurate decision making (Brown 1976:99). Looking at the exam (after it replaces the SLC English exam) from its potential role as a selection device for tertiary education, its predictive power is the one that is most important. If the exam does this job effectively, it both serves the practical purpose and provides evidence for predictive validity.

One of the difficulties in dealing with predictive validity is to find a satisfactory criterion with which the exam results are to be correlated (Guilford 1965:472; Thorndike and

Hagen 1977:61). To establish the predictive validity of an exam, some standard measure of performance must be prespecified. It is this prespecified performance which serves as a criterion.

Brown (1976:107) considers that the most important characteristic of the criterion is relevance. We would like to interpret relevance as the relationship between the skills that the exam calls for and their importance in getting success in the criterion measure. If the skills that the exam requires the examinee to exhibit are the same as the skills the examinee must demonstrate in the criterion measure, the criterion can be justified. The advantage of adopting this procedure for exam validation is a practical one which is useful for an examination like ours. The argument is that since the exam serves the practical purpose, it would be easy to convince the authority concerned that the exam should be used to replace the SLC English exam.

To determine the predictive validity of the exam, the criterion against which the exam scores were validated was the performance of the examinee in the subjects requiring English in tertiary education. The correlation between the exam scores and the performance on the subjects requiring English in tertiary education would be, therefore, interpreted as the measure of the predictive validity of the exam.

The criterion was selected on consideration of the relevance of the skills that the exam calls for and the skills that the examinee needs to have to be successful in the criterion measure. The decision about the criterion selection was largely based upon the practical situation. As discussed in section

2.1.2.1., one of the purposes of teaching English in Nepal is to equip the students with the English they need to cope with tertiary education, and one major source of discontent with the SLC English examination was the poor performance of the SLC passers in tertiary education. Taking these points into account, the performance in the subjects requiring English in tertiary education was adopted.

Despite the fact that academic achievement has been the common criterion for establishing the predictive validity of a test (Anastasi 1982:138), one of the problems in criterion selection is, it is admitted, that we still do not know the extent to which language proficiency determines academic success. Criper and Davies (1988:65) point out that failure in academic achievement may be because of the language component. The tests - the TOEFL, the Michigan Test, the ELTS - are being used as selection devices on the assumption that there is a prominent role for language in academic success.

Moller (1982:56) discusses the predictive criteria against which a proficiency test of English as a Foreign Language is to be validated. He considers that the most common situations that can be used as criteria are those in which the subjects use English for their studies or professional duties. Nevertheless, he also accepts that such measures will not measure language exclusively. Having said that, however, the justification for adopting such a criterion, as he argues, is that providing evidence that a candidate has adequate English proficiency should "enhance a student's chance of academic success and eliminate one source of weakness which could lead to failure".

Furthermore, in the Nepalese ELT situation, failure in tertiary education has always been associated with students' low proficiency in English (see section 2.3.1.3.).

On the basis of these arguments, it was decided that tertiary grades in the subjects requiring English should be used as a criterion.

2.3.2.1.4. Inclusiveness of Construct Validity

Brown (1976:128) (citing from APA 1974:29) writes:

construct validity is implied when one evaluates a test or other set of operations in light of the specified construct.

Accepting this definition would imply that construct validity is important when an examination is designed to measure the attribute or quality or ability that an individual is supposed to possess. Regarding the interpretation of a construct, we agree with Ebel (1979:307) who considers that "any domain of knowledge, any skill or ability can, of course, be called a construct". Perhaps Anastasi (1982:144) has also the same interpretation of construct validity when she writes:

Any data throwing light on the nature of the trait under consideration and the conditions affecting its development and manifestations are grist for this validity mill.

Accepting this argument implies that the concept of construct validity is associated with the validation procedures based on systematic examination of the abilities we want to measure, and evidence of their measurements. This argument is supported by Weir (1983a:92) who refers to Kelly (1978:225) as saying that construct validity is dependent on the prior conception of "appropriate objectives for the test to measure".

To make the argument more explicit, we, first, quote Kelly

Cronbach (1971:443) who maintains:

construction of a test itself starts from a theory about behaviour or mental organisation derived from prior research that suggests the ground work for the test.

It can, therefore, be argued that since, "specification of the behavioural domain sampled by the test serves to define the nature of the construct the test measures" (Brown 1976:171), determining the content validity of the exam will also provide evidence for the construct validity.

Second, if an exam provides evidence for predictive validity, it will also serve as construct validity in the sense that the correlation between the ability the exam measures and the prediction of the abilities in the criterion provide data for the constructs being measured (Brown 1976:133).

On the basis of the discussions presented so far, we agree with Anastasi (1982:153), and Davies (1977:63) who accept the inclusiveness of construct validity. For Anastasi, construct validity is a "comprehensive concept" which includes content and criterion-related validities, and for Davies, content and predictive types of validity are based on construct validity.

By the same token, it can be argued that evidence for content and predictive validity would also provide data for construct validity. In this case, if we accept with Peterson and Cartier (1975:105) that "construct validation presents enormous theoretical and practical problems", establishing content validity, and predictive validity would make the validation job easier.

2.3.2.1.5. Face Validity

Face validity is defined as "what it appears superficially

to measure" (Anastasi 1982:136). Because of its impressionistic quality, Lado (1961), Palmer (1981), and Bachman and Palmer (1981) do not consider it important. We quote Bachman and Palmer (1981) (cited from Weir 1983:72) to present how they underestimate the importance of face validity:

since there is no generally accepted procedure for determining whether or not a test demonstrates this characteristic and since 'it is not an acceptable basis for interpretative inferences from test scores', we feel it has no place in the discussion of test validity.

The importance of face validity cannot be underestimated, however, in the sense that if the examinee does not consider it a valid exam, one would ask: how can we make valid inferences depending on the information elicited by that exam? Anastasi's (1982:136) argument that "face validity should never be regarded as a substitute for objectively determined validity" can be accepted. Nevertheless, how the student feels about the test cannot be ignored in exam validation procedures.

In the present study, to find out how the examinees perceived the exam, they were asked to make comments on the SLC English exam, and the exam under consideration. They were also asked which of the two they think measure their true language ability. Their comments would provide some evidence for face validity of the proposed exam.

2.3.2.1.6. Summary and Comments

In this section, we considered how the validity of the exam would be established. We selected validation procedures from a practical perspective on the assumption that the validity of an exam is a matter of examining whether it elicits the information from the examinee in relation to the desired purpose of the exam.

We came up with the conclusion that as long as an exam serves the purpose of testing, the exam is said to be valid.

Content and predictive validity have been considered to a reasonable extent in the belief that validating the exam following these procedures would serve the main purpose of the proposed exam, as well as providing some evidence for construct validity. Face validity has also been considered important in the sense that no performance can be accepted as valid unless the performer finds the exam valid for its desired purpose for its desired purpose.

2.3.2.1.7. Reliability

Reliability is concerned with examining consistency in the performance of the examinee. In order to establish the reliability of an examination it is necessary to answer the question: how consistent would the examinee's performance be if we asked him to take the same exam at a different time, or another examination which is supposedly similar? (Ebel 1979:67).

With respect to reliability, what we are interested in is the degree of accuracy of measurements. In other words, how accurately an exam measures what it measures. It is assumed that given the measurement is accurate, the performance of the examinee remains more or less the same if the examinee is asked to repeat the same exam or an exam of similar type.

The degree of consistency of measurements is determined by carrying out some statistical analysis. For that purpose, two sets of scores are obtained from the performance of the same sample of examinees. A correlation of the two sets of scores is said to be the correlation coefficient of the examination, which

is interpreted as the coefficient reliability of the exam. Other things being equal, the higher the reliability, the better the exam.

The underlying concept of reliability is that whatever is measured is not only a property of the exam, it is also a property of the examinees' performance. It implies that reliability is not like a permanent quality of an exam which remains the same if it is administered again and again; reliability of an exam may change because of variations in performance. It seems that testing conditions can also have an effect on the reliability of an exam.

Another notion behind reliability is that the observed score obtained from an exam is not without error. Statistical analysis of the observed score segregates it into true and error scores. The correlation coefficient is the measure of true score which is established by reducing error score from observed score. The assumption is that there is a possibility of the occurrence of error in measurement alone.

Basically, two approaches to establishing the reliability of an exam are discussed in the literature (Brown 1976; Weir 1983a): examiner reliability and examinee reliability. The examiner reliability refers to how consistent the examiner(s) is/are in assigning scores to the examinee's performance. This type of reliability is important especially for those examinations which include essay items involving subjective judgement. The examinee reliability is the real reliability of an examination in the sense that it is this type of reliability which tells us how consistent the examinee is in his performance.

To put it another way, this type of information tells us how dependable the measurement is.

Regarding the methods of establishing the reliability of an exam, several methods are discussed (see Brown 1976 and Weir 1983a): inter-intra-marker reliability, test-retest method, split-half method, equal forms method, Kuder Richardson and Cronbach alpha method. The methods to be adopted to establish the reliability must be selected according to the purpose and the type of the proposed exam.

Considering the type and purpose of the proposed exam, our main concern was with determining the inter-intra-marker reliability of the writing part. The rest of the items were objectively scored. This type of reliability was considered important for two reasons. First, as was discussed, the exam must have a positive washback effect. To ensure the positive washback effect of the exam, especially on the composition part, it was necessary to make explicit what language components the examiner/s must emphasise while awarding scores to the examinees' essay items. The assumption was that, after we have made the language components being measured clear to the respective teachers and students, their teaching and learning in the schools would also emphasise those components. Second, after the exam replaces the existing SLC English examination, a great number of school teachers will be involved in marking the examinees' answer sheets. There may be enormous differences in their marking schemes unless we develop a marking system which must be congruent with the washback purpose of the exam.

Taking these points into account, the inter-intra-marker

reliability has been given a reasonable amount of attention. This will be discussed in detail in section 4.2.6..

So far as the methods of establishing examinee reliability are concerned, we found our choices limited because of the heterogeneous nature of the exam-items, and also for some practical reasons. In practice, the exam has to fit into a single administration situation within the whole school leaving examination system in Nepal. Considering this situation, and the prospective role of the exam as a selection device, it was necessary to include heterogeneous test-items. Another practical difficulty was that the prospective subjects had been preparing for their SLC examination when the exam was proposed to be administered. They were already tired of taking examinations. In this situation, we decided to use the Kuder Richard method for internal consistency despite the argument that it underestimates the reliability of an exam if it is composed of heterogeneous items.

2.3.2.1.8. Validity and Reliability: Tension or Compromise?

It was revealed that several factors are responsible for affecting exam reliability: e.g. the length of the exam, homogeneity of items, sufficient testing time. Considering the purpose of the exam, and the practical situation in which the proposed exam was supposed to work, we were bound to sacrifice some reliability in order to achieve the validity of the examination.

It was also found that the decision to sacrifice some reliability in favour of greater validity was not new. In this connection, Ebel (1979:275) argues that "reliability is a

necessary but not a sufficient condition for quality in an educational achievement test". Davies (1965:14) also has the similar opinion:

Reliability is the first essential for any test; but for certain kinds of language test may be very difficult to achieve.

Guilford (1965:481), Davies (1978), and Weir (1983a) hold the view that the construction of a single examination with a high validity and high reliability is impossible. There should not be a tension between the two at the cost of validity. Whenever a choice is to be made, there should be a compromise between the two since validity is the primary factor, as Weir (1983a:77) argues:

If... validity is lost to increase reliability we finish up with a test which is a reliable measure of something other than what we wish to measure.

Considering the purpose of the exam under consideration, we adopted the strategy suggested by Guilford (1965:482) and Weir (1983a:77) to sacrifice some reliability to increase the validity of the exam. However, reliability as an essential quality of an examination should not be underestimated.

2.3.2.1.9. Practicality

In the preceding sections, it has been shown that reliability and validity are the most important aspects of exam efficiency. Another aspect of it, though non-technical, is practicality in the absence of which even a valid and reliable exam can be of no use. Heaton's (1975:158) explanation of practicality is that the exam "must be fairly straightforward to administer". Generally, practicality involves the cost, and ease of administration, and scoring. We would interpret this rather in a vague way, such that an exam must fit in the intended situation

in all respects.

This aspect of exam efficiency is important for the proposed exam because failing to achieve the practicalities would lead us to fail to convince the authority concerned that the SLC English exam should be replaced with the proposed exam which is the crux of the matter. Further explanation is that, after all, the new exam was conceived on the assumption that it would solve some existing problems: e.g. the difficulty that college entrants find in coping with tertiary education. If the proposed exam is likely to create some other problem, so that it did not fit into the given situation, it would be difficult to convince the people who would actually implement the exam.

The whole argument is that, in order to achieve the practicality of the exam, the designer must keep a close look at the situation which the exam is supposed to fit into. Otherwise, the current literature based on sophisticated situations may lead the designer to be highly ambitious, and to forget the practical problems which are likely to occur at implementation.

The point to be emphasised is that we accept the current emphasis on the communicative nature of language testing, the importance of testing listening and speaking skills from the point of view of washback effect, and criterion-referenced testing. However, the proposed exam would show some sort of sacrifice in fully encapsulating the spirit of these points so as to enable the exam work in the Nepalese situation. For example, listening and speaking skills were not included in the proposed exam mainly because of practical reasons; we do not see the possibility of assessing these skills, at least for some years to

come, in the SLC English exam. The designer would have liked to include them. At the design stage of the exam, considerable attention was given to criterion-referenced testing, however, the results obtained from the exam, in practice, would have norm-referenced interpretation. As the proposed English exam is one of the several other exams in the SLC examination, its cutoff point cannot be different from the others. This situation makes it difficult to determine the 'criterial levels of performance' (Hughes 1989:50).

2.3.2.1.10. Conclusion

In this section, we considered the three constituents of the efficiency of an examination. It has been argued that, in order to give an exam a positive washback effect to make it work as a lever for change, it must be efficient. In making an exam efficient, the primary constituent to be considered is validity. Looking at the validation procedures of the proposed exam from the point of view of the purpose for which it is designed, it was agreed that the content and the predictive validity should be established. The face validity of the exam was proposed to be established by asking the subjects to make comments on how they felt about both examinations: i.e. SLC English exam and the proposed examination.

Accepting the potential role of content validity in giving an exam a positive washback effect, and of predictive validity in making an exam a good selection device, it was decided to sacrifice some reliability in order to increase the validity of the proposed exam. To make the exam work in the Nepalese situation, we took a practical perspective.

2.3.2.2. Negative Washback Effect of the SLC English Exam

In this section, we consider the washback effect of the SLC English exam. To exemplify the point, we examine the SLC English exam, 1989 in the light of the discussion presented above:

- i. washback effect of an exam
- ii. the objectives of teaching English at grades- 9-10
- iii. content validity

To recapitulate the washback effect of an exam, we have argued that an exam which is associated with extrinsic values cannot stop influencing teaching and learning. Since a final examination stands as a prominent criterion for determining whether schools have been doing what they are supposed to have been doing, teaching and preparing for the exam cannot be avoided. On these grounds, we have argued that a final examination should be designed such that teaching and preparing for it should facilitate learning, and that those who pass the exam cannot avoid having achieved what they are supposed to have achieved.

In section 2.2.1., we presented the general objectives of teaching English at grades 9-10 in Nepal. It has been said that though they look vague and difficult to evaluate, they clearly emphasise language skills: reading, writing, speaking, and listening.

To put these arguments together, the SLC English exam, to show that it has content validity, must represent the language skills which can be seen in the course objectives. Failing to do that can lead the SLC English exam to have negative washback effect because, while preparing for the exam schools will be doing something which is not in the course objectives.

To examine whether the SLC English exam, 1989, represents a sample of language skills presented in the course objectives, we refer to section 2.2.6.6. 'Analysis of SLC English exam, 1989'. Here we critically consider the exam under separate skills.

Reading.

As was exemplified above, Q.N. 1 (a-n), and Q.N. 2. in the exam were intended to assess the reading ability of the SLC candidates. Since Q.N.1 is entirely based on the prescribed textbook, all the items require the examinee to focus on the content. What these items require of the examinee is to recall the factual knowledge of the text.

Another aspect of these items is that, as has been exemplified, all the items were directly borrowed from either the textbook exercises or previous exam papers. The answers to these items are available in Bazaar notes. The items have high predictability as every year the same type of items appear in the exam. Davies et al (1984:31) have made a similar comment. To show how predictable the items are, we quote Davies et al's example (p.31) which shows that every year, the same question from the passage "The Disappearing Tiger" appears in the exam.

The Disappearing Tiger

1980 Why should we preserve tigers?

1982 State the reasons given by scientists for the preservation of tigers.

1983 Why is the preservation of the tiger considered necessary?

1984 Write down the arguments of people who want to preserve the tiger.

Question No. 2 is the only item in the exam which is an unseen task. The passages in this question were intended to

assess the reading ability of the examinee. On the surface, it appears to assess reading but, as we see the items which follow the passage, it becomes clear that they assess only one reading subskill: i.e. locating information from the passage. So, in practice, these passages do not require the examinee to exhibit other reading subskills: e.g. inference, guessing meaning from context, skimming.

Writing

Question No. 3 is intended to assess the writing ability of the student. As has been exemplified in section 2.2.6.6., the student does not have to answer this question in his own words. The answers to these tasks are available in the notebooks or commercial cribs. It is obvious that the SLC English exam does not assess writing ability at all.

Speaking

Question No. 4 is intended to assess the recognition of phonetic symbols by the student. Perhaps it was intended to assess the speaking ability of the candidate. We would argue that this is not how speaking can, and should, be assessed. From the washback point of view, it does not have any value at all. Another problem with this question is that, as has been exemplified, all the items were borrowed from previous exam papers which also encourage the candidate to memorise from the cribs.

Listening.

No attempt has been made to assess listening in the SLC English Exam.

Grammar

Question No. 5 to 15 were intended to assess the linguistic competence of the candidate. These items are based on

sentence level tasks. So they do not require the examinee to process the language above the sentence level.

Another problem with these items is that most of the items were borrowed, as has been exemplified in section 2.2.6.6., either from previous exam papers or from the prescribed text materials. Even these items encourage the candidate to memorise the exercises available in the source materials.

Overall, it can be seen that the SLC English exam and the general course objectives of teaching English at Grades 9-10 do not have any relationship at all. What the SLC English exam contains is a representative sample of what the students practise in the classroom, and what the students practise in the classroom is what they think will appear in the exam. Then it becomes clear that teaching and testing have isolated what has been conceived in the course objectives. As the exam fails to represent the course objectives, it does not have content validity.

Our argument is that if the SLC English exam does not require students to perform the tasks that require them to process language, then, as the SLC holds a key role in the Nepalese education system, teachers and students do not spend time on the tasks that require the candidate to process language. What the schools do is to prepare the candidate for the SLC English exam by providing them with a variety of answers to the questions which are likely to appear in the exam that year. This process of preparation for the exam does not facilitate learning at all. Therefore, we say that the SLC English exam has a negative washback effect on teaching and learning English in the Nepalese schools.

2.3.2.3. SLC English Exam as a Lever for Change

We have described, in the preceding sections, the ELT

situation and the importance put on the SLC English exam. We have criticised the existing ELT at school in the light of the discussion that even the students who pass the SLC English find it extremely difficult to cope with the subjects that require a higher level of English in tertiary education. To find out the reasons why the SLC passers have difficulty in coping with tertiary education, we examined the existing methods, the textbook materials, and the school leaving examinations used at the school leaving stage in the light of the recent trends in language teaching and testing. It was revealed that all the factors (i.e. methods, materials, and examinations) were unsatisfactory.

Mention has frequently been made that an exam which is associated with some extrinsic values, like our SLC, cannot stop influencing what goes on in the classroom. Seeing the SLC English exam in the light of this, it was argued that the SLC English exam has a great influence on what the English teachers teach in the Nepalese schools no matter what is there in the course objectives and the textbooks. Examining the existing SLC English exam, it was revealed that the exam was isolated from what was envisaged in the course objectives, hence a negative washback effect. Accepting its extrinsic values, it would be difficult to bring any effective change in the existing ELT at school without making the exam encapsulate the spirit of the course objectives.

Then it would appear that any change in the direction of bringing about changes in the Nepalese ELT situation would not be effective while the SLC English exam remains the same. This being so, we entirely agree with Davies et al (1984:50) who have concluded that:

The SLC exams do not test students' ability to function in English. They are unsuitable and require

complete overhaul. This is a major priority if the standard of English teaching and learning in Nepal is to improve.

We have considered the role of an exam as a lever for change. To exemplify the role, we have presented some four case studies in which the exams have been found to have played an influential role on the teaching and learning. Among the cases we have presented, the Turkish case and the Chinese case have shown that innovations through examinations can be effective. Getting insights from those case studies, and the theoretical considerations discussed in Chapter 1, we propose to bring about changes in the Nepalese ELT situation through changing the SLC English exam. It is hoped that changing the SLC English exam in relation to the educational needs of SLC passers would work as a catalyst to bring about changes in the teaching methods and the textbook materials used in the Nepaleses schools.

In the present study, it is not possible to deal with all these factors of ELT. We concentrate only on examinations in this study. However, in practice, we would not feel that the study is complete until the exam brings the desired changes in the whole ELT situation.

2.3.3. Summary

The need for English in Nepal for the development of the nation has been discussed. The existing ELT situation has been presented. The situation was found unsatisfactory.

We examined the SLC English exam, and found that it has a negative washback effect. Evaluating the ELT situation, it was felt necessary to bring about changes in the SLC English exam if we wanted to improve ELT in Nepal. It was argued that, since the SLC English exam has been playing an important role in the whole educational setting, no effective change can be made in the ELT situation unless the exam is changed.

CHAPTER 3: CONTENT VALIDATION OF THE NEW EXAM

3.0. Introduction

This Chapter is devoted to a priori validation of the proposed exam. In the preceding section 2.3.2.1., we have considered what makes an exam efficient. It has been argued that the basic considerations of the efficiency of an examination are: validity, reliability, and practicality of which validity is the primary consideration.

We felt it necessary to carry out the content validation of the new exam before designing it on the assumption that the validation would supply the necessary information about the specification and the format of the proposed exam. Doing this, it was hoped, would also enable the exam to encapsulate the aspects of language neglected by the existing SLC English exam.

We take the view that what is to be tested should be made clear not only to the exam designer, but also to the teacher and the student, especially when an exam is expected to have a positive washback effect. It has already been argued in section 1.5. that teachers and students cannot avoid preparing for the exam, and consequently, the exam cannot avoid influencing the teaching and learning. If the students and teachers have an idea about the aspects of language that are to be tested, then their preparation for the exam, we argue, will certainly lead them to achieving the skills to be tested, especially when not achieving them would be damaging for the students: e.g. low scores or failing the exam.

It will be argued that if we design the SLC English exam in relation to the language skills that the SLC passers need in

order to cope with the tertiary subjects which need a high level of English proficiency, the exam can bring about a desired change in the Nepalese ELT situation.

In this section, we first consider the approaches to language testing adopted over the decades. Secondly we present the guidelines developed in the light of the literature surveyed, and the discussion of the Nepalese ELT situation for the design of the new exam. Finally we consider the nature of reading and writing. On the basis of the discussions, we present the specification, and the format of the new exam.

3.1. Theoretical Considerations: Approaches to Language Testing

3.1.0. Introduction

In this section, we consider different approaches to language testing which have been adopted over the decades. We will discuss those approaches in terms of their contributions to language testing so that insights can be obtained for determining what type of test would serve our purpose, and what guidelines we should follow to achieve what we want to achieve through the exam.

3.1.1. What is to be Tested?

In a language test, it is obvious that the stuff to be tested is language. But the phenomenon of language is so complex that we must have some rationale for deciding what aspects of language are to be tested since we cannot test the whole of a person's language. Jacobovits (1970:75) finds it difficult to determine what is to be tested in language testing because of the complex nature of language. For example, he says:

the question of what it is to know a language is not well understood, and consequently, the language

proficiency tests now available and universally used are inadequate because they attempt to measure something that has not been well defined.

It may be because of the indefinite form of the material to test that the decision on what aspects of language are to be tested has been changing over the decades. What was considered to be tested in language testing in the 1950s changed in the 1960s, which also changed in the 1970s, and again in the 1980s. It can, therefore, be argued that what to test should be considered in terms of what Morrow (1977:13) calls the "spirit of the age". Because the understanding of what to test in second language testing is not independent, the idea about what aspect of language to test changes as the understanding of how a second language is learnt, and how it is taught changes. According to Davies (1968:1-2), the testing of second language proficiency tends to follow teaching methodologies, but Hughes (1989:1) maintains that in most of the cases language tests have a detrimental influence on teaching and learning because of their failure in measuring what they intend to measure. In teaching there have been changes from one approach to another. One method which was strongly recommended has been succeeded by another method with severe attack on the validity of all preceding methods (Farhady 1983:311). This tendency has a strong influence on determining what aspects of language to test when testing second language proficiency.

If we look at what to test from the point of view of the general notion of a final examination, what we test will be seen as the description of achievement of the course objectives. Examination of achievement in a classroom situation has to compare with what the pupils being assessed are supposed to have

achieved and what they have really achieved. In other words, the examination needs to be seen as a measurement of the attainment of course objectives, and the outcomes as the result of instruction: what change in the pupil's behaviour the instruction has brought about.

Then it would appear that what is to be tested should be seen in terms of what we count as valid educational knowledge, and what we want the student to achieve.

In a formal educational setting, what we want the student to achieve can be considered the course objectives of the program. Rowntree (1977:90) describes what he means by objectives as:

the skills, abilities, knowledge and understanding in which the teacher intends that students should improve as a result of his interventions.

According to Bloom (1972:38), skills and abilities are the traits which enable the individual to find "appropriate information and techniques in his previous experience to bring to bear on new problems and situations".

Applying this notion to language testing, then, it is the skills and abilities to use language that are to be tested. In relation to what is to be tested, the Department of Education and Science (1977:2), in its 'Why, What, and How' leaflet, emphasises testing "communication through reading, writing, listening and speaking in a wide range of models to suit the occasion, the purpose and the subject".

3.1.2. What is to be Tested: A changing Concept

What is to be tested has passed through different stages over the years. It has been considered under separate headings: pre-scientific, psychometric-structuralist, and psycholinguistic

-sociolinguistic (Spolsky 1978b); psychometric, psycholinguistic -sociolinguistic, and communicative paradigm (Weir 1988); intuitive, scientific, and communicative (Madsen 1983). However, the demarcation line between them is spurious.

To get insights for determining the format and the specifications of the proposed exam, the different trends in language testing will be considered in terms of what contributions they have made to language testing. They will be discussed under the following headings: pre-discrete point, discrete point, integrative, functional, and communicative testing.

3.1.3. Pre-discrete point Testing

Pre-discrete point testing is used to refer to the language testing situation before the concept of discrete point testing came into existence. The pre-discrete point approach to language testing is based on the assumption that no special expertise is required for testing: e.g. any teacher can do it. Language examination is entirely subjective in the sense that the total assessment is based on the subjective judgement of the examiner. Following this trend, the abilities or skills to be tested, or the aspects of language to be tested are the abilities to translate, and write open-ended essays. The format of the examination takes the form of translation from L1 to L2 or L2 to L1, open ended essays in L2, and grammatical items.

3.1.4. Discrete point Testing

The discrete point approach came as a reaction to the pre-discrete point approach which has been severely criticised for being not reliable and not valid. This trend is based on the

assumption that language testing can be made precise, objective, reliable and scientific (Spolsky 1978b:vi). This assumption rests on the language theory that "knowledge of the elements of a language is equivalent to knowledge of the language" (Morrow 1979:145). Read (1981:ix) considers that this trend is highly influenced by the structural school of linguistics which views language as "a hierarchical system of units - sounds, morphemes, words, grammatical structures" that could provide the basis for the selection of test items.

The discontent with the unreliable judgement following the pre-discrete point trend led to the aim for more reliable judgements or new kinds of tests which would control unreliability. This major concern with reliability led to the development of objective types of test which can be labelled 'discrete structure point' tests (Spolsky 1978b:vii).

So what is to be tested, following this trend, is the ability to use aspects of language - phonology, lexicon, grammar, and syntax - in a mechanistic way. The format of the test is composed of short answers, and multiple choice items. The major contributions of this phase are the concern with reliability, and the construction of objective type tests.

This approach has been attacked on the grounds that language is not merely an arrangement of elements which can be tested in terms of yes/no answers, and that require the examinee to exhibit only the ability to use the language elements in isolation. Rather language should be seen in terms of the total context in which it is being used in an authentic-like setting.

Rea (1978:51) (quoted in Weir; 1988:3) criticises

discrete point testing for not giving the information about how the examinee uses language in real situation. For him, the information produced by discrete point tests is irrelevant and artificial.

Morrow (1979:145) criticises atomistic types of testing on the ground that language is something more than mere language elements. He argues that knowledge of language elements has no importance without the knowledge of language use.

For Spolsky (1985:182), discrete point testing gives only a limited information about the testee, hence the problem for extrapolation.

Oller (1979:212) makes this point more explicit when he argues:

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. The fact is that in any system where the parts interact to produce properties and qualities that do not exist in the part separately, the whole is greater than the sum of its parts. If the parts cannot just be shuffled together in any old order- if they must rather be put together according to certain organisational constraints- those organisational constraints themselves become crucial properties of the system which simply cannot be found in the parts separately.

The attacks on discrete point testing provided an impetus for another development in language testing which will be considered in the section which follows.

3.1.5. Integrative Testing

The emergence of integrative approach to language testing is based on the assumption that "knowledge of a language is more than just the sum of a set of discrete parts", and also on the belief that it is only the integrative test items that gives a

true measure of language ability (Spolsky 1978b:viii). Another point associated with the integrative test is the concern with 'overall language proficiency' popularised by Oller (1979). Spolsky (1978a:69) defines overall proficiency as "ability to operate in a language ...plus a number of specific areas based on experiences, and which will turn out to be either the skills or certain sociolinguistic situations".

Moller (1982:16) considers the notion of general language proficiency the presence of a general factor of language proficiency, which can be indicated by the relatively high interconnectedness obtained between sub-tests.

Spolsky (1985:186) agrees with Oller (1979:37) who advocates the assessment of general language proficiency (i.e. describing language on a single gradient), and believes that it is the integrative test (e.g. cloze) not the discrete point test that does the job. Oller goes on to argue:

The concept of an integrative test was born in contrast with the definition of a discrete point test. If discrete items take language skill apart, integrative tests put it back together. Whereas discrete items attempt to test knowledge of language one bit by at a time, integrative tests attempt to assess a learner's capacity to use many bits all at the same time, and possibly while exercising several presumed components of a grammatical system, and perhaps more than one of the traditionally recognized skills or aspects of skills.

To sum up what has been pointed out, we quote Davies (1978:141) who supports integrative tests on the assumptions that:

1. language is not a set of unrelated bits, that it forms a whole and that the bits must be integrated and tested in combination with one another.
2. language learning is purposeful, that the purpose is always communicative ability, and not formal language.
3. discrete point language tests are too general to be of value and that what are required are specific tests.

For Oller (1975:52, and 1979), cloze and dictation tests

are the valid test for the assessment of overall language proficiency. He considers the strong correlation between them the indication of the validity of such tests.

However, Weir (1988:5) disagrees with Oller, and argues that high correlations between cloze and other measures may mean that they are measuring different skills which are highly correlated among individuals. It does not necessarily mean that, for example, there will be no individuals who do not differ in terms of their performance in the various skills.

Farhady (1983) finds discrete point tests have high correlational evidence as integrative tests. In that case, Oller's correlational validity of cloze and dictation cannot be without suspicion.

Carroll (1983:104) maintains that general language proficiency tests are too general, and suggests that tests should include specific language skills. Skehan (1988:213) finds some empirical support for skill based division of language proficiency.

Moller (1982) criticises cloze and dictation types of test on the grounds that they do not assess the productive skills, and do not require the testee to perform the tasks which can be considered relevant with his future use.

Morrow (1979:22) criticises cloze and dictation for not being communicative.

Despite the above criticisms, integrative approach to language testing led to the conclusion that a valid language test requires the testee to process language in a given context.

Read (1981:x) looks at language from a psycholinguistic

point of view, and finds it a dynamic, creative, functional system. It is, therefore, he argues, difficult "to show that any single linguistic unit is indispensable for communication". Looking at language from a sociolinguistic point of view, he observes that communicative competence is the most important aspect which covers "not only knowledge of rules for forming grammatical sentences but also rules for using those sentences appropriately with different people in different context". The implication is that language testing procedures should require the testee to use the language in an authentic-like setting, and the assessment should be based not only on grammatical correctness but also, more importantly, on contextual appropriateness and communicative effectiveness.

3.1.6. Functional Testing

The debate whether language proficiency can be assessed as a single factor led to a development of the functional approach to language testing. It was argued that "the nature of language knowledge is best captured by detailing the various uses to which the language can be put" (Spolsky 1985:182). The argument is that performance of the examinee may vary from task to task. For example, Alderson and Urquhart (1985) have shown that academic background can have an effect on test scores in reading comprehension. In that sense, collecting information about the testee only with a test based on the overall language proficiency construct (e.g. cloze) may not give a true picture of the ability of the examinee.

The functional approach is based on the assumption that knowledge of language should be seen in terms of language

related functions, not in terms of underlying grammatical structures (Spolsky 1985:185). This approach places importance on performance rather than on the linguistic ability of the examinee. Tests based on this approach are sometime called performance tests.

The purpose of performance tests is to assess the ability of the examinee in terms of job-related tasks. Jones (1985:16) refers to Sanders and Sachse (1977) who maintain that performance tests measure "performance on task requiring the application of learning in an actual or simulated setting".

This approach, then, would appear to approximate testing to real life situations. It also provides insights for direct assessment of the language ability of the examinee in relation to his needs for using the language.

Wesche (1985:1-2) finds performance testing useful for assessing achievement and predictive purposes.

The construction of ELTS raises an interest in functional language proficiency which emphasises "the ability to communicate effectively in specific situations determined by the needs of a particular, restricted, social group" (Moller 1982). On that ground, different sub-tests for different groups of people, for example, Life Sciences, Physical Sciences, Technology, Medicine, Social Studies and General Academic, were constructed. This approach led to the development of English for Specific Purposes proficiency tests.

3.1.7. Communicative Testing

A greater emphasis on communication in language teaching marked a shift in testing communicative ability which places

more importance on language in use than language in usage. In other words, communicative approach to language testing is concerned with assessing effectiveness in communication rather than "the formal elements structures, and systems of the language" (Porter 1983b:190).

This approach is aimed at assessing communicative competence. The widely accepted model of communicative competence is the one that was proposed by Canale and Swain (1980) and further developed by Canale (1983). The model is considered to give the comprehensive description of communicative competence.

Canale (1983) demonstrates that the communicative competence model includes the four components of language competence: i.e. grammatical competence, sociolinguistic competence, discourse competence, and strategic competence. Following this model, grammatical competence is concerned with the knowledge of formal system of language; sociolinguistic competence is concerned with the knowledge of appropriateness of utterances within the sociolinguistic context; discourse competence is concerned with the knowledge of processing language beyond sentence level in different modes; and strategic competence is concerned with verbal-nonverbal strategies for effective communication.

Following this model, the main concern is the nature of test tasks (Porter 1983b:191) which should reflect the criteria "employed in the assessment" (Weir 1988:9). This would also imply that the test must have content validity. For Davies (1988a:13), what makes a language test communicative or non-

communicative is "the problems, the tasks, the skill manipulations". Davies (1988a:4) considers context and purpose the only features relevant to testing of communicative language. According to Morrow (1977), the features that make language tasks communicative are: interactiveness, unpredictability, context, purpose, performance, authenticity, and behaviour-base. Some relevant points of these features will be touched on in section 4.2.1.1..

3.1.8. Summary and Comments

We have considered the different approaches to language testing. It was observed that there have been changes in what to test because of the changes in understanding what it means to know a language, and to teach a language.

It has been argued that what is to be tested has passed through different stages over the decades. For example, following discrete testing, it is the mastery of language elements that is to be tested; following integrative testing, it is the general proficiency that is to be tested; following functional testing, it is the language skills related to a particular job that are to be tested; and following communicative testing, it is the communicative ability that is to be tested. Performance tests and communicative tests thus appear to test similar things.

To get insights for the design of a new exam, we would like to briefly consider these approaches with the perspective of issues like language proficiency, language skills, generalizability, and needs-orientated testing.

The hypothesis that language can be handled as a single

factor has been widely discussed in the testing literature, and eventually rejected. It is argued that language proficiency is composed of linguistic competence and language skills (Carroll 1983, Spolsky 1985, Skehan 1988). On that ground, it can be argued that a test based only on a general language proficiency construct would not give a true picture of the language ability of the testee unless other components that contribute to the language skills are included. Then, following performance testing, language skills are assessed (e.g. ELTS) according to the testee's need for language for a certain purpose. At the same time, a language test has also to satisfy theoretically or empirically the criterion of generalizability. Now the argument is: when we talk about the generalizability of a test, are we interested in generalizing the test to the core of the language or to other tasks of a similar type? The issues of needs orientation and the generalizability of a test appear to be contradictory. If the test is designed to look for whether someone has the ability to do something in particular, how can the ability discovered by this type of test be generalized if we follow Alderson (1989:3) who argues, for example, that "it is misleading ... to say that someone can or cannot understand a test or type of test which s/he has not been asked to read"? In that case, can need-orientated testing and generalizability go together? Or how generalizable should a test be?

Considering these points, it seems that inclusion of items to assess linguistic competence is safer (Spolsky 1985:183) in order to increase the generalizability of the test.

3.2. Guidelines for the Design of the New Exam

On the basis of the survey of the literature on trends in language testing, and the discussion on the Nepalese ELT situation, we developed guidelines for the design of the proposed exam. The following issues were focussed on:

- i. Needs-orientated approach
- ii. Proficiency plus language skills construct
- iii. Washback effect of the exam.

3.2.1. Needs-orientated Approach: A Practical Consideration

It must be clear by now that the design of the main study was conceived in the light of the needs for English in Nepal in general, and the needs for English for tertiary education in particular. The idea was that making the exam reflect the abilities that the tertiary entrants are expected to have achieved would lead the teaching of English at school to change in the desired direction. It should be noted that one of the purposes of suggesting changes in the SLC English exam was to make the school English congruent with the basic needs for college English.

At the design stage, it was not feasible to carry out a needs-survey to determine what language skills and subskills should be included in the exam. So the design of the exam was based on the information obtained from Davies et al (1984), the objectives of the SLC English course, and the researcher's personal experience as a teacher of English in Nepal. However, we also decided to carry out a survey of the needs for English for college students in order to avoid entire dependence on an armchair speculation in further improvement of the exam.

It would be strange to talk about needs-analysis while considering an achievement test. Nevertheless, for the purpose of the present study, we felt that it was necessary to consider the needs for English for tertiary students since the SLC exam functions both ways: e.g. forward looking, and backward looking. Though the focus of the study was on the needs for English for college students in Nepal, it could be argued that the basic language skills which could serve the purpose of tertiary students would also serve the candidates who would like to use English for occupational purposes. Taking this situation into account, we decided to consider, though it may be in a crude form, what level of language proficiency could be expected from a school leaver who spends at least six years on studying English at school level, and what sort of language proficiency is needed for a college entrant in order to determine the content of the exam. We also considered the course objectives of the SLC English which is taught at Grades 9-10 for the content validation of the exam. Another source of information was Davies et al's (1984) study which was carried out for the purpose of improving the teaching of school English in Nepal.

On the basis of the information obtained from the above sources, the expected level of proficiency of a school leaver was determined as follows.

Reading

As was mentioned earlier, the main purpose of teaching English at school is to develop reading ability in the students to enable them to read books either on content courses, or on courses related to general English or ESP in colleges. We have

already said that most books in the faculties - Faculty of Agriculture, Faculty of Medicine, Faculty of Engineering, Faculty of Science, and Faculty of Forestry - are available only in English. The English courses at tertiary level, both general and specific, also require the students to have a high level of reading ability. The stated objectives of the SLC English course (section 2.2.1.) are to develop the essential reading skills in the students. On this ground, we found it reasonable to expect a school leaver to be able to exploit basic reading skills while reading any text suitable for that level.

Writing

The courses which require a high level of reading ability require writing ability as well. The students have to answer questions in English in internal and external exams in these courses. For that purpose they need to take notes from a text, and elaborate them into a summary. The objectives of the course also aim at developing their communicative ability in writing. They need to be able to write formal and/or informal letters for academic and occupational purposes. In short, they ought to be able to express their ideas, and opinions in simple English. They must have, therefore, the ability to identify main ideas, and organise coherent paragraphs.

Listening

Students at this stage ought to have the ability to understand simple English spoken by a competent speaker of English. They should be able to take notes from lectures which are given in English. They should also be able to understand their fellow-students, and directions and instructions in a

classroom setting. The objectives of the course are stated as being to develop the students' oral communication as well.

Speaking

The school leavers ought to be able to speak about themselves, their family and locality and ask others about the same topics. They should also be able to ask questions and present arguments in English because at tertiary level the subjects which are taught in English are conducted in an English speaking environment. Not being able to ask questions in English may restrain their learning.

3.2.2. Assessment of Listening and Speaking: A Practical Problem

It has already been mentioned that in the existing SLC English exam, no items were included to test listening and speaking. But the brief survey of the needs for English for tertiary education revealed that listening and speaking are also necessary for the school leavers. These skills have also been emphasised in the objectives of the SLC English course. It is also necessary to think that in every walk of life, and in every part of school curriculum, oral communication plays as important a role as written. It is, therefore, necessary to teach listening and speaking in schools, and therefore, also to test them.

Any assessment system which has as its main aim the testing of students ability to put a language to use will include an aural/oral category. We would have liked very much to include tests of listening, and speaking in the exam. However, our technical and logistical problems are such that testing oral communication nation-wide for the purpose of certification exam,

like SLC, is not possible, at least, for a few years to come. The explanation is that, as the SLC exam is nation-wide, the number of participants is so great in each exam centre that it would take weeks to give them an oral interview. It would be an insurmountable problem of administration in the Nepalese situation, and technically unfeasible: e.g. required number of trained interviewers, and secrecy of the test. In respect to testing oral/aural, another problem is the difficulty in supplying equipment to all the exam-centres.

In spite of these problems, what is essential is the provision of testing oral communication, somehow, from the washback point of view. For that purpose, we would like to recommend strongly the inclusion of such components (i.e. listening, and speaking) in the qualifying test (a mock SLC) for the SLC. We see the possibility of testing oral communication in the qualifying test because it is locally controlled by the district education office, and the school teachers play the decisive role in it.

From now on, this study deals only with reading and writing skills.

3.2.3. General Proficiency and Language Skills Construct

It has been argued that the SLC English exam looks like an achievement test in the sense that it has a given syllabus. On the other hand, it is also an achievement test in the sense that its score has a face value for college entry. At school level, it cannot be decided which student is going to join which faculty. In this situation, the SLC English must serve the general purpose of enabling those who pass it to cope with the

tertiary courses which require English, to have the basic skills in reading and writing which are applicable to any faculty. The survey of the literature (section 3.3.) reveals that the reading and the writing abilities of a student depend on his linguistic competence as well as the particular language strategies adopted. In other words, to be able to read and write properly, a candidate needs a formal knowledge of language and ability to use that knowledge. To serve this purpose, it was decided that the new exam should include the test-items which reflect overall language proficiency as well as the language skills essential for tertiary education. In this sense, agreeing with Bialystok and Smith (1985), and Spolsky (1985), the study was based on the assumption that language proficiency is a construct composed of linguistic competence and language skills.

3.2.4. Washback Effect of the Exam

As was discussed in section 2.3.2.3., the proposed exam has to play a role as a lever. In that sense, it must have a positive washback effect. It has been frequently mentioned that a communicative test has a positive washback effect (Porter 1983, Swain 1985). Taking this view into consideration, a communicative competence model was found to be suitable for the proposed exam. However, the exam was not called communicative in order to avoid strong resistance for change from people orientated to the structural approach to language testing. And also the test was not fully communicative in the usual way a communicative test is described, since it included a cloze procedure as well for assessing linguistic competence. In any case, what was important was that the exam must have a positive

washback effect. To ensure that the exam has a positive washback effect, we considered the features that make a test communicative at the design phase. We also developed a scoring scheme that encapsulates the spirit of communicativeness of language testing so that the features that are salient for scoring would be emphasised in the teaching of English in schools (i.e. washback).

3.3. Assessment of Reading and Writing

3.3.0. Introduction

Considering the Nepalese ELT situation in the light of the survey of the literature on trends in language testing, we decided to concentrate on the assessment of reading and writing. To determine what to test in reading and writing, it is important to understand their nature. In this section, we consider the nature of reading and writing for the purpose of deciding 'the what' and 'the how' of testing these skills in the proposed exam.

3.3.1. The Nature of Reading

3.3.1.0. Introduction

It is generally accepted that reading is the most essential activity for a student at college level. Reading comprehension is so pervasive and complex that it is difficult to pin down what reading is composed of, and what is necessary to develop this ability in a learner. The same type of difficulty lies in determining what, and how to test reading comprehension.

We do not propose to go into details beyond the capacity of the study about reading comprehension. What we would like to

do is to touch on some relevant points about reading so as to supply a basis for determining the content, and the format of the exam. In this section, first we consider whether reading is unitary competence or a series of subskills, and then we look at it from a testing point of view.

3.3.1.1. Reading Comprehension: Unitary Competence or a Combination of a Set of Subskills?

There have been several studies on the nature of reading comprehension. However, the dispute whether reading is a unitary act or a series of skills seems to be unresolved; some research studies find it a single factor, and some studies find it a composite form of several subskills.

Is reading comprehension a combination of reading subskills? Harrison and Dolan (1979) discuss the studies carried out by Davis (1944 and 1946, Threstone 1946) on the nature of reading comprehension. In these studies, data was analysed by using factorial techniques. Davis came up with results which show the involvement of five factors in the process of reading comprehension. But Threstone found only one factor salient when he analysed the same data. So the argument that reading comprehension is a series of subskills was not beyond suspicion.

Harrison and Dolan (1979) report the study based on the Edinburgh Reading Test. This test was designed by using four passages (600 words) to measure different subskills - word meaning, literal comprehension, inference, metaphor, selecting salient points from a text, and evaluation. The data obtained from the test was analysed by performing different statistical analyses including principal component analysis (Kaiser 1970). The researchers could not find any clear indication so that they could argue that the test measured distinct abilities.

This assumption was further backed by Lunzer et al's (1979) study. They designed batteries of tests to investigate whether reading comprehension is unitary or manifold. They used factorial techniques for data analysis. Their hypothesis that reading comprehension is a set of subskills was unconfirmed.

If we take these arguments into account, it will be problematic to consider reading comprehension as a series of skills.

If so, is reading comprehension unitary competence then?

Looking at the other side of the coin, the focus will be on finding what reading comprehension is composed of. One would argue that if there is any distinct activity that can contribute to improve reading comprehension, then that activity should be considered to account for reading comprehension. In other words, any contributing elements involved in reading can be taken as subskills of reading comprehension no matter what factorial techniques present as the results of the analysis. The information based on the elements that contribute to reading as the total activity would be more convincing than the information entirely based on statistical analysis like factorial techniques.

While defining 'reading', the Department of Education and Science (1975) describes three types of skills involved in reading: Primary, Intermediate, and Comprehension.

'Primary skill' is seen as the responses to the print by recognising the stages of the separate letters, groups of letters and the whole words. The reader must have a reasonable mastery of this process of seeing a letter or group of letters before he can respond to the sequence of words.

'Intermediate skill' is seen as the ability to handle the sequence of letters, words and larger units of meaning. This

skill involves the knowledge of the probability with which sequences occur. In other words, while reading a sequence, a reader has to be able to anticipate what is most likely to follow it. Here, in this way, he can make guesses at the meaning of the familiar words and also specify the meaning of the words in terms of context.

'Comprehension skill' is seen as the way a reader extracts meaning from the printed page: i.e. understanding the writer's intended meaning. It is also seen as the way a reader formulates ideas associating what he understands from the printed page with the ideas he already had about the topics. In other words, it is seen as an interaction between the meanings a reader derives from the passage and his purpose in reading it. In the process of reading, he associates these two types of ideas and modifies them.

On the whole, it is concluded in the report that reading comprehension is the composite result of all these three types of skills (Department of Education and Science 1975).

Hosenfield (1983) has carried out some research studies to uncover the kinds of strategies students use in solving problems of understanding foreign language texts. She (p.233) finds that the skilled readers tend to:

- keep the meaning of the passage in mind,
- read in broad phrases,
- skip inessential words,
- guess from context the meaning of unknown words and have a good self-concept as reader,
- identify the grammatical category of words,
- read the title and make inferences from it,
- examine illustrations,

-use their knowledge of the world.

The implication of what has been mentioned is that reading is not merely a matter of response to letters, and spelling patterns, it is also not only reconstruction of the authors meaning. What it is is 'the perception of those meanings within the total context of the relevant experiences of the reader - a much more active and demanding process' (Department of Education and Science 1975:95).

The survey of the literature presented above was based on several assumptions of what constitute reading comprehension. On the whole, it appears that reading is an activity which can be seen as a complex task which depends on "a multiplicity of perceptual, linguistic, and cognitive processes" (Adams 1980:11); an interactive process involving linguistic, psycholinguistic, and pragmatic information that a reader has at one time; and a strategic process in the sense that understanding of a passage can be achieved if the reading involves purpose, motivation, and real effort of the reader. When we look at reading as an interactive process, it is necessary to observe what happens in a reader when he actually carries out the activity. But from the testing point of view, especially in an exam situation, it is not possible. The impossibility of testing the process aspect of reading comprehension in an exam situation forces us to concentrate on the product aspect of it.

The product aspect of reading comprehension depends on what the reader gets out of the given text which does not say anything about the procedures he follows to arrive at certain decisions. Considering only the outcome aspect of reading comprehension seems to be incomplete in the sense that different readers may follow different routes to arrive at the same

outcome. However, from the testing perspective, we are interested in finding out an indication of the levels of understanding of readers. This argument does not imply that the process aspect is not important. We accept the view of Alderson and Urquhart (1984:xviii) who warn that the product of reading can vary according to the purpose and interest of the reader. In the assessment of reading, these issues are important for designing test-tasks.

Bransford et al (1984:42) conclude that comprehension depends on the linguistic ability of the reader, and how he activates his linguistic ability while doing reading. So the ability to activate one's linguistic ability is vital in comprehending a text.

Gardner (1978:77) finds that adopting reading strategies contributes remarkably to gaining reading comprehension. He (p.68) holds the view that reading comprehension is not merely a function of capabilities within a reader. A reader encounters different kinds of texts in terms of their difficulty level. The difficulty is partly because of the reader's inability to process it, and partly because of the writer's way of making it difficult to read. Gardner argues that reading depends upon how the reader overcomes the difficulty created by the writer. The implication is that reading difficulty can be reduced by adopting reading strategies. For him, therefore, understanding a passage is the function of the intention of the reader and the assumed reading competence.

For Hall et al (1986:91), skilled reading involves building an interpretative structure for a text using the available text base together with one's prior and contextual knowledge.

Gardner (1978:72) reports a study which claims that comprehension can be improved by training in vocabulary and retention exercise. He also reports Berry's (1931) finding about a difference between general comprehension and comprehension of detail, Dewey's (1935) finding about the difference between obtaining facts and inference, Feder's (1938) finding about a distinction between reading for information and reading for inference. These studies lead us to consider reading a set of subskills.

Though Lunzer and Gardner (1979:37) could not find reading comprehension a set of subskills, they admitted that effective reading depends on the use of study skills. For them, skimming, scanning are study skills which, they argue, need to be deployed if any benefit from reading is desired.

Based on the above discussions, it seems logical to deal with reading comprehension by accepting it as an activity which involves several other sub-activities, such as skimming, scanning, working out the meaning of words from context, making inferences. The argument is that if it is evident that doing any particular activity can enhance a learner's comprehension ability, then why not pursue the study that way. The implications for our exam is that if adopting reading strategies - like skimming, scanning, inference, guessing meaning from context - makes the reader able to do his job much better, then including test items that reflect these skills would require the students and teachers to concentrate on them in the classrooms. As a result, the pupils' reading ability will be improved. The aim to make the learners able to read a given passage for certain purposes.

3.3.2. The Nature of Writing

To determine what to test in writing, we consider, first, the nature of writing. Then we look at it from the testing point of view.

Gannon (1985:25) defines writing as;

a way of realising language in those communities which have acquired a writing system. It is characterised by its use of visible signs systematically ordered.

According to Widdowson (1978:62), one way of describing writing is to say '... the use of the visual medium to manifest the graphological and grammatical system of the language' and another 'that writing is the act of making up correct sentences and transmitting them through the visual medium as marks on paper'. He further considers writing an activity of developing a discussion as transferring information of various kinds from the writer's world knowledge to that of the reader's and that linguistic rules facilitate the transference, though mere linguistic rules are not sufficient to do this, knowledge of use is needed.

Looking at writing as a piece of discourse (whether it embodies correctness or appropriateness of style, theme, topic) is to consider it a finished product. Seeing writing from this point of view does not tell us anything about the processes involved in writing. However, writing is not any longer seen as merely a way of putting ideas on paper. Rather it is considered a process of generating meaning as Spencer (1983:77) states:

Complex thought and feelings are crystallised in words and become more directly available for reflection, critique or response.

Taylor (1981:6) also accepts the heuristic role of writing which is used to explore beyond the context we know. This happens when an interaction between content and language takes place. We

agree with Odell (1981:43) who argues that writing is a good activity to improve our understanding of any subject. The assumption is that the knowledge or ideas in a crude form may be refined by trying to put them on the paper. The argument is that writing is an activity which involves the writer in the process of formulating ideas, then testing and confirming them.

A brief survey of the literature suggests that writing is not seen just as one of the language skills to be learned, but as an effective way for a learner to generate words, sentences and even chunks to discourse. This is how the recent studies on writing have shifted the emphasis away from the finished product on to the process involved in writing: e.g. the activities through which complex ideas evolve into a written text. From a pedagogical point of view as well, the process cannot be ignored. As Arndt (1987:257) points out:

... it may be ill-advised, and perhaps even impossible to divorce the processes and products from each other, either in teaching or research. For at the heart of effective writing lie the techniques for successful fusion of thought and language to fit the rhetorical context.... in the fundamental sense of gearing message to audience.

It can, therefore, be argued that teaching of writing should be based on giving students experience in the composing process, rather than upon prescribing forms overly concerned with correctness. For example, Krashen (1984:36) warns:

...providing these students with more information about the product will result in a wooden, awkward approximation to good prose.... it can seriously inhibit their writing process... overteaching can prevent the discovery of new meaning and can seriously disrupt fluency and the flow of ideas on to the page.

Researchers have felt that the problems that students find in writing L2 are the constraints of the foreign language. It has also been found that unskilled writers use inefficient writing strategies and skilled L2 writers use effective strategies of evaluation and text-generation. The explanation is that a skilled

writer revises 'what has been written' several times and writes further while maintaining consonant relationship with this. To put it another way, he keeps revising and reorganising 'what is said' and generates meaning 'what to say next' in order to make the writing organised (Arndt 1987:258).

Perl (1979:324) considers the recursive nature of writing the most salient feature of composing process. She also finds, in her study, a sequence of patterns of pre-writing, writing, and editing (p.328).

Zamel (1983) has carried out a study of six ESL students studying in an American university. On the basis of the study, she presents some strategies that skilled ESL writers follow when writing. Such writers:

- write primarily to express meaning, i.e. are more concerned about the content than the form,
- pay more attention to revision at the discourse level than at the word level,
- constantly read back what they have written for organisational details and reorganise whole chunks of writing as well as discard ideas to incorporate new information, and express original ideas, etc,
- distance themselves from their writing and read it from the point of view of a target reader and supply modification,
- edit and improve their writing till they feel it reflects what they want it to reflect.

The implication of this is that if we really want to develop our students' writing ability, looking only at what they have written is not enough, we must understand how that product

came into being. We also have to try to understand what goes on in the mind of the writer during the act of writing.

What can be suggested, here, is that a writing programme should aim at combining both process and product orientation. In the same way, a real test of writing should assess whether the intended testees have acquired certain writing strategies, as for example, organisation, revision, coherence, etc., and the linguistic rules as well. But the problems in assessing writing ability has another history.

3.3.3. Writing from the Testing Point of View

Despite many research studies on writing, as Spencer (1983:79) points out, learning to write is such a complicated and still inadequately researched process that it is not easy "to identify simple categories of teacher behaviour which indubitably constitute good teaching of writing". In the same way, the Department of Education and Science (1975) expresses the same difficulty in assessing writing. The DES' problems were: how to decide what features of writing should be tested? by what criteria is one to measure them? how are reliability and validity to be ensured? To help with these problems, the report suggests the reconciliation of the marking system and the difficulties in administering the test. In other words, a practical approach (e.g. doing the best whatever possible) can be adopted.

The immediate implication of the suggestion is that writing as such should not be assessed simply by asking students to complete a task or solve a problem within a limited period of time, because it does not happen in a real life situation: (i.e. in a real life situation a student will not have to write a reply to a letter in 10/15 minutes, but in an exam, he will be assessed through such tasks). But what can be done in a situation like the

SLC exam in Nepal, which is used to assess all that a student has achieved at school in English in a 3-hour exam? Perhaps, it may be because of such frustrations, people working in this field have been arguing that assessment systems such as yearly exams, should be abolished and it should be replaced with a system of continuous assessment. We agree with the argument, but it is not practicable in all situations, as we have already argued that a final exam is unlikely to disappear.

Now the question is: what is to be assessed in writing? The answer is: writing strategies; such as the ability of combining information so as to make it an organised piece of writing, the appropriateness of the writing according to the intended reader. Putting it another way, in assessing writing, the important question is not just a matter of seeing how well pupils can use commas and full stops; the important question is of seeing how appropriately, how flexibly, and how confidently the students can use the resources of the writing system for desired purposes.

However, looking at the methods of assessing the writing ability of students, especially in an exam situation like the SLC, when the students are asked to produce pieces of writing in a limited time, it is very difficult, if not impossible, to assess the process aspect. Only the product (i.e. the finished work) is assessed. What can be suggested in this situation, as the research studies show that a skilled L2 writer produces an organised and coherent piece of writing only by exploiting the writing strategies, is that emphasis should be given to the organisational aspect of writing by allocating separate marks to organisation, content, and appropriateness and style on the one hand, and on the other hand, to knowledge of grammatical and

orthographic conventions. Here, organisation involves cohesion and coherence, content involves supporting details for the important points in a piece of writing, appropriateness and style involves the register, and the proper use of words and knowledge of grammatical, and orthographic conventions involve formal rules, and word-spelling. It can be argued that this way of assessment of writing would lead teaching to focus on the process aspect. Further discussion on this point will be presented under section 4.2.6..

3.4.Specifications of the New Exam

3.4.0.Introduction

In Chapter 2, we have already described the Nepalese ELT situation. The chapter included the description of the purposes of teaching English, the objectives of the SLC English course, and the assessment systems (SLC exam). We have also described the expected level of language proficiency, and the target level for the school leavers. As was discussed earlier, the proposed exam has to be general in nature in the sense that no test items would be included which call for specialist types of language skills: e.g. reading engineering type of texts or writing business letters. The reason is that the school leavers from all over the country have to take the same exam no matter which faculty they intend to go to. So the exam has to fit in this situation by including the test items to assess the basic skills applicable to all faculties.

On the basis of the above discussion, we determined the language skills or abilities that the exam should include in order to serve the purpose for which we want it. The detailed specification of the exam will be presented in section 4.2.2..

We have accepted that language proficiency is a construct composed of linguistic competence, and language skills. The implication is that the exam should include both types of items: skill-based, and linguistic competence-based. This section presents the basic skills which should be included in the exam.

3.4.1. Reading

On the basis of the survey of the literature, we came up with the conclusion that reading comprehension depends on reading strategies and linguistic competence. Considering the level of students, we decided to include only basic reading skills. Munby (1979) recommends that the reading skills - reading for information, reading for implied meaning, reading for gist, and reading for required information - should be emphasised when teaching English as a Second/Foreign Language at school level. We found this recommendation applicable to the Nepalese situation as well. On top of these skills, we decided to include also the items that require the examinees to make guesses of the meanings of unfamiliar words. We believe that this skill is also very important for our students because they have to do quite a lot of reading at tertiary level, and they may encounter several unfamiliar words while reading their course materials. Inferring meanings of unfamiliar words from their context is an effective way of improving comprehension (Nuttall 1982:70). If they get this type of training at school, it will certainly speed up their reading.

3.4.2. Linguistic Competence

On the basis of the survey of the literature, we concluded that the reading and writing abilities of the students depends on their linguistic competence, and the strategies they adopt while performing the task. Having accepted that language proficiency is

a construct which is composed of language skills and linguistic competence, we felt it necessary to test linguistic competence separately.

There have been different views about whether formal knowledge of grammar should be taught and tested separately, or whether it should be included within language skills. It has even been argued that grammar need not be taught separately because it is the language skills that are of main interest. In other words, grammar can be included within language skills in the sense that it can be achieved while concentrating on the skills. For example, Beretta and Davies (1985:126), in their Bangalore evaluation project, conclude that "grammar construction can take place through a focus on meaning alone".

However, there have been arguments that grammar is the heart of language, and it has a high level of generalisability (Davies 1978:151). In most language teaching programmes, grammar is taught implicitly or explicitly. It can be argued that if it is the kind of ability to be developed in the students, it has to be taught, and also tested. From this point of view, we decided to include the test items that call for the linguistic competence of the SLC students.

Here again, we faced the problem of the content of the grammar test. Generally, the content of an achievement test is determined by the course objectives and the given syllabus. From this point of view, the content of the test of linguistic competence should be determined only in the light of the grammatical structures presented in the syllabus since nothing has been said about it in the course objectives. The problem with determining the content of the test this way was that the proposed exam was to move away from discrete item testing. We

wanted our exam to be integrative. In that sense, it was difficult to design a test that could represent all the grammatical structures presented in the given syllabus.

So far as the procedures for testing linguistic competence is concerned, Hughes (1989:143) suggests three methods: i.e. paraphrase, completion, and modified cloze. To make the exam fit the integrative model, it was only the modified cloze which could be considered. But Oller (1983:356-7) reports the studies (Taylor 1957; and Oller and Inal 1971) which showed that the standard cloze yields higher reliability and validity coefficients. Johnson (1981:199) also maintains that "the cloze item challenges the linguistic competence in very precise terms". Alderson (1978) concludes that cloze is a test of "core proficiency" (p.292), by dealing with ability to handle syntax and lexis at sentence level (p.395).

Taking the above views into consideration, we decided on a cloze test to assess the linguistic ability of the students.

3.4.3.Note-taking

Note-taking is another skill which is important for tertiary students. Weir (1983a) shows that note-taking skill is important from an educational point of view as well. Taking notes from text helps students understand the passage much better, remember the points for a long time, identify main points, and organise them in order. In other words, this is a kind of skill which can improve reading skills as well as writing (at least at pre-writing stage).

The study carried out by Evanechko et al (1974) reveals a very strong relationship between reading and writing behaviors of foreign language learners. Note-taking, for example, can be an activity to coordinate both skills.

When a SLC passer goes on to tertiary level, it is important that he can take notes from the course materials for either exam purposes or for understanding the passage. That is why we decided to include a test for this skill so that schools would give the SLC passers a basic training in note-taking.

From the testing point of view, assessing the note-taking skill is difficult in the sense that it is the most idiosyncratic skill. It has been claimed that some students do not need to take notes, but even if they do, it varies from person to person. Another problem is that notes are meant for the one who has taken them, not for others, so how can they be evaluated? S/he does not have to communicate through notes. To test the note-taking skill, we, therefore, decided to give students a reading passage followed by a format for taking notes from the passage so that everyone would have to follow the same pattern. The assumption was that training in a study skill like note-taking would not have any harmful effect on learning, but rather the opposite.

3.4.4. Writing

The survey of the literature on writing indicates that writing is not merely the transfer of ideas from mind to paper. It is more a process of refining them. To train students to write, therefore, they should be asked to perform writing by creating a situation in which they find something to say which is familiar to them. In other words, they should be given opportunities to express what they think about an issue related to them or how they feel about an event they have experienced. It can be argued that their expression based on their own experience would give a valid index of their writing proficiency.

To assess their ability to express ideas and feelings, therefore, we decided to include items which require them to

express their own feelings and opinions about events they have experienced, and the topics they are familiar with.

For that purpose, we included letter writing and essay writing tasks in the exam.

3.5. Formats of the New Exam

The exam was intended to assess reading, writing, linguistic competence, and study skills, so we had to use tests of different varieties of format. The details of the format of the new exam will be presented in section 4.2.3.

To present the format briefly, the reading skills were assessed by using two reading passages which were followed by multiple-choice items and supplying a word or a phrase to given gap items. The idea was that the expression of the reading comprehension of the students should not be restrained by difficulty in writing, so that they were not required to write a complete sentence to answer any items related to reading comprehension.

A cloze test with a six deletion rate was used for the assessment of linguistic competence.

A passage followed by a format of note-taking was used for assessing note-taking skill.

To assess writing, two tasks were used. One was letter writing in which the students were asked to reply to a given letter. The other was essay writing, in which the students were asked to write an essay by using given points.

As the exam was intended to replace the existing SLC English exam, the time allowed to complete the exam was three hours in order to make it fit in the SLC framework.

3.6. Summary and Comments

A survey of the literature on what to test, and how to test a foreign language has been presented. The review reveals that there have been different views about what to test, and how to test a foreign language. The concept of 'the what' and 'the how' has changed as a result of a change in the understanding of how a foreign language is best learnt, and how it should be taught. On the basis of the survey, we have decided to adopt the integrative approach to language testing. And we have also decided to take into account the communicative features of a language task in order to make the exam have a positive washback effect.

We have examined what level of language proficiency can be expected from a school leaver, and what is necessary for a college entrant in the light of the Nepalese ELT situation. We have concluded that a school leaver needs all the language skills - reading, writing, speaking, and listening. However, taking into account the practical problem in testing listening and speaking in the SLC exam, we have concluded that the new exam should concentrate only on the assessment of reading and writing.

The nature of reading and writing have been reviewed. We have come up with the conclusion that the teaching and testing of reading and writing can be improved if we accept them as a combination of subskills. At the same time, we have also said that linguistic ability is essential for exploiting the reading and the writing subskills. On that ground, we have decided to include in the test items which call for both abilities: i.e. linguistic, and strategic abilities.

Finally, we have briefly presented the specifications, and the formats of the new exam.

CHAPTER 4: DESIGN OF THE EMPIRICAL INVESTIGATIONS

4.0. Introduction

Chapter four presents the design of the empirical investigation of the study, and the procedures followed in data collection. The chapter has three main sections: predictions of the study, procedures followed in collecting data, and a description of the field study.

First we present the queries the study was designed to answer. Three main issues will be considered: the adequacy of the SLC English, the washback effect of the SLC English exam, and the effect of the different types of teaching of English in the Nepalese schools. The first issue will be addressed by a quantitative analysis of the performance of SLC students, and the rest will be addressed by hypothesis testing.

Then we consider the methods used for data collection. We present the detailed specifications, and the formats of the exam. This section also includes the description of the questionnaires used for collecting data. We also present a marking scheme developed especially for the present study which would serve also after the implementation of the new exam.

Finally, we present the description of the field study carried out for data collection. This section includes the discussion of how the two sets of tests were piloted, and the procedures adopted in the administration.

4.1. Formulation of Predictions

4.1.1. Background to Predictions

To provide a background to the formulation of predictions, we sum up the main points of the preceding 3 chapters.

In Chapter 1, we examined the nature of an exam which is associated with some extrinsic values. It is clear that exams

of that sort have a powerful influence on teaching and learning. On the basis of the discussion, we concluded that in an educational setting, an exam has the potential to take a dynamic lead in the process of improving teaching programmes.

In Chapter 2, we presented the Nepalese ELT situation. Examining of the ELT situation in terms of its course materials, methods, and assessment system against the purposes for which English is taught at school in Nepal, we found that the programme was unsatisfactory. We argued that there was an urgency of bringing about changes in the existing ELT situation. We considered the factors responsible for making an exam good, and examined the SLC English exam within that framework. It was found that the SLC English was likely to have a negative washback effect on the teaching and learning of English at school. Accepting the powerful role that an exam can play in an educational programme (Chapter 1), we argued that no change could be effective in the direction of changing the teaching of English in Nepal without changing the SLC English exam. We, therefore, proposed to bring about a change in the Nepalese ELT situation through a change in the School Leaving English exam.

In Chapter 3, we discussed what skills and abilities a language exam should call for, and how the examinees should be assessed. The idea was to supply a basis for the content validation of the proposed exam so that the exam would have a positive washback effect. It was also hoped that the a priori validation of the exam would make it provide a valid index of language proficiency of the examinees. The purpose behind this was to determine the adequacy of the existing ELT in terms of the performance of school leavers on the new exam.

4.1.2. The Main Research Questions and Procedures to Address them

Based on the above discussion, we designed a research project to answer the main research questions:

To what extent is the present situation of English Language Teaching at school in Nepal adequate to equip students with a working knowledge of English for the purpose for which they study English at school level? And how can the influencing factors be accounted for?

The intentions for the study of the questions were to address the following issues:

1. To investigate the adequacy of the SLC English for the purpose for which English is taught in schools so that explanations could be given of why the students who complete the SLC English representing the different types of schools (i.e. Types - A, B, C, D) still find it difficult to cope with the tertiary courses which require English.

Procedures: This issue will be addressed by a quantitative analysis of the performance of SLC students on the SLC English exam, and on the new exam. It was predicted that the SLC students would find the new exam more difficult than the SLC English exam because the way SLC English was taught and tested did not equip the students with the knowledge of how to put language in use. It was hoped that a quantitative analysis of the performance of the students would provide convincing evidence for whether the existing SLC English is adequate for the purpose for which English is taught in the schools.

Another purpose of the study was to design an exam and validate it so that specifications and format for the exams which would be used for replacing the SLC English exam could be determined.

2. To investigate the effect of the different types of teaching of English which take place in different types of schools on the learning of students so that the teaching of English in type D schools (which represent more than 85% of the total schools) can be accounted for.

Procedures: This issue will be addressed by comparing the performance of SLC students from Types - A, B, C, and D schools, on the new exam.

3. To investigate the effect of preparations for the SLC English exam on the teaching and learning of SLC English. As has been pointed out, in the final year of school the whole focus of teaching is on how to get good results in the SLC exam. Thus, the washback effect of the SLC English exam can be accounted for, and by the same token, a possibility of using exams (which is likely to have a positive washback effect) for educational change can be explored.

Procedures: This issue will be addressed by comparing the performance of the students at the beginning of the final year of school (i.e. grade 10), and the students at the exit level of school studies.

4.1.3. Formulation of Hypotheses

To answer questions raised in the above issues - 2 and 3, the following main and subsidiary hypotheses (in null form) were formulated:

Main Hypothesis One (HO)

There would be no significant difference in performance between Grade 10 and SLC students.

Rejecting this hypothesis would mean that SLC exam preparation in Grade 10 has a marked effect on learning English by SLC students. It has already been pointed out that schools

tend to concentrate on how to get good results in the SLC English exam especially with the students who are in the final year of school studies (Grade 10). If SLC students do not do better than Grade 10 students, it would mean that learning does not take place during the SLC English exam preparation. In other words, the SLC students spend the whole year doing something which does not improve their abilities in using English.

Subsidiary Hypotheses (HO)

To supply more information on the above issue, the following subsidiary hypotheses were formulated:

1. There would be no significant difference in performance between Grade 10 and SLC students from Type A schools.
2. There would be no significant difference in performance between Grade 10 and SLC students from Type B schools.
4. There would be no significant difference in performance between Grade 10 and SLC students from Type D schools.

Main Hypothesis Two (HO)

There would be no significant difference in performance between students from different types of schools (i.e. Types - A, B, C, and D).

Rejecting the hypothesis would mean that the different types of courses the different types of schools offer, and the different methods they follow lead them to produce students with different abilities in using English. We anticipated Type A schools would do better than type B schools, Type B better than Type C, and Type C better than Type D. Providing this sort of evidence would supply a basis to account for the weaknesses of the teaching and testing of the compulsory English course at school. It must be remembered that the schools which come under type D schools offer only the compulsory English.

Subsidiary Hypotheses (HO)

To provide more information on the issues, the following subsidiary hypotheses were formulated:

1. There would be no significant difference in performance between Type A and Type B schools.
2. There would be no significant difference in performance between Type B and Type C schools.
3. There would be no significant difference in performance between Type C and Type D schools.

4.2. Methods of Investigation

In order to answer the questions posed above, the following methods of data collection were used:

- i. a language proficiency test based on the specifications, and the formats presented in Chapter 3, and
- ii. two sets of questionnaires; one set to college students, and another to the people involved in the teaching and testing of SLC English.

In this section, we present a description of how the tests and questionnaires came into being, and how they were administered.

4.2.1. Construction of Two Sets of Tests

Two sets of tests were constructed to assess the reading, linguistic, note-taking, and writing proficiency of the intended students. The idea was to pilot the two sets first, and then to design the final exam on the basis of the pilot results.

Before presenting the detailed specifications of the exam used for data collection, we would like to touch on a few points of principle bearing on the design of the tests.

4.2.2. General Principles of Test Constuction

In the preceding chapters, we have already discussed

several features of test construction, including performance-based, integrative, direct, and communicative testing. We just mention some other features here which were also taken into consideration while designing the tests.

4.2.2.1. Contextualisation

We take the view that each item of the exam should be set in a clearly defined context. Language in real-life occurs in a variety of forms dependent upon the context of situation and social environment, whereas in the exam-situation, language takes place in an artificial situation. It is, therefore, necessary to set test-items in such a way that they are presented within a clear context, such that testees will not interpret an exam item in different ways. Attempts were made, therefore, to contextualize as far as possible the test-tasks so that all the testees would interpret them in a similar way. For us, contextualization includes explicitness of test instructions as well, so that test instructions should allow for no misinterpretations. The implication is that, if necessary, test instructions in some test tasks can be given even in L1 (i.e. Nepali).

4.2.2.2. Authentic-like Materials

It is argued that tasks to be included in a test should reflect the language in the real world as far as possible. It is further said that the language in the tasks should not be of the stimulus - response kind (Davies 1977:40), but be natural, and consist of features of day to day discourse. However, it cannot be ignored that an exam is in itself a simulated environment. Every attendee knows that it is an artificial situation. Anyway, there is a sense in which testees and testers know that the situation is not the real one. What this means is that the sense

of authentic-like materials is relative; it is a matter of fitting authentic materials into possible practice. We, therefore, attempted to use authentic materials in our tests as far as possible (e.g. the tasks which look authentic), but they may not have the language of the real world.

4.2.2.3. Unseen Tasks

Items in the test must be selected from a wide range of unseen tasks with a considerable difficulty level. It has already been noted that the defects of the existing SLC exam were its course-book-based and predictable nature. The situation cannot be improved unless the classroom is changed from an exam-rehearsing centre into a language learning place. For that reason, we decided that the tasks in the exam should be unseen, and new, (not practised by the student as prescribed texts and exercises) so that teachers and students would be discouraged from preparing ready-made answers for the exam. The students should, therefore, be made ready to face any tasks suitable to their level. Preparations for this type of examination would lead the examinees to learn the use of the language the exam requires.

4.2.2.4. Criteria for Text Selection

Following the discussion presented above, we tried to meet the following criteria in text selection for reading, cloze and note-taking tasks.

i. Interest

Considerable attention was paid to finding texts which were interesting (i.e. if the reader enjoys reading a passage) in themselves and texts which children would like to read.

Jordan(1978) finds those texts most popular at school in Nepal which deal with adventure, neighbouring countries, and animals. His study was based upon school children from different

parts of the world as well as Nepal.

Anderson (1976) concludes that interest is closely related to comprehension and hence facilitates reading. Furst (1959:220) also suggests that texts for an exam should be chosen in such a way that students should not be able to answer without reading the passage. But at the same time, it is important that the text should be novel, interesting, and within their grasp.

ii. Familiarity in Terms of Concept

Attempts were made to avoid conceptual problems for understanding reading passages by selecting test-tasks which would be familiar to the testees in terms of content. The explanation is that the topic of a text should be within the grasp of intended readers. That is why those texts were selected which dealt with more or less the same topics as the texts the course book prescribed for the intended students. For example, let us take the topics "Talking to a Tiger" and "The Elephant and the Sky" of the texts selected for the reading tasks. The intended testees should be familiar with these topics because they contained a content similar to the topics "The disappearing tiger" and "RNAC", which appeared in their course book. It is the organisation of the passage, i.e, the use of different types of sentences and vocabulary, which will be different from the SLC materials.

iii. Establishing Difficulty Level

It was difficult to select the texts which exactly meet the difficulty level of the intended students while designing the tests in Edinburgh. To sort out this problem, texts were selected from those source materials which were being used for the same level of students in different countries.

Most texts were taken from the following books.

iii.a. Howe, D.H. (1977) *New Guided English*. OUP.

This book was prescribed for the students who learn English as L2 at Secondary Level in Malaysia.

iii.b. Green, G. (1972) *Read, Think, and Answer*. OUP.

This was a book intended for secondary school students.

iii.c. JEMC (1986) *English Reader, Part II*.

A book for SLC students in Nepal.

iv. Length of the Passages Selected

The question is: how long should a passage to be selected for assessing reading ability be? Actually, there is no established system. However, Heaton (1975:113) and Lee (1981:138) suggest that a text containing 200 to 300 words is appropriate for intermediate level.

Considering the suggestion, those texts were selected which presented a complete picture of an event or a topic using roughly 200 to 300 words.

4.2.2.5. Criterion-Norm Referenced Testing

In the testing literature, one of the major subjects of discussion, though testing experts have not yet arrived at an agreement, is the issue of criterion-referenced tests (CRT), and norm-referenced tests (NRT). It would be relevant to touch on some points regarding these two types of tests in order to make explicit how the results in the proposed exam would be interpreted.

A norm-referenced test has been described as an assessment entailing comparison of one pupil's attainment with that of others, i.e. comparison with some 'normative' group, and establishing his performance on a scale. That means, a pupil's performance is measured against the performance of others who sit for the same exam. A criterion-referenced test, on the other

hand, is described as an assessment that compares a pupil's attainment with a pre-specified criterion. It identifies what the pupil knows or has attained.

Traditionally, a NRT is associated with discrete point, analytical, structural tests, and a CRT with integrative, global, communicative tests. But the case is not necessarily so.

There has been a growing controversy between NRTs, and CRTs for several years, and testing experts still seem to disagree. For example, Davies (1978:141) maintains:

a criterion-referenced test is a use of a norm-referenced test; but the argument is often presented as though criterion-referenced were in itself a method of test construction. It is not.

On the other hand, Bejar (1983) refers to Popham as suggesting that NRTs are essentially worthless. Cziko (1983:293) refers to Denham (1975) as noting that "we may also make norm-referenced interpretations of scores obtained on domain-referenced tests".

It is on the latter assumption that the test was devised. Looking at the guiding principles of exam construction, it would seem that our exam would be criterion-referenced. And it was also true that the exam was not truly based on psychometric procedures of test construction (i.e. selection of test items which maximize differences between individuals), which has been characteristic of a NRT. Despite these characteristics of the proposed exam, the results in the exam would have norm-referenced interpretations. The test designer would have liked very much to use the exam as criterion-referenced, but the Nepalese socio-political situation was such that an introduction of a criterion-referenced interpretation of exam results, like the results of the SLC, would not be accepted by the Government.

The possibility of interpreting results of an exam which

is based on integrative language tests in norm-referenced terms has been discussed in the testing literature. The principle is that integrative language tests seem to be very sensitive to inter-individual difference in language proficiency. And they are also found to correlate quite highly with discrete point tests of language proficiency (Cziko 1983).

On the basis of this, we decided to design the exam using domain-referenced principles though the results in the exam would have a norm-referenced interpretation.

4.2.3. Detailed Specifications of the Exam

Bearing in mind the theoretical and practical considerations discussed in the preceding sections, two sets of tests were constructed.

4.2.3.1. Objectives of the Exam

The objectives of the exam were to assess the following abilities of SLC students in Nepal:

- i. ability to understand written English for reading prescribed books, and course materials, because they would have to read those types of texts at their tertiary level.
- ii. ability to use their linguistic competence while reading or writing; linguistic competence is necessary for reading and writing.
- iii. ability to understand a passage and take notes from it; because the intended students would have to take notes from their course materials in English in different subject areas.
- iv. ability to write some coherent paragraphs which show that they could express their ideas, and experience in English. At tertiary level they would have to write answers inudent

English. They would be expected to be able to express themselves.

To achieve the above objectives, we made the following decisions:

1. Objective i

Students should be asked to demonstrate their ability to understand written English. For that purpose, they should be given some passages written in English and asked to answer the questions which follow. The questions should reflect the following reading sub-skills and study skills.

Reading Sub-skills

1. Skimming--surveying the text for the gist.
2. Scanning--locating specific information in the text.
3. Guessing meaning from context.
4. Inference--understanding meaning of a passage which is not explicitly stated.

Study skills

5. Separating essentials from non-essentials.
6. Organizing--rearranging the given passage.
7. Summary--picking up the main ideas of the passage.

2. Objective ii

Candidates should be asked to demonstrate their linguistic ability. For this purpose, a standard cloze procedure should be used.

3. Objective iii

Candidates should be asked to demonstrate their ability to take notes from a given passage. For that purpose, they should be given a passage written in English. The passage should be followed by a form in which the students would have to fill in extracting the main points from the passage so that every student

would follow the same pattern.

4. Objective iv

Candidates should be asked to demonstrate their ability to write some coherent paragraphs in English. They should be given tasks which provide them with an opportunity to express their ideas and feelings. For that purpose, they should be asked to write a letter, and an essay. In both tasks, as the purpose was to measure writing, students should be provided with clear guidelines in order to avoid problems in terms of content.

4.2.3.2. Contents and Formats of the Exam

Each set of the test had two parts: Part one and Part two. Part one contained 3 tasks.

Task 1.

Skills involved - skimming and scanning.

The students would have to read a passage and answer in 25 minutes the questions which follow. The answer sheet would be collected when the time was over. That means, they would not be allowed to keep the paper for more than the given time. The main reason for doing this was that since Task 1 was meant for assessing whether the testees could skim and scan, they would spend more time than the time they were supposed to spend if the answer sheets were not collected.

Types of test items - multiple choice items, and
supplying a word or a phrase.

Task 2.

Skills involved - guessing meaning from context,
-inference,
-separating essentials from non-essentials,
-organisation and summary.

They would be given a passage which contained some unfamiliar (non-existent in English) words, and questions reflecting the skills mentioned above. The purpose of using unfamiliar (non-existent) words was to make sure that everyone would have to guess the meaning of the given words so that the test would not be a test of vocabulary. They would have to guess the meaning of those words from the context first, then answer the questions which followed. Here, since the purpose of the test was to assess whether they could understand the passage, they were free to supply the guessed meanings even in Nepali if they found it difficult to express the guessed meanings in English.

The task would be followed by a question containing 9 statements in which they would have to separate essentials from non-essentials, organise the selected statements in such a way that they would constitute a summary of the passage. The time allowed for this task was 35 minutes.

Types of test items--unfamiliar words,
 --multiple choice items and
 --statements based on the content of
 the passage.

Task 3

Traits involved-- linguistic ability.

A reading passage containing 34 blanks (every seventh word deleted) would be given. The students would have to fill in the blanks with appropriate words.

Part two.

Part two contained the following: note-taking and writing tasks.

Task 1-Note taking

Skill involved - Note-taking

The testees would have to read a passage, and extract the important points to fill in the form given. The format was given to help students write a similar type of notes.

Task 2-Letter writing

Skill involved - composition

They would be given an informal letter and asked to reply to it, or they would have to read job advertisements and apply for a job they liked. The decision about which task would be used in the exam was based on the pilot study. To help them write the application, guidelines would be provided. The time allowed for the task was 25 minutes.

Task 3-Essay Writing

Skill involved - composition

The testees would have to write an essay in 250 words in 25 minutes. To help them write the essay, guidelines would be given. When the time was over, they would be asked to stop writing. Then they would be given a green pen and advised to supply corrections onto what they had written. They could amend their essay in terms of its grammar and content. The purpose of giving them a green pen each was to make their amendments stand distinctive from the original writing. Extra credit would be given to the corrections. The time allowed for correcting was 10 minutes. Invigilators would be instructed not to allow the testees to use their own pens while amending their writing.

4.2.4. Design of Questionnaires

In addition to the exam, we decided to use two sets of questionnaires for the purpose of collecting some additional information related to the teaching of SLC English: one set was for people involved in teaching and testing of the SLC English, and another was for college students.

4.2.4.1. Questionnaire for People Involved in SLC English

We felt it necessary to gather information about how the people involved in the teaching and testing of the SLC English react to the SLC English exam so that some insights could be obtained for the improvement of the proposed exam. People such as the teachers of English and CTSDC experts play a vital role in the implementation of the new exam. It was hoped that this information would be helpful in understanding their responsiveness to change, and the nature of the change they were interested in. It was also hoped that teachers' comments on the effect of the SLC English exam on their teaching would provide a basis for explaining the difference of the performance of the students on the new exam and the SLC English exam.

The questionnaire had two parts: the first part was intended for all kinds of people involved in the teaching and testing of SLC English, and the second part was intended for the teachers of English at SLC level (see appendix II). The first part contained six questions related to the coverage of the course objectives by the SLC English exam, and the congruence of the SLC English exam with the needs of the students for the tertiary courses that require a high level of English proficiency. It also contained the questions related to whether they found the SLC English exam a language test which measures true language abilities of the students, and whether chief examiners' reports ever existed.

The second part of the questionnaire contained a main question followed by four subsidiary questions related to the effect of the SLC English exam on the teaching of English to SLC students.

4.2.4.2. Questionnaire for College Students

Higher educational institutes are the main consumers of SLC passers. Of these institutes, Institute of Agriculture, Institute of Engineering, Institute of Medicine, Institute of Science, and Institute of Forestry offer the courses which require a high level of proficiency in English. It does not mean that other institutes do not require their entrants to be proficient in English. The difference between the aforementioned institutes and the rest is that these institutes teach English courses as well as most content courses in English, but other institutes, namely, the Institute of Education, the Institute of Humanities, and Social Sciences, the Institute of Law, the Institute of Management teach English in English and content courses in Nepali. It was hoped that collecting information about the adequacy of SLC English for the tertiary courses which require a high level of English proficiency would be helpful in determining the content of the exam, hence the content of the course, which would be used to replace the existing SLC English exam. Furthermore, the exam would have a positive washback effect as well.

For that purpose, we decided to use a questionnaire (given in appendix III) to be administered to those students who had had experience of the adequacy of their school English in the light of the demand the tertiary courses placed on them. We believed that the second year students who had received the results of their first year exam from the institutes which require English quite a lot would be potential candidates for the purpose. It was hoped that they could provide information about whether their school English was adequate for the courses they did, and what sort of language activities they carried out when doing those

courses. In the Nepalese educational setting, Science and Maths are taught as core subjects at school. In Nepali medium schools, these subjects are taught in Nepali, but, at tertiary level, these subjects are taught only in English. So we felt it important to ask the students who were from Nepali medium schools for their comments on the problem.

4.2.4.3. Questionnaire on Needs Analysis

A detailed survey of the needs for English for tertiary students was not possible before the exam was designed. For the purpose of the present study, therefore, the needs were determined in terms of the available sources as in Davies et al's (1984) survey and the objectives of the SLC English course. To avoid entire dependence on armchair speculation for further improvement of the exam specification and the marking scheme used in the present study, we decided to carry out a survey of the needs in English for tertiary students by asking tertiary tutors of English about their opinion of what reading and writing subskills they believe their students need in order to cope with the tertiary subjects which require English. The purpose was to compare the list of language skills popular among the tertiary teachers against the exam specifications. It should be noted that the questionnaire presented in section 4.3.6.2. (i.e. Questionnaire for college students) also yielded some useful information related to the needs for English for tertiary students.

We believed that the college teachers' selection of the types of language activities that their students need to carry out would provide a basis for making decisions about the exam

specification. One of the problems we had, however, was to reconcile the assessment of general ability in language (i.e. core abilitites) and the specialist language activities that the students in different faculties require. For practical reasons, it is not possible for the SLC English exam to include all the different types of reading and writing activities demanded by different faculties: e.g. writing a prescription in the faculty of Medicine, reading an engineering text. From this point of view, it seems that a needs analysis has nothing to contribute to improving the specification of the exam in the sense that it would not be possible to include different types of reading and writing activities which would be reflected by the survey. However, we decided to carry out this survey because it would provide information about the most important reading and writing activities that the tertiary students from different faculties need to perform. It would also provide information about which activities were more important than others, and which activities were common to most faculties. We hoped that collection of this type of information would help enable the exam have predictive utility and, from the washback point of view, content validity after its eventual implementation. It would also validate the marking scheme.

The questionnaire (see appendix IV) has three sections: reading, writing activities, and traits to be emphasised while teaching and testing writing.

The reading section was adapted from Weir (1983b). This section has the following 11 reading subskills (given in Table 3) out of which tertiary tutors were asked to indicate 5 subskills

which they felt the most important for their students. They were also asked to rank them in order of importance.

Table 3: Reading skills included in the questionnaire

-
1. Reference skills, e.g. recognising titles, headings, index
 2. Guessing meaning of unfamiliar words from context
 3. Skimming: i. surveying to obtain gist of a text,
ii scanning the text to locate specifically
required information on a single point
 4. Understanding relations within the sentence
 5. Being able to understand parts of a text which are
connected
 6. Understanding relations between parts of text by
recognising indicators in discourse, e.g. clues,
linking signals, indicators used to introduce or sum up
ideas
 7. Understanding conceptual meaning, e.g. quantity and amount,
direction, condition
 8. Understanding explicitly stated ideas and information
 9. Understanding ideas and information in a text not
explicitly stated
 10. Separating the essential from the non-essential in a text
 11. Interpreting information presented in a non-linguistic
form, e.g. tables, graphs, diagrams
-

Writing skills were considered under two headings: writing activities and traits to be emphasised while teaching and testing writing. In the specification of the exam, we have not been very

explicit about the writing activities we were trying to select. Rather we decided to concentrate on the components of composition like content, organisation, grammatical correctness etc. We found it difficult to specify clearly a set of writing activities to provide a basis for exam construction because the students who pass the SLC exam would go on to different faculties which would require a wide range of writing activities like writing a prescription (Faculty of Medicine), writing a business letter (Faculty of Management). In this situation, we decided to choose to develop an exam based on a theory of language proficiency (e.g. ability to express oneself in writing) and mark students' writing in the light of widely accepted criteria for measuring a composition (e.g. content, organisation, language) for the washback purpose. The purpose behind the survey was to find out the writing activities which were common to most faculties so that, at least, these activities, if not all, could be included in the exam.

To find out the most popular writing activities, we asked the tertiary teachers to indicate, of the following 11 (given in Table 4.) activities, 5 which they think most important for their students. They were also asked to rank them in order of importance.

Table 4: Writing activities included in the questionnaire

-
1. Expressing ideas, opinions
 2. Writing personal essays
 3. Writing/replying to letters
 4. Summarising a passage
 5. Making comments
 6. Describing events, objects, graphs, pictures, processes
 7. Presenting arguments
 8. Reporting events
 9. Persuading
 10. Making complaints
 11. Writing assumptions, speculations, clarifications
-

In our marking scheme, (section 4.2.7.1.) different weightings have been allocated to different components of composition. The reason for this was to lead the tertiary classrooms to concentrate on these components. For that purpose, we decided to ask the tertiary tutors to give their opinion on the importance of the traits to be emphasised while teaching and testing writing. In the marking scheme, we have allocated, out of 20 marks, 4 to content, 4 to organisation, 3 to language, 3 to vocabulary, and 2 to mechanics for assessing an essay. It would be worth investigating whether the tertiary tutors agree with us in ranking these traits in the same way they were ranked in the marking scheme. We have already mentioned that the marking scheme was based on the existing literature in general, and Jacobs et al's (1981) ESL profile in particular. To validate the marking

scheme, following Jacobs et al (1981), we included the following traits (see Table 5) in the questionnaire, and asked the tertiary teachers to rank them in order of importance.

Table 5: Traits to be emphasised in teaching and testing writing

-
1. CONTENT: e.g. knowledge of subject matter, thorough development of argument, relevant to the topic in question
 2. ORGANISATION: e.g. logical sequence, fluency, supporting details, completeness
 3. VOCABULARY: e.g. effective word/idiom, appropriate register
 4. LANGUAGE: e.g. grammatical correctness, effective construction,
 5. MECHANICS: e.g. spelling, punctuation, convention
-

4.2.5. Scoring Procedures

4.2.5.0. Introduction

In this section, we discuss scoring procedures to be followed in marking the students' answer-sheets. For that purpose, we designed a marking system to be used for evaluating the intended subjects' written work and the eventual implementation of the proposed exam in Nepal. As mentioned earlier, for the purpose of the present study, two writing tasks were devised; i.e. those designated Writing Task 2 and Writing Task 3, in the exam.

Other matters to be discussed, though very briefly, are: what composition is (for the purpose of the present study, at least), and different approaches to marking compositions. We also present the design, and justifications of the marking system developed for the purpose of the present study. Designing any marking system is, to a large extent, based on the purpose and nature of the test. Therefore considerable attention will be paid to designing a marking scheme which suits the purpose of the test, the nature of the tasks, and is applicable to the Nepalese situation.

4.2.5.1. Scoring Reading, Cloze, and Note-taking tasks

The reading, cloze, and note-taking tasks in the exam are to be scored objectively.

The reading tasks are followed by either multiple choice items, or supplying one word or a phrase items. Students should not be penalised for spelling mistakes provided that the intended meaning is clear. 23% of the total marks is allocated to the task.

For cloze, 'any acceptable alternative' scoring method will be used. Following Alderson (1978:291) 'any acceptable word' scoring method is the best method of predicting overall proficiency of English as a Foreign Language. 17% of the total marks is allocated to the task.

The note-taking task is to be scored objectively. Students are not required to supply complete sentences to complete the task. However, they should not be penalised for supplying a complete sentence unless they repeat a same sentence from the text. 20% of the total marks is allocated to the task.

4.2.5.2. Marking Compositions

Writing Task 2 and Writing Task 3 are to be marked subjectively. To mark their writing we designed a marking system. It was felt necessary to design a marking system for use in the present study as well as after the implementation of the exam. The design of the marking scheme constitutes the subject matter of the following sections.

4.2.5.3. What is a Composition?

In section 3.3.2., we reviewed some literature on writing. In the following discussion, we would like to use the word, 'composition', following the 'Longman Dictionary of Applied Linguistics, by Richards et al (1985), which defines it as "(in language teaching) writing practice which deals with texts longer than a single sentence, such as paragraphs, essays and reports". For the purpose of the present study, writing a letter (i.e. Writing Task 2) also comes under the same definition.

In a broad sense, 'composition' will be interpreted as a way of expressing ideas or conveying meaning. It also entails thinking. It seems as if they are inseparable. Berthoff (1979:252) views it as:

the mechanism of composing consists in the ability to combine what follows with what precedes, all the while keeping in mind what is already written ...
... in thinking and writing at the same time without having one of these acts interfere with the other

Steel and Talman (1936) are also of the opinion that composing ability is something which deals with the ability to express oneself coherently, lucidly and economically.

4.2.5.4. What We Mean When We Say 'Scoring a Composition'

It is important to say a few words about what we are

proposing to mark before we try to discuss how to mark a composition. When we say 'marking a composition', we generally mean 'evaluating a piece of writing' in terms of the features (e.g. values) it contains. In other words, while marking, we take into account the different components of a composition such as content, and the way it is presented and so on, and then we measure the quality of the composition by seeing how well it fulfils the purpose of the writing task. Then we gauge those prominent components or elements which play an important role in leading a marker to determine what a score to be assigned to a piece of writing. Though it is hard to separate the elements which make up a composition since they are interwoven, they can be loosely separated in the way that a reader focusses on them if these elements are already specified. Then it can be argued that scoring a composition is quantifying a piece of writing in terms of the presence or absence of certain aspects of it, whether specified implicitly or explicitly.

The decision about which elements in a composition are to be focused on, and to what degree, is, to some extent, based on the purpose of the test and the nature of the test tasks.

4.2.5.5. Approaches to Marking Compositions

Approaches to marking compositions has traditionally been discussed under two main headings; e.g. atomistic and holistic (Cooper 1977; Ali 1988; Lloyd-Jones 1977). Atomistic scoring methods, sometimes called 'indirect', or 'frequency-count', as Lloyd-Jones (1977:33) says, "rely on the assessment of particular features associated with skills in discoursing" whereas,

holistic methods "consider samples of discourse" (ibid). Cooper (1977:4) considers the holistic method an evaluation procedure "which stops short of enumerating linguistic, rhetorical, or informational features of a piece of writing". The atomistic approach covers marking methods such as T-unit analysis, cohesion analysis, error-count, etc, and the holistic approach covers essay scale, analytic method, general impression method, primary traits, dichotomous scale, feature analysis, etc. However, Brown (1984) and Hamp-Lyons (1987) prefer to discuss the primary trait method as a separate approach to evaluating a composition.

The approach to be followed in marking any composition is generally determined on the basis of the purposes and the nature of the tasks being used. Looking at the nature of the tasks in our tests (i.e. Writing Task 2 and Writing Task 3, and the purpose of the test) some assumptions can be made such as: judgements should be based on discourse rather than sentence level, the emphasis should be given to conveying meaning, etc. Considering the purpose and the nature of the tasks, marking methods which come under the atomistic heading will not be discussed here because, as Jacobs et al (1981:29) argue, "a composition evaluated by a frequency-count method has been judged not for its communicative effect, but for its number or kinds of elements". They will not be used in the present study. The explanation is that, in the present study, the ability of learners to use language would not be assessed by separating them bit by bit as is done in atomistic methods. A composition has to be evaluated, as the present study intends, in terms of the writer's ability to communicate his/her meaning.

In the following sections, holistic marking methods such as general impression, and analytic method will be discussed, because, it will be argued later in section 4.2.7., the combination of the two will contribute to the formation of a marking system appropriate both for the present study and the eventual implementation of the proposed exam in Nepal. These methods are considered simple and reliable in assessing the writing ability of learners in terms of their ability to communicate. It is assumed that the marking system developed as a combination of these methods (i.e. analytic and general impression) will be suited to the purpose and nature of the test-tasks. Also, from the point of view of practicality, the marking system which will be designed by taking ideas from both methods will fit in the Nepalese situation.

4.2.5.6. General Impression Method

The general impression method is said to be the simplest method of marking composition. Heaton (1975:136) maintains the view that in using this method, a score is assigned to a composition on the basis of "total impression of the composition as a whole". This method does not require a detailed description of the features. This also does not require the marker to award separate scores for separate features which a rater might take into account while evaluating a piece of composition. What a rater needs to do first is to follow the 'rubric' and to read the paper to reach a decision intuitively about the quality of the composition and then award a score on the basis of the compositions available. The rubric will be about the relevance of the answer to the question asked and the content of the answer.

As Cooper (1977:12) says;

the rater scores the paper by deciding where the paper fits within the range of papers produced for that assignment or occasion.

This is, basically, a norm-referenced procedure of marking a composition, though a rater might use some sort of criteria intuitively. The advantage of using this method of scoring is that it is simple and quick.

4.2.5.7. Analytic Method

The analytic method of marking is a process of assigning a score to a composition by awarding separate scores for prominent features or characteristics of a composition. The prominent features are already listed and described in some detail. These features are grouped according to Diederich's model, under 'General merit' and 'Mechanics', and under different levels - Low, Mid and High.

Cooper (1977:7) presents the following sample which was originally developed by Diederich (1974);

	Low		Middle		High
<u>General Merit</u>					
Ideas	2	4	6	8	10
Organisation	2	4	6	8	10
Wording	1	2	3	4	5
Flavor	1	2	3	4	5---
<u>Mechanics</u>					
Usage	1	2	3	4	5
Punctuation	1	2	3	4	5
Spelling	1	2	3	4	5
Handwriting	1	2	3	4	5 ---
					Total ----

The description of the levels - high-mid-low is as follows:

High. The student has given some thought to the topic and writes what he really thinks. He discusses each main point long enough to show clearly what he means. He supports each main point with arguments, examples, or details; he gives the reader some reason for believing it. His points are clearly related to the topic and to the main idea or impression he is trying to convey. No necessary points are overlooked and there is no padding.

Middle. The paper gives the impression that the student does not really believe what he is writing or does not fully understand what it means. He tries to guess what the teacher wants, and writes what he thinks will get by. He does not explain his points very clearly or make them come alive to the reader. He writes what he thinks will sound good, not what he believes or knows.

Low. It is either hard to tell what points the student is trying to make or else they are so silly that, if he had only stopped to think, he would have realised that they made no sense. He is only trying to get something down on paper. He does not explain his points; he only asserts them and then goes on to something else, or he repeats them in slightly different words. He does not bother to check his facts, and much of what he writes is obviously untrue. No one believes this sort of writing - not even the student who wrote it.

What a rater needs to do, while rating a composition using this method, is to read the paper first, then decide the quality of the paper looking at it in terms of established criteria. He assigns separate scores to separate features. He reads a composition several times to decide to what degree the specified features are present in it. Even using the analytic method, there is a great reliance on an impressionistic way of assigning a score to a composition, because, agreeing with Cooper (1977), it can be argued that a rater is not required to count the surface features one by one to assign a score to a piece of writing. In our experience, a rater assigns separate scores to different features impressionistically, anyway.

On the basis of the description of the two scoring methods, questions may be raised such as: What is the difference between the two methods? Does a rater, while using a general impression method, assign a score to a piece of writing wholly on

the basis of impression? Does he also use some sort of criteria implicitly? Or is there any difference between the two methods in terms of the psychological processing that takes place in a rater when he actually decides what score should be assigned to a piece of writing?

We do not propose to answer these questions here. However, we would like to raise some relevant issues.

Looking at these two methods, it would appear that they share two characteristics: subjectivity in judgement, and use of criteria either implicitly or explicitly. What makes them look different is the degree of subjectivity and explicitness of criteria involved.

It has been argued that with the analytic method a score is assigned to a composition on the basis of the presence of the prespecified criteria. Here the issues are: Is the rater totally uninfluenced by his first overall impression when reaching a decision? Is not his decision to assign a score for a composition influenced by the other compositions he scores on the same occasion? It is very likely that the answer is 'no' for the former and 'yes' for the latter question. The researcher has a strong feeling that it is very difficult, if not impossible, to remove the memory of the previously read essay and the score assigned to it while evaluating the next composition. And assigning a score to a composition even on the basis of prespecified criteria is also somehow intuitive.

On the other hand, while using the general impression method, a rater assigns a score for a composition following the rubric to see the relevance of the answer to the question. It is

not convincing to say that while rating a composition that way, he does not use criteria at all, even if implicitly, as Jacobs et al (1981:29) conclude;

We might assume that impressionistic readers are using an implicit, rather than explicit, list of features or qualities to guide their judgement.

The research studies (Diederich et al 1961; Diederich 1974; Freedman 1979; Remondino 1959) show that raters use some sort of criteria implicitly while judging a composition. It is because they use their individual criteria, that they vary in their decisions.

4.2.5.8. Reliability in Holistic Evaluation: Problems and Suggestions

It cannot be denied that there is a problem of reliability in evaluating composition (Charney 1984). Cooper (1977:18) calls it 'an incontrovertible empirical fact'. It is especially on the grounds that scoring a composition is unreliable that inclusion of test items which assess 'direct writing' has been challenged.

Two factors are generally discussed in the testing literature as being responsible for making composition evaluation unreliable: the writer variable and the reader variable.

Writers tend to vary in their performance from topic to topic and occasion to occasion. It is, therefore, (elsewhere) suggested that two types of tasks that require writers to use different modes of discourse should be used in any testing programme (if possible, on different occasions).

Readers tend to vary in focussing on different features of a composition while evaluating them. For instance, Diederich et al (1961) (cited in Gosling 1966) find very low agreement among markers. The median correlation between all markers was .31.

Diederich (1974) finds a correlation of .50 between two markers. In both of the studies, holistic evaluation was used and no list of features were provided.

On the other hand, there are some research studies which give us an impression that using holistic evaluation does not mean that an acceptable level of reliability cannot be achieved.

For instance, Hamp-Lyons (1987:274) has estimated reliability .883 using her revised 'global' assignment scale and the global method from the second version with a single rater, and she has estimated .96 using the same method with three raters. Gosling (1966:47) reports that the correlation between the marks of the chief examiner and the average of the marks of another 10 examiners on a sample of papers on three essays in an English composition exam were .98, .96, and .97. In that study, marks were assigned on a 15-point scale, ranging from a high of A+ to a low of E-, and raters were guided by a list of criteria.

The difference between the studies that show poor reliability and the studies that show an acceptable level of reliability is that in the former type of evaluation no specific criteria and guidelines were provided, whereas in the latter type, specific criteria and their descriptions were provided. Cooper (1977:16) says that a reliability coefficient of .80 is considered high enough for programme evaluation. The implication is that inter-intra-marker reliability can be improved if proper attention is paid to those factors that are responsible for causing raters to vary in assigning a score to a composition. As Cooper (1977:18) argues, improvements in reliability can be achieved to an acceptable level "when raters from a similar

backgrounds are carefully trained". He presents Stalnaker (1934) as demonstrating an increase in intermarker reliability of .30 to .75 before training to a range of .73 to .98 after training.

It would appear that reliability in holistic evaluation is something which can be achieved to an acceptable level, but not to perfection, as Coffman (1971:277) notes;

to some extent,.... differences may be reflecting only the random differences that accompanying any human responses, if so, they are inescapable.

Oller (1979:392) holds the same view, when he argues:

It has long been supposed that subjective judgement were less accurate than more objective scoring methods. however, as we have seen repeatedly, subjective judgements are indispensable in decisions concerning whether a writer has expressed well his intended meaning, and of course, in determining what that intended meaning is. (underlining added).

Based on the above discussions, it can be argued that reliability can be achieved to an acceptable level, (although not perfect i.e. .1.) even when a holistic evaluation is used. To achieve that level, what has to be done is to ensure that every rater involved in marking uses the same standard on all occasions, focuses on the same characteristics of writing using the same scale of scoring (i.e. weighting for separate features), and reacts to the piece of writing in the same way others do. This can be done with the help of guidelines and training.

4.2.6.Design of Marking Scheme

4.2.6.0.Introduction

Devising a marking system was felt necessary for two reasons: first that a marking system suitable for the study will have to be used while marking all the scripts of the subjects, and, secondly, that since the purpose of the study aims to replace the existing SLC examination with an exam

similar to the newly devised test, the proposed marking system could be used after the proposed exam actually replaces the present SLC exam.

Keeping these views in mind, a marking system was designed in the light of the following criteria:

- i theoretical considerations,
- ii. the purpose and the nature of the tasks set in the test, and
- iii. practicality.

Below, we consider them briefly.

4.2.6.1. Theoretical Considerations

In the light of the theoretical considerations presented in sections 3.3.3., it would appear that the reasons why composition tasks were found to be an unreliable way of assessing learners' writing ability were: the writer variable and the reader variable. It follows that improvements in the reliability of that type of assessment is possible by minimising the effect of the writer and reader variables.

The effect of the writer variable can be reduced by providing the intended subjects with more than one task, structuring the tasks in such a way that they would find the tasks in a clear context; i.e. writing purpose, audience, hints about the message to be conveyed and so on. While devising the tasks for the present tests, considerable attention was paid to reducing the effect of the writer variable by presenting two tasks in a context.

So far as the reader variable is concerned, on the basis of the discussion above, it can be argued that intra-inter-marker

reliability can be achieved to a satisfactory level if readers use the same criteria to evaluate a composition, minimising use of their own criteria and maintaining the same standard. This can be done, as was argued in the literature, by appealing to all raters to focus on the same features of composition with the same standard. That means that if the intended readers are trained to focus on prespecified criteria, their scores will be close enough to establish an acceptable level of reliability.

4.2.6.2. Purpose and Nature of the Test-tasks

The purpose of the test and the nature of the tasks set in the test largely determine the method to be used in marking compositions. Applying this argument to the present study, since one of the purposes of the test was to discover the washback effects of the SLC English exam and on the basis of that, replace it with a similar type of exam to the one proposed, the marking system to be adopted must have washback potential. In other words, the features to be focused on in the marking system should be features of composition-writing which we would wish to be taught in schools, as Jacobs et al (1981:34) suggest to:

include aspects of composition which are at least minimal ingredients for teaching and learning skills in written discourse.

The test-tasks are set for assessing the intended learners' ability to communicate in written discourse. Therefore the judgement must be based on effectiveness in conveying meaning (ideas). That means the marking system must accommodate features related to focus on meaning.

4.2.6.3. Practicality

It is not worth designing a marking system, whatever the quality, unless it can be used in the situation which it is intended for. In other words, the marking system to be developed here should be applicable to the Nepalese situation. The practicality criterion means that we must consider the following constraints: the number of the SLC participants and time.

As has been said earlier, a large number of students sit for the SLC exam every year. The number of SLC candidates (including private) was 107,029 in 1986, and the number is very likely to increase every year in the future because of natural growth in population and interest in education. So the marking system to be developed should be workable with great numbers.

As was discussed in section 2.2.6., the SLC candidates cannot go on to tertiary level until they pass the SLC exam. It would not be reasonable to keep them waiting for the SLC results for a long time. That means a method by which a composition can be marked quickly is preferable to a method that takes a long time. This type of time constraint also means that it is not practicable, in the Nepalese context, to use double marking (by which we mean having a composition marked by two markers and establishing the mean of the two as a true score). The implication of this is that an analytic method will not be practicable in this situation, though it is considered more reliable than a general impression method. Gosling (1966:25) reports that a certain number of compositions (no number was given) were marked in 2 hours 7 minutes using a general

impression method, but it took 5 hours 37 minutes when an analytic method was used.

4.2.7. Focused Impression Method

Gaining insights from the above discussion, a "focused impression method" (our terminology) of evaluating composition has been proposed. For the purpose of the proposed study, the focussed impression method is interpreted as 'an impressionistic method of scoring composition focusing on prespecified criteria'. Following the proposed method, a score for a composition is assigned impressionistically: however, raters are provided with a list of criteria to be focussed on while evaluating it. The raters do not have to award scores for separate features separately, but they must take account of the listed features and the weightings allocated to them while deciding a score for a composition.

4.2.7.1. Criteria and Weighting

The following are listed as the criteria, and weightings for marking Writing Task 2 and Writing Task 3.

Writing Task 2

Total Mark- 20

- i.Format -----2
- ii.Content-----5
- iii.Organisation--4
- iv.Language use--3
- v.Vocabulary----3
- vi.Mechanics-----3

Writing Task 3Total Marks- 20

- i.Content-----6
- ii.Organisation---5
- iii.Language use---3
- iv.Vocabulary-----3
- v.Mechanics-----3

The terms used here, apart from the term 'format', are borrowed from the ESL composition profile developed by Jacobs et al (1981). A copy of the profile is in appendix V. Even the description of the terms is adopted as they appear in the ESL composition. The term format is used to refer to the general conventions (e.g. writer's address, date, way of addressing, and closing) of writing an informal letter.

4.2.7.2. Rationale for Inclusion of the Criteria

There are two main reasons why the above criteria are selected as the prominent features of a composition, and why they are in the order in which they appear. The reasons are: washback effect, and some research findings. A quick glance at the list of the selected features will give an impression that emphasis is given to 'what to say' (content), 'how to say it' (organisation), and 'effectiveness in saying it' (e.g. language use, vocabulary and mechanics). The intention of including them and ordering them in such a way is that this way of evaluating writing will attract teachers and materials designers to emphasise these features. Another reason is that the research studies (Harris 1977; Freedman 1979) report that raters tend to focus on the features as listed above while evaluating a composition. Freedman's (1979)

study shows that the raters evaluated compositions focusing on content most, then on organisation and then on mechanics. Harris (1977) reports that the order of the features the raters in his study were emphasising on is this: content, organisation, appropriateness in expression, sentence structure, and mechanics. Jacobs et al's (1981) study also shows the same pattern.

4.2.8. Summary

A holistic approach, especially focusing on the analytic and general impression methods, has been discussed. A critical review of holistic evaluation has been presented regarding reliability problem in using it. Some suggestions are put forward.

A focused impression method of scoring compositions has been proposed for the present study, and also for the purpose of scoring answer sheets after the implementation of the proposed exam in Nepal.

Rationale for including the criteria have been given.

4.3. Description of Field Study

4.3.0. Introduction

This section of the chapter will be devoted to a description of the field study which was carried out between 1-1-1989 and 20-3-1989 in Nepal. It will also include the description of the piloting of the two sets of tests here in Edinburgh before piloting them in Nepal. The description of the field study covers piloting the test in Nepal, insights gained from the piloting of the test, preparation for the final version of the test, administration of the final version

of the test, and problems encountered while administering the test.

The section which deals with piloting the test covers a short description of the subjects who took part in the piloting, and the visits that the researcher paid to some SLC exam centres when the SLC exam was being administered. The section which deals with the administration of the test gives the details of the subjects who took the test, and the selection of schools and colleges for administering the test.

4.3.1. Piloting Tests in Edinburgh

As no group of students who had exactly the same level of proficiency in English as the intended subjects was found here in Edinburgh, a group of students (3 Japanese) who were learning English as a foreign language at the Institute for Applied Language Studies were asked to take the tests, assuming that they would serve as the next best group for the purpose of the study. They were also requested to give suggestions so that the tests could be improved. Their suggestions were found useful and some improvements were made in the tests.

4.3.2. Piloting Tests in Nepal

The two sets of tests were also piloted with the type of students which the tests were really intended for, in Nepal. For that purpose, the researcher had to go to Nepal in the last week of December, 1988, before the SLC candidates got engaged in taking the SLC exam, which is administered every year in January/February, otherwise it would have been difficult to get hold of them because they did not have to

come to their schools after they had completed their school studies.

4.3.2.1..Subjects

To recapitulate, the purpose of the study was to present the new test with the intention of using it to replace the present form of the SLC English exam. It is, therefore, true that the focus of the study was on the SLC exam, although students at other Grades such as Grade-10 and college level were also involved in the final study. That is why it was decided that the two sets of tests should be piloted only with the students who were preparing for the SLC exam for the year 1989, and would then be administered, after amendments, to the other levels. Altogether, 116 students representing all types of schools (i.e. A B C D) took the tests in the pilot. Table 6 shows the numbers of students who took the test from different types of school.

TABLE 6: Students who took part in the pilot

Type of school	No. of students
Type A	15
Type B	18
Type C	30
Type D	53
Total	116

4.3.2.2. Procedures of the Pilot Study

The procedure followed in the pilot study was as

follows. The students mentioned above were attending coaching classes for the SLC exam at the time, so it was not difficult to give them the tests.

As the purpose of the pilot was to make decisions on the selection of test tasks, their answer sheets were not scored for the purpose of statistical analysis. Their scripts were examined only to get an impression of how well they did, so that an insight could be gained as to whether the test tasks were workable or not. Furthermore, apart from evaluating the tests on the basis of the students' performances (in the test), the subjects were also asked questions such as, "which tasks did you find interesting?", "Did you find any of the tasks boring?", "Did you find any of the instructions confusing?", after they had finished taking the tests. They were also asked to give suggestions for making the final version of the test easier to understand.

The decisions on the selection of tasks and the amendments in the instructions were made taking insights from:

1. the performance of the testees in the test,
2. the way the testees reacted to the test while they were engaged in taking the test,
3. the individual conversations/interviews the researcher had with the testees after they had completed the test (impressionistic observations).

Attention was also paid to whether they were able finish the test-tasks within the time allowed.

During the pilot study, the researcher felt the necessity of testing the test also from the point of view of

its practicality.

It has previously been discussed that the reading ability of the intended subjects would be assessed using multiple choice tests. However, previously, the use of objective tests in the SLC exam was stopped on the grounds that this type of tests made cheating easier. It was, therefore, necessary to try out the tests to discover whether they could be administered in a way which would prevent the students from copying answers from one another.

In order to attempt to administer the multiple choice test in the existing situation while minimising cheating, some 20 SLC candidates from 3 different schools were gathered in a school. They were asked to take their seats in such a way that everyone would sit beside someone from another school who he was unlikely to be acquainted with. They took the tests there. After they had completed the test, they were asked if they had copied answers from others. In reply, they told the researcher that they would not copy answers from others until and unless they were sure that they were copying the right answers, or they knew that their neighbours knew better than them.

It seems obvious that what has to be done in order to reduce the chances of copying in the exam hall, even if the m/c test is administered, is to plan seating arrangements for examinees in such a way that their neighbours would be unfamiliar to them. This way of administering examinations is possible in Nepal, because it is unlikely that the SLC candidates from one school would know the SLC candidates from other schools. Another advantage with the English exam is that

it has become a tradition to administer the SLC English exam on the first day of the examination. It is very unlikely that the SLC participants would know who knew better than who at the first meeting. The researcher was encouraged to visit some more exam centres with the insights gained from that pilot.

Then arrangements were made to visit some SLC exam centres to have a look at how and in what circumstances the SLC exam was administered.

4.3.2.3. Visiting SLC Exam Centres

After obtaining permission to visit SLC exam centres from the Controller of the SLC examination, the researcher visited some 10 centres in Kathmandu and Chitawan districts. It was observed that exams were fair, ie. candidates were not allowed to use unfair means in the exam hall, and seats for the examinees were planned in such a way that the gap between two students was reasonable. The result was that it was not easy for them to read others' answers. As was said above, it was also observed that the SLC candidates in every exam centre were from different schools.

Based on the direct observation of the SLC exam centres, the researcher does not see any reason why a multiple choice test cannot be administered to the SLC students in Nepal. We see every possibility of administering the new exam in the existing situation.

4.3.3. Insights Gained from Piloting Tests

The necessity of making some amendments to the original version of the tests was realised when the two sets of tests were piloted with 3 Japanese learners of English as a

Foreign Language. Amendments were made before the two sets were piloted in Nepal. The piloting of the tests in Nepal revealed that the final version of the test needed further amendments. The subjects who took the tests found some instructions confusing, some test-tasks complicated, and some test-tasks boring. For instance, in Reading Task 2, they found item (20) complicated (see the exam in appendix V), so the instruction 'to summarise' was deleted. Although test instructions were given in Nepali where it was anticipated that students would have difficulty in understanding, and especially where the Japanese students had had problems, the students in the pilot found the tasks unfamiliar which made them difficult to understand. The nature of the tasks was explained in Nepali, especially in the Nepali medium schools. Then it must be admitted that the test was administered in a Nepali speaking environment; i.e. students were free to ask for anything in Nepali. The case was not the same in English medium schools.

While devising the original version of the test, an attempt was made to include an item to assess examinees' writing ability from a process perspective (see Writing Task 3). In this the intended subjects were asked to supply some amendments to their essays. The subjects who took part in the pilot found it boring and the researcher also found it impracticable. The subjects reported that they found it boring to read their own essay and most of them found nothing to amend. The researcher's difficulty was that it was not practicable to take away the pens they were using and supply

them with ones with a different colour of ink to make sure that everyone's amendments were distinct from his/her first draft. It was also realised that the idea of supplying pens with a different colour would be more difficult and impracticable in a real exam situation. Then it was decided that the activity of supplying correction or reorganising the essay should be removed from Writing Task 3.

4.3.4. Final Version of the Exam

The final version of the test was prepared taking into account how the testees who took part in the pilot study reacted to the tests. That is to say, among the two sets of test-tasks, those tasks were selected which the subjects found more interesting than the others. A sample of the final version of the test is available in appendix VI.

4.3.5. Administration of the Exam

The final version of the exam was administered to the Nepalese students who were studying English as a Foreign Language at different levels in different schools and colleges in Nepal.

4.3.5.1. Selection of School

English medium schools were selected on the basis of their SLC results and the social recognition those schools had in Nepal.

The following were selected among the schools that belonged to Type A;

1. Budhanilkantha school, Narayanthan, and
2. St. Xaviers school, Lalitpur.

The following were selected among the schools which

belonged to Type B;

1. Siddhartha Banasthali school, Balaju, and
2. Bhanubhakta Memorial Madhyamik Vidhyalaya,

Apart from comparing the performance of the English medium students with non-English medium students, there were other reasons for selecting the schools mentioned above. One of the reasons was that they would provide the best type of performance of the students at school leaving stage. Another reason was that people working in the field of education have the impression that it is Budhanilkantha or St. Xaviers which provide the best education in Nepal, so that these schools should obtain the best results in the SLC. But what has been the case is that it is Siddhartha Banasthali school which has been out-performing all the schools in Nepal in terms of SLC results for several years. Bhanubhakta school maintains the same standard.

Schools which belonged to Type C and D were also selected taking their previous SLC results into account. It should be noted that the main focus of the study was on Type D schools.

The following were selected among Type C schools.

1. Kanya school, Lainchair, and
2. Janasewa Madyamik Vidhayalaya, Panga.

The following were selected among the Type D schools.

1. Janajagriti school, Pithuawa,
2. Panchakanya school, Panchakanya,
3. Chitawan school, Chitawan,
4. Jhuwani school, Jhuwani,

5. Nepal High school, Ratna Nagar,
6. Adarsha Madhyamik Vidhyalaya, Gita Nagar,
7. Narayani Vidhyamandhir, Shiva Nagar,
8. Khairahani High school, Khairani,
9. Nepal Rastriya Madhyamik Vidhyalaya, Nepaltar,
10. Alka Tar Madhyamik Vidhyalaya, Lamjung,
11. Janata Madhyamik Vidhyalaya, Danda, and
12. Gokundayswor Madhyamik Vidhyalaya, Dhankuta.

4.3.5.2. Selection of College

Two types of colleges were selected; English medium and Nepali medium, bearing in mind that the environment outside classrooms in any college in Nepal is Nepali speaking.

The following colleges were selected.

- i. Birendra Multiple Campus, Bharatpur, and
- ii. Shree Commerce Campus, Tandi, Ratna Nagar.

4.3.5.3. Selection of Subject

The subjects who were at Grade 10, and college level were selected randomly. All the students who were present that day in the chosen groups were asked to take the exam, which they did but the SLC students were not present in the schools. They had to be informed in some other way. Schools were requested to gather them. All the students who came on the days they were requested to come were asked to take the exam, which they did. So in practice, what happened was that only the students who were informed and interested in taking the exam came and took the exam. However, it would not have any effect on the results because the participants represented all varieties of performance ranging from top to bottom.

4.3.5.4. Type of Subject

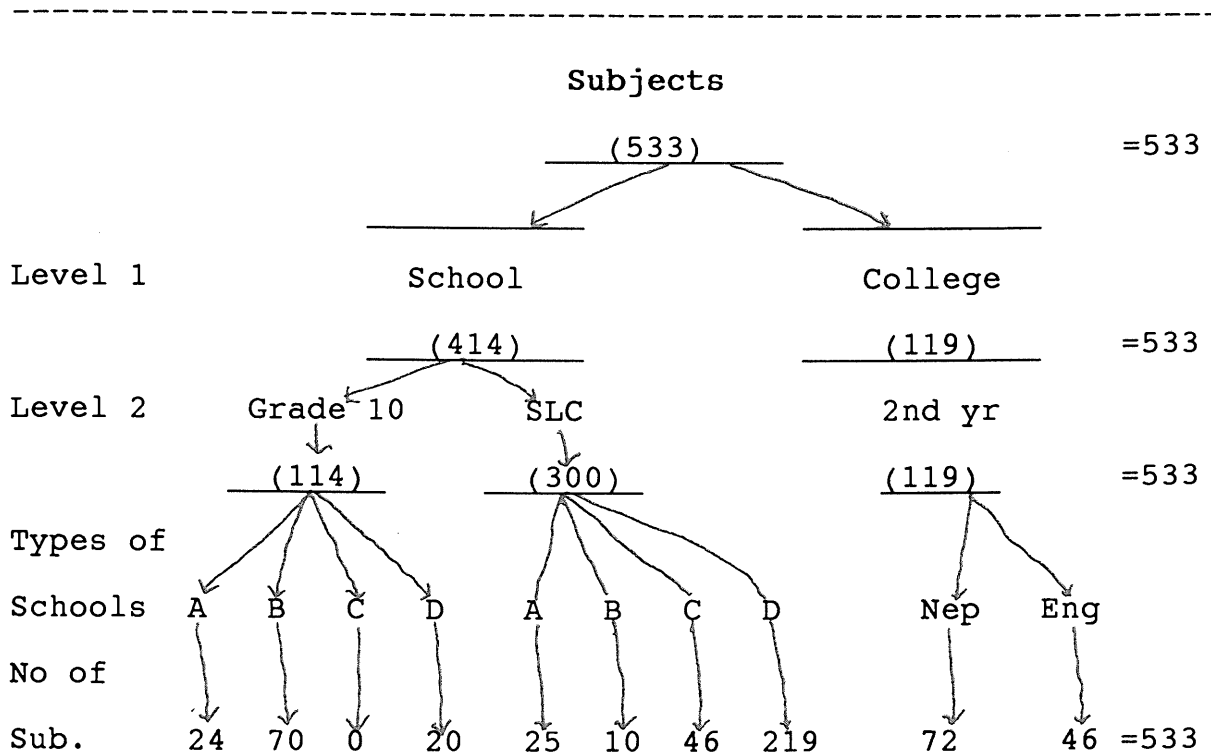
The test was administered to three different levels of students. The subjects who took the test were 533 Nepalese learners of English as a Foreign language, studying at different levels in different types of schools and colleges. From the point of view of their L1, all the students were homogeneous. Most of them had Nepali, the National language, as their L1, and some who did not have it as L1, nevertheless, had no difficulty in using it for any purpose.

For the purpose of the present investigation, the subjects were grouped in two ways;

- i. according to their level of education, and
- ii. according to the types of schools and colleges they represented.

We would like to present the types of school and the different levels of education they represented in diagram form.

Figure 2: Types of subjects who took the exam



Level 1 = Grouping the subjects into school and college level

Level 2 = Grouping them into Grades

Grade 10= Students studying at Grade 10

SLC = SLC Group

2nd yr = Students studying 2nd year tertiary level

Nep = Nepali medium

Eng = English medium

The numbers within the brackets indicate the number of students.

Grade 10

The Grade 10 students who took the exam had 9 years of school education. Those who were from English medium schools had studied English for 9 years and those from Nepali medium had studied English for 6 years. Their average age was 16.

SLC

These students had completed their school education and took the SLC exam in January/February 1989. Those students who were from English medium schools had studied English for 10 years and those from Nepali medium schools had studied English for 7 years. The average age of those students was 17.

2nd yr

This group of students was doing second year tertiary level courses. Their average age was 20. The English group were doing courses in the Faculty of Science, and the Nepali group were doing courses in the Faculty of Management and Business Studies and the Faculty of Humanities and Social Sciences.

4.3.6. Administration of Questionnaires

Two types of questionnaire, as mentioned in section 4.2.4. were administered.

4.3.6.1. Questionnaire for People Involved in SLC English

The people who took part in this questionnaire were: 36 school teachers of English at secondary level, the ELT adviser in the British Council in Kathmandu, the CTSDC expert, two teacher trainers. Of the 36 school teachers of English, 2 teachers were from Type A school, 3 from Type B, 6 from Type C, and 25 from Type D. The teachers from type D represented different parts of the country, and the rest of the teachers were from Kathmandu. On the whole, the people who took part in the questionnaire represented a wide range of people involved in the teaching and testing of English at school level.

4.3.6.2. Questionnaire for College Students

All the students who took part in the questionnaire were doing their second year tertiary education in different faculties. Since the study concerned the problems of tertiary students in coping with the courses that require a high level of English proficiency, we decided to administer the questionnaire to the students at the faculties which offer most courses in English.

We were also interested in finding out whether the students from Nepali medium schools and the students from English medium schools had had the same type of problem, so the students were grouped according to the medium of their schools. The number of students who took part in the questionnaire representing different faculties are given in Table 7.

Table 7: Students who participated in the questionnaire

	Nepali.	English	Total
1. Institute of Medicine	6	5	11
2. Institute of Science	10	11	21
3. Institute of Agriculture	11	--	11
4. Institute of Engineering	9	6	15
5. Institute of Forestry	2	-	2
Total	38	22	60

4.3.6.3. Questionnaire on Needs Analysis

A wide range of tertiary teachers of English participated in the investigation. 63 questionnaires were

filled in by college and university lecturers. Among them, 59 were tertiary teachers and four were university teachers. The purpose of involving university teachers was to involve the people responsible for testing tertiary English also. The 59 college teachers represented the faculties shown in Table 8.

Table 8: Tertiary tutors who participated in the questionnaire

Faculties	Number of Participants
Humanities and Social Sciences	11
Education	12
Science	7
Law	7
Management	11
Engineering	5
Agriculture	3
Medicine	3
Total	59

Of the 63 questionnaires we administered, four were found to have been carelessly filled in; the tutors ranked 2/3 subskills under the same rank. Those questionnaires were rejected. The results obtained from the analysis of the rest of the questionnaires will be reported in section 5.2.6.3..

4.3.7. Problems Encountered while Administering the Exam

The researcher admits that several ideas he had in mind while devising the tests here in Edinburgh regarding the help he would get from people in Nepal turned out to be over-optimistic when he actually started the field work. The enthusiasm of the researcher might have blinded him, preventing him from being able to predict what he would encounter in the actual situation. The researcher had believed that since the proposed study was meant to bring about changes in ELT at school in Nepal, there would not be any difficulty in obtaining cooperation from all sectors, irrespective of who they were, whether they were teachers or headmasters or administrators. Apart from some headmasters from English medium schools, most of the people he met were helpful; however, some problems were encountered while administering the exam.

The greatest disappointment was the reluctance of headmasters of the English medium schools to cooperate in gathering the SLC candidates from their schools for the exam. The researcher got the impression that the schools which had been successful in obtaining excellent results in the SLC exam were frightened of being exposed when they knew something about the nature and the purpose of the present study. It appears that they did not want their students to be compared with students from other schools when put in the situation which the new exam would create, i.e. assessing students' true language abilities. This situation made it very difficult for the researcher to gather students from Type B schools. With

the cooperation of some teachers of those schools, it was possible to get a few students to take the test although not possible to get 60/70 SLC candidates as had been desired. However, there were no such problems with the students who were studying at Grade 10 in those schools.

All the subjects reported that the test was long. But it was not possible to shorten the test, mainly because it had to replace a 3-hour exam. The students at Grade 10 from the English medium schools reported that they enjoyed taking the exam.

One of the practical problems with this research was that the students at Grade 10 and college level had to miss 3 hours of their regular classes.

Altogether some 700 students took the test, but only 533 students returned their answer sheets to the researcher. Even of those 533 returned scripts, some were incomplete. It was observed that a large number of tertiary level students who were doing courses in Nepali medium returned the scripts incomplete.

Motivation was another problem encountered. The students who were asked to take the test inquired about the purpose of the test. They were told that the purpose of the test was to assess their true language abilities in the hope that this would motivate them to exhibit their true language abilities, as we thought students always want to know their proficiency. The fact that they were not going to get their result there and then made them less motivated. Some of the SLC candidates also seemed worried about whether the new exam

would have any effect on their SLC results.

4.3.8. Arrangement for Collecting SLC Results

It was hypothesised that students' performance in the SLC English exam and the new exam would be different. It was, therefore, necessary to have the SLC results of the candidates who took the new exam. On our written request, the Education Secretary who was also the head of the Board of the SLC Examination, and the Controller of the SLC Examinations agreed to send us (here in Edinburgh) the SLC results of the candidates who took the new exam.

4.3.9. Summary and Comments

While devising the tests and designing the procedures to be followed in administering the exam, the researcher had certain preconceptions about the sample which would be available and the way the exam would be administered. In practice, the preconceptions proved to be ill-founded, and in need of modification.

The main problems encountered were: difficulty in gathering the SLC students from Type B schools and the idea of assessing the writing ability of the students from a process perspective.

The difficulty in gathering the SLC type students was because of the non-cooperative attitude of the school administrations. This problem was to some degree solved, though not to the extent that the researcher would have liked, by persuading individual teachers in those schools to gather at least some SLC candidates to take the test by visiting the students at home. This is how 13 SLC candidates were assembled

to take the test. A good number of Grade-10 students from these schools took the test however.

The difficulty of assessing writing from a process perspective would seem to be the limitation of any test which is administered in a 3-hour exam situation, like the SLC exam in Nepal.

Problems such as students' difficulty in understanding the test-tasks was solved by supplying them with explanations of the test-tasks in Nepali, and by allowing the subjects to ask in Nepali for explanation whenever they were not clear about anything in the exam.

In spite of the problems, the field work was carried out according to the broad framework envisaged in the proposal. It was hoped that all the relevant information that the researcher had proposed to tap had been tapped. That information constitutes the data which will be analysed in the following chapter.

CHAPTER 5: RESULTS

5.0. Introduction

Chapter 5 is devoted to the presentation of the results obtained from the different instruments we used for data collection. It also includes the results derived from validating the new exam. The description of the instruments, and the description of the administration of the instruments, have already been presented in the preceding chapters. The results are presented under the four main headings: i.e. reliability, descriptive data, hypothesis testing, and exam validation.

The reliability of the exam will be reported in the reliability section.

The descriptive data covers the description of the results of the new exam and the SLC English exam obtained by grade 10, SLC, and college level students. The descriptive data also includes a description of the information obtained from the analysis of the responses to the questionnaires we used.

The hypothesis testing presents the results obtained from testing the first two main hypotheses and the subsidiary hypotheses we presented in section 4.1.3.. Analysis of variance, comparison of means, and Scheffe tests were performed.

The exam validation includes the presentation of the results derived from the investigation of the new exam in terms of the performance of the SLC candidates who took it. Though the main subjects for this study were the SLC candidates, the performances of the grade 10 and college students were also exploited. In this section, we also present the results obtained from the investigation of the prediction which was carried out in terms of a quantitative analysis of the performance of the SLC candidates on the new exam and the SLC English.

5.1. Reliability

5.1.0. Introduction

We report below the reliability of the exam as established by different methods. It must be remembered that our choice of methods was limited, partly because of the heterogeneous nature of the test-items, and partly because we found it difficult to use a parallel form of exam for the present purpose. Internal consistency is the basic procedure we followed in establishing the reliability of all the tests of the exam. The intra/interraters' reliability of the writing test has also been determined.

5.1.1. Internal Consistency

Internal consistency of all tests has been established. We have followed Flanagan's method of determining the discrimination index of an item for the reading and cloze items.

To determine inter-component consistency of writing tests (i.e. letter and essay) a sample of 25 answer sheets was marked using our marking scheme as an analytic scoring method. The results are given in Table 9.

Table 9: Item Analysis: Discrimination Index

Discrimination Index	Reading	Cloze	Note-taking	Writing
80-89	-	3		
70-79	5	11		
60-69	3	11		
50-59	3	5		
40-49	5	4		
30-39	3	-		
N	19	34		
	KR-20=0.89	KR-20=0.75	Alpha=.9	Alpha=.97

Estimating the mean score for the internal consistency of all subtests, it was $\alpha=.82$.

As can be seen, the two writing tasks have a high degree of inter-component consistency. It is very high for a writing task to have $\alpha .97$. This must have been the result of using an analytic scoring method.

5.1.2. Intermarker Reliability

In the process of determining intermarker reliability of the writing tasks, we asked an experienced teacher of English in Nepal to mark a sample of some 20 answer sheets of the students, and a native speaker of English in Edinburgh to mark a sample of 25 answer sheets. The procedures followed in marking the answers were the same as described in section 5.3.3.5. Before they marked the answer sheets, we discussed the procedures to follow while assigning a score to a piece of writing. Then we established the

correlation coefficient of those scores with ours. We found that the correlation coefficient of the relationship between the scores by the native speaker and ours was .92, and the correlation coefficient of the relationship between the scores of the teacher of English in Nepal and ours was .68. One of the possible explanations for the discrepancy is that the Nepalese teacher was trained following a traditional method of teaching and perhaps was concerned very much with the notion of grammatical correctness. That is why the sharing between the two markings was 46%. In this case, the native-rater would have shared our opinion: the sharing between the two marking was 85%.

5.1.3. Single-Rater Reliability

It should be remembered that, in practice, the writing tasks would be marked by a single rater. Therefore, we found it important to establish the single-rater reliability of the exam. For that purpose, a sample of 25 answer sheets was double marked. The correlation coefficient of the relationship between the scores obtained from double marking was .91. In comparison to Hamp-Lyons (1987:204) who reports the average reliability of the aggregate score for 3 raters as .865, our result is quite high. The high correlation between the marking of the single rater may be the contribution of the marking system we used.

5.1.4. Test-Retest Reliability

To determine the test-retest reliability of the new exam, we asked a group of 50 SLC students to retake the exam 3 days after they first took the exam. When they discovered that it was the same exam that they had taken before, their response to the exam was unexpected. They did not seem to be interested in taking the exam twice. They asked us the purpose of asking them to retake the exam. When we told them what the purpose was, they

told us that they were not interested in taking the exam because they knew what they would find in the exam. So we left them free to choose whether to take or not to take it a second time. Only eight students out of fifty chose to take it. Later, we asked the people who took the exam twice why they took it, they told us that they wanted to help us. So, they took the exam only because they thought it would have been a problem for us had they not. We got the impression that the test-retest method was not the proper method to be used for determining reliability of a test mainly because, according to our experience, testees do not accept the test in the same manner when they take it a second time. Establishing the test-retest reliability of the exam based on those eight students' performance, we found $r=.87$.

While talking about test-retest reliability, Alderson (1989:3) sets as a criterion that "learners will have learned nothing, either from the test administration the first time or have learned nothing in the time interval between the two tests". Agreeing with him, we would like to add testees' interest in taking the test a second time as a criterion to be considered when test retest reliability of a test has to be established. He also talks about variation in performance of a testee in terms of the way he/she encounters a task, and variation in terms of test taking occasions. In this context, the question arises as to how the testees' interest can be ignored in taking the test a second time when the testees are not the same on two occasions. Contrary to these arguments, Criper and Davies (1988:51) present a case of the test-retest reliability of the ELTS established in a gap of 8 months. Nothing is said about the interest of the people who took the test a second time. Since there was a gap of eight months it is hard to believe that the testees were at the same ability

level as they were when they took the test first time; they must have learned or unlearned something over a period of eight months related to the traits tested. Taking this as an example, we report below another evidence of the test-retest reliability of the exam based on the data we collected from the second administration of the exam to the grade 10 students in December 1989 (section 5.1.1.2.). The correlation coefficient of the relationship between the scores of the first and second administration of the exam obtained by the Grade 10 students was .83. However, we do not recommend this way of establishing test-retest reliability of a test.

5.2. Descriptive Data

5.2.0. Introduction

In this section, we present the descriptive data from the results of the new exam and the SLC English exam, and the information collected from two sets of questionnaires. The procedures which were followed in data collection have already been mentioned in the preceding chapter. The results which will be presented here will be further analysed, and used in subsequent sections for testing hypotheses.

The raw scores from which the descriptive statistics of the two exams were derived are given in appendices VII and VIII. The statistical analyses were carried out by using SPSS and Minitab statistical packages.

The descriptive statistics are presented under different headings so that they can be referred to directly, and used in relation to inferential statistical analyses showing the results of the testing of hypotheses. The descriptive statistics include histograms, cumulative frequency distributions, standard deviations, and means of the scores obtained by the different

groups of students on different exams.

Responses to the questionnaires were analysed quantitatively. The summary of the relevant information obtained from the analysis of the questionnaires will be presented.

5.2.1. Results of New exam: Gradewise (Grade 10 and SLC)

In this section, the results of the new exam obtained by Grade 10 and SLC students will be presented.

The new exam was administered to Grade 10 students twice: first at the beginning of their Grade 10 studies, and second when they were about to finish them.

5.2.1.1. Grade 10: First Administration in January 1989

The frequencies of the scores obtained by the Grade 10 students in the new exam are presented in Table 10. The raw scores are given in appendix VII. The description of the subjects who took the exam has already been given. It is important to note that this group of students had spent one year less than the SLC group studying English at school.

Table 10: Frequency of scores obtained by Grade 10 students

Midpoint	Count	
10	1	*
20	2	**
30	8	*****
40	8	*****
50	9	*****
60	31	*****
70	46	*****
80	9	*****

The descriptive statistics are given in Table 12.

5.2.1.2. Grade 10: Second Administration in December 1989

The same exam was administered again in December 1989 to the Grade 10 students who had taken the exam in the first administration in January, 1989. The idea was to investigate their development of language proficiency after they started the Grade 10 studies. This group of students was preparing for the SLC English exam, 1990.

The grade 10 students who participated in the second administration of the exam were only from Type D schools. Other types of schools (Type A and B) were reluctant to allow us to use their students for this study. From this point of view, the data may not be adequate to reach any conclusion. However, it was hoped that it would give some information about the effect of SLC preparation on the improvement of language proficiency. Thirteen students from Type D schools took the exam. Calculating their scores, the mean score was found to be 40.07, and the standard deviation to be 14.58.

5.2.1.3. Results Obtained by SLC Students

The description of the subjects who took the exam has already been presented. The following analysis includes the results of only 281 candidates out of 300 as described in section 4.3.5.4.. The reason is that the students who took the new exam had also taken the SLC English exam the same year. We were interested in looking at their performance on the new exam in relation to their performance on the SLC English exam. Keeping this in mind, we decided to include only those students' results whose results in the SLC English exam were available. The results of the SLC English exam were collected from the Office of the Controller of the SLC examinations. We could not collect some 19 students' results of the SLC English exam for further analysis,

so we had to omit 20 students from the initial list of the students who took the exam. The raw scores are given in appendix VIII. Table 11 presents the frequency distributions of the scores obtained by the SLC candidates in the new exam.

Table 11: Frequency of scores obtained by SLC students

Midpoint	Count	
0	1	*
10	65	*****
20	80	*****
30	47	*****
40	36	*****
50	9	*****
60	14	*****
70	11	*****
80	18	*****

The descriptive statistics of the results are given in Table 12.

Table 12: Descriptive statistics of the scores obtained by Grade 10 and SLC Students

	Grade 10	SLC
Mean	59.78	30.04
S.D.	14.75	24.00
N	114	281

Figure 3 presents the results graphically.

Comparison of Frequency Distributions of Scores obtained by Grade 10 and SLC

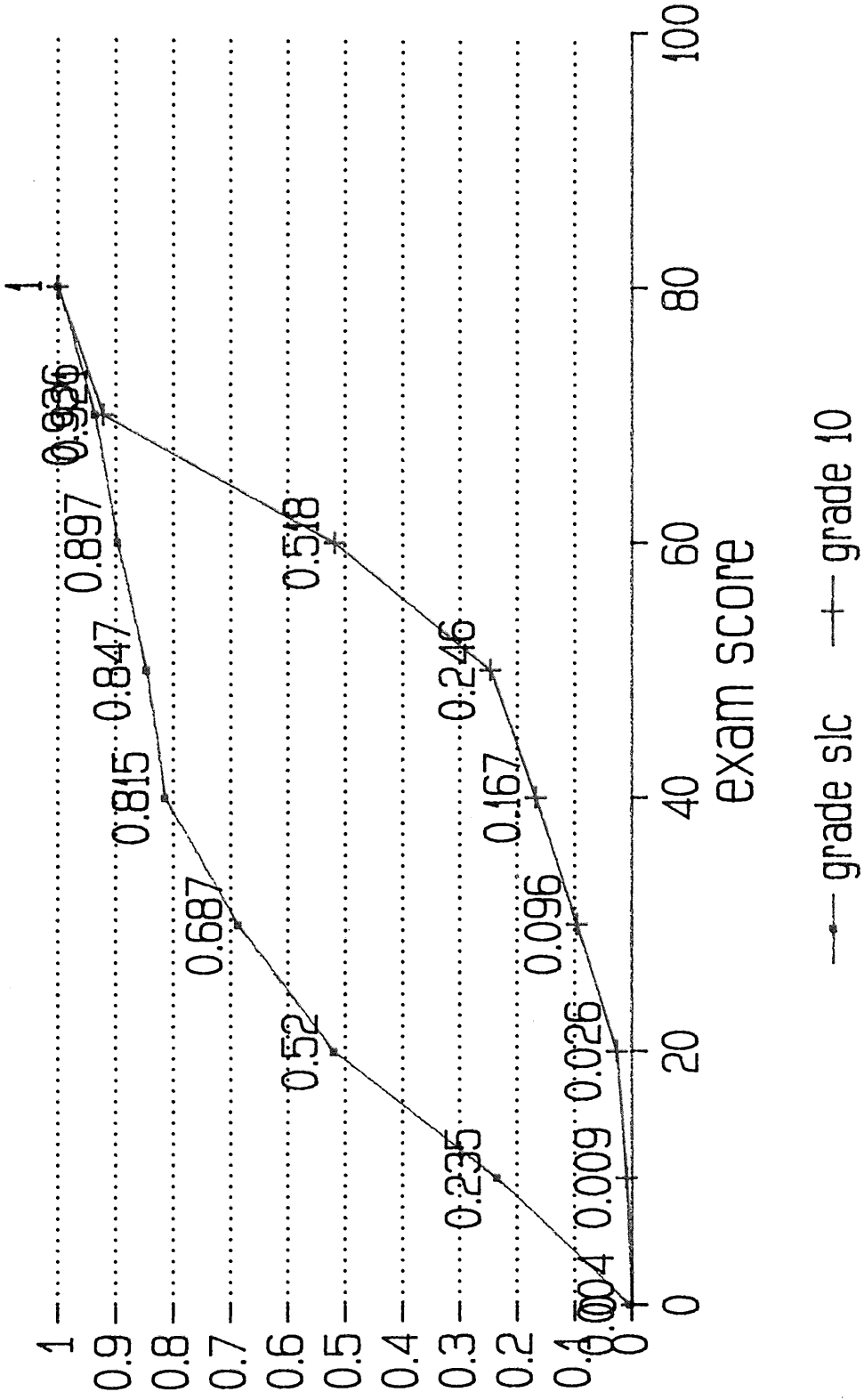


Figure 3

Comments

Looking at Table 12, it appears that the Grade 10 students who had studied English one year less than the SLC students have done better than the SLC students. It is necessary to find out whether the difference is statistically significant. It would also be interesting to see whether their performances would follow the same pattern if we compared the results of the grade 10 and the SLC students according to their type of school. One possible reason for the difference may be the difference in the number of students who represented the two groups of students from different types of schools. For example, the number of students who participated in the administration of the new exam representing Type B school at Grade 10 level was higher than the students who represented the school at SLC level. Furthermore, the number of students who represented Grade 10 and SLC from Type D schools also varied considerably. So we were encouraged to look at the data by dividing it into school type. For that purpose, the scores were analysed (see below) according to the type of school they represented.

5.2.2. Results of New Exam: Schoolwise

Table 13 presents the results of the new exam obtained by Grade 10 and the SLC students according to school type. Figure 4 presents the comparative results of the Grade 10 and SLC students graphically.

Comparison of Scores Obtained by SLC and Grade 10 Students on New Exam According to School Type

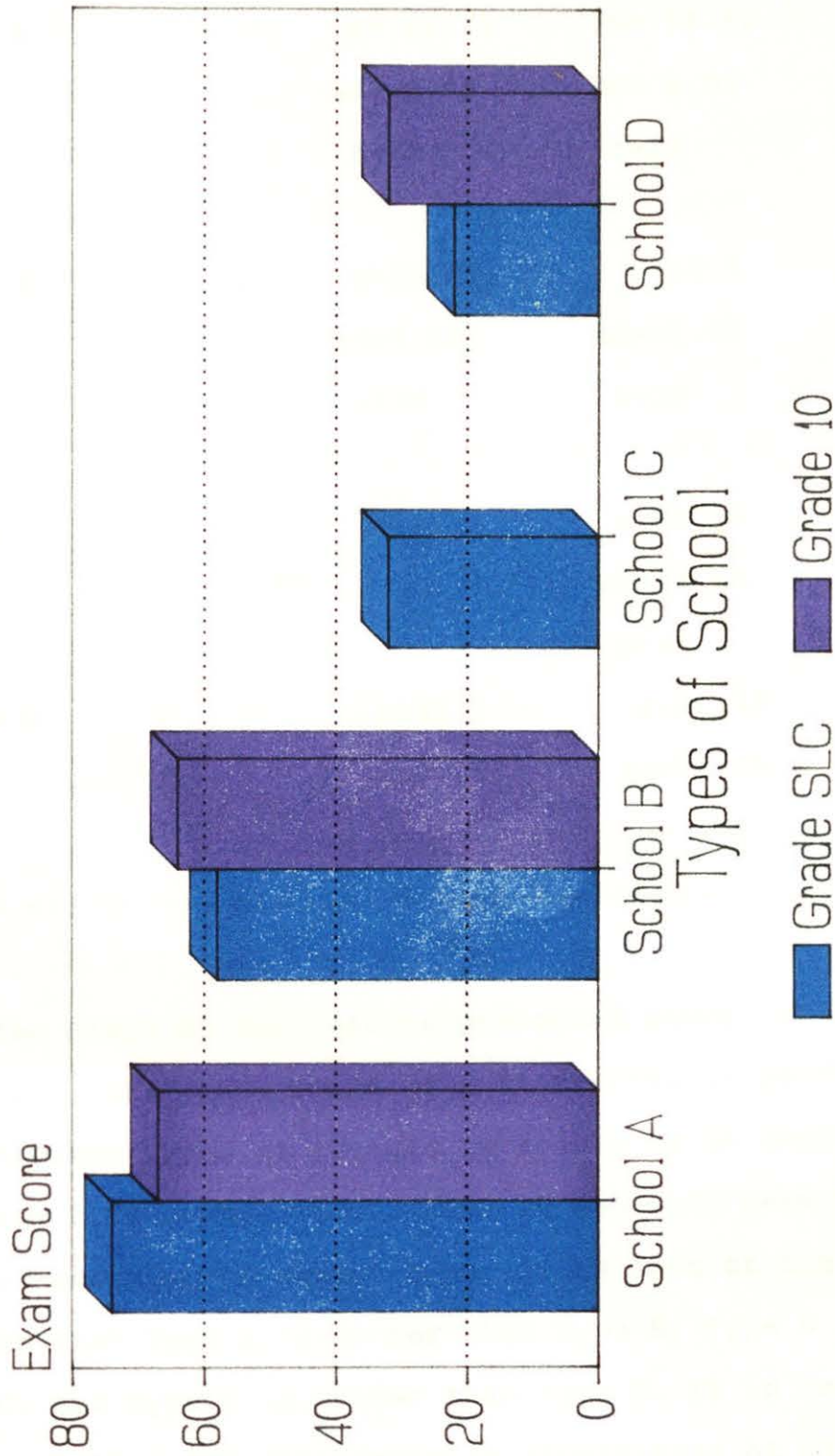


Figure 4

Table 13: Description of the scores according to school type

Type of School	Grade 10	SLC
Type A	x= 67.75 sd= 6.96 n=24	x= 74.88 sd= 6.02 n=25
Type B	x=64.81 sd=7.16 n=70	x=58.5 sd=11.82 n=10
Type C	x=32.96 sd=14.26 n=46
Type D	x=32.1 sd=9.73 n=20	x=22.34 sd=13.06 n= 200

Comments

On the basis of the results presented above, it seems that our prediction that there would be a difference in performance between different types of schools is likely to be confirmed. Students from both groups (i.e. Grade 10 and SLC) have performed differently according to school type. If we look at their means, we will find that Type A is better than Type B, Type B is better than Type C, and Type C is better than Type D. It is necessary to investigate whether the difference is significant at $p < .05$. This type of result will be reported under 'hypothesis testing'.

5.2.3. Results of the SLC English Exam

It has already been mentioned that the SLC students who took the new exam had also taken the SLC English exam the same year. The SLC results were collected from the Office of the Controller of the Board of the SLC Examinations. The raw scores are given in appendix VIII. Table 14 presents the descriptive statistics of the results according to school types.

Table 14: Results in the SLC English exam according to type of school

	Mean	Std Dev	Number
Entire group	47.65	17.49	281
TYPE A	72.72	8.65	25
TYPE B	65.80	9.38	10
TYPE C	45.76	14.23	46
TYPE D	44.05	16.2	200

Comments

The descriptive statistics obtained from the performance of the SLC students on the SLC English exam show that the schools are different and the hierarchy of the schools is the same as was reported before. This result also supports our prediction that the schools were different. However, the difference between Type C and Type D on the SLC English Exam is quite small. We still do not know whether the difference is real or not. Whether they are significantly different at .05 level of significance or not will be shown in section 5.3..

5.2.4. Comparison of Performance of SLC Group in New Exam and SLC English Exam

One of the questions we raised in the present study was about the adequacy of the present ELT at school in Nepal in relation to the needs of SLC passers for tertiary education. In section 4.1.1., we said that we would address this question by analysing performances quantitatively. In the process of investigating whether the existing English language teaching at SLC level was adequate in relation to the purpose for which English is taught, we proposed to find out the minimum level of English proficiency required for passing the SLC English exam. The argument is that determining first what level of English proficiency is adequate for passing the SLC English exam, and comparing it to the demand college courses which require a high level of English proficiency place on entrants will provide a basis for discussion as to whether the teaching of the existing SLC English is adequate. More precisely, we proposed to investigate what was sufficient (i.e. criterial levels) for college entry, and what was expected from college entrants in order to provide a basis for determining the adequacy of the teaching of the SLC English.

It must be noted that it is hard to pin down what skills the SLC English exam measures. We analysed the exam paper for 1989, (section 2.2.6.6.) and concluded that the SLC English exam measures whether the students have memorised the ready-made answers and the exercises which they practise during their school studies. We did not have access to the SLC answer-sheets of the students who took the new exam to determine an index of their language proficiency. Therefore, we decided to measure the adequacy of SLC English by using another measure (i.e. the new

exam) in order to determine an index of language proficiency of SLC passers so as to examine how adequate a SLC passer's English was in the light of what he/she was expected to have achieved as a college entrant.

The new exam represents the basic language skills that a college student needs to have achieved to be able to cope with the tertiary courses which require a high level of English proficiency. The idea was that those who perform well in the new exam would be considered to have achieved an expected level of English proficiency for tertiary level, though it must be admitted that there was no definition of an expected level of proficiency. We predicted, since the SLC students were not prepared for a test based on language skills, that they would find the new exam more difficult than the SLC English exam. In other words, their scores in the SLC English would be higher than their score in the new exam. It was hoped that, by analysing their performance this way, it would be possible to account for what the SLC passers could do and what they could not do with their English. It was also hoped that this would provide a basis for accounting for the proficiency level of SLC passers from different types of schools.

The descriptive statistics of the performance of the SLC candidates on the new exam and the SLC English exam show that their performance on the SLC Exam ($x=47.65$) is higher than their performance on the new exam ($x=30.04$). It supports our prediction that there would be a difference in performance of SLC students on both exams. We were also interested in finding out whether their performance would differ according to school type as well. Figure 5 presents the graphic representation of the results schoolwise, and Figure 6 presents the cumulative probability

Mean Score For SLC and New Exam According to School Type

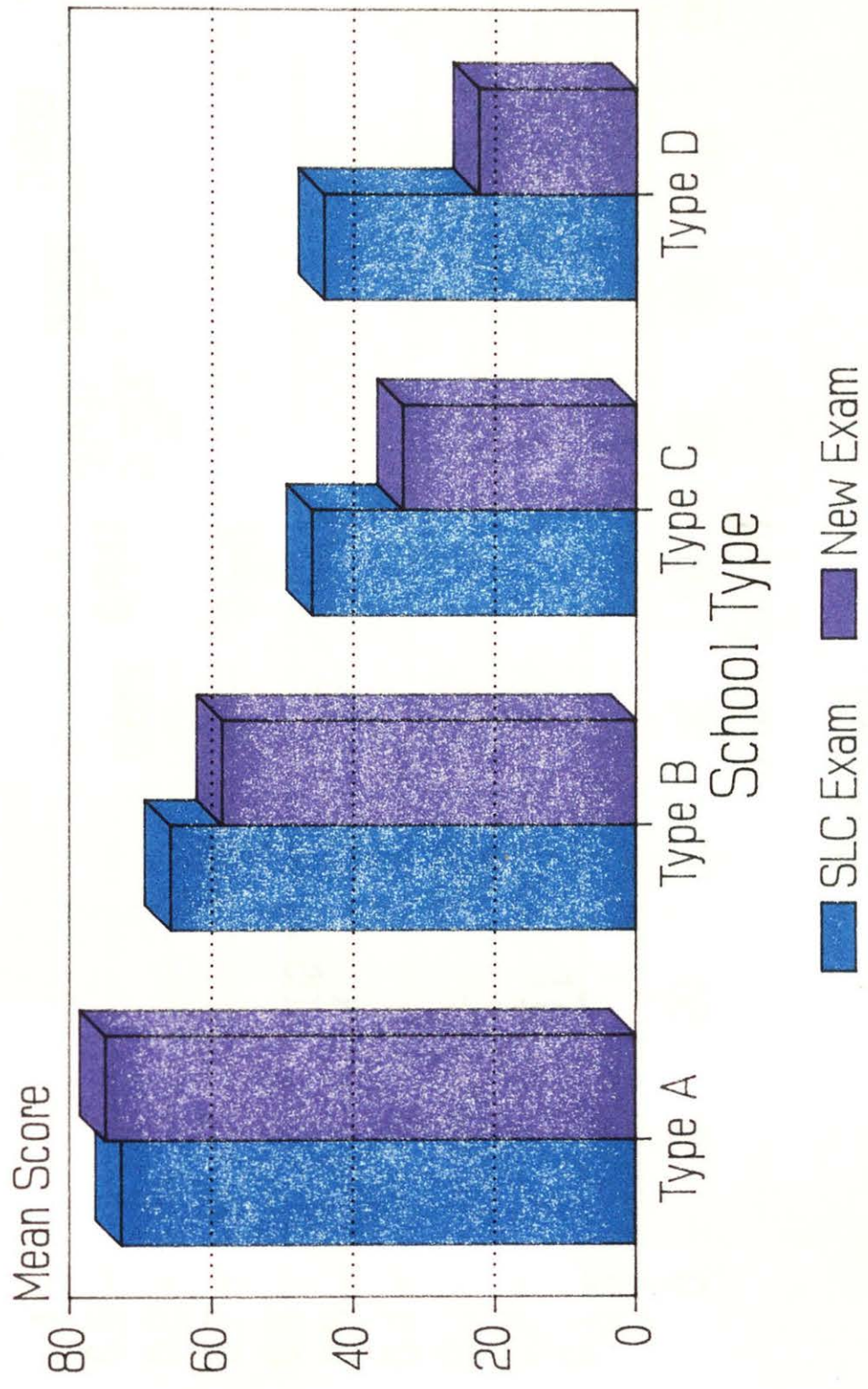


Figure 5

Comparison of Frequency Distributions of Scores in the two Exams

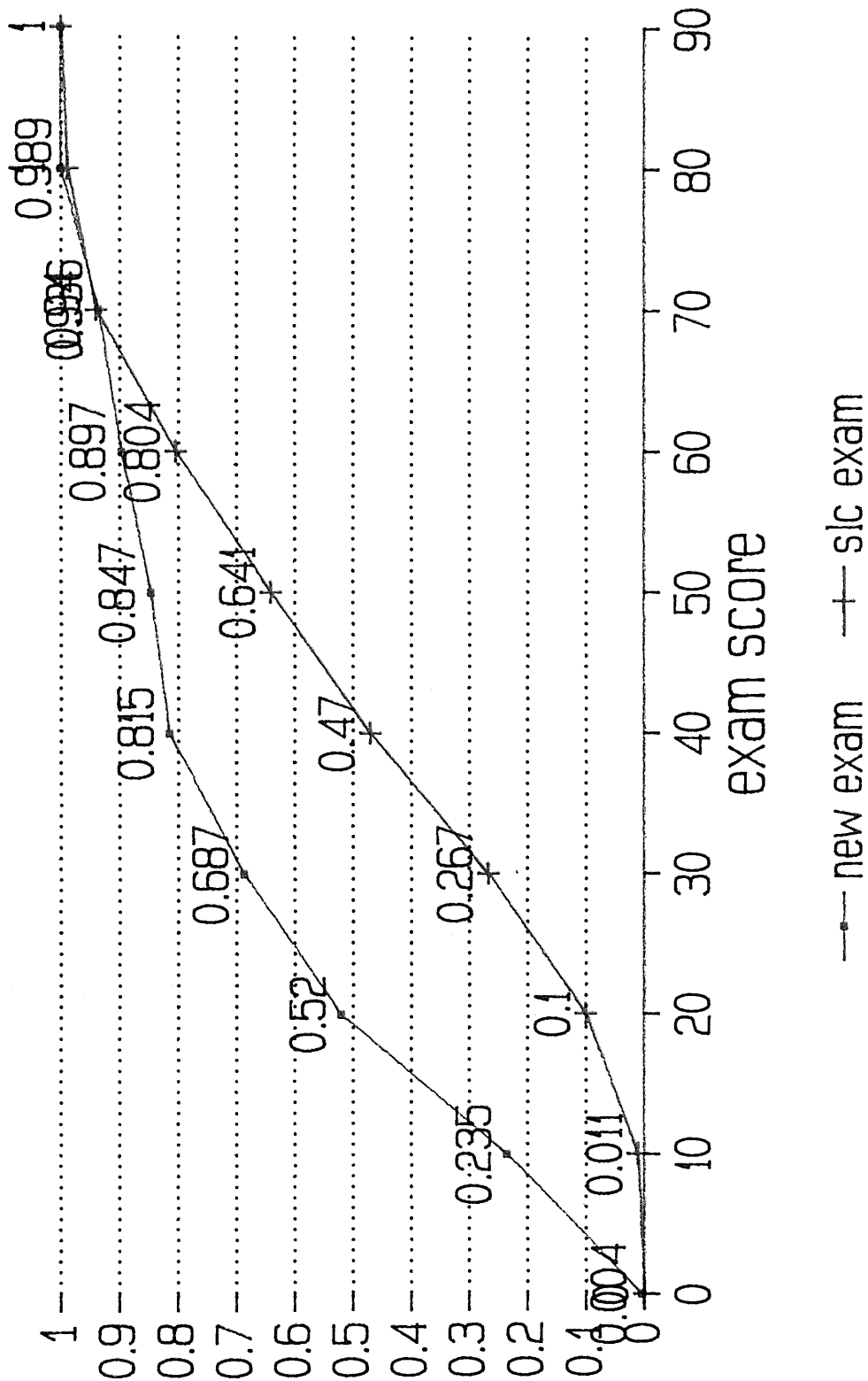


Figure 6

Table 15 shows that the percentage of students who reached different cutoff points varies from one exam to another. Another remarkable difference is in the number of students who fall under 32-44, 45-59, and 60 and above cutoff points: e.g. 45, 16, 36 in the new exam and 88 68 and 81 in the SLC English exam. Only 13% of students reached the 60 and above cutoff in the new exam, whereas in the SLC English exam the percentage of students who reached 60 and above was 29%. This is also the case with the cutoff points 45-59, 32-44, and 0-31. This clearly shows that the performance of the students varied from one exam to another. Looking at the results, it appears that students had found the new exam more difficult than the SLC English exam. This is what we predicted. For example, in the new exam 66% of the students obtained 0-31 whereas in the SLC English exam only 16% of the students students obtained 0-31. This is a notable difference.

Another important point that emerges from the results is the nature of the performance differences of the various types of schools. Table 15 shows that the performance of the students from Type A and B schools on the two exams did not differ in the same way as did the performance of the students from Type C and D schools. This point will be more explicit if we look at the following table:

Table 16: Percentage of students who exhibited equal performance on both exams

Cutoff	A	B	C	D
0-31	-	-	6	19
32-44	-	-	6	2
45-59	-	-	4	1
60 above	96	60	2	1
Total	96%	60%	19%	23%

Table 16 shows the number of students who obtained the same cutoff point on both exams. Of the students from Type A schools 96% did equally well on both exams and 60% of the students from Type B schools displayed equal performance on both exams. The difference between the total scores on both exams, of the remaining students from Type B schools, was very small. However the percentages of the students who showed equal performance on both exams from Type C and D schools is remarkably low (e.g. 19 % in Type C and 23% in Type D).

There appears to be a problem with Type C and D schools where students' performance varied enormously from the new exam to the SLC English exam. In Type A and B schools, the difference was not salient.

To find out the students whose performance differed considerably from one exam to another, the results were organised in the following table:

Table 17 shows the percentage of the students who fall under the 0-31 cutoff in the new exam but 32 and above in the SLC English exam.

Table 17: Percentage of Students who reached 0-31 cutoff in new exam, and 32 and above in the SLC exam

Cutoff	A	B	C	D	Total
32-44	-	-	33	33	
45-59	-	-	9	15	
60 above	-	-	-	11	
			42%	59%	49%

Table 17 shows that the students from Type C and D had performed differently on the two exams: 42% of the students from Type C and 59% from Type D passed the SLC English exam with different divisions but reached 0-31 in the new exam. This was not the case with the students from Type A and B schools. We were also interested in finding out whether this was the case with the students who obtained 0-31 in the SLC, and 32 and above in the new exam. This type of information is presented in the following table. On the whole, we found that 49% of the total students who obtained 0-31 in the new exam passed the SLC English exam with 1st, 2nd, and 3rd divisions.

Table 18: Percentage of students who obtained below 32 in the new exam but 32 and above in the SLC English Exam

Cutoff	A	B	C	D
32-44	-	-	-	-
45-59	-	-	6	-
60-above	-	-	-	-

Table 18 shows that only 6% of the students (3 students) from Type C schools who reached 0-31 cutoff in the SLC English exam obtained 45-59 cutoff in the new exam. This shows that the number of students who found the SLC English exam more difficult than the new exam is insignificant.

Looking at Table 15, we find that only 16%, 05%, and 13% of the students had achieved 32-44, 45-59, 60 and above cutoff points respectively in the new exam. In other words, only 34% of the total students reached 32 and above in the new exam. Our main interest, in this context, was to ascertain what the 32-44 passmark in the new exam means in terms of language ability.

What appeared to be worth investigation was the difference between these 34% of students and those who could not reach 32 and above cutoffs (i.e. 66% of the students out of which 49% could do quite well in the SLC English exam). To analyse the performance of the students who reached 32-44, 45-59, and 60- and above cutoffs on the new exam in terms of language ability requires a long time and adequate space. We realised that an analysis of the performance of the students who reached 32-44, and 0-31 cutoffs would serve the purpose.

To determine an index of language ability for the students who reached cutoff point 32-44, we decided to look at the performance of the students who fell under the 32-44 cutoff. Our argument was that students who obtained cutoff 45-59, and 60 and above were mostly from A and B school types and performed equally well on both exams. Since we could not analyse the performance of the students on the SLC English exam, we decided to use their performance on the new exam as an index of their language proficiency. It was hoped that the 32-44 cutoff would provide us with useful information about what 'passmark' in new exam means: the criterial level of performance.

Comments

1. In general, the SLC students found the new exam more difficult than the SLC English exam (Table 15).
2. 49% of the total students had passed the SLC English exam who reached 0-31 cutoff in the new exam (Table 17).
3. 96% of the students from Type A, and 60% of the students from Type B performed equally well on both exams, but 42% from Type C and 59% from Type D performed well in the SLC English exam, but badly in the new exam (Tables - 16 and 17).

5.2.5. Results of New Exam Obtained by College Students

5.2.5.1. Administered to the Students at the Institute of Science

In the process of predictive validation of the new exam, we administered the new exam to second year college students who were doing the courses which require English proficiency. The idea was to compare their performance in the first year courses to their performance in the SLC English exam, and their performance in the new exam to their performance in the second year exams. We administered the exam to the students attending

the Institute of Science, Bharatpur.

We also collected the scores obtained by these students in the SLC English exam, the I.Sc. first year English exam, and the I Sc second year English exam. The idea was to investigate the predictive validity of the SLC English exam and the new exam as to the performance of these SLC passers on the tertiary courses. Below, we present the correlations of their scores in all the above exams.

Table 19: Correlations of the scores in the SLC English exam, the new exam, I Sc first and second year English exam obtained by the students from the Institute of Science, Bharatpur.

	I Sc. 1st Yr	I Sc. 2nd Yr
SLC English exam	.3	.32
New exam	.51	.47

5.2.5.2..Administration of the New Exam to the students at the Institute of Humanities and Social Science, and the Institute of Management

Altogether 72 students were asked to take the exam. Their answer sheets were not worth analysing statistically, because almost all of the students had returned them incomplete. They had seemed to be willing to take the exam, but were not compelled to take it, so the incomplete tasks showed that either they were not interested in it or they found the exam very difficult.

5.2.6. Questionnaire Results

5.2.6.1. Questionnaire for People Involved in SLC English

Table 20 presents the summary of responses to the questionnaire used for collecting information from the people involved in the teaching and testing of SLC English.

Table 20: Summary of responses to questionnaire for people involved in SLC English

Questions	Yes	No
1. To what extent the existing SLC English exam covers the course objectives of SLC English? not appropriate		
2. Does the SLC English exam assess the true language abilities of the candidates?	10%	90%
3. Have you had chief examiner's report or comment on the SLC English exam?		100%
4. Does the SLC English exam look like a language test?	28%	69%
5. Does the SLC English exam assess the activities that an SLC passer would need to do after he passes the SLC exam?	23%	76%
6. Any comment on the SLC English exam?	not appropriate	
Teachers' responses		
i. Are you free to teach what you think good for the students?	50%	50%
ii. Do you have to prepare answers for the possible questions to be asked in the SLC English exam?	80%	20%
iii. Are you selective about what have already been asked and what will be asked in the coming year in the SLC English exam?	75%	25%
iv. Do you have to spend a considerable amount of		

time on preparing the SLC candidates for the SLC
 English exam apart from their prescribed course? 67% 33%

The overall impression of the people who responded to the questionnaire was that they were not satisfied with the existing English language testing situation at school in Nepal. Almost all indicated that the SLC English exam did not cover the course objectives, in that it did not assess listening and speaking skills at all, and that the SLC English exam assessed only reading and writing to a satisfactory extent. On the basis of their responses, it was revealed that they did not consider the SLC English exam to be a language test that measures the true language abilities of the students, and believed that whatever it assessed was not congruent with the language skills that the SLC passers need to cope with tertiary education. All the respondents indicated that they had never read a report by the chief examiner. The general comments they made on the SLC English exam can be summarised in the following points:

- the SLC English exam has an adverse effect on teaching and learning,
- the SLC English exam should be more skill-orientated rather than grammar-orientated and assess all the four skills,
- it does not reflect the course objectives,
- it should be congruent with the tertiary requirements,
- some suggested that the existing exam should be changed.

On the whole, there was a great deal of readiness in the respondents to change the existing SLC English exam, and the expected change was in the direction of making it skill-orientated and relating it to the course objectives and tertiary

education. We concluded that the people involved in SLC English would be responsive to measures to bring about changes in the SLC English exam in the direction of making it include the skills that are congruent with the tertiary courses.

The teachers' responses on the effect of the SLC English exam on the way they teach SLC English to their SLC students indicate that the SLC English exam had a constraining effect on their teaching. Half of all the teachers said that they felt free to teach whatever they thought beneficial for their students but half said that they were not free, in other words, that their whole teaching was confined to the SLC English exam. Three quarters of teachers said that they had to prepare answers for the SLC English exam for their students. Almost all teachers said that while teaching SLC English they were selective according to what had been asked before and what would be asked that year. Almost all teachers said that they spent a considerable amount of time preparing students for the SLC English exam.

This information supported the following predictions:

- people were not satisfied with the existing SLC English exam,
- the SLC English exam had an adverse effect on the teaching of SLC English at school,
- the SLC English exam had a strong influence on what goes on in the classroom,
- an exam cannot avoid influencing what goes on in the classroom.

5.2.6.2. Questionnaire for College Students

Below we summarise the information obtained from the analysis of the responses of the students to the questionnaire.

The courses they studied in English were:

1. Institute of Medicine

Physics, Surveying, Botany, Zoology, Medicine, Surgery, Pharmacology, Chemical Pathology, ESP

2. Institute of Science

Biology, Physics, Chemistry, Technical Math, Zoology, English.

3. Institute of Agriculture

Physics, Zoology, Botany, Chemistry, English.

4. Institute of Engineering

Physics, Chemistry, Hydraulics Mechanics, Materials, Surveying, Maths, Soil Mechanics, Estimating and Costing, English.

5. Institute of Forestry

Survey Engineering, Soil and Water Conservation, Management, English, Botany.

Analysing their responses, it was found that all four language skills were necessary for a student to cope with the courses though reading and writing were the most important. They clearly indicated that listening and speaking were also necessary.

All students, irrespective of medium of instruction, found the school English inadequate for speaking. Students from Nepali medium schools (Type C and D) found the school English inadequate for reading and writing purposes as well. This was not the case with the students from English medium schools (i.e A and B).

In response to the question whether they found any courses difficult only because their English was not adequate for that purpose, they indicated that their main difficulty was to express themselves in writing in English. In other words, they found writing in English the most difficult task.

All of the students from Nepali medium schools were of the opinion that Science and Maths should be taught in English.

They were asked whether they would have done better than they did if their English was better. All of them indicated that they would have done much better if their English was much better. They placed a great importance on English proficiency for success in the courses they were doing.

The major points which emerged from their general comments on the issues can be summarised as:

- the bad effect of the rote-learning type of teaching of English at school,
- the inadequacy of SLC English in relation to the demand of English proficiency their tertiary courses made on them,
- the necessity of teaching Science and Maths in English

They were also asked to supply us with their scores in the SLC English exam and their overall performance in the first year exam. Upon looking at their performance in the first year exam, it was found that eight students, all from Nepali medium schools, had failed in English. Correlation between the two sets of scores was .48.

5.2.6.3. Questionnaire on Needs Analysis

The questionnaires were analysed by assigning scores 5,4,3,2,1 to the first, second, third, fourth and fifth ranked responses respectively. The raw score obtained by each item is given in Table 21.

Table 21: Raw scores obtained by each item

Reading subskills

1. Skimming	-----	240
2. Understanding ideas and information in a text not explicitly stated	-----	122
3. Understanding explicitly stated ideas and information	---	106
4. Guessing meaning of unfamiliar words from context	-----	101
5. Understanding relations between parts of text by recognising indicators in discourse	-----	89
6. Understanding conceptual meaning	-----	80
7. Interpreting information presented in a non-linguistic form	-----	65
8. Understanding relations within the sentence	-----	61
9. Being able to understand parts of a text which are connected	-----	58
10. Reference skills	-----	26
11. Separating the essential from the non-essential in a text	-----	19

Writing activities

1. Expressing ideas, opinions	-----	251
2. Describing events, objects, graphs, pictures, processes	---	197
3. Summarising a passage	-----	135
4. Writing/replying to letters	-----	124
5. Presenting arguments	-----	69
6. Writing personal letters	-----	46
7. Reporting events	-----	45
8. Making comments	-----	29
9. Writing assumptions, speculations, clarifications	-----	26
10. Persuading	-----	2

11. Making complaints -----	2
Traits to be emphasised while teaching and testing writing	
1. Content -----	253
2. Language -----	206
3. Organisation -----	172
4. Vocabulary -----	179
5. Mechanics -----	90

Below we present the comparison between the exam specification and the questionnaire results.

Reading subskills

In the exam specification, we have allocated 10 marks to skimming and scanning, five marks to guessing meaning from context, four marks to inference, and four marks to organisation and summary. Comparing this list with the questionnaire results, it seems that they are more or less similar. Nevertheless, there are some points to be made.

One important point is that the college teachers emphasised the subskill "understanding ideas and information in a text not explicitly stated" more than it was emphasised in the exam specification. This skill, therefore, needs to be given more weighting than it was given before. The skill "guessing meaning from context" was found to be as important among the tertiary teachers as it was in the specification. So far as the skill "understanding explicitly stated ideas and information" is concerned, it needs to be included in the specification of the exam.

Another point is that the college teachers emphasised "understanding discourse markers" as one of the most important reading subskills. In our exam specification, we have included an

item for testing discourse markers (see item No.18). Taking this investigation into account, it seems that this subskill needs to be given more weighting than it was given before for the further improvement of the specification.

Subskills like "summary and organisation" which were included in the specification do not seem to be popular among tertiary teachers. It can be suggested that these skills should be taken away to leave spaces for the skills like understanding discourse markers, understanding explicitly stated ideas and information.

Writing activities

As was said earlier, we have not been explicit about the types of writing activities we were trying to include in the exam. The specification should be seen in terms of what has been described as the objectives of testing writing (ability to express ideas and opinions) as well as the test tasks presented in the specification. In the specification, we included two tasks: writing an essay on the basis of given outlines and replying to a letter. The whole idea was to assess students' abilities to express themselves in writing, rather than their abilities to write specialist pieces of writing. Looking at the results, we found that "expressing ideas, opinions" was the most important subskill in writing. Writing a letter also appeared to be important. Another point revealed by the investigation was that descriptive and argumentative writing subskills were also popular. These skills were not considered in the existing exam specification. It is, therefore, recommended that these subskills should be incorporated in the specification.

It should be appreciated that it is not possible to include more than two writing tasks in the SLC English exam.

However, if the tasks are designed carefully, the underlying ideas behind argumentative and descriptive writing can be encapsulated in the tasks we use. For example, a letter can be argumentative and an essay can be descriptive. And at the same time, these tasks can yield enough room to allow students to express themselves.

Traits to be emphasised while teaching and testing writing

Following different research findings (see section 4.2.7.2.) and Jacobs et al (1981), we decided, for the purpose of the present study, to develop a marking scheme emphasising content, organisation, vocabulary, language and mechanics as the important components of a composition. It was found that the way these traits were ranked in order of importance was more or less the same as they were ranked in the marking scheme. The only difference was that that the tertiary tutors appeared to emphasise the component "language" (i.e. grammatical correctness) more than we did. This result did not surprise us because, unlike us, the tertiary teachers were oriented to traditional methods of language teaching and learning. Unfortunately, they believed that grammatical correctness was the most important quality of a piece of writing. We do not see this as a problem because after the teachers are exposed to the communicative approach to teaching and testing they will begin to realise the importance of communicative value over grammatical correctness in a piece of writing. It does not deny the importance of grammar.

The overall impression of the validation study was that our exam specification and the marking scheme were valid to a reasonable extent. The specification and the marking scheme appear to have captured the most popular reading and writing abilities that the tertiary tutors think important for their

students. However, it does not mean that the specification and the marking scheme are perfect. They need to be improved. They can further be improved by taking insights from this investigation. For that purpose, the skills, for example, understanding discourse markers, and descriptive and argumentative writing can be incorporated in the specification.

5.2.7. Summary

The descriptive data obtained from the administration of the new exam, and two sets of questionnaires to different groups of students and people involved in SLC English were presented. The scores of the SLC English exam obtained by the students who had taken the new exam were also presented. The data was presented qualitatively and quantitatively so as to make it possible to refer to it when using descriptive statistics for testing our predictions.

5.3. Hypothesis Testing

5.3.0. Introduction

This section shows the results received from testing the two main hypotheses and the subsidiary hypotheses presented in section 4.1.3.. The research questions on which the present study is based have already been presented and the rationale behind them discussed.

Analysis of variance, comparison of means, and post-hoc Scheffe comparison of group means will be performed on the original data. The information obtained from inferential statistics will be used to test the main and subsidiary hypotheses. .05 level of significance has been set as evidence of chance factors.

5.3.1. Main Hypothesis 1 and Subsidiary Hypotheses (1a-1c)

Tested

Main hypotheses 1 (H1)

There would be a significant difference in performance between Grade 10 and SLC students on the new exam: The Grade 10 group would do significantly better than the SLC group.

Subsidiary hypotheses (H1)

- 1a. There would be a difference in performance between SLC group and Grade 10 group students from Type A schools: the Grade 10 group would do significantly better than the SLC group.
- 1b. There would be a difference in performance between SLC group and Grade 10 group students from Type B schools: the Grade 10 group would do significantly better than the SLC group.
- 1c. There would be a difference in performance between SLC group and Grade 10 group students from Type D schools: the Grade 10 group would do significantly better than the SLC group.

Main hypothesis 1, and the subsidiary hypotheses arose from the discussion presented in Chapter 1 and Chapter 2. We have argued that a final examination, like the SLC English exam, which is associated with some sort of extrinsic values cannot avoid influencing classroom teaching and learning. From the washback point of view, it has been argued that what is not tested can be ignored while teaching in the situation which we have described. In other words, it has been argued that if a final exam does not require the candidates to demonstrate the skills and the abilities the objectives of the course intend to develop in the

students, teaching of that course may fail to motivate the students to achieve them. In that case, teaching for that exam can have a harmful effect (i.e. negative washback) on learning.

We proposed to investigate to what extent the argument is true in the light of the Nepalese ELT situation. The objectives of the SLC English course have been presented. The SLC English exam for 1989 has been analysed. It has been shown that the SLC English exam fails to assess the language skills that the SLC English course intends to develop in the students.

We were primarily interested in finding out the effect of the preparation for the SLC English exam on the teaching and learning of the SLC English. We have argued that in most of the Nepalese schools teachers and students tend to spend most of their time on exam preparation especially with the students at the school leaving stage. So, whatever they do from the beginning of Grade 10 study until they sit for the SLC English exam will be considered preparation for the SLC English exam. It has been argued that the SLC English exam, because of its textbook and previous-exam-paper-orientated nature, does not encourage the students and teachers to focus on language skills entailed in the course objectives. In this situation, language learning may not take place after the students embark on Grade 10 studies. We proposed to discover whether there was any effect of the preparation for the SLC English exam on the teaching and learning of English at that level. It has been argued that if the SLC students do better than the Grade 10 students on the new exam, it would mean that the nature of their language learning process follows a natural path: the longer they study, the higher their proficiency. The issue as to whether or not their proficiency level is adequate for the purpose for which English is taught is

different. It would be natural for SLC students to do better than Grade 10 students because of their extra one year of training, provided that the situation is favourable for language learning. But if SLC students do not do better than Grade 10 students, one of the possible explanations would be that no learning whatsoever took place at the Grade 10 level, hence the negative washback.

Main hypothesis 1 was tested by carrying out a t-test on the descriptive statistics presented in Table 12.

Table 22: T-Test for the performance of Grade 10 and SLC

tobs	tcrit	df	Prob at .05
14.94	1.64	393	Sig.

On the basis of the above result, the alternative form of Main Hypothesis 1 was accepted: there was an effect of the preparation for the SLC English Exam on the teaching and learning of English. As was predicted, the Grade 10 students did significantly better than the SLC students. The superior performance of the Grade 10 group showed that the SLC group of students had not developed their language abilities during their Grade 10 studies; rather, they lost their previously achieved level of language proficiency.

To test the subsidiary hypotheses, we computed t-tests on the possible group means.

The following results were obtained.

Table 23: T-Tests for the comparison of group means obtained by Grade 10 and SLC groups from different schools

Types of Schools	tobs	tcrit	df	Prob at .05
Type A	3.83	1.67	47	Sig.
Type B	1.64	1.66	78	No Sig.
Type D	5.61	1.64	212	Sig

On the basis of these results, subsidiary hypotheses 1a, and 1c were accepted, and subsidiary hypothesis 1b was rejected.

The confirmation of main hypothesis 1 was further reinforced by the result obtained by readministering the new exam to the Grade 10 students at exit level (see section 5.2.1.2.). Although, as was explained above, the data is not adequate to draw any conclusion, it serves as supporting evidence for whatever we found in testing Main Hypothesis 1. We ran a matched t-test on the data (i.e. first and second administration of the exam to Grade 10 students from Type D school). The result is as follows:

Table 24: Matched T-Test for the comparison of group means obtained by Grade 10 students from Type D schools on the two administrations of the new exam

tobs	tcrit	df	Probability at .05
1.77	1.78	12	No sig.

There was no significant difference in the performance of Grade 10 students on the new exam at the entry and exit level. This reinforces the evidence for the argument that no learning of English took place after the students embarked on Grade 10.

studies.

5.3.2. Main Hypothesis 2 and Subsidiary Hypotheses (2a-2c) Tested

Main hypothesis 2 (H1)

There would be a difference in the performance of the students from different types of schools on the new exam.

The Subsidiary Hypotheses (2a, 2b, and 2c) (H1)

- 2a. There would be a difference in the performance of SLC students from Type A school and Type B school on the new exam: students from Type A schools would do significantly better than students from Type B schools.
- 2b. There would be a difference in the performance of SLC students from Type B school and Type C schools on the new exam: students from Type B schools would do significantly better than students from Type C schools.
- 2c. There would be a difference in the performance of SLC students from Type C schools and Type D schools on the new exam: students from Type C schools would do significantly better than students from Type D schools.

Main hypothesis 2 and subsidiary hypothesis 2a, 2b, and 2c arose from the discussion presented in Chapter 2. It has been said that in the Nepalese ELT situation, on top of the compulsory English, different schools offer different courses in English. On that ground it was predicted that schools would come up with different ranges of performance on the new exam because of their differences in terms of the courses offered and the way they prepared their students for the SLC English exam.

We were interested mainly in schools characterised as Type D, because these schools represent 85% of the total schools. The

purpose of including all types of schools in the present study was to attempt to account for the effect of the different types of teaching that took place in different types of schools. At the same time, it was assumed that this type of information would be helpful when making suggestions about changes in ELT. However, it does not mean that other types of schools do not need further improvement in their teaching.

It was predicted that the schools differed in terms of their teaching of English at the school leaving stage. To see the effect of the different types of teaching of English in different schools, a Oneway ANOVA was performed on the original data which was obtained from the performance of SLC students on the new exam. The result is given in Table 25.

Table 25: Anova for the difference between school types at SLC level

	df	sum of sq	mean sq.	F
Between groups	3	70614.63	23538.21	144.03*
Within groups	277	45265.93	163.41	
Total	280	115880.56		

*=Significant at $p < .05$

Table 25 shows that there was an effect of the 'school type variable' on the students' performance. The difference was significant at $p < .05$. On the basis of this result, the alternative form of main hypothesis 2 was accepted: different types of schools produced students with different language abilities.

Post-hoc Scheffe comparisons of all the possible groups were performed on the original data to test the subsidiary hypotheses. The following results were obtained:

Table 26: Post-hoc Scheffe comparison of the performance of SLC students according to type of school

	Type of School			
	D	C	B	A
Type D				
Type C	*			
Type B	*	*		
Type A	*	*	*	

* Indicates significant at $p < .05$

Table 26 shows that all groups significantly differed at $p < .05$. On the basis of the results, the three subsidiary hypotheses were accepted. In other words, every type of school appeared to be significantly different from the others in terms of the way teaching of English took place.

We were interested also in finding out whether the different types of schools were different at entry level for Grade 10 studies as well. It was assumed that the evidence of the effect of the 'school type variable' would further be reinforced if the performance of Grade 10 students from those schools supported the distinction previously made between the types of schools.

For that purpose, a Oneway ANOVA was performed on the original data obtained from the performance of Grade 10 to find out whether the schools were different at entry level for Grade 10 studies. Performing a Oneway ANOVA, the following results were

found:

Table 27: Oneway Anova for performance of Grade 10 students according to type of school

	df	sum of sq	mean sq.	Ratio p.
Between groups	2	18621.36	9310.68	160.30*
Within groups	111	6446.88	58.08	
Total	113	25068.25		

*=Significant at $p < .05$

Table 27 shows that the difference is significant at $p < 0.05$. This means that the schools were different even at entry level for Grade 10. Then we decided to carry out post-hoc Scheffe comparison of possible group means to find out whether all the types of schools were different at this stage. It was assumed that if all types of schools were significantly different at Grade 10 level, the confirmation of the subsidiary hypotheses would be reinforced.

We made post-hoc Scheffe tests on the original data. The results are described as follows:

Table 28: Post-hoc Scheffe comparison of the performance of Grade 10 students according to type of school

	Type of school		
	D	B	A
Type D			
Type B	*		
Type A	*		

* Indicates significant at $p < .05$

Following the results presented in Table 28, we found subsidiary hypotheses 2b and 2c reinforced but subsidiary hypothesis 2a was not reinforced. What this means is that Type A and Type B schools were not different at entry level for Grade 10 but became different at the exit level of Grade 10 studies.

5.3.3. Summary of the Hypotheses Tested

In this section, we tested our two main hypotheses, and six subsidiary hypotheses. The summary of which hypotheses were accepted and which were rejected is presented in Table 29. The meaning of their acceptance and rejection will be discussed in Chapter 6.

Table 29: Summary of the hypotheses tested

Hypothesis (H1)	Main	Subsidiary	P<.05	Results
	1		Sig.	Accepted
		1a	Sig.	Accepted
		1b	No sig.	Rejected
		1c	Sig.	Accepted
	2		Sig.	Accepted
		2a	Sig	Accepted
		2b	Sig	Accepted
		2c	Sig	Accepted

5.4.Exam Validation

5.4.1.Test Validation: A Rationale

Why validation? We have long had problems with the question of why we were validating the exam we used.

Validation of a test is time consuming and expensive. It is also a kind of study which is done as a continuous process in the direction of improving a test. For example, there have been some validation studies of the ELTS Test, and every time attempts have been made to improve it. A validation study of a language test also examines the test in the light of the recent trends in language teaching and testing. In that sense, it helps to update the test. The importance of validation of a test which can be used again can be seen because of its very reusability. In that sense improvement of a test can be beneficial because validating the test also helps to validate the test-tasks. On the other hand, validating an exam like the present one is difficult to justify. One wonders why it is necessary to validate the new exam since it is not the same exam that will be used in the real situation. The new exam is simply a sample of the exams which will be designed later.

Cronbach (1971:443) considers validation "the process of examining the accuracy of a specific prediction or inference from a test score". Following this definition, validation can be seen as checking the test in relation to what we intend to do with it. Does this mean that validation of a test is only an investigation of the prediction made on the basis of the scores obtained from a test or it is also an investigation of the procedures, followed in making predictions and examining them?

Following Cronbach, validating a test is to examine whether the test has done what we intended. In other words, it is

an examination of the prediction we make from our results. But at the same time, we all know that whatever decision has been made on the basis of the test, it cannot be changed. For example, if we validate the SLC English exam after we administer it, whatever the validation study reveals cannot change the decision: those who pass the exam will go on to the tertiary level and those who fail will not be permitted to do so.

Then, one would ask what are we interested in when we validate a test?

The simple answer is that when validating a test we are interested in examining how likely it is that our prediction would turn out to be true if we were to use the test again on a similar group. If we validate an exam which is used to make decisions about people, we will be providing some explanation of what the test can do and what it cannot do. If the test does not do what is expected, we will also look for a possible explanation of why it has not done the job we want it to do. This is how we can improve the test as to reduce the chances of making wrong decisions.

Looking at test validation from this point of view (i.e. descriptive), we find a validation study helpful for improving the exam if any fault is uncovered. The argument is that when we validate an exam (for example, the new exam), we do not do so assuming that we will be using the same exam again because the exam has to be different anyway, especially in the case of exit exams; otherwise there will be no secrecy. Then, the point is what is the use of validating an exam which is not going to be used again and where the decisions which are made on the basis of the old exam are not meant to be changed? All we can do is design a parallel test based on the insights obtained from the

validation.

Another argument is that if we say that the information we get from validation can be used for improving the exam, then the problem is that every exam is bound to be different, and the people who participate in the validation are bound to be different from the people who will actually be taking the exam which will be devised after the validation. Alderson (1988:2) also makes a similar point when he talks about a true score. He finds it difficult to say "what any given test is testing with any degree of certainty", because, he further argues, learners vary in the way they encounter test-tasks and the route they follow to reach a given response. On the whole, it appears that in language testing, we just make a rough estimation of language abilities of testees in terms of the tasks they are exposed to, and this is bound to include some sort of uncertainty. In the same way, validation of an exam cannot escape uncertainty.

We find, therefore, measuring language abilities and validating a test a matter of approximation rather than accuracy of predictions made and procedures followed in making them. We should look for possible benefit which can be obtained from the study. In other words, we should look for what we can do to improve an exam so that we can reduce the chances of making wrong decisions. Alderson (1988:220), for example, hopes to improve content validity of a test by validating the specification, which, he hopes, will improve the concurrent, predictive, and construct validity of the test. In that sense, for him validation includes validating procedures of testing, not merely the criterion. Applying this notion to our validation study, we see validation as to the procedures followed in the design and administration of the test as more important than validation of

the prediction we have made; it is the content validity, hence the washback validity, that is most important.

Another issue that a validation study has to deal with is the purpose of the test. In other words, we validate an exam from the point of view of whether it has done the job we want it to do. Examining a test from this point of view is a big task in itself. It is not possible to provide this type of information in all validation studies. All that we can do is provide some information (a sort of evidence) related to that or some sort of logic to support the argument. The argument is that when we talk about validation of a test, it is not always the case that we have enough evidence for claiming what the test can do, or/and what it can not do. In other words, we can have some evidence to say that the test is likely to serve the purpose for which we want it. To exemplify this, we would like to present the case of the present study.

We proposed in the study to design a test on the assumption that a parallel exam will be used to replace the existing SLC English exam so that the proposed exam would have a positive washback effect on teaching and learning. Then as a result, the students who pass it would have achieved the language skills which are described in the course objectives. Because this type of exam will lead the schools to teach according to the course objective, those who pass this exam will not have difficulty in coping with the tertiary courses which require a high level of English proficiency. They will have achieved the necessary language skills for the tertiary courses.

On the whole, the purpose behind the exam is to provide basic structures for designing exams which can work as catalysts to change the whole teaching-learning situation.

Now the question is how to examine an exam as to whether it can do this job or not. More precisely, how do we examine whether our exam can have that effect? and how do we find this information? The straight forward way to do this is to use the exam in the place of the SLC English exam and determine whether it works the way we want it to work. But it is not so simple. First of all, we cannot see whether the exam can have a positive washback effect until it replaces the SLC English exam. To see its real effect, therefore, we need to use it and see how it works at the real situation. At the same time, we have to present some evidence that it works for the purpose for which we want it in order to convince the people concerned that it is a valid test. People want to see convincing evidence for a positive washback effect of the exam.

In this situation, it is obvious that the real effect of the exam cannot be examined at this stage. All that we can do is present some sort of logical evidence that, if an exam is designed following certain procedures, it is most likely that the exam has a positive washback effect after it is introduced into the real situation.

On the basis of the above discussions, we would like to argue that in the case of an exam like ours, it is the procedures of test design, the specifications, and the formats which need to be validated. Providing evidence for the efficiency of the exam would mean that the procedures followed in the whole study have the potential to produce an exam as efficient as the present one.

5.4.2. Problems in Test Validation: Case of the New Exam

We encountered several problems in the validation of the procedures of the new exam. It has already been said that the real effect of the new exam cannot be investigated without

introducing it into the real situation. However, there were other things which could have been done in the direction of the validation had the situation been favourable. In this section, we present the problems we faced in the validation study.

One of the problems was the length of time needed for checking the predictive validity of the exam. We have said that the students who would do well in the new exam would also do well in the tertiary courses that require a high level of English proficiency. To examine the test against this criterion it would have been necessary to wait until those SLC passers who took the new exam completed their tertiary education. Then only could we have studied their performance in tertiary education in the light of their performance on the new exam. Doing this was impossible for us for two reasons: one is that for some socio-political reasons, the SLC Exam Board could not bring the SLC exam-results out within the usual time (e.g. 3-4 months after the examination) which caused a delay for the SLC passers in going on to tertiary level, and the other is that, for the same reason, the tertiary students had not taken the exam by the time of the writing of this section. We still do not know when they will sit for the exam. In this situation, it was impossible, for the purpose of the present study, to wait until the intended students complete their tertiary level so that the predictive validity of the exam could be determined.

Taking this situation into account, we decided to establish the predictive validity of the new exam using teachers' judgment on the students language proficiency. It has been a tradition to use teachers judgement for predictive validation of a test. For example, we can note the ELTS validation study carried out by Criper and Davies (1988) and Moller's (1982).

study.

Another problem was that some schools were scared of letting their students be used for this study. Colleagues in Nepal wrote to us saying that all the schools, which allowed us to use their students the first time, except type D, rejected our request the second time. We proposed to administer the test a second time in December 1989 to those Grade 10 students who took the new exam in February 1989 and were preparing for the SLC English exam in 1990, so that we could ascertain the effect of the SLC preparation on their learning of English in a 10 month gap. Unfortunately, we could not do that in Type A and B schools.

We encountered a problem of money and time as well. Since we had to collect data from far away, we had to depend on the information given by the Headmasters or college teachers and colleagues, instead of collecting the necessary information ourselves. It would obviously have been better if we could have ourselves collected the necessary information.

Another problem arose from the nature of the study. We proposed to study the predictive validity of the exam. We found it difficult to find the whereabouts of those SLC passers who had also taken our exam and had been doing tertiary education. So, the number of students who we contacted is small.

Despite the problems we managed to collect some relevant information on the validation of the new examination which we report in the following sections.

5.4.3. Validation of the New Exam

5.4.3.0. Introduction

In this section, we report the results we have received from the validation of our study. We present first the practicality of the exam in terms of its use in the Nepalese

situation especially in relation to its applicability in the place of the SLC English exam. We then report the item statistics of different subtests. Throughout the whole study, validity has been the main concern. We have analysed the exam against face, content, predictive, and construct validity criteria.

In this section, we also examine our prediction that the SLC students would find the new exam more difficult than the SLC English exam. This examination will also provide some evidence for the construct and content validity of the new exam.

5.4.3.1. Traits Tested

To recapitulate what the exam was intended to measure, we present, below, the traits proposed to be tested by the exam.

Subtests	Items	Traits tested	Item type
Reading	1-3	Skimming	Multiple choice
	4-10	Scanning	" "
	11-15	Guessing Meaning from context	Supplying an acceptable word
			m/c
	16-19	Inference	m/c
20	Summary & organisation	m/c	
Cloze	21-34	Linguistic competence	Cloze procedure with 6th word deletion rate
Note-taking		Ability to take a note from a given passage	Reading a given passage and taking notes from it
Writing		Ability to express ideas and feelings	Replying to a letter
		Ability to express ideas and feelings	writing an essay

5.4.3.2. Practicality

How the exam worked in practice has already been reported in section 4.3.. Throughout the whole study, we have considered the practicality of the exam very important mainly because if the exam does not fit in the real situation, it would not be worth suggesting that an exam similar to the new exam should replace the existing SLC English exam.

As was reported in section 4.3.3., the exam is satisfactory for the Nepalese examination-situation. From the point of view of the length of the exam, we observed that most of the students from different types of schools completed the exam in between two 2 and 1/2 - 3 hours. There were some students who completed the exam in between 1 - 1 and 1/2 hours, but almost all of them were from Type A schools. We also found some students unable to complete the exam even after three hours. Our explanation of this was that they found the exam very difficult, not that they did not have enough time to do the tasks. Looking at the average student's performance, we came to the conclusion that the exam as a whole stands as a task for three hours for an SLC candidate.

The exam is practicable from an administrative point of view too. It placed no more demand than the SLC English exam on the people involved in administering the SLC Examinations in terms of invigilators and seat-planning. It also does not involve an extra amount of money to get the exam printed and answer-sheets marked.

The markers' reliability in the writing section of the exam will be reported in the section which follow. It was said earlier that double marking is not possible in the SLC English exam, at least in the present situation. This does not seem to be

a problem with the new exam in the sense that the exam has been found to be reasonably reliable even with a single rater. What is important, of course, is that raters need training.

5.4.3.3. Item Statistics

To find out how difficult/easy the intended group of students found the exam, items of reading and cloze, and notes, letters, and essays have been analysed using the total sample who took the exam as SLC candidates. The notes, letters, and essays have been analysed by categorising them as good, fair, and poor.

5.4.3.4. Item Analysis: Facility Values: Reading and Cloze

On the basis of the results obtained from item analysis for reading and cloze tests, we have established the facility values for their items. The facility values for the reading and cloze items are summarised below.

Table 30: Item Analysis: Facility Value

FV	Reading Test items	Cloze Test Items
70-72	4	-
60-69	1,2,7	-
50-59	17,16,19	21,28,42,23
40-49	6,18,8,5,9	32,36,37,44,39,34
30-39	3,10,12,15,13,14	31,54,26,30,35,38,43, 49,24,50,53,41,33,46,48
20-29	11	51,25,27,47,45,29,40, 52,22
N	19	34
Mean F V	47	35
S.D.	12.54	9.20

The items fall within the range of 72 to 20. Oller (1979)) holds the view that items within facility values which fall between 15 to 85 can be accepted. According to this, all the items of the reading subtest appear to have worked at an acceptable level.

5.4.3.5. Comments on Reading Tests

It would be interesting to investigate what made one item more difficult than others. A detailed analysis of this kind is beyond the capacity of this study. However, a brief discussion of the items by grouping them into different ranges of facility values will throw light on what factors contributed to making some items easier. In other words, if items which fall within a range follow a pattern, it will provide a basis for an explanation of why some items were found difficult and some easy.

For the purpose of analysis, we grouped the items of reading and cloze tests into easy, moderate, and difficult categories.

Easy items: The items which were answered correctly by nearly two thirds (facility value 60 and above) of the students were considered to be easy.

Moderate: The items which were answered correctly by nearly fifty percent (facility value 40-59) of the students were considered to be moderate.

Difficult: The items which were answered by nearly one third (facility value 39 and below) of the students were considered to be difficult.

Considering the reading items under this framework, items 4, 1, 2, and 7 were found to be easy. Items 17, 16, 19, 6, 18, 8, 5, 9, 3, 10, 12, 15, 13, 14, were found to be moderate, and item

11 was difficult.

Matching the easy items against the given context, a pattern emerges: the items which have a direct bearing (e.g. verbatim and transformed verbatim types) on the passage were found to be easy. For example, the word 'reward' in item No. 4 would have given the testees a clue to pick up 'dollar 25,000 prize'; the phrase 'first flight across the Atlantic' in item No. 1 has direct bearing on the phrase 'the first man to fly the Atlantic'; Charles Lindbergh in item 2 would have been answered on the basis of the knowledge required for item 1, item 7 has the word 'far' which would have given a clue to pick up '3,600 miles'. Another way of interpreting them is that answering these items may not have required the testees to have understood the whole passage.

Examining the moderate items, items 17, 16, 19, and 18 were intended to assess the ability of the students to understand the meaning of a passage not explicitly stated. In that case, these items required the examinees to understand the implicit meaning associated with the items. In case of item 5, students seemed to have become confused with the word 'time' because most students answered the item by writing 'May 20th, 1927' instead of 'ten to eight'. Items 6, 8, 9, and 10 do not have direct bearing on the passage: the testees have to move their eyes forward and backward to locate the right answers to the questions. In short, the items which require the examinees to look for the meanings in the passage which are not explicitly stated were found to be moderate.

To answer items 11, 12, 13, 14, and 15, since these items were intended 'to assess the ability of the students to guess the meaning of a word from the given context', the testees were

required to understand the given passage. In other words, there was not an explicit cue for any item. Item No 3 does not have an explicit clue. On the whole, it appears that items which required the testees to understand paragraphs or the whole passage, and supply an appropriate word or expression were found to be difficult.

On the basis of this analysis, what appears to be emerging is that the items which require an understanding of a paragraph or more than a sentence level were be more difficult than the items which have some sort of explicit cues and which do not require examinees to understand a combination of several sentences. What also appears, on the basis of this analysis, is that the completion type of items are more difficult than multiple choice items. One of the possible explanations of this pattern can be that not all the distractors of the multiple choice items are attractive.

Item No. 20 has not been included in the above analysis. The reason is that this item was difficult to analyse statistically.

There are 9 statements in the item. Students are to select four of them and put them in an order so that the four statements make a summary of the given passage. The item was intended to assess study strategies-'summarising' and 'organising'. Out of 2 marks allocated to the item, 0.5 was allocated to each sub-item. The item was included for the backwash purpose of the exam.

On the whole, no student answered all sub-items correctly. In that sense, it was the most difficult item in the whole test. However, we were not quite sure whether it was because they could not understand the passage, or because the

item was badly designed. All that we can say at this stage is that the item is important from an washback point of view, but it needs to be improved.

5.4.3.6. Response Frequency Distributions

A sample of 150 answer-sheets representing upper group, middle group, and lower group, was chosen to analyse the multiple choice items- 1, 2, 3, 16, 17, 18, 19. The following result was found from the analysis of the response frequency distributions of the distractors of the items:

Table 31: The percentage of responses attracted by each distractor

No of item	A	B	C	D	Correct answer
1	6	10	80	0	C
2	12	6	75	3	C
3	45	15	3	36	A
16	8	62	16	12	B
17	70	6	17	2	A
18	53	22	15	5	A
19	38	20	10	48	D

According to this table, the distractors of items 1 and 2, were not attractive. In item 1, distractor D did not attract anyone, and distractors A and B were chosen only by 16 candidates. In item 2, distractors B and D were not attractive. In the same way, distractor C in item 3, A in item 16, and D in 17 and 18 were unattractive. The distributions of response frequency of the distractors in item 19 appear to be reasonably good. The distractors which were not attractive apparently increased the chances of getting right answers to

these items. In other words, there was a great chance of answering these items correctly.

5.4.3.7. Comments on Cloze Test

It has already been said that items in cloze procedures were marked accepting any acceptable word as correct. Items were analysed and the following figures were found.

Looking at the discrimination index of the items, we find that the items fall within a range of .82 to .41. As was discussed above, accepting .35 as a valid index of item discrimination, every item is valid.

The facility values of the items fall within the range of 56 to 20 which means that they are all acceptable. Here again, for the purpose of throwing light on why some items were found to be more difficult than others, items can further be grouped according to easy, moderate, and difficult. In the cloze test, no item was answered correctly by two third of the students (i.e. no item was categorised as an easy item).

Items 21, 28, 42, 23, 32, 36, 37, 44, 39, 34, 31, 54 were moderate. The rest of the items were difficult. It was impossible to carry out a detailed analysis of these items from the point of view of what factors would have made some items more difficult than the others. However, we would like to discuss briefly these items in terms of the contextual clues available in the passage, and the way they were marked so that an explanation can be given as to why some items were difficult and some easy.

The expected answers to the moderate items were- 'going' (21), 'I' (28), 'the' (42), 'I' (23), 'the' 32, 'I' (36, 'madman' (37), 'the' (44), 'her' (39), 'the' (34), and 'did' (54). It would appear that most of these moderate items are ones which require 'I' and 'the' choices. In the case of the items which require

choices like 'going' and 'madman', we accepted choices like 'coming', 'passing', and 'thief', 'badman', 'man', 'woman' as correct answers. We found it difficult to explain why items which expect 'her' as a correct answer were found to be difficult. In general, it seems that the candidates could handle definite articles, personal pronouns, and proper nouns to some extent.

Items which require main verbs and prepositions turned out to be difficult. However, this does not imply that the context provided by the given sentence did not have any effect on making items difficult or easy. A detailed analysis of the whole context would provide an explanation but this is beyond the limit of the present study.

5.4.3.8. Test Analysis: Difficult/Easy: Note-taking, Letter Writing, and Essay Writing

The Notes, Letters, and Essays written by all the students from SLC group were analysed by grouping them as follow:

Good: The notes, letters, and essays are categorised 'good' if they are worthy of 60% and above of the total marks allocated to the task.

Fair: The notes, letters, and essays are categorised 'fair' if they are worthy of 32% to 59% of the total marks allocated to the task.

Poor: The notes, letters, and essays are categorised 'poor' if they are worthy of below 32% of the total marks allocated to the task.

The results are summarised in Table 32.

Table 32: Percentage of notes, letters, and essays categorised as good, fair, and poor

Tasks	Good	Fair	Poor	Unattempted
Note	9	33	55	3
Letter	13	16	61	10
Essay	10	14	62	14
Total	11%	21%	59%	9%

On the whole these tasks were difficult for the students. Among the tasks, essay writing seems to have been the most difficult, followed by letter writing. What this means is that writing tasks were more difficult than note-taking tasks.

5.4.4. Validity

5.4.4.0 Introduction

Validity has been the main concern of the exam. Due to changes in the Nepalese circumstances, the idea of establishing the predictive validity of the exam became impossible.

In this section, we report the face and content validity and the statistical evidence of the construct and concurrent validity of the new exam. The basic sources of information about the construct and concurrent validity of the exam were the relationships of the new exam with the SLC English exam, Cambridge O' Level Language test, College teachers' judgement of students' English proficiency, and the interrelationships between subtests.

5.4.4.1. Face and Content Validity

For a language test it is important that the testees accept that the test is valid, otherwise their performance may

not be without suspicion. To gather information about how the SLC candidates accepted the new exam, we asked them to make a comment on the new exam after they had completed it. Mention has already been made that the SLC candidates who took the new exam had just taken the SLC English exam that year. They still had a fresh memory of the demand the SLC English exam made on them. So it was hoped that they could compare the two exams in terms of the demands the two exams placed on them.

They were also asked to make comments on the new exam in the light of the questions: which of the two exams would you like to take as your school exit examination? Which of the exams do you think measures your true language abilities?

We wanted their honest responses. Considering their language problem we, therefore, allowed them to make comments on the exams in Nepali or English. Most of the students from Type D schools wrote their comments in Nepali.

Their responses indicated that they prefer the new exam to the SLC English exam. Not a single student gave his/her opinion in favour of the SLC English exam. Most of them seemed to have liked the new exam in comparison to the SLC English exam, finding the SLC English exam mechanical and based on rote-learning. They said that they liked the new exam because they found in it the types of tasks that required them to think and express their views. For example, one student wrote "I think, this type of exam (the new exam) is the appropriate for the SLC candidates in the sense that we can express for our opinion and not imagging [imagining perhaps] the text book". Several other students wrote that they wanted the new exam, not the existing SLC English exam type which encouraged rote-learning and cheating.

During the administration of the exam and the questionnaires we also asked school teachers to comment on the new exam. Almost all the people we met gave us the impression that the new exam was acceptable as an exit English exam at SLC level. From the information obtained from the questionnaires, it appears that the exam has a high face validity which in Davies' (1988a:30) view is helpful in convincing people that the test is valid.

Considerable attention was paid to the a priori validation of the new exam (Chapter 3). To get an impression about the content validity of the exam we asked a group of experts and English teachers in Nepal for their reaction to the content of the exam in relation to its specifications. According to them, the exam has high content validity.

5.4.4.2. Construct and Concurrent Validity

5.4.4.2.0. Introduction

In this section, we report evidence of the construct validity of the new exam established by following some statistical and logical procedures.

The statistical evidence of the construct and concurrent validity of the exam was determined by establishing interrelationships between subtests, and the relationships of the new exam with the SLC English exam and Cambridge O' Level. Principal Component Analysis was also performed.

Logical evidence of the construct validity of the exam was established by examining test-items in relation to the performance of different groups of SLC students: e.g. have the test-items required the testees to carry out the activities which the items intended?

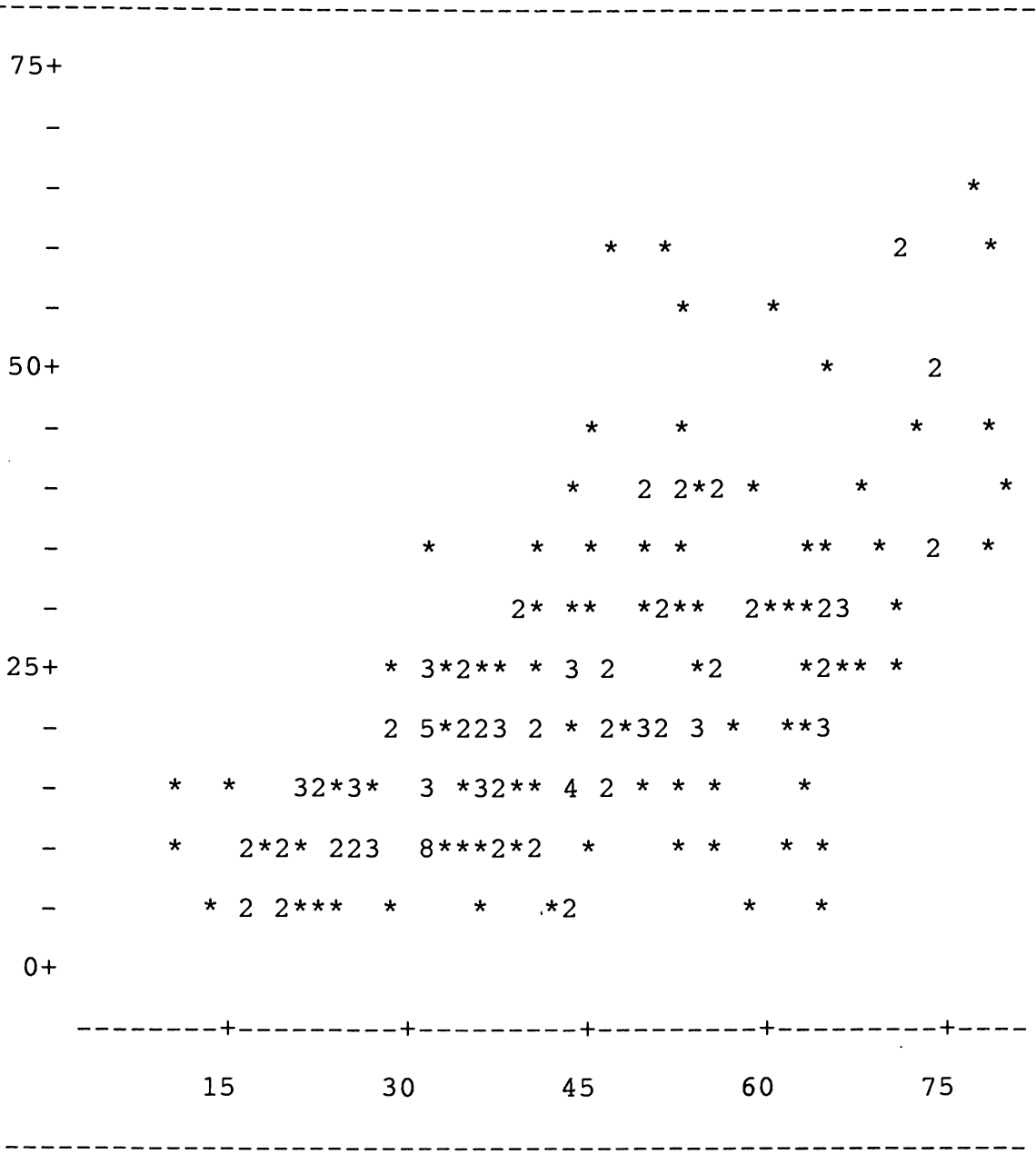
5.4.4.2.1. Relationship Between the New Exam and the SLC

English Exam

The correlations between the new exam and the SLC English exam as a whole, and the breaking down of the results into school-types, were computed. The correlation coefficient of the relationship between the scores in the two exams was found to be $r.72$. This correlation was not beyond our expectation since both exams are language exams; they must measure something in common.

Figure 7 presents the relationship between the two exam scores graphically.

Figure 7: Scattergram of the scores in the new exam and the SLC English exam



The correlation between the two exams was also established by breaking down the scores according to the types of schools. The following relationships were obtained:

Table 33: Correlations between the SLC English exam and the new exam according to school type

Type of school	Correlation
Type A	-0.12
Type B	0.73
Type C	0.62
type D	0.63

This table shows that the performance of the students from Type A schools on the two exams correlated negatively, and the students from Types B, C, and D schools correlated moderately. It was difficult to explain the negative correlation of the performance of the students from Type A schools on the two exams. One of the possible explanations is the low range of the distributions of the scores (sd=6.02 in the new exam, and 8.65 in the SLC, whereas in other schools, standard deviations of the scores were much greater).

Looking at the relationship between the exams from the point of view of the cutoff points presented in section 5.1.4., we came up with another result. Referring back to Table 9, we find that students from Type A schools had performed equally well on both exams: 96% of the students from Type A reached the same cutoff point in both exams. From this point of view, the correlation between the two exams would have been very high if there was no effect of the distributions of the scores. On the other hand, the students from Type C and D performed well in the SLC English exam but badly in the new exam. In this case, one would expect the relationships between the scores in the two exams to be low but as we can see, the correlations are moderate. One explanation is, as the scattergram shows, that the relationship between the two exams is not linear.

5.4.4.2.2. Unique Variation Between the New Exam and the SLC English Exam

Spearman (1904) (cited in Horst 1966:312) gives a formula for estimating 'true variation between two measures'. The formula is called "coefficient of correlation corrected for attenuation". The assumption is that the correlation estimated by using the formula is "the ratio of the observed correlation between two measures divided by the geometric mean of their estimated reliabilities coefficient" (Horst 1966:312).

We used this formula to estimate the true variation of the SLC English exam, and the new exam. If we look at section 5.1.4.1., we find the overall reliability of the exam as $r = .82$. Though we do not have a measure of the reliability of the SLC English exam, as a traditional discrete point test with objective marking, and reasonable length, we have good reasons to assume that it is as reliable as the new exam. The correlation

coefficient of the SLC English exam and the new exam was estimated $r = .72$.

Calculating the correlation coefficient corrected for attenuation of the two exams by using Spearman (1904), we found $r = .77$. In other words, $3/4$ of their true variation is common, and nearly $1/4$ is unique to each.

5.4.4.2.3. Relationships between New Exam, SLC English Exam, and Cambridge O' Level Language Test

It has been possible, though the data is not adequate, to establish the relationships between the new exam, SLC English exam, and the Cambridge O' Level language test. A sample of the Cambridge O' Level language test is given in appendix XI. As was mentioned earlier, the SLC students from Budhanilkantha school (one of the schools under Type A) were also about to take the Cambridge O' Level language test at the time when the new exam was administered. They had already taken the SLC English exam. To see the relationships between the three exams, we decided to use their scores on the three exams. The scores of the students in the Cambridge exam were collected from the school.

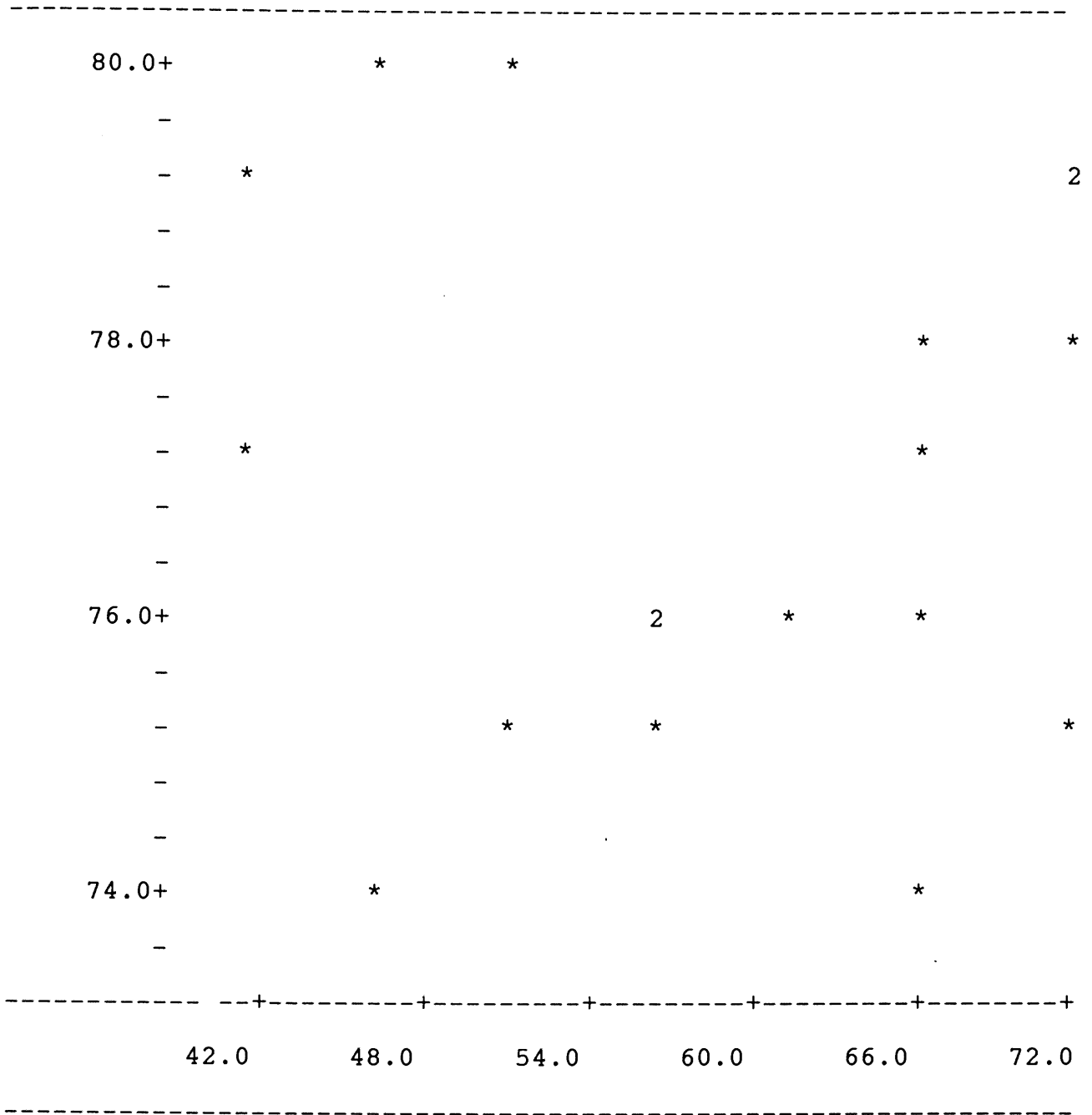
Computing the correlation coefficient between the exams the following results were obtained. Table 34 presents the coefficient of their correlations.

Table 34: Correlation between New exam, SLC English exam, and
Cambridge O' Level language test

	New exam	SLC English exam
Cambridge O' Level	-0.06	.15

The results were quite different from what we expected. A high correlation was anticipated because the new exam and the Cambridge O' Level were based on more or less on the same principles of test construction. A weak correlation was anticipated between the SLC and the Cambridge O' Level for the same reason. It was difficult to explain a negative correlation between the new exam and the Cambridge O' level. As can be seen, the sample has most of the factors responsible for affecting correlation, such as a restricted range of score as the sd of the new exam was 1.97, an extreme group of students in the sense that this group of students was the most proficient group in the sample. Then we decided to look at the data by plotting it in a scattergram.

Figure 8: Scattergram of the scores in the new exam and the Cambridge O' Level language test

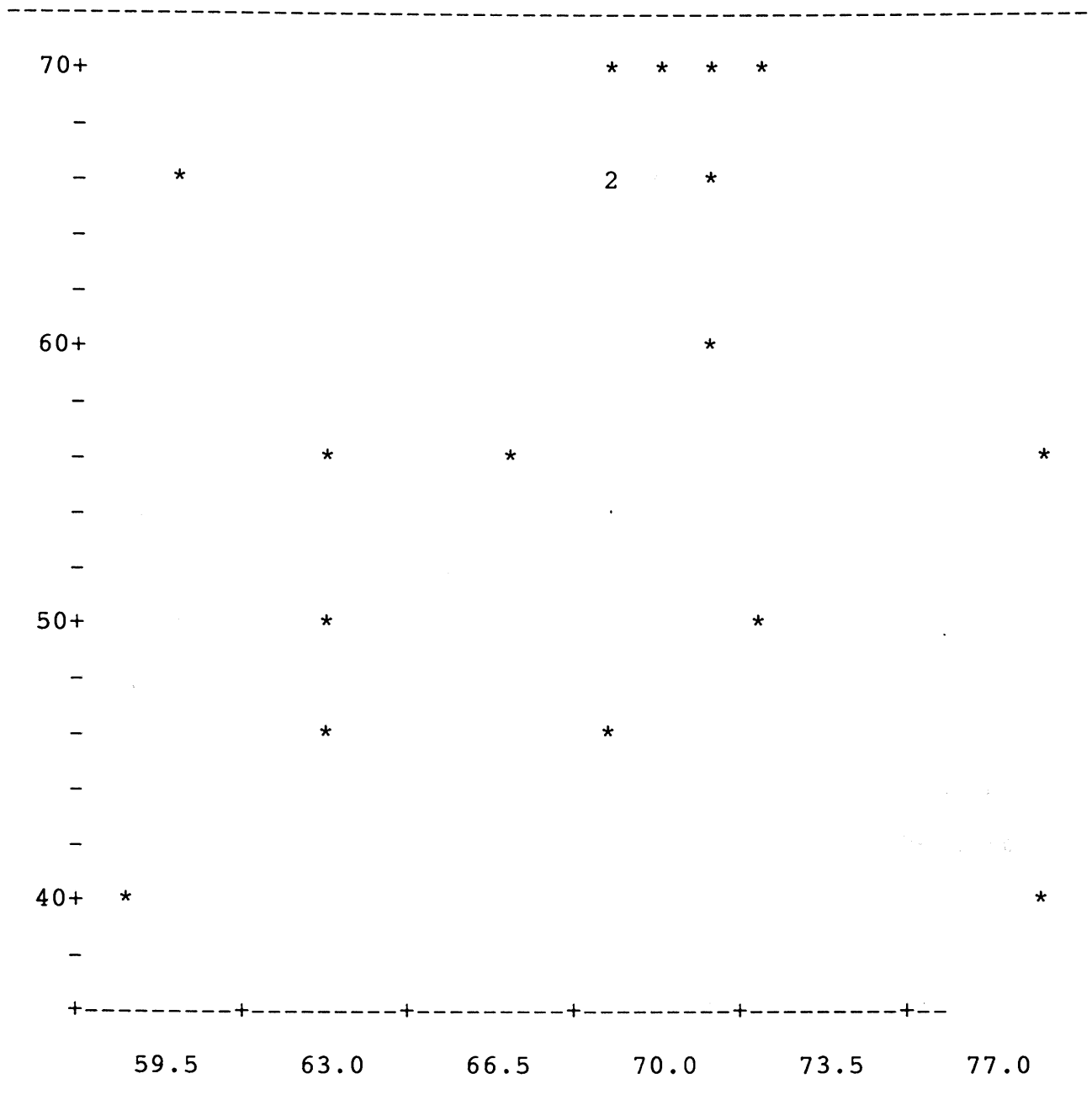


As can be seen, the overall picture is linear. The weak relation may be the result of the outliers which may not belong to the population. Though the data is quite small and we do not have any logical reason to remove the outliers, it is still important to see the relation established in absence of those outliers. We thus computed the correlation between the two sets of scores removing the four outliers. The relationship was found

to be $r.64$. This gives an impression that doing such a comparison with a reasonable sample would yield a strong relation between the two exams.

We also tried to see the picture of the correlation between the SLC English exam and the Cambridge O' Level. The scores on the two exams were plotted in a scattergram.

Figure 9: Scattergram of the scores in the SLC English exam and the Cambridge O' level language test



This picture gives the impression that there is no systematic relation between the two exams. Nothing can be predicted on the basis of the results. Though the students were taught for both exams by the same teachers, and in the same environment, the relationship between the two exams appears to be difficult to establish.

5.4.4.2.4. Relationships between New Exam, SLC English

Exam, and College English teachers' Judgement of their Students' English

The difficulty in carrying out a follow up study on how the students who passed the SLC exam and went on to college were doing in the tertiary courses has already been pointed out. Though their performance on the courses which require English at tertiary level was set as criterion for the predictive validity of the exam, collecting this type of information turned out to be impossible for practical reasons. In difficult circumstances, we requested some college teachers, representing both English and Nepali medium colleges, to rate the English proficiency of our students by filling in a questionnaire (a modified version of a questionnaire used in Criper and Davies (1988)) in the light of the adequacy of their proficiency for the college courses. A sample of the questionnaire is given in appendix IX. 46 questionnaires were returned to us which were only from Nepali medium colleges. The results obtained by correlating the three measures are presented.

Table 35: Correlationships between New exam, SLC English Exam, and Colege Teachers' Judgement on their Students' English

	New Exam	SLC English Exam
College Teachers' Rating	.07	.67

The correlation coefficients between the three sets of scores show that college teachers' rating correlated moderately with the SLC English exam, but weakly with the new exam. Here again, we encountered a problem in data collection procedures. One possible explanation for the low relationship of college teachers' rating with the new exam, and moderate relationship with the SLC English exam could be that despite the fact that teachers were instructed not to take into account any previous assessment of the students' proficiency while rating, they appear to have been highly influenced by the SLC scores. This was apparent as the SLC scores for each candidate appeared on the top of the questionnaires, and from the pattern they followed in filling them in.

According to our informants, the teachers found it difficult to rate the proficiency of the students because of the short time period since the students had entered the colleges. It is also likely that in the Nepalese situation teachers do not know all the students and their proficiency levels until they assess them.

It was difficult to interpret the results because when comparing the college students' SLC scores and their performance on the college first year, the correlation was very low. It has been shown that SLC scores have no predictive validity. The low correlation between the teachers' judgement on the

proficiency of the students and their performance on the new exam was considered to be the result of a methodological problem.

5.4.4.2.5. Relationships Between Subtests

Intercorrelations between the subtests were computed and the following results were found:

Table 36: Correlations between subtests

	READING	CLOZE	NOTE	WRITING
READING		.78	.59	.74
CLOZE			.60	.78
NOTE				.63
WRITING				

The correlations between the subtests ranged from .59 to .78. All the correlations were positive and highly significant. The high correlation among themselves show that they all measured something similar. In that case, it would appear that, from the construct point of view, the exam measured general language proficiency. However, the range of difference appears to be .19. This indicates that each subtest contributed something unique to the total measurement.

Reading, cloze, and writing have high correlations between themselves. Note-taking appears to have relatively lower relationships with them.

5.4.4.2.6. Principal Component Analysis

Bachman and Palmer (1981:138) find principal component analysis a major procedure to be used for construct validation of a test. Cureton (1983:297) also prefers this procedure for finding the contribution of a variable of a test to the total measurement. In simple words, PCA has been an effective

statistical method for determining distinct abilities of exam measures.

Our correlational studies of the new exam showed that the different subtests were measuring something similar. From the washback point of view, every subtest is important. There is no point in reducing the number of subtests whether they measure the same construct or not. From the very beginning, we have been arguing that the main purpose of the exam is to exert influence on teaching and learning by requiring the teachers and students to focus on the different skills which are important for tertiary education. However, we felt it important to validate the new exam also from a construct point of view. We were keen to find out whether the subtests which show a range of .19 unique contribution to the total measurement measured the same ability or different abilities. High correlations between scores on the four subtests gave us the impression that they were not measuring four distinct abilities but we were not sure how many abilities - one or two or three - were being measured.

For that purpose, we performed principle component analysis using the SPSS package. The results are given in appendix X.

Hatch and Farhady (1982:281) suggest that it is rotated Factor matrix that has to be interpreted. However, Woods et al. (1986) and Cureton (1983) point out that, considering the difficulty in interpreting results obtained from PCA, it is safer to adopt different methods to interpret PCA results. Their argument is that if different methods give the same results, we can be more confident about accepting the salient components measured by a test.

The summary of the final statistics obtained by performing PCA on the original data is given in Table 36

Table 37: Principal components and eigenvalues for four subtests by extracting four components (n=281)

Component Matrix				
	1	2	3	4
Reading	.59950	.46248	.44592	.03214
Cloze	.71226	.48871	.27978	-.01227
Note-taking	.34250	.56669	.28261	-.02922
Writing	.55632	.66486	.21879	.08907

Eig.value	2.81607	.06794	.03337	.00688
PCT of Var	70.4	1.7	.8	.2
CUM PCT	70.4	72.1	72.9	100.0

Table 37 shows that the reading, cloze, and writing subtests have high loadings, and the subtest note-taking has a low loading on component 1; on component 2, all variables have high loadings; and only the subtest reading has a high loading on component 3. No variable has a high loading on component 4.

We found the results difficult to interpret so component 4 was discarded, and PCA was again performed on the original data extracting 3 components only. We came up with another result which is summarised in Table 38.

Table 38: Principal components and Eigenvalues for four subtests by extracting three components (n=281)

Component Matrix			

	1	2	3
Reading	.86412	-.12472	.08772
Cloze	.89082	-.10425	-.06836
Note-taking	.69876	.15237	.08273
Writing	.88239	.10672	-.08240

Eig.value	2.80714	.06103	.02600
PCT of Var	70.2	1.5	.7
CUM PCT	70.2	71.7	72.4

Table 38 presents a different picture from the one in Table 37. Table 38 shows that all the variables have high loadings on component 1, and very low loadings on components 2 and 3. On component 2, Reading and Cloze contrast with note-taking and writing, and on component 3, cloze and writing contrast with reading and note-taking.

Looking at the eigenvalues, component 1 has 2.80714, 2 has .06103, and 3 has 02600. As can be seen, of the total variance, components 1,2,3, account for 70%, 1.5%, and 7% respectively. Interpreting the results in terms of loadings of the variables on each component, we find that only component 1 is salient: all the variables have high loadings on it. Components 2 and 3 have very low loadings from all the variables. If we interpret the results in terms of the eigenvalues we find the same results.

Kaiser (cited in Cureton 1986:161) recommends retaining as

salient factors those corresponding to eigenvalues greater than unity, and discarding all those with lower eigenvalues. The caution is that it is dependent on the number of variables very much, especially when the number of variables is very small or very large. That is why, for Cureton (1983:161), the critical eigen value varies according to the number of variables.

Another way of solving this problem is to follow Woods et al (1986:283) who suggest a rule of thumb. The rule is: "if the original data has p dimensions, assume that components which account for less than a fraction $1/p$ of the total variance should be discarded" (p.283).

In this case, following Woods et al (1986), we have 4 variables: i.e. the subtests-reading, cloze, note-taking, and writing. The critical eigenvalue will be .25 ($1/4$). Looking at Table 38, we find only component 1 which has eigenvalue greater than .25. That means only component 1 is salient.

It appears on the basis of the analysis that the total test measured only one distinct ability. In a wider sense, it can be interpreted as a measurement of general language ability in English. It can generally be accepted that students who have higher general ability in English will perform better than students who have lower general ability in English. Nonetheless, if we look at Table 38, we find that the first component accounts for only 70% of the total variance. In other words, 30% of the total variance is left unexplained if we accept only component 1 as a salient component. It would not be enough to ascribe 30% of the variance to error variance. Something is missing here, but we find it difficult to explain. This result led us to agree with Bachman and Palmer (1981) who say that PCA is biased towards general language proficiency.

5.4.5. Is the Existing ELT at School in Nepal Adequate?

5.4.5.0. Introduction

This section addresses the question of whether or not the existing ELT at school in Nepal is adequate to serve the purpose for which English has been introduced. It is hoped that the discussion will also provide logical evidence of the construct validity of the exam.

In section 4.1.2., we said that our question, (i.e. Is the existing ELT at school adequate?) would be addressed by performing a quantitative analysis of the performance of the SLC students on the new exam. The intention behind the quantitative analysis is to determine an index of language proficiency of the SLC students in terms of their performance on the new exam. Then the index of language proficiency can be used to determine a minimum level of language proficiency required in order to pass the SLC English exam. We felt it necessary to establish a minimum level of language proficiency for an SLC passer in order to compare that level of proficiency with the purpose for which English is taught at school. We hope to investigate whether or not a SLC passer as a college entrant goes on to college level with adequate English.

5.4.5.1. Meaning of Passmark in New Exam: An Index of Language Proficiency

To determine the criterial level of performance in the new exam, we decided to examine the performance of the students who reached cutoff 32-44 in the new exam. Table 13 shows that 16% of the total sample of the SLC students fall under this cutoff.

Items of Reading and Cloze subtests were analysed using only the performance of the students who scored cutoff 32-44 in the new exam. The figures are given below.

Table 39: Item Analysis: Facility Value

Facility Value	Reading items	Cloze items
90-	7	-
80-89	6	28
70-79	4	21, 23, 36, 39
60-69	1, 9, 2, 3, 19	34, 42, 32, 44, 37
50-59	18, 5, 17	31
40-49	16, 8, 15,	24, 30, 53, 38, 33, 47, 35, 26,
30-39	14	27, 48, 49
20-29	13, 12	46, 40, 43, 51, 50, 41
10-19	11	25, 22, 45, 52
0-9	-	29

Comments

The average score obtained by this group of students in this subtest was 47% of the total marks allocated to the task. That means, the performance of this group on this task was above cutoff 32-44. Looking at the facility values of the items, it can be understood that more than two-thirds of the students had correctly answered item Nos. 7, 6,, 4, 1, 10, 9, 2, 3, 19, nearly fifty percent of the students had correctly answered item Nos. 18, 5, 17, 16, 8, and nearly one third of the students had correctly answered item Nos. 8, 15, 14, 13, and 12. No. 11 was answered correctly by a small percentage of the students. Putting it in terms of language ability determined by analysing items the students appeared to have achieved reading abilities - skimming, scanning, guessing meaning from context, inference - to

an acceptable level of proficiency.

Cloze task

The overall performance of this group of students on the cloze task was 35%, which is also above the minimum pass percentage (i.e. 32%). Analysing the cloze items using only this group of students, we found the following figures.

Comments

Cloze items were found to be more difficult than the items of the reading subtest for the whole group of students: i.e. no item was easy according to our categorisation. It is not surprising then that this group of students had found the cloze items difficult. Of the cloze-items established on the basis of this group of students, items 28, 21, 23, 36, 39, 34, 42, 32, and 31 were correctly answered by nearly two thirds of the students; 24, 30, 53, 38, 33, 54, answered correctly by fifty percent; items 47, 35, 26, 27, 48, 49, 46, 40, 43, 51 answered correctly by nearly one-third; and 50, 41, 25, 22, 45, 52, and 29 by a few of them. Matching this result with the facility values established for the total sample, it would appear that this group of students had done reasonably well in the cloze task; more than fifty percent of the students correctly answered 50% of the cloze items. The items which this group found difficult were difficult for the whole group as well. In terms of language ability it can be argued, therefore, that they had achieved an acceptable level of language proficiency.

Note-taking Task

The average score obtained by this group of students on the note-taking task was 37% of the total mark allocated to the task. The notes prepared by this group of students were examined and were categorised into three types: poor, fair, and good.

Poor- A note was considered poor if it was worthy of below 32% of the total marks allocated to the task.

Fair- A note was considered fair if it was worthy of 32% to 44% of the total marks allocated to the task.

Good- A note was considered good if it was worthy of 45% and above of the total marks allocated to the task.

Of the notes prepared by those 45 candidates within this framework, 22% of the notes were poor, 29% were fair, and 49% were good. On the whole, 78% of the students were capable of taking notes from the given passage.

Writing Tasks

Writing tasks were also analysed to determine an index of the proficiency of the students who reached the 32-44 cutoff in the new exam. Their average score on the writing subtest was 32% of the total marks allocated to the tasks.

Examination of the Letter-writing Task

For the purpose of the study, all the letters written by the SLC candidates were categorised into three types: poor, pass, and good, though the categorisation was arbitrary.

Poor- A letter was considered poor if it was worthy of below 32% of the total marks allocated to the task.

Fair- A letter was considered fair if it was worthy of 32% to 44% of the total marks allocated to the task.

Good- A letter was considered good if it was worthy of 45% and above of the total marks allocated to the task.

Of the letters written by those 45 candidates within this framework, 11% of the letters were poor, 44% fair, and 42% good. One candidate left the task unattempted.

Comments

On the basis of this analysis, the following points emerged:

1. It is possible to reach cutoff 32-44 in the new exam without attempting the letter writing task, or performing the task badly. For example, one candidate who did not do the task, and 5 candidates who did the task badly reached the cutoff.
2. However, most students (87%) who reached cutoff 32-44 had done the task to an acceptable standard. Then it can be argued that the students who reached cutoff 32-44 were most likely able to write a fair letter, at least of this kind.

Examination of the Essay Writing Task

For the purpose of the present study, all the essays written by the SLC candidates were categorised into three types: poor, fair, and good.

Poor- An essay was categorised as poor if it was found to be worthy of below 32% of the marks allocated to the task.

Fair- An essay was categorised as fair if it was found to be worthy of 32% to 44% of the total marks allocated to the task.

Good- An essay was categorised as good if it was found to be worthy of 45% and above of the total marks allocated to the task.

Evaluating the essays written by the group of students within this framework, 25% of the essays were poor, 49% of the essays were fair, and 14% were good. 5 candidates (i.e 11%) left the task unattempted.

Comments

1. To achieve 32-44 in the new exam, it was not necessary to complete the essay writing task. For example, the 6 students who did not do the task and the 11 students who did the task badly reached the cutoff.

2. However, 63% of the students who reached cutoff 32-44 were found to be able to write an essay to an acceptable standard. On this ground, it can be argued that most students who reached this cutoff (i.e. 32% to 44%) could write a letter and an essay of these varieties. In other words, they could fairly express their ideas and feelings in writing.

5.4.5.2. Meaning of Passmark in SLC English Exam: The Minimum Level of Language Proficiency Required for the SLC English Exam

In our analysis, we found a large number of students (49%) who passed the SLC English exam with 1st, 2nd, and 3rd division but reached 0-31 cutoff in the new exam (see Table 10). This clearly shows that the SLC candidates found the new exam more difficult than the SLC English exam. One of the possible explanations of why the SLC candidates did better in the SLC English exam than in the new exam is that they were trained for the SLC English exam but not for the new exam. If we look at their performance in terms of type of school, we find that 96% of the students from Type A, and 60% from Type B schools have done equally well in both exams. But the percentage of students from Type C and D who have exhibited equal performance on both exams is very low (e.g. 19% from Type C, and 23% from Type D). This sort of evidence led us to investigate why a large number of students, especially from Type C and D schools, performed quite well in the SLC English exam but badly in the new exam. Investigating this would also provide a basis for an explanation of what passmark in the SLC English exam means in terms of language ability so that it can be determined what the minimum level of English proficiency is which an SLC candidate must have in order to get through the exam.

For that purpose, the performance of the members of the SLC group who have obtained 0-31 in the new exam, and 32 and above in the SLC English exam, (see Table 11) have been analysed.

Table 40 presents the results of the analysis of reading and cloze items based on the performance of only this group of students.

Table 40: Item Analysis:Facility Value:Reading and Cloze

Facility Value	Reading items	Cloze items
90-	-	-
80-89	-	-
70-79	-	-
60-69	-	-
50-59	7,1,2	-
40-49	4,6,17,	21,
30-39	18,16,19,3	23,28
20-29	8,9	42,44,32
10-19	10,12,5,14,15,13	54,34,50,36,37, 29,53,25,24,31
0-9	11	51,49,30,41,33,38 43,46,48,39,47,40 26,35,27,52,45,22

Comments on Reading Test

The average score of this group of students in this task was 29% of the total mark allocated to the task.

If we look at the facility values of the items presented above, we will find that nearly 50% of the students have correctly answered the most easy items - 7, 1, 2, 4, 6, and 17. One third of the students have correctly answered the moderate items 18, 16, 19, 3. Very few of the students have correctly answered the rest of the items. Expressing their performance in terms of reading ability, it can be said that they could understand the meanings of a passage which were explicitly stated. They could not understand the implied meanings in the passage. In other words, they have not achieved the ability to understand a passage which requires an ability to understand implied meanings.

Comments on Cloze Test

The average score obtained by this group of students in this task is 15% of the total mark allocated to the task which is quite low in terms of criterial level of performance (i.e. pass percent 32%).

If we look at the facility values of the cloze items established for this group of students, only item 21 has been correctly answered by 50% of the total students. Items 23 and 28 have been correctly answered by only 1/3 of the students. The rest of the items were extremely difficult for this group of students.

Note-taking, Letter-writing and Essay-writing Tasks

The performance of this group of students on note-taking, letter-writing, and essay writing tasks has been examined. The framework within which their notes, letters, and essays have been examined is the same as the one we have used for examining the performance of the group of students who scored 32% to 44% in the new exam.

Note-taking Task

The average score of the whole group in this task was 22% of the total marks allocated to the task.

Of the notes, we found 68% poor and 26% fair. 6% of the students had not attempted the task.

Comments

Only 1/5 of the students were found to have reached an acceptable level for taking notes from a given passage. In other words, 74% of the students passed the SLC English exam without being able to take notes of this kind.

Letter Writing Task

The letters written by this group of students have been examined within the framework we used in section (5.3.6.1.).

Examining the letters, 63% of the letters were found to be poor and 18% were found to be fair. 19% of the students did not attempt the task.

Comments

82% of the students were passed the SLC English exam without being able to write a letter of this kind. In other words, of the 141 students who passed the SLC English exam, only about 1/5 of the students achieved an acceptable level of letter writing.

Essay Writing Task

The essays written by this group of students were examined.

Examining the essays, 48% of the essays were found to be poor, 4% were found to be fair. 48% of the students left the task unattempted.

Comments

Based on this analysis, it was revealed that 96% of the

students passed the SLC English exam without having achieved the ability to write an essay of an acceptable standard.

5.4.5.3. Summary and Comments

To sum up this section, we have examined the performance of the group of students who performed well in the SLC English exam but badly in the new exam. We have found that most of them were not able to understand the implied meanings of the passages. They were not able to take notes from a given passage or write a letter and an essay using the given context. Nearly 50% of the total students passed the SLC English exam without being able to understand the implied meanings of a passage and to express their ideas and feelings in writing. Putting it another way, it was not necessary to be able to comprehend a reasonably difficult passage and to be able to express ideas and feelings in writing to get through the SLC English exam.

Going back to the question as whether the existing SLC English teaching in Nepal is adequate, we have to examine it in relation to the purpose for which English is taught. If we look at the educational purposes of English teaching at school in Nepal, we can see different subjects taught in English in tertiary education in section 5.1.6.1. in which students need a high level of English proficiency at tertiary level. We can also see the activities (in terms of language skills) which the students at tertiary level are expected to carry out in English. Now the question is: can the SLC passers who have a minimum level of English proficiency (in this case, below 31 in the new exam) carry out these activities? On the basis of their performance in the new exam, we claim that they cannot carry out those activities. The SLC English is not adequate for equipping the SLC passers with a knowledge of English necessary for tertiary

education.

5.4.5.4. Logical Evidence of the Construct Validity of the New Exam

Following Jacobs et al (1981:76) and Cronbach (1971:463), one way of carrying out construct validation of a test is to examine whether students with superior ability demonstrate higher performance than students with lower ability because of their advanced level of proficiency. The argument is that if a test does not allow test-taking strategy as a factor to gain scores in a test, students with higher ability will perform better than students with lower ability.

In our study, we classified schools into four types (i.e. A B C D) in terms of the way English learning opportunities were created in those schools. We predicted that students from Type A would do better than students from Type B, students from Type B would do better than students from Type C, and students from Type C would do better than the students from Type D because of their different advanced level of proficiency. We have presented the result that the schools were significantly different and the order of their performance was the same as we categorised. This indicates that the chances of gaining a high score without being able to do the tasks is less in the new exam. From this point of view, the exam has construct validity.

Another piece of evidence as to the construct validity of the exam can be produced by examining the characteristics of the test-tasks and the ways the performance of the students were evaluated. Criper and Davies (1988:119) present a list of critical features of a writing task which make it have construct validity. They are: authentic tasks, communicative tasks, neutral tasks, reasonable tasks, tasks which balance freedom and

constraint, tasks which are linguistically and rhetorically controlled, sufficient reading, thinking and 'writing' time, and communicative evaluation criteria. We find these features equally applicable to reading, and note-taking tasks as well. Evaluating the exam within this framework, we can argue that the test-tasks were designed keeping these features in mind. In other words, the exam tasks encapsulate the spirit of these criteria.

Evidence of the construct validity of the exam can further be reinforced by examining the way students have performed these tasks. We have examined the performance of the students at different levels of performance (e.g. students who reached cutoff 32-44, and 0-31) on these tasks. We have also examined their performance in relation to what the test tasks were intended to measure: e.g. reading test for measuring their reading ability, writing tests for measuring their writing ability. We have also carried out items analysis for the reading and cloze tests. We have examined the notes, letters, and essays of the students. We have found that the tasks on the whole were good enough to stimulate the reading and writing ability of the students who have such ability. Here again we found students with higher ability to have performed better than the students with lower ability on all tasks.

To make this point more explicit, we present the results based on the performance of the students who reached 32-44 cutoff and the students who reached cutoff 0-31. The average scores (in terms of the percentage of the mark allocated to the total task) obtained by the two groups of students on the four tasks are given in Table 41.

Table 41: Average score on the four tests obtained by the two groups of students: cutoff 32-44 and 0-31

	32-44	0-31
	cutoff group	cutoff group
Reading	47	29
Cloze	35	15
Note-taking	37	22
Writing	32	22

Looking at the results, we find that students with higher proficiency have done better than the students with lower ability on all tasks.

5.5. Summary and Comments

This chapter set out the results obtained from the investigation described in the preceding chapters. The results were presented under descriptive data, hypothesis testing, and exam validation headings.

Under descriptive data, the statistical and graphical descriptions of the data were presented. The results obtained from the new exam and the SLC English exam were presented qualitatively and quantitatively in order to make it possible to make direct reference to them in the subsequent sections while examining our predictions. A summary of the information obtained from the administrations of the questionnaires was also presented in this section.

We tested the hypotheses and subsidiary hypotheses using Anova and post-hoc scheffe comparison of group means. T-Tests were also made on the possible group means. On the basis of the

inferential statistics of the exam, we accepted the two main hypotheses, and all but two subsidiary hypotheses.

The exam validation section set the results obtained from validating the new exam. We addressed the question 'Is the existing SLC English teaching adequate?' in this section. We concluded that a high percentage of the students who passed the SLC English exam managed to pass it without having achieved an expected level of English proficiency required for tertiary education. From this point of view, we found the existing teaching and testing of SLC English inadequate.

CHAPTER 6: DISCUSSION AND IMPLICATIONS

6.0. Introduction

This chapter is devoted to discussions of the findings presented in the preceding chapter. It also includes discussion of the possible implications of the findings for language teaching and testing in general and the Nepalese ELT situation in particular.

We discuss first what the findings mean. Then the implications of the study for language testing will be discussed in terms of the power of a final exam to influence teaching and learning, the relationship between exams and the course objectives, and the relationship between formal instruction and learning. The efficiency of the new exam will also be assessed.

Next we consider the implications of the study for the Nepalese ELT situation. We will argue that the findings of the investigations provide evidence that if we want to make SLC English teaching serve the purposes for which English is taught at school in Nepal, it has to be changed, and the change has to be through the SLC English exam. The new exam is a good start for the change.

Finally, the methods and procedures of this study were limited. We make a note of them so that any subsequent study of this kind could benefit from our experience.

6.1. Meaning of the Findings

In section 5.3., we have presented the summary of which hypotheses were accepted and which hypotheses were rejected. In section 5.4.5., the findings based on the quantitative analysis of the data have been presented. This section deals with the meanings of those findings.

6.1.1. Washback Effect of SLC English Exam on Teaching and Learning of the SLC English Course

Main hypothesis 1 was formulated for investigating the relationship between the teaching focused on an exam based on prescribed textbook facts and discrete-point items and its contribution to learning. It was hypothesised that since the SLC English exam did not require the students to demonstrate the abilities to use the language in life-like situations, teaching for this exam would be educationally detrimental. The statistical evidence shows that hypothesis 1 (Alternate H1) can be accepted. What it means is that there was evidence for a negative relationship between preparation for the SLC English exam and its contribution to the students' learning of English.

Subhypotheses 1a-1c were formulated to provide more information about the washback effect of the SLC English exam on the teaching and learning of the SLC English according to school types. The statistical analysis shows that alternative subhypothesis 1c (i.e. the negative relationship between teaching for the SLC English exam and learning in type D schools) can be accepted. The statistical evidence reinforces the negative relationship between the preparation for the SLC English exam and learning. As the statistical evidence for alternative subhypothesis 1a (i.e. the negative relationship between teaching for the SLC English exam and learning in Type A schools) was not in favour of the negative relation, it needs further explanation. A possible explanation of the superior performance of the SLC group against the Grade 10 group from Type A schools is that the schools could have been immune from the negative washback effect of the exam. As has been mentioned earlier, these schools did not focus on preparing the SLC students for SLC English exam at the

cost of language skills. Further explanation is that the finance of these schools does not depend on the SLC results because these schools have national as well as foreign support. The economic situation may have lessened their worries about SLC results, and as a result, they may have not been influenced by the deleterious effect of the SLC preparation. And also one of the two schools (i.e. Budhanilkantha) which comes under Type A has introduced Cambridge O' Level exams which means that the students and the teachers from this school do not have to worry about the SLC results since they will not determine whether or not students should be allowed to go onto the next level. However, as the SLC holds a system value within the existing system in the sense that every office either public or private accepts its face value, the students take the SLC exam, and naturally want to get good marks.

The statistical evidence for alternative subsidiary hypothesis 1b (i.e. the negative relationship between teaching for the SLC English exam and learning) was not in favour of the negative washback effect of the SLC English exam either. However, the rejection of the subhypothesis was not considered counter evidence for the negative relationship between the SLC English exam and learning. It needs further interpretation. The explanation is that if Grade 10 and SLC students from Type B schools represented the same ability level despite the one extra year's study of the SLC group, it could mean that the SLC group had not educationally benefited from the Grade 10 studies. That means the SLC preparation had no educational contribution at all.

The relationship was further backed up when Type A and Type B schools were not found to be different before they embarked on the preparation for the SLC English exam (i.e. Grade 10 entry level), but different after they completed grade 10

studies (i.e. SLC: Grade 10 exit level). The difference could be ascribed to the ways the two types of schools prepared their students for the SLC English exam; Type A continued teaching as before, but Type B started teaching for the SLC English exam. Another interesting thing revealed by the researcher's personal communication with the school teachers and students from Type B schools was that these schools shifted the medium of instruction for Science and Maths from English to Nepali during SLC preparation (i.e. during Grade 10 studies). Since the students were not going to study Science and Maths in Nepali at college level, no convincing explanation for why they did so was given except saying that they did it for intensifying the SLC preparation. The result was that by the time the two types of schools completed the SLC preparation, they appeared to perform differently.

The negative relation was further supported by the performance of the Grade 10 students (representing Type D schools) who took the new exam twice. It was evident that this group of students had made no significant improvement in English proficiency during their Grade 10 studies.

Analysis of the responses of teachers and students to the questionnaires also supported the negative relation; teachers said that considerable time was spent on preparing students for the SLC English exam in spite of the fact that they knew that the exam was not a true measure of the language ability of their students. The students also indicated that they did not believe that the SLC English exam was a true measure of language ability; it merely led them to memorize ready-made answers and practise exercises.

On the whole, the findings show robust evidence that the washback effect of the SLC English exam on teaching and learning was negative. No counter-evidence for the negative relation was found.

6.1.2. Relationship between Teaching in different Types of Schools and Learning

Main hypothesis 2 was formulated for investigating the effects of the types of teaching offered by different types of schools on the learning of their students. The alternative hypothesis was accepted. That means there was evidence for a strong relationship between the types of teaching of English in those schools and the learning of English of their students: students from the four types of schools were found to be different in terms of their ability to use English. To provide supporting evidence, subsidiary hypothesis 2a (i.e. the effect of the types of teaching in Type A and Type B schools on the learning of their students), subsidiary hypothesis 2b (i.e. the effect of the type of teaching in Type B and Type C schools on the learning of their students), and subsidiary hypothesis 2c (i.e. the effect of the type of teaching in Type C and Type D schools on the learning of their students) were formulated. The statistical evidence shows that all the alternative subhypotheses can be accepted. The results reinforced the relationship between the types of teaching that the different types of schools used and the learning of their students. It was clear that every school was different from the others in terms of the language-using ability of their students.

There are several factors responsible for the significant difference in performance of students from the four types of schools. However, as the schools were classified, for the purpose

of the present study, in terms of the extra courses they offered on top of the compulsory English, the medium of instruction, and the emphasis they put on passing the SLC exam, only these variables were taken into account in interpreting the findings. Other factors, like school environments, types of teachers, home environments, rural and urban schools would have contributed to some extent to the performances of their students but they were not taken into account in the present investigation. The interpretation then is that the students who had extra course/s in addition to the compulsory English appeared to have benefited: students with extra course/s outperformed the students without it/them. It was also evident that the students who did courses in the English medium appeared to have done significantly better than those who did them in the Nepali medium. Since all four types of schools were statistically different, no further statistical analysis was necessary for comparing Nepali medium schools and English medium schools.

What it means is that, because of the differences in terms of the teaching different types of schools provided for their students, the schools produced students with different ability levels. As was hypothesised, the rank order of the schools was-A> B> C> D. This rank order was maintained by the performances of the Grade 10 groups as well. At both levels, students who had received only the compulsory English course appeared to be the least proficient group. The rank order also suggested the negative washback effect of the SLC English exam.

6.1.3. Adequacy of the SLC English Course

There was clear evidence (see Table 17) that a large number of students passed the SLC English exam without having achieved the minimum required level of English proficiency for

the tertiary level. Most students who did well in the SLC English exam but badly in the new exam were the students who had received only the compulsory English course. This means that the schools which depended entirely on the compulsory English course (i.e. Type D) produced students below the expected level of proficiency for the tertiary courses which require English. The explanation of why the students from Type A, B, and C schools could demonstrate the criterial level of English proficiency could be attributed to the amount of extra instruction they received.

The students from Type D schools, from a language skills point of view, were found unable to understand implied meanings of a reading passage and express their ideas and feelings in writing. In this situation no one would be surprised if these students found difficulty in coping with the tertiary courses which require English.

The evidence of inadequacy of the SLC English was further reinforced by the performance of the college students who had been educated in Type D schools. The students who had received high scores in the SLC English exam were found to have difficulties in coping with the tertiary courses. On the basis of the information we collected through questionnaires, it was revealed that many of them had failed the tertiary level because of English, and those who had passed the tertiary level got low scores which would have prevented them from stepping up to the next level. For example, let's take the performance of the Science group. These students did quite well in the SLC English exam and were enrolled in the Faculty of Science. But their scores in their first year exam at tertiary level revealed that several of them had failed their English course and other content

courses. Their total performance was also very low. Another group of students who went to a Nepali medium college found the new exam very hard. It shows that their ability to use English for reading and writing was quite low.

At this point, the compulsory English course appears to be inadequate for the purpose of enabling the students to cope with tertiary courses. If these students cannot read and write for the purpose of communication, by the same token, they will not be able to use English for occupational purposes either.

6.2. Efficiency of the New exam

In section 2.3.3.2.1. we have argued that only the exam which is designed in relation to efficiency criteria and the purpose for which it is used can be educationally beneficial. The efficiency criteria of an exam were set as reliability, validity, and practicality. To recommend that the new exam should be used for replacing the SLC English exam, it is necessary to articulate how efficient the exam is. The practicality of the new exam has been discussed in section 5.4.3.2. There is no need to repeat that here. Below we summarise the reliability and validity of the exam.

6.2.1. Reliability

The reliability of the new exam as a whole and of its different components was established. The overall internal consistency of the total exam ($\alpha=.82$) was high. The test-retest reliability ($r=.87$) of the exam was also high. The inter/intramarker reliability of the writing section of the exam was also very high ($\alpha=.94$). On the whole the exam appeared to have high reliability.

An analysis of the distractors of the reading tests showed that some items needed to be improved. Involvement of

subjectivity in scoring writing would call for a training for the raters. Careful revision of the distractors, and the training for rating would also increase the reliability of the exam.

6.2.2. Validity

Qualitative and quantitative analyses of the performances of the students on the new exam, and analysis of the questionnaires were carried out in order to establish the validity of the new exam. Statistical and logical validity of the exam and its components were established.

6.2.2.1. Statistical Validity

Intercorrelationships between the subtests of the exam showed that they were highly correlated. That means the subtests were measuring something similar to a great extent. The same thing was confirmed by performing principal component analysis on the original data. In this sense, the exam appears to be a test of general language proficiency in terms of language construct.

Concurrent validity of the exam was established by correlating the exam with the SLC English exam and Cambridge O' Level Language Test. The correlation between SLC and the new exam showed that the two exams were measuring 2/3 in common. The rest of the variance the new exam measured was attributed to the basic language skills the exam was intended to measure. The coefficients of correlation between the scores in the new exam and the Cambridge O' Level Language test and the new exam and the college teachers' judgement were weak. However, since the data was inadequate, this was interpreted as flaws in the data collection procedures. Plotting the scores in a scattergram and computing the coefficient of correlation by removing the outliers gave us an impression that the exam is likely to have a strong relation if it is administered to a reasonable sample by

minimising the factors which affect correlations. For certain practical reasons (e.g. schools and colleges in Nepal remained disturbed for a long time), no predictive validity of the exam was established.

6.2.2.2. Logical Validity: Face, Content, and Washback

The exam has high face validity in the sense that the SLC students who took it and the people who were involved in the SLC English found the exam a true measure of language proficiency. Efforts have been made to make the exam have content validity by carrying out a priori validation. Experts' reaction to the exam was that the exam was valid in terms of its content. The exam specification was found to be reasonably valid.

Analysing the performances of the students quantitatively, the exam appears to be a valid measure. As was argued earlier, the validation of the exam was not merely the validation of the exam; it was also the validation of the procedures followed in designing the exam. The point is that if an exam designed on the basis of the prespecified procedures provides valid information about the examinees, then it can be assumed that other exams which are designed following the same procedures will be as valid as the exam validated. In that sense, the procedures have the potential to produce valid measures of language ability. In other words, as the new exam appeared to be a valid measure, the specifications, the formats of the new exam and the basic principles which were used in the design of the exam have the potential to produce other valid exams.

There are positive aspects of the exam. The most important, and the basic concern of the exam is its washback effect. The quantitative analysis of the performance showed that the exam required the students at least to read and write. From

this point of view, preparing students for this exam will be educationally beneficial in the sense that to be able to attempt the exam, students will have to be able to read unseen texts of similar difficulty level and express their ideas in writing.

The comparison of the performance of the SLC students on the two exams gave us the impression that their scores on the SLC English exam were not above suspicion. It was necessary to examine how they scored so high in the SLC English exam, but so low in the new exam. On the basis of the analysis, we came to the conclusion that rote-learning was enough to score high in the SLC English exam. Then one would ask what the scores in the SLC would mean. Can they be good predictors of their college performance?

This study could not provide empirical evidence for the washback effect of the new exam for the reason that teaching for this exam was not possible before the completion of the present investigation. Its washback can be empirically studied after it replaces the SLC English exam or an exam of similar type. However, it can be argued that the exam could come up with a positive washback effect after it gains the power the SLC English exam has been associated with, and the total ELT situation changes accordingly. Here we agree with Hughes (1988b) and Li Xiaojun (1989) who could demonstrate positive washback effects of the tests designed following a similar approach, i.e. integrative and communicative in nature, and future use oriented in terms of needs.

6.3. Implications of the Study for Language Testing and Teaching

As the reported findings can be generalised to SLC students studying English in Nepalese schools, the study also supports some previous findings and assumptions. The research could profitably be replicated in other places and with other

languages as well. Below we point out some points which emerged from the investigation.

6.3.1. Exams as Powerful Instruments for Gearing up Classroom Activities in a Desired Direction

The study provides evidence that a final exam has a great power to attract classroom activities. This investigation supports, and further extends, the claims made by Li Xiaoju (1989) and Hughes (1988b) that exams have a great influence on teaching and learning.

The findings of the investigation also back up the assumption made by Wong (1969), Wiseman (1961) and Wrightstone (1963) that no matter what teachers think good for the students, their work is influenced by the exam. In the present study, teachers made it clear that they did not find the SLC English exam a true measure of language ability. They made it clear, however, that they worked for the exam. The implication is that students and teachers work for obtaining good marks in the exam regardless of whether or not the exam they are working for is a true measure, provided that the exam has extrinsic values.

The argument which follows, then, is whether or not working for the exam is desirable. This study was not designed to investigate the issue in question; however, it can be argued on the basis of the findings that working for the exam is not necessarily harmful. If any preparation for an exam is harmful, it may be because the particular exam may have been isolated from the overall aims of the programme. For example, preparation for the SLC English exam was educationally harmful because it failed to reflect the purposes of the SLC English course. On that ground, we would like to argue that preparation for an exam like

the new exam would not be educationally detrimental. We would like to ask if there is any harm in teaching for an exam based on the specifications and format of the new exam.

To make this point more explicit, we take the performance of SLC students at Budhanilkantha, a Type A school, as an example. These students were preparing for the Cambridge O' Level language examination as well as the SLC English exam and considered the former to be the more important qualification. Since the skills and abilities required for the Cambridge O' Level are different from those required for the SLC English exam, students would presumably have concentrated their efforts on acquiring the skills and abilities required for the exam that had serious consequences for their future. They therefore continued to enhance their English proficiency because they were preparing for an exam which calls for language skills rather than the recalling of discrete linguistic items and textbook facts. As the results show, they did quite well even in the SLC English exam. It could mean that their high proficiency and familiarity with the exam content led them to obtain high scores even in the SLC English exam. This does not provide counter-evidence for the negative relationship between the SLC English exam and its contribution to learning because the exam-preparation of this group of students was not strongly influenced by the SLC English exam.

What determines whether the effect of exam preparation is positive or negative is, therefore, the design of the exam: if the design is in consonance with the purposes of the teaching programme it can have a positive washback effect; if not it is bound to have a negative washback effect. On the whole, it should be stressed that people who work for the exam - teachers,

students, material producers, crib writers - work according to what they think the exam they work for requires. The way in which they work and the traits they work for can be changed as their expectations of what the exam calls for change. This power of the exam can be used for bringing about changes in an educational programme.

6.3.2. Relationship between Formal Instruction and Learning

This investigation supports the argument that students' learning depends on the amount of instruction they receive (Yalden 1987:15) provided that the instruction is conducive to language learning. Students with different ability levels provided a basis for an explanation that every extra course that the schools offered made some contribution to learning. The explanation is that the variety of the courses and the amount of exposure in terms of language use caused the different schools to come up with different proficiency levels. It shows that there is a strong relationship between the amount of instruction and learning.

This finding is in favour of Upshur (1968) who sees "the positive relation between amount of material taught in the course and mastery of the material" and Porter (1982a:40) who argues "the greater a person's experience of a particular area of language activity, the fuller their competence". Krashen (1976:167) also finds that extra courses contribute to the learning of the students.

Given the above, it must be asked: why could the SLC students from Type B and D schools who had an extra year's study of English, including extra hours of instruction and at least one extra book, if properly written, not improve their proficiency?

The finding is in contrast with, for example, Krashen et al (1978) who find years of instruction a good predictor of English proficiency.

Examination of the SLC English exam 1989 (section 2.2.6.6.) and the prescribed textbook for Grade 10 (section 2.2.5.) showed that preparation for the SLC English exam does not engage students in using language for meaningful purposes. If, as Fathman (1975) concludes, instruction based on grammar and drills provide no comprehensible input, the instruction intended for the SLC English exam does not provide relevant linguistic data for language learning either. This argument confirms the hypothesis that the total linguistic environment is not important in language learning; what is important is what input has been given, and how the input has been given. In that sense, our finding agrees with Burt and Dulay (1973:256) who argue that to activate the language learning process students should be exposed to a natural communication situation, (e.g. as found in authentic types of language activity). Regarding this issue, there appears to be a strong interaction between the types of material students are exposed to and their learning. It is a fertile area for further investigation. As has been argued before, since the SLC students were not engaged in the language developing activities, their English proficiency did not develop.

The implication of the above argument is that, if an exam uses authentic-like materials, then preparation for the exam requires the examinee to use authentic linguistic data, which in turn can enhance his/her proficiency in the language. The same notion is applicable to materials writing.

6.3.3. Relationship between Course Objectives and Exam Content

This investigation supports the argument that if there is

an incongruent relationship between what is tested and what is supposed to be tested, it will have an educational disadvantage. It has been shown that the exam influences teaching and if teaching aims to achieve the course objectives, it is obvious that the exam has to reflect the course objectives. Failure to design an exam accordingly would contribute to a negative washback effect because the exam, given that it is extrinsically valid, may fail to motivate the student to achieve the course objectives, but it may not fail to attract the student and the teacher to work for it. For example, we would argue that the time and labour spent on the SLC English exam preparation would have contributed to the students' achievement of language skills had the SLC English exam been representative of what has been described as the course objectives of the SLC English. One of the reasons why the SLC English exam had a negative washback effect was its failure to represent the spirit and intent of the course objectives.

The argument is that, if the purpose of the course is set, for example, as enabling students to read and write, the measurement of whether or not the objectives of the course have been achieved has to be referenced to whether or not the examinees can read and write. In other words, the validity of a teaching and testing programme depends on whether or not the learners are required to do the activities encapsulated in the course objectives.

The implication is that the content of a final exam has to be derived from the course objectives, not from the course syllabus, in case there is a mismatch between what has been taught and what ought to have been taught. If the exam content accords with the course objectives, what is taught can be

reshaped according to what ought to be taught. In this context, a final exam moves away from the usual definition of an achievement test, and reaches the area of a proficiency test. If a final exam has to contribute to making judgements on the examinee, it seems reasonable (agreeing with Hughes 1988b) to make it like a proficiency test which, following Davies (1985:8), can eventually work as a teaching syllabus. This is possible if the content of the course objectives is adequately reflected in the content of the final exam.

6.3.4. Direct and Indirect Testing

The investigation provides evidence for the argument that ability to recall and mechanically manipulate discrete linguistic facts about phonology, syntax, morphology and textbook facts is different from the ability to operate effectively in life-like language use situations. It further extends the view that once a student has got mastery over linguistic competence he/she will automatically be able to use the language in real-life situations (Schulze 1977:95). The superior performance of the SLC students on the SLC English exam over the new exam was evidence for the argument.

The investigation supports Schulze (1977) who concludes that linguistic proficiency and memorization and communicative proficiency are not identical constructs as they call for different traits and skills. The implication, then, is that an indirect measure of language proficiency may not give true information about the testee's language ability, and any decision based on that measure is not without suspicion. On the other hand, a direct measure appears to give a true picture of the language ability of the testee. Thus, it seems obvious that if we are interested in measuring the testee's true language ability, a

direct measure is preferable to an indirect measure.

6.4. Implications of the Study for the Nepalese ELT Situation

English language teaching at School in Nepal will be the main beneficiary of the present investigation. As the main study was based on the Nepalese ELT situation, the immediate benefit of the investigation will be received by it. The findings of the study are applicable to other situations as well where an exit exam, externally controlled, is used to make decisions about the examinees.

The most important thing revealed about the existing ELT situation was the negative washback effect of the SLC English exam. There is evidence that whatever the students did in the name of the preparation for the SLC English exam has not contributed to their English proficiency. Rather it appeared that the preparation caused a loss in the previously achieved level of proficiency. This is a matter of great concern for a nation in the sense that the exam which has the largest number of participants and which has been exercising the most decisive role in making decisions about school leavers' higher education has been found to be working in the wrong direction. How can the amount of money and the manpower involved in it be justified? The nature of the assessment system (based on textbook and previous exams) and the materials (based on structural approach) used, indicate that this exam may even be detrimental to students' education.

A year for the students at that stage is extremely valuable. What surprises us is the fact that in several parts of the world there are language teaching programmes, for example, TEFL in the UK, which are based on the assumption that a Foreign Language can be learnt in 6 months or 1 year for certain

purposes. On the contrary, in the case of Nepal, the students who had spent at least seven years studying English in a formal setting appeared to be unable to read and write it. For the purposes of teaching English at school in Nepal, it is strongly recommended that if we want the students to achieve the expected level of English proficiency, it is very important that the present situation should change to accord with the recent trends in teaching and testing.

Analysis of the SLC English course and the exam has shown that they are not adequate to create a proper situation for language learning. The importance of exposure in terms of materials and methods is another important thing revealed by the present study. Extra (lesson) time, if spent carefully, has contributed successfully to the English proficiency of the students. There was a strong relationship between the amount of exposure in terms of extra course/s and learning. It is necessary to look at how the students may have spent the whole year during their SLC preparation. The analysis of the course revealed that the textbook was old-fashioned. The texts and the exercises were of a mechanical type in the sense that students did not have to process language in the real sense. This is not how a foreign language is best learnt (Morrow 1977). The basic foundation of the theory (i.e. structural approach) on which the book and the SLC English exam were based has been severely attacked. The materials should, then, be changed. Exposure to the use of language and authentic materials have been shown to be of such value to language learning (Krashen 1976) that it can only be concluded that authentic materials, suited to the proposed exam, should be introduced in the schools.

The responses of the teachers, the students, and the experts indicated that they are ready for change. It has also been indicated that the students, teachers, and experts in Nepal are interested in making the teaching and testing of English communicative in nature. The college students' responses indicated that a high level of English proficiency is needed for the college entrant. On the basis of their responses, it is also necessary to make the SLC English congruent with the college requirements.

Finally, but importantly, how should the change take place? This investigation provides robust evidence that an exam is the most powerful component of an educational programme, and it can take a dynamic lead if used properly. It has been evidenced that any change in an educational programme cannot be effective if this component is ignored. And it has also been evidenced that accepting this component as a powerful instrument can accelerate the movement of change by keeping all the relevant components including methods and materials on the track. On the whole, a change including all the aspects of a language teaching programme, not a piecemeal operation, is demanded.

The argument is that there is no reason to change the course book or methods of teaching if the objective of the teaching is to pass the exam by memorising the ready-made answers. This will happen if the existing SLC English exam remains the same. An education system which encourages the students to think creatively, which makes them expressive and critical, is needed, not one that makes them slaves of the prescribed books and mechanical rules. It is the thinking process that has to be changed. As the exam appears to be an influential component in an educational setting, changing the exam may lead

students to change their expectations, and hence the way they prepare for it. If the exam calls for thinking, the students will be prepared for that. Their dependence on ready-made answers should be reduced so that the whole preparation would be conducive to language learning. Preparation for this type of exam has potential for enhancing language proficiency.

The present form of the SLC English exam is old anyway. Every exam needs change as time passes in order to incorporate recent trends in testing and teaching. The SLC exam has worked well for 18 years. Now the time has come for a change. The basic assumption behind a structural approach has been ineffective in Foreign Language teaching and testing.

6.5. Methodological Limitations

Many problems were faced in carrying out the investigation. The problems faced in the field study and the validation of the exam have already been described. In addition to these, there are some confessions to make.

It must be admitted that the study was carried out in difficult circumstances in the sense that the potential schools and colleges remained disturbed for a long time because of the socio-political situation in the country. The time for the field study was limited. The researcher has the impression that ten weeks' time is insufficient for this type of study, which involves a great number of people from different parts of the country. Piloting the tests, administering the exam and questionnaires to an adequate number of students and people concerned, and contacting Government people were unpredictably difficult. Living in Britain and collecting necessary information from Nepal became much more difficult than previously anticipated. Due to these difficulties, it became extremely

difficult to obtain the information we wanted. At some stage, we had to ignore the previously collected data due to the lack of supporting evidence. For example, we administered the exam to college students from different faculties (e.g. Science, Humanities, Management, and Agriculture) on the assumption that their performance on the exit exam would be available. But the collected data was not very useful because we could not collect the scores of the students in the exit exams.

The nature of the study was another source of difficulty. As the main subjects were at a crucial year of their school leaving stage, it was not sensible to administer the exam when they were at the peak of their SLC preparation. Later, since they did not have to come to the schools after the completion of the SLC exams it became difficult to get hold of them.

Until the time the exam had to be designed, the researcher was influenced by the literature (e.g. Anderson 1972) in which arguments that cloze is a test of overall reading ability prevailed. However, we agreed with the literature reviewed later (e.g. Alderson 1978) which argues that cloze is not a test of reading ability but a test of linguistic competence. The implication is that if we are to repeat the experiment, we would like to use a modified/rationale cloze for assessing linguistic proficiency. Assessing note-taking was another problem. We would have liked to ask the students to develop the notes they had prepared into a complete piece of writing rather than to depend entirely on the notes. This proved to be too difficult for practical reasons.

Despite these problems, the researcher has a strong belief that, as the participants represented a wide range of ability levels, and most regions (e.g. urban-rural, hilly-plains) of the

country, the basic findings have not been affected by the flaws in selection of the students and the schools.

6.6. Summary

The Meanings of the findings have been discussed. The present study has found some evidence that an exit exam can be very powerful to determine what goes on in the classroom. It has even been argued that the washback effect of an exam is not limited to classrooms. Evidence has been produced that if an extrinsically valid exam fails to measure what it is supposed to measure, it can have a deleterious effect on education. An exam which captures the sentiment of the course objectives in terms of its content, and authentic-like communication in terms of its format can have the potential to produce valid measures of language ability with beneficial washback.

Looking at the Nepalese ELT situation from the point of view of what has been found in the study, it has strongly been recommended that the existing SLC English exam should be changed to make it congruent with the purposes for which English is taught at school in Nepal. The new exam provides a starting point in this direction.

CHAPTER 7: CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

7.1. Conclusions

On the basis of the preceding discussion, it is possible to draw a series of conclusions.

7.1.1. Washback: An Inherent Quality of an Exam

One major conclusion is that washback is inherent in an exam: an exam is bound to influence teaching and learning. The possible reasons why the exam is influential on education should be seen in terms of its functions. The functions of the exam can be summarised under its forward looking and backward looking purposes though they are not exclusive. The backward looking functions of the exam have to do with a sense of achievement and evaluation of the effectiveness of teaching. The forward looking functions are related with making decisions about the examinees; for example, selection and certification. In a practical world, the procedures of selection and certification are necessary to establishing communication between educational institutes and receiving agencies.

Putting the arguments in the Nepalese situation, we can see the forward and the backward looking functions of the SLC English exam. As has been discussed, the SLC has major functions to perform: e.g. sense of achievement, licence for prestigious work, and permit for higher education. Whatever form in terms of content and quality the SLC English exam takes, it will be used as an established way of communication, because it has face value between schools and the receiving agencies, such as employers and colleges. Then the use of the exam-results for forward looking purposes naturally influences the procedures for achieving what is considered evidence of valid achievement (that is, a sense of achievement). Students obviously want to obtain it and teachers

are there to help them have it.

This power of an exam is always in control of the classroom. The exam retains this power while it exists, and exams in general are most unlikely to vanish, at least in Nepal. No matter whether we want it or not, whether the exam calls for the necessary skills and abilities or not, students and teachers are bound to work for the exam.

It can, therefore, be concluded that the washback is inherent in the exam.

7.1.2. Ingredients of the Exam Determine Whether the Washback is Negative or Positive

The traditional assumption that an exam is destined to produce negative washback is false. It is not the exam itself as a measuring instrument but its ingredients which determine whether the washback is negative or positive.

The ingredients of the exam are the crux of the washback. The exam is a means of communication between the exam-giving agencies and exam-taking agencies. 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 are in accordance with the purpose of the whole teaching programme, the people responsible for preparing students for the exam and the students themselves can work towards that. In other words, the exam shapes the expectations of the people who work for it. Whatever we want them to work for, one way of telling them the message is the exam. This can be done by making the specifications and the formats of the exam explicit to the people concerned. It can, therefore, be said that if the specification and format of the exam correspond to the purpose of the education programme, working for the exam will not have negative washback.

Therefore, it can be concluded that a final exam has an undoubted influence on education but the influence may not necessarily be negative. It is the design of the exam (i.e. the ingredients it is composed of) which determines whether it can have negative or positive washback.

It can also be concluded that language testing calling for the knowledge of the language in terms of grammar and textbook facts can produce negative washback. The skills and abilities which the knowledge of grammar and textbook facts involve are not the same as the skills and abilities required for using language for communication. In that sense, it can be concluded that testing communicative ability is likely to produce positive washback. And if the teaching course is aimed at developing communicative ability in the students then this is what the exam has to measure in order to produce desired outcomes.

7.1.3. Teaching for a Final Exam: Inevitable or Desirable?

It has been concluded that teaching for the exam is inevitable. Whether or not teaching for the exam is educationally beneficial depends on the design of the exam.

The power of the exam undoubtedly accelerates the preparation work for it: teachers' and students' work can be speeded up because of the implicit and explicit incentives attached to the exam they work for. If the teaching for the exam involves the language activities necessary for achieving the purposes of the course, teaching for an exam of this sort can be educationally beneficial.

For example, if the SLC English exam contains the ingredients that the SLC English course intends to develop in the students, then that is what the schools have to teach. If the course is intended for developing students' reading and writing

proficiency, then the assessment should be based on the principles of testing reading and writing. Given that an exam of this type exists, is teaching for this exam harmful?

Therefore, it can be concluded that teaching for the exam is not only inevitable, it is desirable as well.

7.2. Suggestions for Further Research: Washback Validity: Need to Innovate

For Morrow (1986), washback validity is the first criterion for a public exam. In the present study, we came to the conclusion that washback effect (either negative or positive) is an inherent quality of a final exam. It has been argued that, if exploited properly, the positive aspect of the washback can make the exam work as a lever for a desired change in an educational setting.

The studies carried out by Hughes (1988b) and Li Xiaojun (1989) have produced some empirical evidence that if an exam is designed carefully for the purpose of achieving the positive washback, it can achieve it. Hughes (1989) discusses how to design a test to achieve positive washback. Alderson (1986) proposes innovations through testing, and Davies (1985) suggests that teaching should follow the lead of testing.

In most testing literature, as can be seen in the above studies, there is an element of the washback interwoven in the whole business of pedagogy. Nevertheless, neither has the concept of washback been explored adequately so as to determine the areas it covers, nor have the elements involved in it been identified so as to make it easy to measure the washback separately. This being so, it becomes difficult to determine how to measure it.

Much research is needed to understand the nature and the coverage of the concept. Is it limited to classroom teaching and

learning or is it beyond that? What method/s can be adopted to measure it? The very idea is expressed in a global term: e.g. the effect of an exam/test on teaching and learning. The terms, 'teaching' and 'learning' cover such a wide range of activities that it becomes difficult to account for the effect of the washback because it is difficult to separate it from other factors.

We have held the view that whatever is done with the intention of helping students get through the exam is its washback effect, including any sort of motivation, incentives or pressures created by the exam. In this context, the washback would appear as a force compelling the people involved in it to work for the exam. This way of looking at the washback does not make it easy to isolate the effect of washback from the effect of other factors on teaching and learning either. The benign idea behind washback still remains crude. It needs, therefore, further refinement so as to reduce the antagonistic attitudes towards a final examination system.

In the present study we found it difficult to pin down what washback validity actually is in terms of a concrete concept. Rather it appears to be incorporated with face and content validities as well as other appealing forces of the examination. In terms of the area it covers, it is not confined to school walls in the sense that an exam controls not only the classroom but also the student's study at home. Since parents are not immune to exam results, washback can have an effect on how they create study environments at home for their children. For example, if an exam includes test items from local newspapers or radio broadcasts, parents try to make such resources available for their children at home. In the same way, people involved in

writing materials, either supplementary or crib, would produce materials required for the exam.

This study does not offer adequate evidence for, or a concrete method of measuring, the washback effect of an exit examination. However, if research of this type is the attempt to identify relationships of one variable with others, then the study provides groundwork for studying an exam from the washback point of view: e.g. how educationally influential an exam can be, what happens if the exam is not educationally beneficial, and what steps can be taken to make it educationally beneficial. What is most lacking in this area of testing is the method of measuring washback, which merits further investigation. From the pedagogical point of view, the washback appears to be potential. It is essential to develop methods of measuring washback.

Some suggestions can also be made towards the next step of the study in the Nepalese context. One desirable piece of research would be the replication of the study in a more controlled situation in terms of the number of subjects, types of schools and their representation. The exam should also be improved by gaining insights from the present study.

Another desirable piece of research is the piloting of the new exam in the real situation. A study should be carried out by using the new exam to replace the SLC English exam. It would be possible to do that in two or three districts in Nepal in the preliminary stage of the exam. At the same time, other factors, like the materials required for the exam, and the training for the teachers according to the new exam should also be considered. Such a study can be the real validation of the new exam in terms of its washback validity. It is important that the students and teachers know the specifications and formats of the exam. If the

exam comes up with the desirable results, the new exam should replace the SLC English exam nationwide. Necessary changes in the materials and methods should also be made. It can be suggested that different methods, e.g. evaluation of teaching and learning opportunities provided by the introduction of the new exam, and assessment of students' performance, should be used to measure the washback validity of the exam.

BIBLIOGRAPHY

- Abdrabou, A. (1984) Achievement Testing and Foreign Language Teaching. In: Proceedings of the Third Annual Linguistics Conference (Irbid, Jordan, April 1-3, 1984) ED 27010.
- Adams, M.J. (1980) Failures to Comprehend and Levels of Processing in Reading: In Spiro et al (eds.) (1980):11-32.
- Agazzi, A. (1967) The Educational Aspects of Examinations. Strasbourg. Council for Cultural Cooperation of the Council of Europe. Education in Europe. Series II. General and Technical Education-No. 10.
- Alderson, J.C. (1978) A Study of the Cloze Procedure with Native and Non-native Speakers of English. Unpublished Ph.D. Thesis. Department of Applied Linguistics, University of Edinburgh.
- Alderson, J.C. (1986) Innovations in Language Testing. In Portal (ed.) (1986):93-105.
- Alderson, J.C. (1988) New Procedures for Validating Proficiency Tests of ESP? Theory and Practice. Language Testing. 5/2:220-246.
- Alderson, J.C. (1989) Bands and Scores. Paper Presented at the Language Testing Colloquium - Bournemouth 17-19 November, 1989.
- Alderson, J.C., and Hughes, A. (eds.) (1981) Issues in Language Testing: ELT Document 111. London. The British Council.
- Alderson, J.C., and Urquhart, A.H. (1984a) Introduction: What is Reading? In Alderson and Urquhart (eds.) (1984):xv-xxviii.

- Alderson, J.C., and Urquhart, A.H. (1984b) Reading in a Foreign Language: a Reading Problem or a Language Problem. In Alderson and Urquhart (eds.) (1984):1-24.
- Alderson, J.C., and Urquhart, A.H. (eds.) (1984) Reading in a Foreign Language. London. Longman.
- Alderson, J.C. and Urquhart A.H. (1985) This Test is not Fair: I am not an Economist. In Hauptman et al (eds.) (1985):25-43.
- Ali, H.A. (1988) A Process Based Approach to Teaching Written English to First-Year Students in Lebanon. Unpublished Ph.D. Thesis. Department of Applied Linguistics, University of Edinburgh.
- Allen, J.P.B., and Davies, A. (eds.) (1977) Testing and Experimental Methods: The Edinburgh Course in Applied Linguistics. Vol.4. Oxford. Oxford University Press.
- American Psychological Association (1974) Standards for Educational and Psychological Tests. Washington, D.C. American Psychological Association.
- Anastasi, A. (1982) Psychological Testing. Fifth Edition. New York. Macmillan Publishing Co., Inc.
- Anderson, J. (1976) Psycholinguistic Experiments in Foreign Language Testing. University of Queensland Press.
- Arndt, V. (1987) Six Writers in Search of Tests. In ELT Journal 41/4:259-267.
- Atiyeh, N.M. (1969) Examinations: Trends and Prospects. In Lauwerys and Seanlon (eds.) (1969):375-391.
- Audrey, H., and Nicholls, H. (1983) Developing Curriculum: A Practical Guide. London. George Allen and Unwin.
- Bachman, L.F., and Palmer, A.S. (1981) The Construct Validation of the FSI Oral Interview. Language Learning. 3/1:67-86.

- Barruta, R. (1967) A Close Look at the Importance and Place of Testing in the Foreign Language Program. Diagnostic Database. Ed 016439.
- Bejar, I. (1983) Achievement Testing: Recent Advances. California. Sage Publications Inc.
- Beretta, A., and Davies, A. (1985) Evaluation of the Bangalore Project. ELT Journal 39/2:121-127.
- Berthoff, A.E. (1978) Tolstoy, Vygotsky, and the Mankind of Meaning. College Composition and Communication. 29:249-155.
- Berry, B.J. (1931) Improving Freshman Reading Habits. English Journal. 20/1:824-828.
- Bhadra, S., and Yadav, Y.P. (1988) Report on Failures in Compulsory English at the University Level. Kirtipur. Research Division. Tribhuvan University.
- Bialystok, E., and Smith, H.S. (1985) Interlanguage is not a State of Mind: an Evaluation of the Construct for Second Language Acquisition. Applied Linguistics. 6:101-117.
- Black, H.D., and Devine, M.C. (1986) Assessment Purpose: A Study of the Relationship between Diagnostic Assessment and Summative Assessment for Certification. Pamphlet. The Scottish Council for Research in Education.
- Black, H.D., and Dockrell, W.B. (1988) New Experiments in Educational Assessment. Monograph Series No. 3. Scottish Academic Press.
- Bloom, B. (1972) Taxonomy of Educational Objectives. London. Longman.
- Boisson (1887) Dictionnaire de Pedagogie.
- Bowe, R., and Whitty, G. (1984) Teachers, Boards and Standards: The Attack on School-Based Assessment in English Public Examinations at 16+. In Broadfoot (ed.) (1984):179-198.

- Bransford, J.D., Barry, S.S., and Sheltron, T. (1984) Learning from the Perspective of the Comprehender. In Alderson and Urquhart (eds.) (1984):28-44.
- Briere, E.J. (1971) Are we Really Measuring Proficiency with our Foreign Language Tests? Foreign Language Annals. 4:385-386.
- British Council, The (1978) The Teaching of Comprehension: ETIC Occasional Paper. English Teaching Information Centre.
- Broadfoot, P. (1984) From Public Examinations to Profile Assessment: The French Experience. In Broadfoot (ed.) (1984):199-220.
- Broadfoot, P. (ed.) (1984) Selection, Certification and Control: Social Issues in Educational Assessment. London. The Falmer Press.
- Brown, F.G. (1976) Principles of Educational and Psychological Testing. New York. Holt, Rinehart and Winston.
- Brown, J.D. (1984) A Categorical Instrument for Scoring Second Language Writing Skills. Language Learning. 34/4:21-36.
- Brumfit, C. (ed.) (1988) ESP in the Classroom: Practice and Evaluation: ELT Document 128. Modern Language Publications.
- Brumfit, C., and Johnson, K. (eds.) (1979) The Communicative Approaches to Language Teaching. Oxford. Oxford University Press.
- Burgess, T., and Adams, E. (1980) Outcomes of Education. Macmillan Education.
- Canale, M., and Swain, M. (1980) Theoretical Basis of Communicative Approaches to Second Language Teaching and Testing. Applied Linguistics. 1/1:1-47.
- Carroll, J.B. (1973) Foreign Language Testing: Will the Persistent Problems Persist. Paper Presented at the ATESOL Conference (Dublin, Ireland, June 26-29, 1973).

- Carroll, J.B. (1983) Psychometric Theory and Language Testing. In Oller (ed.) (1983):80-107.
- Centre for Research and Innovation Development (1983) School Facilities and SLC Examination Results. In Sharma et al (eds.) (1983):109-153.
- Charney, D.C. (1984) The Validity of Using Holistic Scoring to Evaluate Writing: A Critical Overview. Research in the Teaching of English. 18/1:65-81.
- Clark, J.L. (1987) Curriculum Renewal in School Foreign Language Learning. Oxford. Oxford University Press.
- Cliff, J.C., and Imrie, B.W. (1981) Assessing Students, Appraising Teaching. London. Croom Helen Ltd.
- Coffman, W.E. (1971) Essay Examinations. In Thorndike (ed.) (1971):271-302.
- Cooper, C.R. (1977) Holistic Evaluation. In Cooper and Odell (eds.) (1977):3-31.
- Cooper, C.R., and Odell, L. (1977) Evaluating Writing: Describing, Measuring, Judging. Urbana, III. National Council of Teachers of English.
- Cooper, M., and Leiter, M. (1980) Teachers on Testing. In Stalford (ed.) (1980).
- Criper, C., and Davies, A. (1988) Edinburgh ELTS Validation Project. Final Report. London. The British Council.
- Cronbach, L.J. (1971) Test Validation. In Thorndike (ed.) (1970):443-507.
- Curiton, E.E., and D'Agostino, R.B. (1983) Factor Analysis: An Applied Approach. London. Lawrence Erlbaum.
- Curriculum and Examinations in Secondary Schools (1943) Report of SSEC Appointed by the President of the Board of Education (The Norwood Report). London. HMSO.

- Curriculum Development Centre (1978) The First National Convention of Tribhuvan University Teachers of English. Kirtipur. Tribhuvan University.
- CTSDC (1980) Secondary Education Curriculum. Sanothimi. JEMC.
- Cziko, A.G. (1983) Psychometric and Edumetric Approaches to Language Testing. In Oller (ed.) (1983):289-308.
- Davies, A. (1965) Proficiency in English as a Second Language. Unpublished Ph. D. Thesis. University of Birmingham.
- Davies, A. (1968) Introduction. In Davies (ed.) (1968):1-18.
- Davies, A. (ed.) (1968) Language Testing Symposium: A Psycholinguistic Approach. London. Oxford University Press.
- Davies, A. (1977) The Construction of Language Tests. In Allen and Davies (eds.) (1977):38-104.
- Davies, A. (1978) Language Testing. In Kinsella (ed.) (1982):127-159.
- Davies, A. (1983) The Validity of Concurrent Validation. In Hughes and Porter (eds.) 1983:141-146.
- Davies, A. (1985) Follow my Leader: Is That What Language Tests Do? In Lee et al (eds.) (1985):3-13.
- Davies, A. (1988a) Communicative Language Testing. In Hughes (ed.) (1988):1-15.
- Davies, A. (1988b) Operationalizing Uncertainty in Language Testing: An Argument in Favour of Content Validity. Language Testing. 5/1:32-48.
- Davies, A., Glendinning, E.H., and McLean, A.C. (1984) The English Language Teaching Survey of Nepal, 1983-1984: A Report Presented by a British Council/ODA Survey Team and Addressed to the Ministry of Education and Culture, HMG/Nepal.

- Davis, F.B. (1944) Fundamental Factors of Comprehension in Reading. Psychometrika. 9:185-197.
- Davis, F.B. (1946) A Brief Comment on Thurstone's Note on a Re-analysis of Davis' Reading Tests. Psychometrika. 11:249-255.
- Dechert, H.W. (1983) How a Story is Told in a Second Language. In Faerch and Kasper (eds.) (1983):175-204.
- Denhem, G.H. (1975) Criterion-referenced, Domain-referenced, and Norm-referenced Measurement: A Parallax View. Educational Technology. 15:9-12.
- Department of Education and Science (1975) A Language for Life (The Bullock Report). London. HMSO.
- Department of Education and Science (1977) Why, What, and How. Pamphlet. Assessment of Performance Unit, Information Division.
- Department of Education and Science (1979) Aspects of Secondary Education in England: A Survey by HM Inspectorate of Schools.
- Diederich, P.B. (1974) Measuring Growth in English. Urbana, III. National Council of Teachers of English.
- Diederich, P.B., French, J.W., and Carlton, S.T. (1961) Factors in Judgements of Writing Ability (ETS Research Bulletin RB-61-15). Princeton: Educational Testing Service.
- Dewey, J.C. (1935) The Acquisition of Facts as a Measure of Reading Comprehension. Elementary School Journal. 35:346-348.
- Dulay, H., and Burt, M. (1973) Should We Teach Children Syntax? Language Learning. 23:245-258.
- Ebel, R.L. (1979) Essentials of Educational Measurement. Third Edition. Englewood cliffs, New Jersey. Prentice-Hall Inc.
- Egger, E. (1971) Secondary School Leaving Examinations: Looking Back- Looking Ahead. Strasbourg.

- Evans, L.D. (1942) The Essentials of Liberal Education. Boston. Ginn.
- Evenechko, P., Ollila, L., and Armstrong, K. (1974) An Investigation of the Relationships Between Children's Performance in Written Language and their Reading Ability. Research in the Teaching of English. 8/3:315-326.
- Fabian, P. (1982) Examinations - Why Tolerate their Paternalising? In Heaton (ed.) (1982):24-27.
- Faerch, C., and Kasper, G. (eds.) (1983) Strategies in Interlanguage Communication. London. Longman.
- Farhady, H., (1983) The Disjunctive Fallacy between Discrete-point and Integrative Tests. In Oller (ed.) (1983):311-322.
- Findley, W.G. (ed.) (1963) The Impact and Improvement of School Testing Programs: The Sixty-Second Yearbook of the National Society for the Study of Education. Part II. Chicago. The University of Chicago.
- Freedman, S.W. (1979) Why do Teachers Give the Grades they do? College Composition and Communication. 30:161-164.
- French, S., Slater, J.B., Vassiloglon, M. and Willmot, A.S. (1988) The Role of Descriptive and Normative Techniques in Examination Assessment. In Black and Dockrell (eds.) (1988):15-32.
- Furst, E.J. (1958) Constructing Evaluation Instrument. London. Longman.
- Gannon, P. (1985) Assessing Writing: Principles and Practices of Marking Written English. London. Arnold.
- Gardner, K., (1978) Reading Comprehension. In British Council (1978):65-81.
- Gosling, G.W.H. (1966) Marking English Composition. Victoria. Australian Council for Educational Research.

- Gow, D. (1989) The Test Case. In The Guardian: National Curriculum Special Part 2. 5 September, 1989:23.
- Gradman, H.L., and Spolsky, B. (1975) Reduced Redundancy Testing: A Progress Report. In Randall and Spolsky (eds.) (1975):59-70.
- Guilford, J.P. (1965) Fundamental Statistics in Psychology and Education. Fourth Edition. McGraw-Hill.
- Gurung, H. (1987) Enrolment Pattern in Higher Education in Nepal. In Mainali and Pradhan (eds.) (1987):32-62.
- Haertal, E. (1985) Construct Validity and Criterion-referenced Testing. Review of Education Research Spring. 55/1:23-46.
- Hall, W.S., White, T.G., and Guthrie, L. (1986) Skilled Reading and Language Development: Some Key Issues. In Orasanu (ed.) (1986):89-112.
- Hairstone, M. (1982) The Winds of Change: Thomas Kuhn and the Revolution in the Teaching of Writing. College Composition and Communication. 33/1:76-88.
- Hamp-Lyons, E.M. (1987) Testing Second Language Writing in Academic Setting. Unpublished Ph. D. Thesis. Department of Applied Linguistics, University of Edinburgh.
- Harris, W.H. (1977) Teacher Response to Student Writing: A Study of the Response Patterns of High School English Teachers to Determine the Basic for Teacher Judgement of Student Writing. Research in the Teaching of English. 11:175-185.
- Harrison, A. (1983) Profile Reporting of Examination Results. Schools Council Examinations Bulletin 43. Methuen Educational.
- Harrison, C., and Dolan, T. (1979) Reading Comprehension: A Psychological View Point. In Mackay et al (eds.) (1979):13-23.

- Hatch, E., and Farhady, H. (1982) Research Design and Statistics For Applied Linguistics. Rowley, Mass. Newbury House Publishers, Inc.
- Hauptman, P.C., LeBlanc, R., and Wesche, M.B. (eds.) (1985) Second Language Performance Testing. Ottawa. University of Ottawa Press.
- Heaton, J.B. (1975) Writing English Language Tests. London. Longman.
- Heaton, J.B. (ed.) (1982) Language Testing. Oxford. Modern Language Publications.
- Heywood, J. McGuinness, S., and Murphy, D. (1980) The Public Examinations Evaluation Project. The Final Report. School of Education, University of Dublin, Trinity College.
- Heywood, J., and O'Connor, S.J. (1974) Public Examinations Evaluation Project. Report No. 1. The School of Education, Trinity College, Dublin.
- Himmelweit, H.J. (1967) Towards a Rationalisation of Examination Procedures. Universities Quarterly. (1967) 21/3:359-372.
- Horst, P. (1966) Psychological Measurement and Prediction. California. Woodsworth Publishing Company, Inc.
- Hosenfeld, C. (1984) Case Studies of Ninth Grade Readers. In Alderson and Urquhart (eds.) (1984):231-144.
- Hughes, A. (1986) A Pragmatic Approach to Criterion-referenced Foreign Language Testing. In Portal (ed.) (1986):31-40.
- Hughes, A. (1988a) Achievement and Proficiency: The Missing Link. In Hughes (ed.) (1988):36-44.
- Hughes, A. (1988b) Introducing a Needs Based Test of English for Study in an English Medium University in Turkey. In Hughes (ed.) (1988):134-146.

- Hughes, A. (ed.) (1988) Testing English for University Study
ELT Document 127. Oxford. Modern Language Publications.
- Hughes, A. (1989) Testing for Language Teachers. Cambridge.
 Cambridge University Press.
- Hughes, A., and Porter, D. (eds.) (1983) Current Developments in
Language Testing. London. Academic Press.
- Hutt, M. (1986) Diversity and Change in the Languages of
 Highland Nepal. In Malla et al (eds.) (1986):1-24.
- Independent, The (1989) GCSE will Bring Back Tradition of Final
 Exam. 9 August, 1989:1.
- Ingenkamp, K. (1977) Educational Assessment. NFER Publishing Co.
 Ltd.
- Ingram, E. (1968) Attainment and Diagnostic Testing. In Davies
 (ed.) (1968):70-97.
- Issac, D.J. (1964) Teachers and Examinations. New Era.
 45:132-136.
- Jacobovits, L.A. (1970) Foreign Language Learning: A
Psycholinguistic Analysis of the Issues. Rowley, Mass.
 Newbury House.
- Jacobs, H. L., Zinkgraf, S. A., Wormuth, D.R., Hartfiel, V.F.,
 and Hughey, J.B. (1981) Testing ESL Composition: A Practical
Approach. Rowley, Mass. Newbury House.
- Jeffery, G.B. (1958) External Examinations in Secondary Schools:
Their Places and Functions. London. Harrap.
- Johnson, K. (1981) Questioning some Assumptions about Cloze
 Testing. In Read (ed.) 1981:177-206.
- Johnson, K., and Porter, D. (1982) Perspectives in Communicative
Language Teaching. London. Academic Press.
- Jones, R.L. (1985) Second Language Performance Testing: An
 Overview. In Hauptman et al (eds.) (1985):15-24.

- Jones, R.L., and Spolsky, B. (eds.) (1975) Testing Language Proficiency. Arlington. Center for Applied Linguistics.
- Jordan, R.R. (1978) The Reading Interests of Lower Secondary School Children in Africa and Asia. ELT Journal. 32/2:143-151.
- Kaiser, H.F. (1970) A Second Generation Little Jiffy. Psychometrika. 35:401-415.
- Kellaghan, T., and Maduas, G.F., and Airasia, P.W. (1982) The Effects of Standardized Testing. London. Kluwen:Nijhoff Publishing.
- Kelly, R. (1978) The Construct Validation of Comprehension Tests: An Exercise in Applied Linguistics. Unpublished Ph. D. Thesis. University of Queensland.
- Khaniya, T.R. (1988) A Comparison Between a Course Book Text and A Real Text. Unpublished M Sc. Discourse Studies. Portfolio. Department of Applied Linguistics, University of Edinburgh.
- Kingdon, M., and Stobart, G. (1988) GCSE Examined. London. The Falmer Press.
- Kinsella, V. (ed.) (1982) Survey 1: Eight State of the Art Articles on Key Areas in Language Teaching. Cambridge. Cambridge University Press.
- Krashen, S.D. (1976) Formal and Informal Linguistic Environments in Language Acquisition and Language Learning. TESOL 10:157-168.
- Krashen, S.D. (1984) Writing: Research, Theory, and Applications. Oxford. Pergamon.
- Lado, R. (1961) Language Testing: The Construction and Use of Foreign Language Tests. London. Longman.
- Lauwerys, J.A., and Seanlon, D.G. (1969) The World Year Book Education 1969. London. Evans.

- Lee, Y. P. (1981) Investigating the Validity of the Cloze Score. In Lee et al (eds.) (1985):137-147.
- Lee, Y.P., Fok, A.C.C.Y., Lord, R. and Low, G. (eds.) (1985) New Directions in Language Testing. Oxford. Pergamon.
- Lister, H. (1958) The Effects of External Examinations on the School. In Jeffery (ed.) (1958):55-77.
- Li, Xiaoju (1989) (Forthcoming) How Powerful Can a Test be - the MET in China. Paper Presented at the BALL 22nd Anniversary Annual Meeting at Lancaster.
- Llyod-Jones, R. (1977) Primary Trait Scoring. In Cooper and Odell (eds.) (1977) 33-66.
- Lunzer, E., and Gardner, K. (eds.) (1979) The Effective Use of Reading. Heinemann.
- Lunzer, E., Waite, M., and Dolan, T. (1979) Comprehension and Comprehension Tests. In Lunzer and Gardner (eds.) (1979):37-70.
- Mackay, R., Barkman, B., and Jordan, R.R. (eds.) (1979) Reading in a Second Language. Rowley, Mass. Newbury House.
- Madsen, H.S. (1976) New Alternatives in EFL Exams or, 'How to Avoid Selling English Shorts'. In ELT Journal 30/2:135-143.
- Madsen, H.S. (1983) Techniques in Testing. Oxford. Oxford University Press.
- Mainali, M.P. and Pradhan, G.S. (eds.) (1987) Education and Development. Kathmandu. CERID.
- Mainali, M.P., Giri, J.R., Pande, B.D., Lamichhane, S.R., and Pradhan, G.S. (eds.) (1988) Secondary Education in Nepal. Kathmandu. CERID
- Malla, K. (1975) English Teaching in Nepal: An Investigation. In Malla (1977):9-20.

- Malla, K. (1977) English in Nepalese Education. Kathmandu. Ratna Pustak Bhandar.
- Malla K.P. (1977) English for Nepalese Education: A Case for Coordination. In Curriculum Development Centre (1978):1-13.
- Malla, K., Tuladhar, N.M., Vajracharya, C.D., Dahal, D.L. and Mishra, C. (eds.) (1986) Contributions to Nepalese Studies. 14/1.
- Moller, A. (1982) A Study in the Validation of Proficiency Tests of English as a Foreign Language. Unpublished Ph. D. Thesis. Department of Applied Linguistics, University of Edinburgh.
- Morris, B. (1972) Objectives and Perspectives in Education: Studies in Educational Theories. London. Routledge and Kegan Paul.
- Morrow, K. (1977) Techniques of Evaluation for a Notional Syllabus. The Royal Society of Arts.
- Morrow, K. (1979) Communicative Language Testing: Revolution or Evolution. In Brumfit and Johnson (eds.) (1979):143-158.
- Morrow, K. (1986) The Evaluation of Tests of Communicative Performance. In Portal (ed.) (1986) 1-13.
- Munby, J. (1979) Teaching Intensive Reading Skills. In Mackay et al (eds.) (1979):142-158.
- Nisbet, J.D. (1969) The Need for Universities to Measure Achievement in the Assessment of Undergraduate Performance. London. Report of Conference. Committee of Vice-Chancellors and Principals and Association of University Teachers.
- Northey, W.B. and Morris, C.J. (1929) The Gurkhas. London. The Bodley Head.
- Nuttall, C. (1982) The Teaching of Reading Skills in a Foreign Language. Heinemann.

- Nuttall, D.L. (1984) Domsday or a New Down? The Prospects for a Common System of Examining at 16+. In Broadfoot (ed.) (1984):163-178.
- Odell, L. (1981) The Process of Writing and the Process of Learning. College Composition and Communication. 31:42-50.
- Oller, J.W.Jr. (1978) Pragmatics and Language Testing. In Spolsky (ed.) (1978):39-57.
- Oller, J.W.Jr. (1979) Language Tests at School: A Pragmatic Approach. London. Longman.
- Oller, J.W.Jr. (ed.) (1983) Issues in Language Testing Research. Rowley, Mass. Newbury House.
- Oller, J.W.Jr., and Inal, N. (1971) A Cloze Test of English Prepositions. TESOL Quarterly. 5:315-326.
- Oller, J.W.Jr., and Khan, F. (1980) Is There a Global Factor of Language Proficiency. In Read (ed.) (1981):3-40.
- O'Meara, J.F. (1944) A Critical Study of External Examinations and of their Influence on Secondary Education. Unpublished Master's Dissertation. University of College Cork.
- Oppenheim, A. N., Jahoda, M. and James, R.L. (1967) Assumptions Underlying the Use of University Examinations. In Universities Quarterly. June, 1967: 341-345.
- Orasanu, J. (ed.) (1986) Reading Comprehension: from Research to Practice. Hillsdale, New Jersey. Lawrence Erlbaum.
- Pande, B.D. and Shrestha, N., and Sigdel, L. (1984) Linkage Between Secondary School Curriculum and Proficiency Certificate Level Curriculum: Report of a Research Study. Kirtipur. CERID.
- Palmer, A.S. (1981) Measurement of Reliability and Validity of Two Picture Description Tests of Oral Communication. In Palmer et al (eds.) (1981):127-139.

- Palmer, A.S., and Bachman, L.F. (1981) Basic Concern in Test Validation. In Alderson and Highes (eds.) (1981):135-151.
- Palmer, A.S., Groot, P.J.M., and Trostler, G.A. (eds.) (1981) The Construct Validation of Tests of Communicative Competence. Washington. D.C. TESOL.
- Pearson, I. (1988) Tests as Levers for Change. In Brumfit (ed.) (1988):98-107.
- Perl, S. (1979) The Composing Process of Unskilled College Writers. Research in the Teaching of English. 13/4:317-336.
- Peterson, C.R., and Cartier, F.A. (1975) Some Theoretical Problems and Practical Solutions in Proficiency Test Validity. In Jones and Spolsky (eds.) (1975):105-118.
- Pilliner, A.E.G. (1969) Multiple Marking: Wiseman or Cox? British Journal of Educational Psychology. 39:313-315.
- Pilliner, A. (1973) Assessment- Principles and Practice With Special Reference to Education in Pakistan. The British Council.
- Portal, M. (ed.) (1986) Innovations in Language Testing. London. NFER-NELSON.
- Porter, D. (1982a) Prepared Comments by Don Porter. In Johnson and Porter (eds.) (1982):38-43.
- Porter, D. (1982b) Assessing Communicative Proficiency: the Search for Validity. In Johnson and Porter (eds.) (1982):189-204.
- Ranson, S. (1984) Towards a Tertiary Tripartism: New Codes of Social Control and the 17+. In Broadfoot (ed.) (1984):221-144.
- Rea, P.M. (1978) Assessing Language as Communication. M.A.L.S. Journal. New series No. 3. University of Birmingham.

- Read, J.A.S. (1981) Introduction. In Read (ed) (1981):ix-xvii.
- Read, J.A.S. (ed.) (1981) Directions in Language Testing.
Singapore University Press. SEAMEO RELC.
- Remondino, C. (1959) A Factorial Analysis of the Evaluation of Scholastic Compositions in the Mother Tongue. British Journal of Educational Psychology. 29:242-251.
- Richards, J., Platt, J. and Weber, H. (1985) Longman Dictionary of Applied Linguistics. London. Longman.
- Rogers, J. (1982) 'The World for Sick Proper'. ELT Journal. 36/3:144-151.
- Rowntree, D. (1977) Assessing Students: How Shall We Know Them ? Harper Row, Ltd.
- Sanders, J., and Sachse, T. (1977) Applied Performance Testing in the Classroom. Journal of Research and Development in Education. 10(3):92-104.
- Scarth, J. (1984) Teachers' Attitudes to Examining: A Case Study. In Broadfoot (ed.) (1984):83-102.
- Schulze, R.A. (1977) Discrete-point Versus Simulated Communication Testing in Foreign Languages. The Modern Language Journal. LXI/3:94-100.
- Seaton, I. (1985) Issues in the Validation of the English Language Testing Service (ELTS) 1976-1983. In Hauptman et al (eds.) 1985:111-121.
- Sharma, S.R., Kasaju, P.K., and Pradhan, G.S. (eds.) (1983) Education and Development. Kathmandu. CERID.
- Shrestha, R. (1979) Workshops, Seminars and Conventions for Campus Level Teachers of English. Curriculum Development Quarterly (1979):22-24.
- Sinclair, J., Fox, G., Moon, R. and Stock, P. (1987) English Language Dictionary. London. Collins.

- Skehan, P. (1988) Language Testing. Parts I & II. Language Teaching Abstracts. Cambridge. Cambridge University Press.
- Spearman, C. (1904) The Proof and Measurement of Association Between Two Things. American Journal of Psychology. 15:72-101.
- Spencer, E. (1983) Writing Matters.:Across the Curriculum. The Scottish Council of Research in Education.
- Spiro. R.J., Bruce, B.C., and Brewer, W. (eds.) (1980) Theoretical Issues in Reading Comprehension: Perspectives from Cognitive Psychology, Linguistics, Artificial Intelligence, and Education. Hillsdale, New Jersey. Lawrence Erlbaum.
- Spolsky, B. (1978) Introduction: Linguists and Language Testers. In Spolsky (ed.) (1978):v-x.
- Spolsky, B. (ed.) (1978) Advances in Language Testing Survey-2: Approaches to Language Testing. Virginia. Center for Applied Linguistics.
- Spolsky, B. (1985) What Does it Mean to Know How to Use a Language? An Essay on the Theoretical Basis of Language Testing. Language Testing. 2.2:180-191.
- Stalford, C.B. (ed.) (1980) Testing and Evaluation in Schools: Practitioner's Views. Washington, D.C. National Institute of Education, U.S. Department of Education.
- Stalnaker, J.M. (1934) The Construction and Results of a Twelve Hour Test in English Composition. School and Society. 39:218-224.
- Steel, J.H., and Talman, J. (1936) The Marking of English Composition. Nisbel.
- Stevenson, D.K. (1985) Pop Validity and Performance Testing. In Lee et al (eds.) 1985:111-118.

- Summer, R. (1987) The Role of Testing in Schools. London NEFR-NELSON.
- Swain, M. (1985) Large Scale Communicative Language Testing: A Case Study. In Lee et al (eds.) (1985):35-46.
- Taba, H. (1962) Curriculum Development: Theory and Practice. New York. Harcourt, Brace and World.
- Taylor, B.P. (1981) Content and Written Form: A Two Way Street. TESOL Quarterly 15/1:5-14.
- Taylor, H.P., and Richards, C.M. (1985) An Introduction to Curriculum Studies. Second Edition. London. NEFR-NELSON.
- Taylor, W.L. (1957) "Cloze" Readability Scores as Indices of Individual Differences in Comprehension and Aptitude. Journal of Applied Psychology. 41:19-26.
- Tharu, J. (1981) Examinations Reforms: Towards an Integrated Perspective. All India Hyderabad Seminar on Examination Reforms.
- Thorndike, R.L. (ed.) (1971) Evaluational Measurement. Second Edition. Washington D.C. Council of Education.
- Thorndike, R.L., and Hagen, E.P. (1977) Measurement and Evaluation in Psychology and Education. Fourth Edition. New York. John Wiley and Sons.
- Thyne, J.M. (1974) Principles of Examining. University of London Press Ltd.
- Tibble, J.W. (1969) The Educational Effects of Examinations in England and Wales. In Lauwerys and Seanlon. (eds.) (1969):347-354.
- Turner, L.R. (1928) The People and their Languages. In Northey and Morris (eds.) (1928).
- Tyler, R.W. (1969) Educational Effects of Examinations in the United States. In Lauwerys Seanlon (eds.) (1969):342-346.

- Upshur, J. (1968) Four Experiments on the Relation between Foreign Language Teaching and Learning. Language Learning. 18:111-124.
- Verma, Y. and Pande, B.D. (1988) Causes of Failures in English in SLC Examinations: A Study Report. Kathmandu. CERID.
- Weir, C. J. (1983a) Identifying the Language Needs of Overseas Students in Tertiary Education in the United Kingdom. Unpublished Ph.D. Thesis. University of London.
- Weir, C. J. (1983b) The Associated Examining Board's Test in English for Academic Purposes: An Exercise in Content Validation. In Hughes and Porter (eds.) (1983):147-153.
- Weir, C.J. (1988) Communicative Language Testing: With Special Reference to English as a Foreign Language. Exeter. University of Exeter.
- Wesche, M.B. (1985) Introduction. In Hauptman et al (eds.) (1985):1-12.
- Widdowson, H.G. (1978) Teaching Language as Communication. Oxford. Oxford University Press.
- Wilkinson, A. (1968) The Testing of Oracy. In Davies (ed.) (1968):117-132.
- Wiseman, S. (1961) The Efficiency of Examinations. In Wiseman (ed.) (1961) 133-164.
- Wiseman, S. (ed.) (1961) Examinations and English Education. Manchester University Press.
- Wong, R.H.K. (1969) Educatinal Effects of Examinations on Pupils, Teachers, and Society. In Lauwerys and Seanlon (eds.) (1969):361-367.
- Woods, A., Fletcher, P., and Hughes, A. (1986) Statistics in Language studies. Cambridge. Cambridge University Press.

- Wrightstone, J.W. (1963) The Relation of Testing Programs to Teaching and Learning. In Findley (ed.) (1963):45-61.
- Yalden, J. (1987) The Communicative Syllabus: Evolution, Design and Implementaion. Englewood Cliffs, New Jersey. Prentice-Hall International.
- Zamel, V.C. (1983) The Composing Process of Advanced ESL Students: Six Case Studies. TESOL Quarterly. 17/2:105-1.

MATERIALS USED

- Gautam, B. (1988) Gautam Guide to English Reader (I&II) and SLC Exam Courses for IX and X SLC (New and Old). Kathmandu. Gautam Prakashan.
- Green, G. (1972) Read, Think, and Answer. Oxford. Oxford University Press.
- Howe, D.H. (1977) New Guided English. Oxford. Oxford University Press.
- JEMC (1986) English Reader.: Grade 9-10, Part I. Ministry of Education, Nepal
- JEMC (1988) English Reader: Grade 9-10, Part II. Ministry of Education, Nepal.
- Ryan, B.F., Joiner, B.L., and Ryan, Jr.J.A. (1985) Minitab:Handbook. 2nd Edition. Boston. Pws-kent Publishing Company.
- Sharma, P.K. (1988) Guess Paper with High Surety for SLC 2045 English (Part I & II). Kathmandu. D.R. Shakya.
- Surya Prakashan (1988) Gautam Super Lucky SLC Guess Paper.

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—1*)
- (d) How does a steam engine work?
(*Transport—1*)
- (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.

- (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?
- (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?
3. Pretend that you are an angry farmer. Write a newspaper article of about 150 words explaining why the tiger must be killed off.

Or

Write a paragraph on "Tribhuvan Highway" with the help of the following notes :

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.....
(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.....
(first, second)
- (f) The pair "....." has the same pronunciation.
(fool, full; hut, hot; sit, seat; son, sun)

Group "A"

Attempt any four : 6 x 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 they had to play.
The teacher made.....

- (b) The teacher always gets us to write down the notes.
The teacher always makes.....
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 isuseful 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. (i)
- (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 x 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 : 4
 (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) 4
 (a) explode (b) discover (c) ignorance (d) move (e) life (f) beauty

- 14. (i) Give the opposites of any two of the following words : 2
 (a) confuse. (b) lure (c) remote
 (ii) Give the opposites of any two of the following words : 2
 (a) dull (b) truth (c) dangerous

- 15. Choose the correct word from within the brackets to fill in the blank : (any four) 4
 (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)

Questionnaire for People involved in SLC English

1. The general objectives of teaching English at Grades 9-10 have been mentioned in the curriculum. To what extent, the existing SLC covers them?

- a. Speaking
- b. Listening
- c. Reading
- d. Writing

2. Does the SLC English exam assess the true abilities of the SLC candidates?

Yes/No

3. Have you had chief examiners' report or comment on the SLC English exam?

Yes/No

4. Does SLC the English exam look like a language test ?

Yes/No

5. Does SLC English exam assess the activities that a SLC passer would need to do after he passes the SLC English exam?

Yes/No

6. Any comment on SLC English exam.

(To be filled in by teachers only)

What effect does the SLC English exam have on your classroom teaching at Grades 9-10?

i. Are you free to teach what you think good for the students?
Yes/No

ii. Do you have to prepare answers for the possible questions to be asked in the SLC English exam?
Yes/No

iii. Are you selective about what have already been asked and what will be asked in the coming year in the SLC English exam?
Yes/No

iv. Do you have to spend a considerable amount of time on preparing the SLC candidates for the SLC English exam apart from their prescribed course?

Yes/No

Name.....
Position.....
Institute.....

Questionnaire for College Students
(Translated Version)

1. Which of the subjects you have been studying at this level require the knowledge of English?
Please write the names of the subjects.....

2. Which of the following activities you need to carry out in the subjects knowledge of English require?

- i. To read books written in English.
- ii. To understand lectures in English.
- iii. To take notes or write answers in English.
- iv. To ask/answer in English.

3. All of us have studied some English at school. Comparing the English you studied at school to the English you need at tertiary level, do you think that the compulsory English at school is adequate?

Yes/No

If you think it is inadequate, which of the following activities do you find difficult to carry out?

- i. Listening
- ii. Speaking
- iii. Reading
- iv. Writing

4. Of all the subjects you have been studying, do you think there is /are subject/s which you find difficult only because you do not have enough English?

If so, which of the activities as abovementioned in Nos. 2 and 3, do you find difficult to carry out?

5. Do you think that you would have done better than you did in the first year courses if you had had better English than you had that time?

Yes/No

6. In what medium, did you study Science and Maths at school? English/Nepali.

If it was Nepali, how difficult have you found them to study now in English at tertiary level?

Not difficult/ Difficult/Very difficult.

7. What is your opinion of teaching, at least, Science and Maths in English?

Correct/ Incorrect

8. Do you feel that had you been taught science and maths in English, you would have done better than you did at your present study?

9. Any comment.

Your school.....

Score in English paper in the SLC.....

Overall performance in the first year.....

University Lecturers' Questionnaire

Below are some of the reading and writing skills necessary for Intermediate students who have to do courses in English. Of the skills listed below, some may be more important than others. Indicate from each list the five skills which you think are the most important for your students. Rank them in order of importance by putting numbers (1,2,3,4,5) in the spaces provided.

Reading

- () Reference skills, e.g. recognising titles, headings, index
- () Guessing meaning of unfamiliar words from context
- () Skimming: i. surveying to obtain gist of a text,
ii scanning the text to locate specifically
required information on a single point
- () Understanding relations within the sentence
- () Being able to understand parts of a text which are connected
- () Understanding relations between parts of text by recognising indicators in discourse, e.g. clues, linking signals, indicators used to introduce or sum up ideas,
- () Understanding conceptual meaning, e.g. quantity and amount, direction, condition
- () Understanding explicitly stated ideas and information.
- () Understanding ideas and information in a text not explicitly stated
- () Separating the essential from the non-essential in a text
- () Interpreting information presented in a non-linguistic form, e.g. tables, graphs, diagrams

Writing activities

- () Expressing ideas, opinions
- () Writing personal essays
- () Writing/replying to letters
- () Summarising a passage
- () Making comments
- () Describing events, objects, graphs, pictures, processes
- () Presenting arguments
- () Reporting events
- () Persuading
- () Making complaints
- () Writing assumptions, speculations, clarifications

Traits to be emphasised in teaching and testing writing

- () CONTENT: e.g. knowledge of subject matter, thorough development of argument, relevant to the topic in question
- () ORGANISATION: e.g. logical sequence, fluency, supporting details, completeness
- () VOCABULARY: e.g. effective word/idiom, appropriate register
- () LANGUAGE: e.g. grammatical correctness, effective construction,
- () MECHANICS: e.g. spelling, punctuation, convention

Name of the Faculty:

ESL COMPOSITION PROFILE				
STUDENT	DATE	TOPIC		
	SCORE	LEVEL	CRITERIA	COMMENTS
CONTENT	30-27		EXCELLENT TO VERY GOOD: knowledgeable • substantive • thorough development of thesis • relevant to assigned topic	
	26-22		GOOD TO AVERAGE: some knowledge of subject • adequate range • limited development of thesis • mostly relevant to topic, but lacks detail	
	21-17		FAIR TO POOR: limited knowledge of subject • little substance • inadequate development of topic	
	16-13		VERY POOR: does not show knowledge of subject • non-substantive • not pertinent • OR not enough to evaluate	
ORGANIZATION	20-18		EXCELLENT TO VERY GOOD: fluent expression • ideas clearly stated/ supported • succinct • well-organized • logical sequencing • cohesive	
	17-14		GOOD TO AVERAGE: somewhat choppy • loosely organized but main ideas stand out • limited support • logical but incomplete sequencing	
	13-10		FAIR TO POOR: non-fluent • ideas confused or disconnected • lacks logical sequencing and development	
	9-7		VERY POOR: does not communicate • no organization • OR not enough to evaluate	
VOCABULARY	20-18		EXCELLENT TO VERY GOOD: sophisticated range • effective word/idiom choice and usage • word form mastery • appropriate register	
	17-14		GOOD TO AVERAGE: adequate range • occasional errors of word/idiom form, choice, usage <i>but meaning not obscured</i>	
	13-10		FAIR TO POOR: limited range • frequent errors of word/idiom form, choice, usage • <i>meaning confused or obscured</i>	
	9-7		VERY POOR: essentially translation • little knowledge of English vocabulary, idioms, word form • OR not enough to evaluate	
LANGUAGE USE	25-22		EXCELLENT TO VERY GOOD: effective complex constructions • few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions	
	21-18		GOOD TO AVERAGE: effective but simple constructions • minor problems in complex constructions • several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions <i>but meaning seldom obscured</i>	
	17-11		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, deletions • <i>meaning confused or obscured</i>	
	10-5		VERY POOR: virtually no mastery of sentence construction rules • dominated by errors • does not communicate • OR not enough to evaluate	
MECHANICS	5		EXCELLENT TO VERY GOOD: demonstrates mastery of conventions • few errors of spelling, punctuation, capitalization, paragraphing	
	4		GOOD TO AVERAGE: occasional errors of spelling, punctuation, capitalization, paragraphing <i>but meaning not obscured</i>	
	3		FAIR TO POOR: frequent errors of spelling, punctuation, capitalization, paragraphing • poor handwriting • <i>meaning confused or obscured</i>	
	2		VERY POOR: no mastery of conventions • dominated by errors of spelling, punctuation, capitalization, paragraphing • handwriting illegible • OR not enough to evaluate	
	TOTAL SCORE	READER	COMMENTS	

Name SLC Symbol No.

School (Full Address)

Date Marks on the qualifying test: (English paper only)

A SECONDARY SCHOOL ENGLISH TEST.

This test contains two parts: Part 1 and Part 2:

Part 1: Reading (1.30 hrs) Part 2: Writing (1.30 hrs)

Part 1 contains 3 Reading Passages.

Part 2 contains the following exercises;

Passage 1: Q. 1 to Q. 10,

Exercise 1: Note taking

Passage 2. Q. 11. to Q. 20.

Exercise 2: Letter writing

Passage 3. Q. 21. to Q. 55.

Exercise 3: Essay writing

Answer all Questions.

You are strongly advised to allow time approximately as follows:

Reading:

Writing:

Passage 1. (only 25 minutes)

Note taking 25 (minutes)

Passage 2. 35. (minutes)

Letter writing (25 mins).

Passage 3. 25. (minutes)

Essay writing (35 mins)

You are also advised to spend 10 minutes on reading instructions;

January/February 1989

Reading task 1.

'I ask you to drink to his health as a young man full of the spirit of adventure who has lit up the world with a flash of courage'. With these words the British Minister of Air turned and raised his glass to the young man who sat beside him—a young man, only a month before, was completely unknown. Yet on that summer day in 1927 his name was on the world's lips—Charles Lindbergh, the first man to fly the Atlantic alone.

He had been an air mail pilot, flying back and forth between Chicago and the city of St. Louis. Determined to win the \$25,000 prize offered by a fellow-American for the first flight from New York to Paris, Lindbergh had persuaded a group of St. Louis businessmen to finance the building of a special plane for him.

The news that Lindbergh intended to fly the Atlantic alone was received with disbelief. The plane would never fly, people said. It would run out of fuel. It had only a single engine. Lloyds refused to insure the flight. Men called Lindbergh the 'flying fool'.

But on May 20th, 1927 just after ten to eight in the morning, Lindbergh's 'Spirit of St. Louis', heavily laden with fuel, struggled into the air from a New York airfield. For several hours the weight of the petrol prevented the young pilot from flying more than a few feet above the wave-tops. Night came and thick fog covered up the stars. Lindbergh flew steadily on hoping that his course was the right one. He struggled to keep awake, checking the fuel all the time to keep his mind active. Throughout the next day the 'Spirit of St. Louis' flew on over the seemingly limitless sea. Then a fishing boat appeared, and, an hour later, land. It was Ireland. Lindbergh set a compass course for Paris.

By ten o'clock the lights of France's capital were shining beneath him. Tired, unshaven, suddenly hungry, the 'flying fool' came down to Le Bourget airport, and landed in front of a huge crowd of wildly cheering people. After 34 hours of continuous piloting, the flight of 3,600 miles was over.

(From Gordon Green : Read, Think and Answer)

(✓) Tick the best Answer.

1. What is the passage about?

a. The British Minister's speech.

- b. The diplomatic relationship between New York and Paris.
 - c. The first flight across the Atlantic.
 - d. A holiday trip to Paris.
2. Who is the 'Hero' of the passage?
- a. A fellow-American.
 - b. The British Minister of Air.
 - c. Charles Lindbergh.
 - d. A group of St. Louis businessmen.
3. In paragraph 3, there are six sentences. Which of them best gives the main idea of the paragraph?
- a. First
 - b. Third.
 - c. Fifth.
 - d. Sixth.

Now read the following questions first, then find out the specific answers from the passage. Answer in a word or phrase.

4. What was the reward for the first flight from New York to Paris?

.....

5. What time did Lindbergh leave New York for Paris?

.....

6. How long did the flight take?

.....

7. How far is Paris from New York?

.....

8. What was the name of the plane, Lindbergh used?

.....

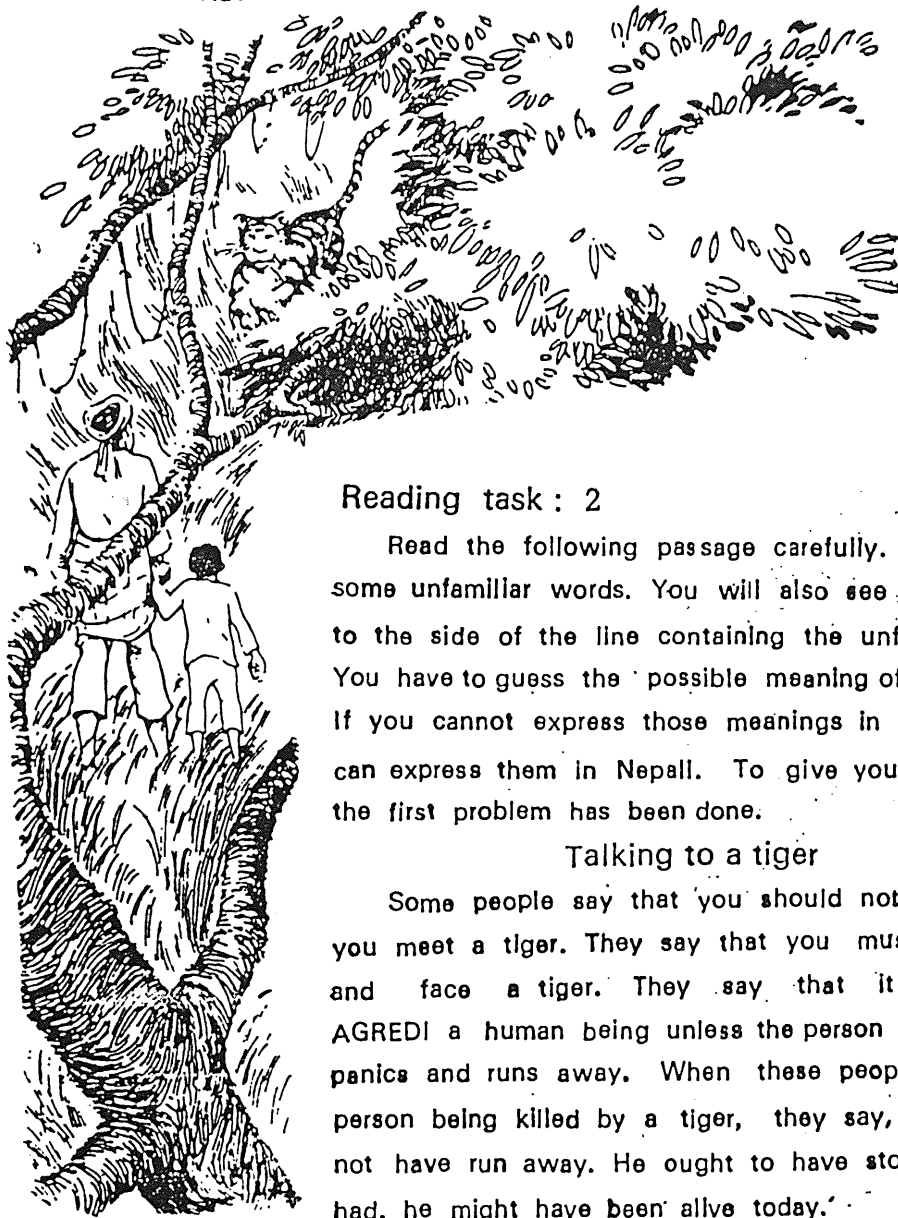
9. Who paid to build his plane?

.....

10. What was Lindbergh's job?

.....

Reading task : -



Reading task : 2

Read the following passage carefully. You will find some unfamiliar words. You will also see a blank box to the side of the line containing the unfamiliar word. You have to guess the possible meaning of those words. If you cannot express those meanings in English, you can express them in Nepali. To give you an example, the first problem has been done.

Talking to a tiger

Some people say that you should not run away if you meet a tiger. They say that you must stand still and face a tiger. They say that it will not AGRED a human being unless the person attack panics and runs away. When these people hear of a person being killed by a tiger, they say, 'He should not have run away. He ought to have stood still. If he had, he might have been alive today.'

In the olden days in Malaysia, they used to call this 'talking' to the tiger. 'When you meet a tiger,' they used to say, 'you should speak to it. You should tell it that there is plenty of food in the jungle and that it need not eat you';

Unfortunately most people who were attacked by a tiger did not get an BOSSILIDADE to speak to it! The tiger crept up behind the person | _____ | that it intended to attack. However there is a story of an old man who used this method to avoid death.

He was returning home, after a visit to a town, to his house at Selita. He was passing through a stretch of road which was thickly surrounded by jungle. He had his little child with him, a boy of seven or eight years, who was contentedly chewing some jajung. Suddenly the man looked up and saw a tiger crouching in front of him. It was apparently preparing to spring.

The old man remembered the old saying He stood still and managed to GRIT out a few sounds. The | _____ | tiger did not move but the old man was anxious not to risk the life of his son. He therefore moved slowly backward to a tree which he remembered passing. As he retreated always facing the tiger the animal advanced towards him step by step. At last his back touched the tree.

'Now,' he said to his son; 'you must climb up the tree slowly and quietly. You must not alarm the tiger. I shall keep on talking to it'.

When his son was safely up the tree, the brave old man APAN his parang. Then he began to | _____ | advance on the tiger. As he did so, he spoke to the tiger' He pointed out that it would be better for them to part without quarrelling. He pointed out that one of them might be FERIDO or killed. | _____ |

'You needn't fight me,' he said. 'There is plenty of food for you in the jungle. You must be able to see how sharp this parang is. You are wasting time with me. You could have killed something else in this time. If you had not stopped to talk to me, you might have caught a fat deer.'

He advanced in this way for about fifty yards, talking all the time, while the tiger retreated.

Finally the tiger turned round and ran off into the jungle. I do not know whether it was VENCIDO by the man's bravery or by his arguments, but the | _____ | man and the boy were both safe. They might both have been killed and eaten.

(From Howe. D. H. : New Guided English)



Now read the passage again. Then answer the the following questions.

(✓) Tick the best answer

16. The man moved backward to.....

- a) get away from the tiger.
- b) keep his son out of sight of the tiger.
- c) get near the tiger.
- d) climb up the tree.

17. As he moved backward, the tiger... ..

- a) came nearer to him.
- b) moved away from him.
- c) kept looking at him.
- d) began to climb up the tree.

18. What does 'who' refer to in paragraph 4 line 4 ? (who ले केलाई संकेत गर्दछ)

- a) The son.
- b) The father.
- c) Both father and son.
- d) The tiger

19. He drew the parang because he wantedto

- a) fight the tiger.
- b) kill the tiger
- b) kill a deer for the tiger.
- c) advise the tiger not to attack.

20. The following statements are based on the passage presented above. Some of them are important and some are non important.

(तल लेखिएको भनाईहरू मध्ये केहि महत्वपूर्ण र केहि महत्व हिन छन)

- i) choose the 4 important expressions, and
- ii) organise them as they appear in the passage.

To show their order of occurrence, write down 1, 2, 3 and 4 in the space provided.

(Passageमा ती भनाईहरूको उपस्थितिको क्रम देखाउन तिनीहरूको अगाडि १, २, ३, ४ लेख्नुस्)

- () The old man saw the tiger preparing to attack him.
- () The eight years old child was eating happily.
- () There is a saying in Malaysia that a tiger does not necessarily attack a man.
- () The old man was from a place called Sallta.
- () The old man advised the tiger to eat a deer.
- () The tiger could not overcome the old man, so it ran away.
- () The tiger wasted time in following the old man.
- () As the old man moved backward, the tiger followed him.
- () People in Malaysia used to talk to the tiger.

Reading task : 3

Instructions : Read the following story very carefully. 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. Do not get afraid to guess. No marks will be taken off for wrong spelling. Try every blank.

As an example, one blank has been filled in for you.

A Meeting with Death

One day a rich merchant of Baghdad sent his servant to the market to buy food. The servant returned very quickly and rushed into his master's room. His eyes were wide and his whole body was shaking with **fear**

"Master!" he cried, "As I was.....through the crowded market, a womanagainst me and nearly fell. When.....turned round to help her, I..... that it was Death and she.....her hand and was about to.....me. I ran from that evil and did not look back. Now.....must escape from this town at....."

"But how did you know thiswas Death?" asked the merchant.

"Igood reason to know her," repliedservant. "Two years ago when I..... a fever, I was alone with....., face to face for six daysnights. And once, only a month ago,looked straight at me when athrew a knife at me and..... cut me an inch above my.....Please lend me your fastest horse.....I will ride like the wind.....Samarra where I will be safe.".....merchant lent the poor man his..... and watched him gallop away to.....north in a cloud of dust..... the merchant walked to the market... looked around for the woman. At..... he saw a woman who was...stranger to him. He greeted hertold her who he was. "Why.....you lift your hand to takeservant?" he asked.

"I did notto take him then," she replied. "... raised my hand in surprise I..... not expect to see your servant in Baghdad today, because I arranged long ago to meet him in Samarra tonight."

(From JEMC: English Reader)

Note taking—writing task: 1

Read the following passage carefully and make notes on it using the form provided. Try to be precise. It is not necessary to write a complete sentence.

China Today

The most extraordinary thing about China is its size. China is so big that England could get lost in it. It is about fifty times bigger than England. When I went by train from Peking to Canton, the trip took forty-six hours and I covered almost 1,500 miles. Yet I was crossing only a small corner of the country.

Nobody knows exactly how many people there are in China, but the number is certainly getting bigger all the time because there are far more births than deaths. The present population is believed to be about 750 million. This is four times the population of the United States and about thirteen times more than that of Britain.

About one-fifth of the world's population lives in China. Compared with the people of Europe and North America most of the people are very poor. This is because China is mainly an agricultural country and they are short of machinery. In my whole journey to Canton I never once saw a tractor, though I was travelling through farmland most of the way. So almost all the hard work must be done by hand.

The main crop is rice. Water buffalo are used for ploughing. At harvest time the rice is beaten out, grain by grain into woden boxes in the fields because there are very few machines. Production in China could certainly be doubled and probably trebled if more modern methods were used.

There are almost no luxuries in China. For instance, in the whole of Peking there are only about half-a-dozen private cars. But the Chinese people are much better off than they used to be. Many factories have been built to make the goods needed, especially goods such as machinery. At least nobody is hungry.

For thousands of years most of the Chinese were little better off than slaves. Since the revolution, China has become a communist country. There are many people in the west who say that this is just another kind of slavery. But conditions are improving. It would be foolish to deny this.

(From Gordon Green : Read, Think and Answer)

1. China... ..

i. size:

ii. population:

2. Economy... .. poor:

The reasons are,

i.....

ii.....

iii.....

3. Situation improving; because.....

i.....

ii.....

Continue

Letter writing-writing task : 2

Instruction: Read the following letter and reply to it.

Department of Applied Linguistics
14, Buccleuch place
University of Edinburgh
U K

Date.....

Dear

..... School

I am a student at the University of Edinburgh in Britain. I have been studying here since 1987.

I am writing to you to tell that I could not celebrate Dasain, as you did there. The reason is, that most people in Britain are christian and do not celebrate Dasain. However, we-the Nepalese students in Edinburgh got together on the day of Tika. We ate 'BHAT' and 'MASHU' and sang some Nepalese songs that evening.

I am sure you celebrated Dasain, especially the three days-Austhami, Nawami and Tika-with a variety of activities. You must have got new clothes as well. I would like to know how you celebrated it. Could you please write and tell me how your Dasain was. I am looking forward to hearing from you. Good luck with your S.L.C. exam.

Thank you.

Yours

(Tirth Raj Khanly)

Writing task : 3

Your school is going to celebrate its anniversary. One of the activities included in the programme is an essay competition. Only the students who have completed their study this year can participate in the programme.

There will be prizes for the 3 best essays.

The topic that has been chosen for the essay is 'My School',

Write an essay in 250 words approximately. The following points may help you to organise your writing. However, you are free to leave out any points you want and you are also free to include your own points

The points are;

1. a. Location.
 - b. The buildings.
 - c. Physical facilities; play ground, equipment etc.
2. a. The subjects taught.
 - b. The school family: teacher-student relationship.
 - c. Its academic performance.
3. Any suggestions for improvement.

You are advised to spend 5 minutes to plan your writing.

Plan here.

Column 1= Serial No.

Column 2= School Type: 1= Type A
2= Type B
3= Type C
4= Type D

Column 3= Grade: 2 = Grade 10

Column 4= Score in Reading

Column 5= Score in Cloze

Column 6= Score in Note-taking

Column 7= Score in Writing

Column 8= Total Score

282	1	2	20	12	09	30	71
283	1	2	19	13	09	22	63
284	1	2	15.5	13.5	11	28	68
285	1	2	17.5	14.5	13	24	69
286	1	2	20.5	16.5	12	33	82
287	1	2	22.5	14.5	09	27	72
288	1	2	15.5	12.5	11	30	69
289	1	2	19	15	11	31	76
290	1	2	18.5	14.5	09	30	72
291	1	2	15.5	12.5	11	28	67
292	1	2	20	14	11	30	75
293	1	2	14.5	14.5	09	30	68
294	1	2	17.5	10.5	10	28	66
295	1	2	20.5	15.5	11	28	75
296	1	2	09.5	05.5	09	28	52
297	1	2	20	13	10	27	70
298	1	2	19.5	13.5	11	28	72
299	1	2	19	13	12	17	61
300	1	2	14.5	14.5	09	23	61
301	1	2	17	13	09	26	65
302	1	2	17.5	10.5	09	18	55
303	1	2	20.5	16.5	10	27	74
304	1	2	19.5	13.5	11	20	64
305	1	2	18.5	08.5	09	23	59
306	2	2	12.5	10.5	11	23	57
307	2	2	10.5	09.5	09	20	49
308	2	2	16.5	07.5	07	19	50
309	2	2	13.5	14.5	11	26	65
310	2	2	14.5	03.5	07	25	50
311	2	2	14.5	16.5	05	24	60
312	2	2	12	17	11	23	62
313	2	2	20	15	05	29	69
314	2	2	16	16	11	26	69
315	2	2	12.5	16.5	11	15	55
316	2	2	19	16	11	30	76
317	2	2	19	16	09	25	69
318	2	2	08	15	15	14	52
319	2	2	16.5	15.5	09	21	62
320	2	2	16.5	10.5	09	22	58
321	2	2	12.5	15.5	05	18	51
322	2	2	16	12	05	21	54
323	2	2	15.5	12.5	05	23	56
324	2	2	20	16	11	22	69
325	2	2	20.5	16.5	11	29	77
326	2	2	08	17	11	29	65
327	2	2	20	13	11	22	66
328	2	2	20	16	07	27	70

329	2	2	18	17	09	29	73
330	2	2	18	16	11	26	71
331	2	2	18	17	09	18	62
332	2	2	19	15	11	31	76
333	2	2	16	17	11	30	74
334	2	2	16	17	09	25	67
335	2	2	20	16	09	31	76
336	2	2	19	17	11	20	67
337	2	2	13.5	16.5	09	28	67
338	2	2	15	09	11	23	58
339	2	2	19	13	11	24	67
340	2	2	15.5	15.5	09	30	70
341	2	2	13.5	16.5	09	20	59
342	2	2	12.5	16.5	09	31	69
343	2	2	20	15	10	29	74
344	2	2	15	16	09	28	68
345	2	2	15	11	09	25	60
346	2	2	18	09	11	28	62
347	2	2	21	13	11	27	72
348	2	2	20	12	11	27	70
349	2	2	14.5	16.5	13	26	70
350	2	2	15	11	11	26	63
351	2	2	14.5	11.5	11	27	64
352	2	2	20	16	11	27	74
353	2	2	16.5	14.5	11	25	67
354	2	2	17	16	11	16	60
355	2	2	14.5	10.5	11	26	62
356	2	2	18	12	11	12	53
357	2	2	17	17	11	26	71
358	2	2	10	16	11	26	63
359	2	2	11.5	15.5	11	28	66
360	2	2	16.5	16.5	11	25	69
361	2	2	14.5	09.5	11	17	52
362	2	2	16	15	11	27	69
363	2	2	17	14	11	31	73
364	2	2	18	15	11	27	71
365	2	2	13.5	15.5	11	24	64
366	2	2	18.5	14.5	05	24	62
367	2	2	15.5	13.5	06	26	61
368	2	2	17.5	16.5	09	27	70
369	2	2	13.5	16.5	13	24	67
370	2	2	16.5	15.5	13	19	64
371	2	2	17	15	13	22	67
372	2	2	18	16	11	16	61
373	2	2	19	17	09	19	64
374	2	2	16	09	07	28	60
375	2	2	19.5	16.5	13	28	77
376	4	2	13.5	03.5	08	15	40
377	4	2	11.5	04.5	06	16	38
378	4	2	06	01	08	17	32
379	4	2	04	02	06	16	28
380	4	2	11	05	06	13	35
381	4	2	10.5	03.5	06	16	36
382	4	2	13.5	05.5	06	12	37
383	4	2	09.5	00.5	06	12	28
384	4	2	06.5	00.5	06	14	27
385	4	2	09	01	09	16	35
386	4	2	07.5	00.5	06	18	32
387	4	2	06	00	07	15	28
388	4	2	13	05	09	14	31

389	4	2	07.5	01.5	09	15	33
390	4	2	14	07	07	14	42
391	4	2	14.5	09.5	09	24	57
392	4	2	04.5	02.5	09	18	34
393	4	2	02	00	07	14	23
394	4	2	02	00	02	05	09
395	4	2	01	00	08	08	17

Appendix VIII

Column 1= Serial No.

Column 2= School Type: 1= Type A
 2= Type B
 3= Type C
 4= Type D

Column 3= Grade: 2 = SLC

Column 4= Score in Reading

Column 5= Score in Cloze

Column 6= Score in Note-taking

Column 7= Score in Writing

Column 8= Total Score

Column 9= Score in the SLC English Exam

001	1	2	20	16	13	25	74	70
002	1	2	20.5	16.5	13	30	80	73
003	1	2	19.5	16.5	10	29	75	64
004	1	2	21.5	16.5	11	33	82	77
005	1	2	19.5	15.5	12	27	74	70
006	1	2	20	16	11	30	77	61
007	1	2	20.5	14.5	08	32	75	64
008	1	2	21	16	14	25	76	79
009	1	2	22	15	11	28	76	72
010	1	2	20	15	13	28	76	72
011	1	2	20.5	15.5	12	27	75	70
012	1	2	22.5	15.5	11	27	76	68
013	1	2	20.5	15.5	11	30	77	60
014	1	2	24	16	11	28	79	73
015	1	2	20.5	16.5	13	28	78	72
016	1	2	19.5	15.5	13	30	78	70
017	1	2	20	15	13	31	79	79
018	1	2	21	16	12	31	80	64
019	1	2	20	14	14	31	79	71
020	1	2	18	15	09	31	73	83
021	1	2	16.5	14.5	09	13	53	79
022	1	2	18.5	16.5	13	28	76	91
023	1	2	21.5	14.5	09	27	72	86
024	1	2	15	14	12	29	70	90
025	1	2	16.5	13.5	00	32	62	60
026	2	2	18	08	08	24	58	76
027	2	2	15.5	15.5	12	30	73	75
028	2	2	17.5	16.5	08	21	63	65
029	2	2	17	16	07	25	65	72
030	2	2	16	17	08	22	63	71
031	2	2	18	17	08	25	68	75
032	2	2	15.5	15.5	05	06	42	57
033	2	2	16	09	08	10	43	49
034	2	2	06.5	13.5	05	17	42	59
035	2	2	15	15	09	29	68	59
036	3	2	24	04	09	17	54	54
037	3	2	14	06	09	11	40	41
038	3	2	10.5	11.5	05	07	34	49
039	3	2	08.5	00.5	08	06	23	39
040	3	2	04	00	08	10	22	24
041	3	2	04.5	01.5	11	11	28	35
042	3	2	05	04	11	02	22	38
043	3	2	08.5	04.5	07	05	25	49
044	3	2	07.5	02.5	08	09	27	32

045	3	2	04	02	07	09	22	24
046	3	2	18	15	10	26	69	72
047	3	2	14.5	12.5	07	27	61	58
048	3	2	10	09	09	07	35	28
049	3	2	10.5	09.5	12	10	42	58
050	3	2	09	10	09	06	34	24
051	3	2	09.5	10.5	11	12	43	35
052	3	2	08.5	09.5	09	08	35	35
053	3	2	13	12	10	21	56	75
054	3	2	08.5	13.5	07	20	49	65
055	3	2	09.5	12.5	07	15	44	57
056	3	2	20	02	09	18	49	51
057	3	2	16.5	10.5	09	21	57	62
058	3	2	19	12	05	21	57	66
059	3	2	08	03	05	15	31	42
060	3	2	06.5	05.5	07	05	24	35
061	3	2	08.5	09.5	10	11	39	26
062	3	2	13.5	01.5	09	00	24	31
063	3	2	07	04	05	17	33	68
064	3	2	05.5	06.5	05	02	19	54
065	3	2	05.5	01.5	05	07	19	33
066	3	2	09.5	01.5	09	09	29	40
067	3	2	11	01	09	14	35	70
068	3	2	11	01	09	18	39	55
069	3	2	12	01	09	18	40	51
070	3	2	02	00	05	09	16	43
071	3	2	05.5	01.5	02	02	11	33
072	3	2	06	00	03	08	17	33
073	3	2	07	00	05	13	25	58
074	3	2	07	00	05	09	21	35
075	3	2	04.5	00.5	01	01	07	35
076	3	2	09.5	02.5	09	16	37	47
077	3	2	11	00	09	09	29	59
078	3	2	05	00	05	02	12	44
079	3	2	08	02	07	04	21	45
080	3	2	02	00	07	10	19	33
081	3	2	13.5	01.5	09	17	41	64
082	4	2	15.5	16.5	08	22	62	47
083	4	2	13.5	15.5	04	02	35	50
084	4	2	04.5	00.5	02	02	09	32
085	4	2	11	05	06	18	40	52
086	4	2	12.5	08.5	10	05	36	78
087	4	2	11.5	07.5	00	20	39	59
088	4	2	03	01	02	15	21	32
089	4	2	10	09	10	15	44	78
090	4	2	05.5	01.5	02	14	23	28
091	4	2	19.5	11.5	11	20	62	78
092	4	2	04	01	01	06	12	35
093	4	2	09	05	04	11	29	61
094	4	2	09	00	05	15	29	39
095	4	2	06.5	01.5	08	06	22	36
096	4	2	08	07	09	16	40	79
097	4	2	06.5	03.5	02	06	18	40
098	4	2	11.5	11.5	09	13	45	53
099	4	2	06	01	07	13	27	34
100	4	2	08.5	00.5	02	04	15	32
101	4	2	11	11	06	23	51	65
102	4	2	11.5	07.5	08	15	42	49
103	4	2	03	00	00	03	06	64
104	4	2	04	00	02	02	08	64

105	4	2	07	05	02	05	19	47
106	4	2	13	14	06	25	58	70
107	4	2	04	03	00	06	13	46
108	4	2	01.5	01.5	00	05	08	32
109	4	2	00	01	00	05	06	59
110	4	2	10	02	05	08	25	33
111	4	2	11.5	04.5	06	08	30	66
112	4	2	08	05	02	08	23	54
113	4	2	12	03	04	10	29	51
114	4	2	04	00	03	05	12	39
115	4	2	14.5	04.5	00	02	21	64
116	4	2	15.5	08.5	02	09	35	45
117	4	2	15	06	03	02	26	64
118	4	2	08	07	11	05	31	66
119	4	2	08	06	08	05	27	71
120	4	2	07.5	04.5	04	03	19	51
121	4	2	09	05	05	15	34	63
122	4	2	11.5	00.5	04	02	18	36
123	4	2	02	01	00	05	08	32
124	4	2	10	09	07	05	31	70
125	4	2	15.5	09.5	10	17	52	73
126	4	2	11	07	03	08	29	60
127	4	2	09.5	02.5	04	07	23	64
128	4	2	10	04	02	08	24	55
129	4	2	07.5	01.5	02	00	11	61
130	4	2	06	01	04	08	19	64
131	4	2	09	07	04	12	32	66
132	4	2	09.5	11.5	08	20	49	74
133	4	2	09.5	06.5	06	07	29	50
134	4	2	12	03	06	02	23	47
135	4	2	06	00	02	01	09	37
136	4	2	13	07	04	07	31	41
137	4	2	14	08	04	12	38	49
138	4	2	02	01	03	06	12	40
139	4	2	06	06	02	03	18	51
140	4	2	05.5	00.5	03	02	11	36
141	4	2	08.5	02.5	07	09	27	63
142	4	2	14.5	00.5	06	06	27	66
143	4	2	11.5	00.5	08	07	27	43
144	4	2	10	00	02	02	14	36
145	4	2	13.5	00.5	04	18	36	73
146	4	2	08.5	02.5	07	05	23	47
147	4	2	03	01	02	03	09	21
148	4	2	07	01	02	00	10	33
149	4	2	15	09	06	11	41	55
150	4	2	13	06	09	09	37	32
151	4	2	15.5	07.5	06	08	37	41
152	4	2	07	00	03	10	20	38
153	4	2	15.5	08.5	08	11	43	45
154	4	2	04	02	02	02	10	32
155	4	2	07	02	04	03	16	21
156	4	2	08	02	03	09	22	28
157	4	2	02.5	00.5	08	04	15	21
158	4	2	09.5	06.5	07	08	31	39
159	4	2	03	00	07	05	15	15
160	4	2	05	02	04	02	13	10
161	4	2	08	01	05	02	16	37
162	4	2	08	01	02	06	17	53
163	4	2	09	05	03	09	26	43
164	4	2	10	01	04	01	16	44

165	4	2	09	02	02	07	20	28
166	4	2	02	02	05	05	14	26
167	4	2	04.5	00.5	06	07	18	33
168	4	2	06	00	04	05	15	36
169	4	2	08.5	00.5	02	02	13	27
170	4	2	03	00	04	04	11	27
171	4	2	06.5	04.5	07	13	31	45
172	4	2	05.5	00.5	01	00	07	14
173	4	2	03.5	00.5	08	08	20	54
174	4	2	09	02	04	06	21	32
175	4	2	04.5	03.5	02	06	16	40
176	4	2	16	09	06	24	55	60
177	4	2	03.5	00.5	01	00	05	17
178	4	2	07	00	06	07	20	54
179	4	2	07	00	00	03	10	27
180	4	2	07.5	01.5	04	07	20	46
181	4	2	06	00	04	08	18	40
182	4	2	06	01	00	04	11	56
183	4	2	07	02	04	00	13	43
184	4	2	05.5	00.5	02	00	08	24
185	4	2	08	02	07	01	18	63
186	4	2	09.5	02.5	08	00	20	32
187	4	2	04.5	00.5	06	01	12	32
188	4	2	06.5	00.5	05	01	13	44
189	4	2	09	05	08	14	36	74
190	4	2	04	04	03	02	13	39
191	4	2	04	01	06	01	12	41
192	4	2	05	01	02	01	09	32
193	4	2	13.5	03.5	05	07	29	63
194	4	2	11.5	02.5	03	04	21	65
195	4	2	03.5	02.5	04	05	15	49
196	4	2	08	03	04	05	20	34
197	4	2	05	00	04	01	10	45
198	4	2	05.5	02.5	07	03	18	57
199	4	2	07	04	08	15	34	69
200	4	2	08	05	06	11	30	64
201	4	2	10	07	06	16	39	44
202	4	2	07	03	00	08	18	50
203	4	2	06.5	03.5	06	08	24	36
204	4	2	05.5	00.5	07	13	26	43
205	4	2	08.5	05.5	07	12	33	52
206	4	2	07	07	07	10	31	53
207	4	2	04	00	02	05	11	38
208	4	2	05	06	07	05	23	32
209	4	2	05	01	04	04	14	32
210	4	2	09.5	00.5	02	02	14	32
211	4	2	08.5	06.5	08	09	32	44
212	4	2	09.5	03.5	04	03	20	32
213	4	2	04.5	05.5	10	09	29	51
214	4	2	13.5	04.5	10	16	42	54
215	4	2	04	05	04	06	19	43
216	4	2	10.5	00.5	06	08	25	35
217	4	2	10.5	09.5	02	06	28	65
218	4	2	05	00	02	09	16	43
219	4	2	02	01	02	10	15	46
220	4	2	07.5	00.5	08	04	20	37
221	4	2	11	03	07	10	31	59
222	4	2	09	01	07	08	25	32
223	4	2	07.5	07.5	03	10	28	59
224	4	2	05	01	02	20	28	54

225	4	2	07	00	02	09	18	50
226	4	2	08	00	04	03	15	21
227	4	2	03	00	03	02	08	16
228	4	2	02	00	02	02	06	42
229	4	2	08	00	03	03	14	36
230	4	2	04.5	00.5	04	00	09	32
231	4	2	03	00	04	01	08	26
232	4	2	05	00	00	04	09	25
233	4	2	12	00	00	04	16	63
234	4	2	05.5	00.5	05	07	18	50
235	4	2	06	00	04	02	12	18
236	4	2	06	00	09	10	25	68
237	4	2	04.5	00.5	07	03	15	25
238	4	2	06	00	04	07	17	25
239	4	2	16.5	09.5	09	24	59	70
240	4	2	06.5	00.5	02	00	09	20
241	4	2	06	00	04	00	10	20
242	4	2	02	00	02	01	05	24
243	4	2	06	00	07	09	22	54
244	4	2	08.5	00.5	05	09	23	32
245	4	2	10.5	04.5	07	11	33	64
246	4	2	06	00	04	02	12	27
247	4	2	05.5	00.5	10	07	23	55
248	4	2	02	00	07	06	15	35
249	4	2	05.5	02.5	02	14	24	40
250	4	2	07	00	03	05	15	56
251	4	2	00	02	02	02	06	20
252	4	2	03	00	00	00	03	36
253	4	2	04	00	02	08	14	24
254	4	2	04	01	08	07	20	38
255	4	2	11	01	02	01	15	22
256	4	2	14	02	08	00	24	38
257	4	2	02	01	01	08	12	17
258	4	2	12.5	01.5	00	00	14	22
259	4	2	01	00	06	03	10	11
260	4	2	01	00	02	02	05	20
261	4	2	17	07	05	15	44	72
262	4	2	11	00	02	09	22	48
263	4	2	05	00	06	00	11	52
264	4	2	02	03	06	09	20	61
265	4	2	03	00	00	02	05	21
266	4	2	05	00	05	09	19	32
267	4	2	03.5	01.5	06	01	12	32
268	4	2	10	08	02	18	38	68
269	4	2	12	02	05	21	40	55
270	4	2	04	02	06	00	12	24
271	4	2	03.5	00.5	02	00	06	23
272	4	2	03.5	00.5	01	00	05	28
273	4	2	02	00	00	03	05	44
274	4	2	03.5	01.5	02	00	07	16
275	4	2	06	00	01	00	07	43
276	4	2	16	09	12	25	62	51
277	4	2	11	09	10	23	53	52
278	4	2	05.5	04.5	02	09	21	34
279	4	2	06	04	11	21	42	52
280	4	2	03.5	00.5	05	06	15	37
281	4	2	16	16	08	26	66	77

Appendix X

INITIAL STATISTICS:

VARIABLE	COMMUNALITY	*	FACTOR	EIGENVALUE	PCT OF VAR	CU
		*				
READING	.65937	*	1	3.06727	76.7	
CLOZE	.70279	*	2	.46910	11.7	
NOTE	.43875	*	3	.25673	6.4	
WRITING	.67786	*	4	.20690	5.2	
0 PAF EXTRACTED. 4 FACTORS. 2 ITERATIONS REQUIRED.						

FACTOR MATRIX:

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
READING	.86386	-.12074	.10498	-.03617
CLOZE	.89461	-.12373	-.08373	.04392
NOTE	.69793	.15002	.08794	.04262
WRITING	.88451	.12469	-.08724	-.04272

FINAL STATISTICS:

VARIABLE	COMMUNALITY	*	FACTOR	EIGENVALUE	PCT OF VAR	CU
		*				
READING	.77317	*	1	2.81607	70.4	
CLOZE	.82458	*	2	.06794	1.7	
NOTE	.51917	*	3	.03337	.8	
WRITING	.80735	*	4	.00688	.2	

VARIMAX ROTATION 1 FOR EXTRACTION 1 IN ANALYSIS 1 - KAISER
NORMALIZATION.

ROTATED FACTOR MATRIX:

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
READING	.59950	.46248	.44592	.03214
CLOZE	.71226	.48871	.27978	-.01227
NOTE	.34250	.56669	.28261	-.02922
WRITING	.55632	.66486	.21879	.08907

FACTOR TRANSFORMATION MATRIX:

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
FACTOR 1	.66980	.64640	.36444	.02669
FACTOR 2	-.58523	.75960	-.27640	.06419
FACTOR 3	-.45279	-.01596	.87353	-.17796
FACTOR 4	.06204	.07014	-.16653	-.98158

Read the following passage carefully, before you attempt any questions.

Answer all the questions. You are recommended to answer them in the order set.

Mistakes in spelling, punctuation and grammar may be penalised in any part of the paper.

Drove Roads

- 1 Many years ago the inhabitants of Europe drove large herds of cattle along broad grassy tracks which came to be known as 'drove roads'. Sometimes they travelled a long way in search of fresh pasture when their own land was not rich enough for their cattle to be fattened on. Sometimes they were going to distant markets when there were not enough rich people locally to buy their meat. The earliest routes followed the dry, high land, avoiding swamps and the danger of surprise attack. Later, the tracks ran along lower ground when people had settled at those levels and where springs were to be found. 5
- 2 Although large numbers of men and boys helped to drive the cattle, the true drover was the man with the overall responsibility to the owners. He had to take complete charge of the animals for the length of the journey. He alone was made answerable for large sums of money which came from the sale of the cows, and which he would eventually hand over to the owners. He was also answerable for any misbehaviour of his men; for example, he prevented them from being cruel to the cattle, an offence which could mean imprisonment. 10 15
- 3 This movement of animals took place over vast distances—often hundreds of kilometres—and required much organisation, which was the sole responsibility of the drover. A constant worry was providing sufficient quantities of food for the hungry cattle, which needed enormous amounts of fodder to sustain them throughout their journey. It was like feeding an army—except that the drover had no waggon train loaded with supplies to feed his 'army' of cattle; he had to rely on the food that could be found along the way. He also showed great skill in manoeuvring these herds which sometimes numbered as many as 200 animals, often spread out over considerable distances. Frequently, he had to ride ahead on horseback to check the wandering instinct of the cattle, redirecting them along the right track, or follow behind to urge them on or bring back stragglers which had got left behind. He had to be particularly careful that the cattle did not stray into private land to feed. Such trespass could be punished by the cattle being driven into a fenced enclosure by the land-owner until the drover had paid for any damage they had done. Then, too, he had to be on the look-out for farms which lay on his route and blow a horn to warn the farmer of his approach, so that the farmer would have ample time to prepare. Imagine the confusion if the farmer's own cattle were to get mixed up with the travelling herd! Many hours would be spent in sorting them out and quarrels would break out over disputed ownership. 20 25 30
- 4 Controlling a large herd was sometimes a dangerous job. Cattle are easily frightened by sudden loud noises. For instance, a herd crossing a wooden bridge in which the planks rattled noisily would break away in uncontrollable panic and it would take hours to calm and reassemble them. The drover also had to cope with bands of robbers who might violently attack him for his money or drive off some of his cattle. 35 40
- 5 But it was the actual handling of the cattle that called for his greatest skill. He had to have an intimate knowledge of the temperament of his animals in order to judge how far and how fast to drive them. They must not be overdriven, for it was essential that they reach their market in prime condition. Rest days would therefore

be necessary, because there would be keen competition between drovers arriving at an important market to get the best prices from the butchers, who could quickly assess the condition of the animals. 45

6 A drover needed close knowledge of the country through which he travelled, some of which was very wild and gave cover to robbers. If possible, he avoided the hard roads, which injured the cattle's feet, and their toll-gates where travellers had to pay money to pass through. He chose the broad green tracks. Along these the herd could move in safety, sometimes foraging through the scrub on either side, or sometimes stopping to graze. Above all a drover had to have honesty, endurance, patience, and courage. The progress of the cattle could be wearisomely slow, mainly because of their enormous appetites, and normally they travelled only about ten kilometres a day. At night, whatever the weather, the drover usually slept rough with the cattle. 50 55

7 The droves of cattle passing along the same routes every year influenced the life and industries of the land they crossed. At various points, tanneries could be found where hides were prepared for the tradesmen working in leather. As the animals were fitted with 'shoes' to ease the wear on their soft hooves, blacksmiths set up their forges to make thousands of iron shoes ready for the cattle that were to come their way. As droving was 'thirsty work', many villagers living along the route found that selling drink became a profitable business. 60

8 At times, the men would drink too much and offend the disapproving country-folk, but generally they were popular and their arrival would be welcomed, for they brought letters or news—sometimes of great events like famous victories or natural disasters. Their highly-coloured tales of the outside world were especially welcome in isolated communities and encouraged some of the bolder spirits to leave in search of fame and fortune. Sometimes the drover would pay rent for a field in which his men and cattle could stay for the night. The grass was usually closely bitten down by morning but, as a result of being heavily manured, the field was soon very fertile. Farmers willing to put up herds for the night sometimes planted clumps of high trees close to their farmhouses at a point visible a long way off. Local people, attracted by the shouting of the men and the bellowing of the animals, would often help to guide in the cattle; they were also, no doubt, encouraged to do so because the head drover usually bought helpers plenty of beer. 65 70 75

9 The drovers' lives were not all hardship. There was variety in their journeys. They could satisfy their curiosity and their love of being constantly on the move. There were frequent changes of scenery and unexpected adventures on the way. Sometimes, too, they would arrive in a village where a celebration was in progress. And always they were conscious of their special skills which they brought to their task and which set them above ordinary villagers. The cattle were an impressive sight as the huge herds moved slowly over the hillsides—and a beautiful one when, scattered over a large field, they grazed peacefully in the fading evening light. 80 85

Answer all the questions.

You are recommended to answer them in the order set.

From paragraph 1:

- 1 (a) What were the two problems which made it necessary for men to drive their cattle long distances? [2]
- (b) The word 'rich' is used twice in the paragraph with different meanings.
 - (i) What does 'rich' (line 3) mean? [2]
 - (ii) What does 'rich' (line 5) mean? [2]
- (c) Why would the drovers fear 'swamps' (line 6)? [1]
- (d) Why would a high route prevent a 'surprise attack' (line 7)? [1]
- (e) The tracks ran along lower ground when 'people had settled' (line 7) and 'where springs were to be found' (line 8). What two benefits would the lower tracks therefore provide for the drovers? [2]

From paragraph 2:

- (f) The drover was responsible for discipline. Quote the 10 words that tell you this. [1]
 - (g) The drover was responsible for financial matters. Quote the 10 words that tell you this. [1]
- [Total: 10]

From paragraph 6:

- 2 (a) The drovers preferred 'the broad green tracks' (line 51) to the 'hard roads' (line 50).
 - (i) Why did being broad make the track safer? [1]
 - (ii) The tracks were green. What two advantages did this give the tracks over the 'hard roads'? [2]
 - (iii) What other reason did the drovers have for preferring the tracks? [4]
 - (b) 'sometimes foraging through the scrub...or sometimes stopping to graze' (lines 52-53). What is the difference between 'foraging' and 'grazing'? [2]
 - (c) A drover needed 'honesty, endurance, patience, and courage' (lines 53-54).
 - (i) Which one of these qualities would he need when the 'progress of the cattle' was 'wearisomely slow' (line 54)? [1]
 - (ii) Which other of these qualities would he need when he 'slept rough with the cattle' (lines 56-57)? [2]
 - (d) Why would the 'enormous appetites' (line 55) make progress slow? [1]
- [Total: 9]

From paragraph 8:

- 3 (a) To the people in 'isolated communities' the drovers' 'highly-coloured tales' were especially welcome (lines 68-69).
- (i) What is meant by 'highly-coloured tales'? [2]
 - (ii) Why would these tales appeal to people in 'isolated communities'? [2]
- (b) Using the material from line 70 to the end of the paragraph give three further reasons why the people on the drovers' route would welcome drovers. Number your answers 1, 2 and 3. [3]
- (c) What form of advertisement did the farmers use to attract the drovers? [1]

From paragraphs 7, 8 and 9:

(d) Choose FIVE of the following words or phrases. For each of them, give one word or short phrase (of not more than seven words) which has the same meaning as it has in the passage.

- | | |
|-------------------------|--------------------------|
| 1. set up (line 61) | 5. frequent (line 80) |
| 2. offend (line 65) | 6. in progress (line 81) |
| 3. attracted (line 74) | 7. conscious (line 82) |
| 4. constantly (line 79) | 8. fading (line 85) |

[5]

[Total: 11]

4 The passage describes the difficulties of driving cattle and the drovers' skills in controlling them. Write an account of the problems the drovers met and what they did to deal with them.

USE ONLY THE MATERIAL FROM LINE 16 TO LINE 47.

Your account, which should be in continuous writing, must not be longer than 160 words, including the 10 words given below.

Begin your summary as follows:

The drovers had to overcome the problems of the long ... [20]

Part Two

Begin your answer on a fresh page.

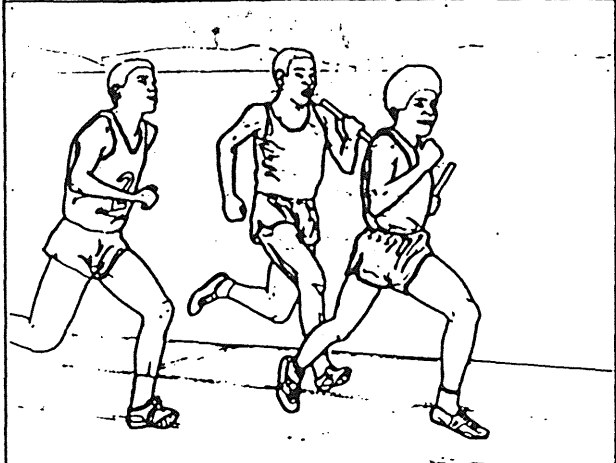
You are advised to spend about 30 minutes on this part of the paper.

Write an account based on one of the picture sequences.

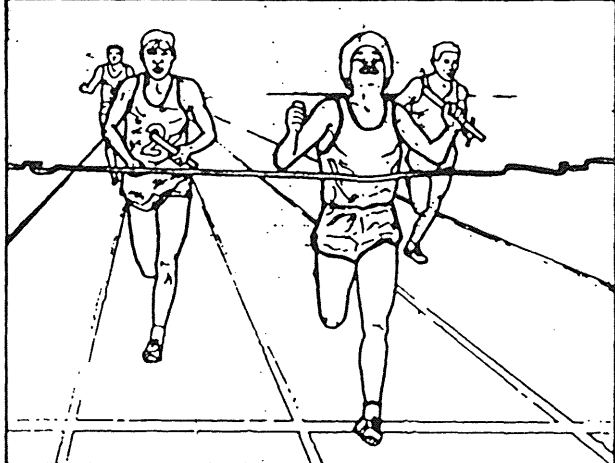
Either below (a) Collecting a prize.

Or on the next page (b) Obtaining a job.

1 RUNNING THE RACE



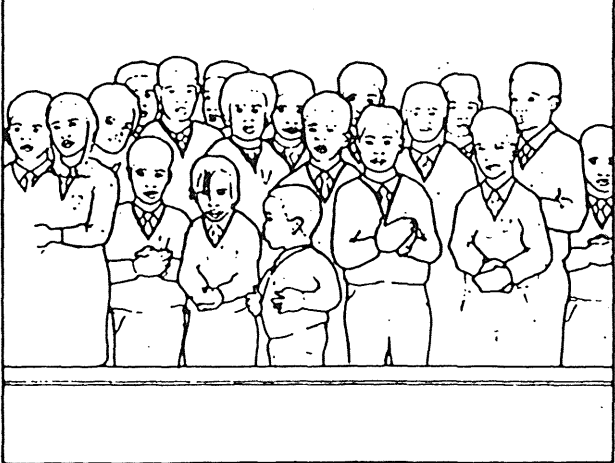
2 WINNING



3 PRINCIPAL ADDRESSING ASSEMBLY



4 APPLAUSE



5 PRESENTING THE CUP



6 CONGRATULATIONS

