

Open Research Online

The Open University's repository of research publications and other research outputs

Reading Strategies in the Nepalese School Leaving Exam: Establishing Construct Validity

Thesis

How to cite:

Dawadi, Saraswati (2016). Reading Strategies in the Nepalese School Leaving Exam: Establishing Construct Validity. MRes thesis The Open University.

For guidance on citations see [FAQs](#).

© 2016 The Author

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the [policies page](#).

oro.open.ac.uk

**Reading strategies in the Nepalese School Leaving Exam: Establishing
Construct Validity**

Saraswati Dawadi

Supervisors

Prithvi Shrestha, Steven Hutchinson and Maria Fernandez-Toro

Submitted in partial fulfillment of the requirements for the

Degree of Master in Research (Education)

Centre for Research in Education and Education Technology

The Open University

2015

DATE OF SUBMISSION: 30 SEPTEMBER 2015

DATE OF AWARD : 29 JANUARY 2016

ProQuest Number: 13834781

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13834781

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Acknowledgements

It gives me an immense pleasure to thank a number of people and institutions, who played a major role in making this study a success. First of all, I would very much like to express my sincere thanks to my supervisors, Dr. Prithvi Shrestha, Dr. Steven Hutchinson and Dr. Maria Fernandez-Toro, who incessantly supported and encouraged me throughout the project. Without their constant encouragement and belief in me, this dissertation would not have come to this form. My sincere gratitude goes to my all inspirational tutors, Dr. Martyn Hammersley, Dr. John Richardson and Dr. Stephane Taylor for enriching my researching skills. I am very much thankful to Mrs. Anne Foward and Mrs. Paula Piggott who continuously supported me with administrative matters.

I am very much indebted to The Open University, Milton Keynes, England for providing me the CREET scholarship without which I would not have been able to study MRes at the university and carry out this research. My sincere thanks also go to Tribhuvan University, Nepal for supporting my study by allowing me sabbatical leave during my study at the Open University.

My hearty thanks go to all my participants and the head teachers of the schools who constantly supported me even in such an adverse situation.

It would be a great injustice not to mention my family members back home in Nepal who always supported and encouraged me to accomplish my goal. I am very much grateful to my husband Mr. Ram Krishna Panthi and father in law Mr. Shreedhar Panthi for their continuous support and encouragement. My son Mr. Rashin Panthi and daughter Ms. Shiksha Panthi deserve many thanks for their patience. Lastly, but not the least, my sincere thanks are due to my late mother Mrs. Jamuna Dawadi, father Mr. Damodar

Dawadi, sister Mrs. Ganga Dawadi, brother-in-law Mr. Anil Neupane, brother Mr. Nabaraj Dawadi and all my family members, relatives and well wishers for their inspiration and encouragement.

Abstract

There are a number of studies on construct validity of a language test in general, and a reading test in particular; most of them are quantitative in nature. As quantitative studies may not capture the cognitive processes the test takers follow to respond to a test item, the importance of collecting verbal reports of test takers has been recommended in order to establish the construct validity of a test (Alderson, 1990b). Additionally, no research to date, has investigated the construct validity of the reading test of the school leaving examination in Nepal. Therefore, the present study was designed to fill this gap i.e. to investigate whether the reading strategies, the test claims to test, are measured by the test. In order to investigate the issue, eight grade 10 students were asked to take the reading test and think-aloud when responding to the test items. Immediately after they completed the test, they were asked to take part in a retrospective debriefing. Additionally, in order to triangulate the data, the views of seven language testing experts on the reading strategies tested by the test were collected. The experts were first asked to make judgements individually with regard to the skills tested by the test. Then, they were asked to take part in a focus group discussion. The findings provide grounded insights into the response behaviours prompted by the reading tasks on the test and indicate some threats to the construct validity of the test as the test does not seem to be measuring all the reading strategies it claims to measure. The study also indicates a very low level of agreement among the subject experts regarding the strategies tested by the test. Consequently, there was a big gap between the strategy use and expert judgements. Therefore, the usefulness of expert judgements to predict which reading strategies the test items are testing is questioned and the need for further research concerning the methodologies used in the study has been suggested. More importantly, the study highlights the need to make the test more valid.

List of Tables and Diagrams

Table 1: Categories of the coding scheme	26
Diagram 1: Total Frequencies of strategy use	28
Diagram 2: Total frequencies of strategy judgements	37
Diagram 3: Overall comparison between strategy use and expert judgements	40
Diagram 4: Comparison of strategy use and strategy judgment for True/False items	41
Diagram 5: Comparison of strategy use and strategy judgements for Gap Filling Items	42
Diagram 6: Comparison of strategy use and strategy judgements for Finding Similar Words	43
Diagram 7: Comparison of strategy use and strategy judgements for Ordering Sentences	44
Diagram 8: Comparison of strategy use and strategy judgements for Matching Items	45
Diagram 9: Comparison of strategy use and strategy judgements for Short Answer Questions	46

Table of Contents

Acknowledgements	i
Abstract	iii
List of Tables and Diagrams	iv
Table of Contents	v
Chapter 1: Introduction	1
Chapter 2: Literature Review	4
2.1 Theories of reading	4
2.1.1 Terminological inconsistencies between reading skills and strategies.....	5
2.2 Frameworks used in reading assessment studies	6
2.3 Construct validity.....	8
2.4 Verbal reports and test validation.....	10
2.5 Expert judgements in test validation	12
2.6. Research questions	14
Chapter 3: Methods of data collection and analysis	15
3.1 Methodological framework.....	15
3.1.1 Access to the field work and ethical considerations.....	15
3.1.2 Location of data collection.....	15
3.1.3 Negotiating access for field work.....	16
3.1.4 Recruiting participants	17
3.1.5 Timing of the study.....	19
3.1.6 Ethical considerations	19
3.2 Data collection and analysis.....	19
3.2.1 Verbal reports.....	20
3.2.1.1 Think-aloud and RDs procedure	21
3.2.1.2 Data analysis	22
3.2.2 Expert Judgements.....	23
3.2.2.1 Procedures.....	24
3.2.2.2 Data analysis	25
Chapter 4: Results	26

4.1 Reading strategies used by the students	26
4.1.1 Reading for understanding of an argument.....	29
4.1.2 Reading for finding specific information.....	29
4.1.3 Reading and matching similar sentences.....	30
4.1.4 Reading and matching lexical items.....	31
4.1.5 Reading and making a guess	31
4.1.6 Reading for general understanding.....	32
4.1.7 Reading for detail understanding.....	33
4.1.8 Reading for understanding themes and ideas in the text.....	33
4.1.9 Reading for understanding the gist.....	34
4.1.10 Reading for understanding texts organisation and appreciating literary text.....	34
4.1.11 Reading for inferring the meaning of unfamiliar lexical items and matching grammatical structure.....	35
4.1.12 Reading and making own interpretation	36
4.2 Expert judgements on the reading strategies.....	37
4.3 Consistency between the strategy use and the expert judgements.....	39
4.3.1 True/false items	41
4.3.2 Gap filling items.....	42
4.3.3 Finding similar words	43
4.3.4 Ordering sentences	44
4.3.5 Matching items.....	45
4.3.6 Short answer questions.....	46
Chapter 5: Discussion	49
5.1 Research questions	49
5.1.1 Research Question 1	49
5.1.2 Research question 2	53
5.1.3 Research question 3	55
5.2 Construct validity of the test with regard to the reading strategies.....	56
5.3 Methodological considerations.....	58

Chapter 6: Conclusion	63
References.....	66
Appendix A: Information sheet and parents' consent form.....	73
Appendix B: Information sheet and experts' consent form	76
Appendix C: Questionnaire for students.....	80
Appendix D: Expert judgement instrument	84
Appendix E: Test Paper.....	88

Chapter 1: Introduction

This study is concerned with investigating the English reading test of the school leaving exam in Nepal, called the School Leaving Certificate (SLC, henceforth) exam. The SLC is a national level standardized exam controlled by the Ministry of Education (MOE), Nepal. The exam is regarded as the most crucial one for a candidate in the Nepalese context because passing it means that the students are allowed to study at higher education institutions. It is the sole factor that dictates one's career path, as its scores decide which course a student can study in higher education (Shrestha, 2003). The MOE has been conducting the exam every year with a steadily increasing number of candidates from a few hundreds to 405,338 in 2014/15 (MOE, 2015).

It should be noted that the SLC test takers come from two different kinds of schooling background: public and private. The students in private schools get an opportunity to practice English as they use English both inside and outside the classroom in their school. English is used as a medium of instruction in the private schools where as Nepali is used as a medium of instruction in most public schools in Nepal. The public schools teach English as a subject about four hours a week. Therefore, it is generally assumed that the graduates from the private schools have better English language proficiency than the graduates from the public schools. Sash (2006) argues that the majority of students in public schools fail English and feel apprehensive of this subject.

To make any exam fully operational, it is very important to establish its validity. A test is said to be valid, if it "measures accurately what it is intended to measure" (Hughes, 2003: 26). A valid test relies, as much as possible, on empirical evidence collected from test takers and offers meaningful information about the test takers' ability (Brown and

Abeywickrama, 2010). Thus, a valid test should guarantee that the test performance provides an accurate picture of the underlying abilities or constructs it is attempting to measure (Weir, 2005).

With regard to the SLC exam, to my experience, significant debate and discussion has gone into deciding what to test but there has been no discussion in how to test it. Consequently, although several efforts have been made to improve the exam since its beginning, the exam has been criticised for lacking reliability and validity (Khaniya, 2010). However, it should be noted that the claims related to the validity issues seem weak as no research has been carried out to investigate the validity of the whole SLC exam in general and the validity of the English reading test of the exam in particular. Therefore, the rationale for choosing this study starts with the need to address the validity issue of the English reading test of the exam. As Mesick (1988) maintains, if the validity of high-stakes tests is not known, this might have some undesirable consequences for the society at large.

The rationale for this research continues by considering the gap in research in test validation within the relatively new but growing body of research with test takers. The literature indicates that language testing research is dominated by quantitative methods. But, as Song (2008) points out, quantitative approaches are problematic as they do not capture how test takers answer an item; a qualitative approach is necessary to reveal which skills test takers really engage in when solving reading test items. Therefore, this study aims to collect information regarding the cognitive processes test takers follow to respond to the test items.

The main purpose of the study is to investigate whether the reading strategies specified in the SLC English curriculum are employed by test takers. By addressing this issue, the

study hopes to contribute to the establishment of the construct validity of the test as the issue has not been addressed in other research to date, despite the fact that the test has a quite long history and plays a crucial role in the test takers' lives.

It should be noted that there is no practice of developing test specifications before writing test items of the SLC exam. Therefore, the SLC test item writers are asked to design the test on the basis of the SLC curriculum. As the curriculum has clearly specified the reading strategies that need to be tested, this study is seeing whether or not the test actually looks at whether or not these strategies are used or need to be used to respond to the test items. Test takers' verbal reports (the accounts of their mental processing) are regarded to be very insightful for the purpose (Green, 1998). Therefore, this study aims to collect verbal reports of Grade 10 students studying in Nepal, the prospective students of the SLC exam. Additionally, it aims to use expert judgements in order to shed more light into the issue. Using these qualitative methods of data collection, this study might offer a great potential for the development of new knowledge and understanding to the existing body of knowledge currently in the field.

The study has been divided into six chapters. This introduction part is followed by the review of literature on the related areas. Chapter three discusses the methodology used in the study and chapter four presents the results, which is followed by the discussions on the results. The final chapter draws a conclusion.

Chapter 2: Literature Review

The literature in the field of reading and reading assessment is vast and complex (Moore, Morton and Price, 2012). Only key studies that have particular relevance to the current study are reviewed. Basically, three broad areas have been identified: reading skills and strategies along with theories of reading, construct validity, and usefulness of verbal reports and expert judgements in test validation. The following section presents a review of the literature in each area.

2.1 Theories of reading

A basic theory of reading suggests that reading is an act or a process in which a reader translates graphic symbols of a text into meaning (comprehension); however, it is very difficult to understand how the reader makes sense of those graphic symbols (Kamil and Pearson, 1979) as the processes the reader follows to comprehend a text are not easily detectable. Different assumptions are made regarding the reading process. As a result, different theories and models of reading have been introduced.

The reading process has been broadly defined from two different perspectives: cognitive and social. Reading from the cognitive perspective is a process where a reader activates various cognitive processes, such as parsing sentences through an analysis of the grammatical constituents, interpreting clauses and making inferences in order to comprehend a written text whereas reading from the social perspective, is “a situated activity that takes place in a specific context with specific task demands; therefore, the social context within which students read influences their reading behaviors by determining how they engage with reading tasks” (Katalayi and Sivasubramaniam, 2014,

p.848). On the basis of the purpose of the study, reading has been seen from the cognitive perspective in this study.

It is equally important to define reading construct i.e. the skills involved in reading. Researchers have different opinions regarding the skills involved in reading. Some researchers (e. g Alderson 1990b; Weir, Yang and Jin 2000; Wu, 2014) view reading consisting of different sub- skills while others (e. g. Rosenshine, 1980; Schedl, Gordon, Carey and Tang, 1996) view reading as a monolithic entity and argue that discrete skills do not exist in reading comprehension. The current study focuses on the divisibility of the reading construct as the SLC curriculum itself has divided reading into different sub- skills and it is thought to be “possible to link particular types of item or task to specific sub- skills that they are said to tap into” (Khalifa and Weir, 2009: 39).

The literature on reading indicates some inconsistencies in the use of the terms ‘reading skills’ and ‘reading strategies’. The following section discusses some of those inconsistencies.

2.1.1 Terminological inconsistencies between reading skills and strategies

There is considerable terminological confusion between reading skills and strategies. Inferring, for example, is a skill for Davis (1968) but a strategy for Olshavsky (1977). Some studies have also used the terms in an overlapping manner. It might be because of the difficulty to distinguish one from another as they are very much interconnected to each other and have a fuzzy relation (Urquhart and Weir, 1998). However, some scholars have come up with different definitions, helping to reduce the confusion.

Reading skills are “automatic actions that result in decoding, comprehension and fluency usually without the readers’ awareness of the components” while reading strategies are

“deliberate, goal-directed attempts to control and modify the readers’ efforts to decode text, understand words and construct meanings out of texts” (Afflerbach, Pearson and Paris, 2008:15). In other words, skill is an acquired ability which operates largely subconsciously, but a strategy is a conscious procedure which a reader carries out in order to solve a problem (William and Moran, 1989). Thus, whether a reader’s actions are under automatic or deliberate control is a key difference between skills and strategies. Reading skills operate without the readers’ deliberate control or awareness of work but reading strategies are under the conscious control of readers.

2.2 Frameworks used in reading assessment studies

Test validation requires a clear frame work, within which data can be gathered and interpreted. Most of the studies that explore the nature of reading in a test situation have relied on “taxonomies that seek to divide reading practices into a variety of skills and sub-skills” (Moore, et al., 2012: 126). However, the literature indicates that reading researchers have different views regarding the number of skills and sub-skills used in reading. For example, Carver (1997) mentions five basic skills of reading whereas Grabe and Stoller (2002) argue for seven skills. However, as mentioned in chapter one, the SLC curriculum divides reading into 12 different strategies such as reading for detailed understanding, reading for general understanding, reading for specific information (scanning), reading for gist (skinning) and so on.

Of the various taxonomies, having covered the important dimensions on reading, the most useful for the current study is the one proposed by Urquhart and Weir (1998), which is constructed around two dimensions of differences: reading level and reading type. Reading level focuses on the reading processes or the level of engagement. A distinction is made between the reading processes focused on a text at a local level and those

operating at a more global level. Local level of engagement (comprehension) refers to “the understanding of propositions at the level of micro structure” like individual words, phrases, clauses or sentences whereas global level of engagement (comprehension) refers to “the understanding of propositions beyond the level of micro structure” (Khalifa and Weir, 2009: 45-46). Thus, the local level of engagement is for basic comprehension while global level of engagement is for comprehending larger textual units like paragraphs or the whole text(s).

However, Urquhart and Weir (1998) do not discuss the two main types of processing: top down and bottom up processing in reading, which most of the reading researchers agree with (e.g Alderson, 2000; Cohen and Upton, 2006, 2007, Khalifa and Weir, 2009). In bottom-up processing (or decoding), readers process smaller units into larger units like letter- by-letter, word -by- word or sentence- by- sentence in a linear fashion, whereas in top- down processing, “larger units affect the way smaller units are perceived”(Khalifa and Weir, 2009: 41). Top down processing (inferencing) gives emphasis on context and background knowledge of readers, and assumes that a reader has some goals and expectations; s/he samples information from a text to confirm or reject his/her expectations. In other words, in order to sample the information efficiently, the reader looks for the most likely places in the text to find useful information and also the readers make some inferences from the context. However, in order to comprehend a text, there needs to be a constant interaction between top-down and bottom-up processing.

For reading types, two basic types of reading, expeditious and careful reading, have been proposed (Khalifa and Weir, 2009; Urquhart and Weir, 1998). Expeditious reading refers to quick and selective reading, for instance, skimming, scanning and search reading; “priority is given on speed of fulfilling a specific purpose at the expense of thorough text

processing” (Shiotsu, 2010: 128). The main purpose of careful reading, in contrast, is to understand all the information, implicitly or explicitly mentioned (Khalifa and Weir, 2009). Thus, comprehending sentences and/or overall text, making inferences etc. are the examples of careful reading.

Reading activities are either more local or more global in their orientation; for example, the act of scanning (i. e locating specific information) has a more local focus; on the other hand, the act of ‘skimming’ (obtaining an overview of a text) is necessarily a more ‘global’ form of reading (Moore et al., 2012). Thus, the analytical framework for the current study will be the one that focuses on the level of engagement and reading types as the reading strategies specified in the SLC curriculum are well covered by these two dimensions. Therefore, in the framework for the current study, following Khalifa and Weir (2009) and Wu (2014), the reading processes and skills are grouped into four categories:

- a. Expeditious reading at the local level (for example, scanning a text for specific information)
- b. Expeditious reading at the global level (for example, skimming for the gist)
- c. Careful reading at the local level (for example, understanding or inferencing lexical meaning)
- d. Careful reading at the global level (for example, understanding explicitly stated main ideas, understanding the overall meaning of the text, understanding the pragmatic meaning of the text, making inferences for the text etc.)

2.3 Construct validity

The notion of construct validity emerged in the early 1950s out of efforts made to address the adequacy of psychological tests by the American Psychological Association (Cronbach,

1988). Construct validity concerns “the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs” (Bachman, 1990: 255). Messick (1992), further states that construct validity measures how much of something an individual displays or possesses. Thus, validating a test means making an investigation on whether the test measures what it is supposed to measure i.e to investigate how accurate are the inferences made on the basis of a test performance (Hughes, 2003).

In the case of a reading test, construct validity is “a measure of how closely a test reflects the model of reading underlying the test”(Moore et al., 2012: 121). More specifically, construct validity in reading is “the ability we wish to test”(Alderson 2000: 1). Therefore, a reading test is said to have construct validity, if the test measures all and only the reading abilities it intends to measure (Hughes, 2003). Hence, construct validity is concerned with the match of a test with what is being tested; the perceived abilities of the readers in relation to the reading task set.

To illustrate with a simple example, imagine we intend to assess reading and our theoretical conceptualization of what it is to be able to read stipulates that reading is a skill consisting of two sub- skills: reading for main ideas and reading for specific information. If we design a reading test, with items that only tap into reading for main ideas we are not assessing the construct (the definition of reading) in a comprehensive manner. This is construct under-representation as we are measuring only one part of the reading construct. But, if we design a test with items that require the test takers to comprehend the text in terms of main ideas, specific information and also requires some mathematical calculations to complete the test items, we are testing the intended construct (i.e reading) but also something that is not part of the construct (i. e

mathematical ability). In that case, we are measuring the construct plus something that is irrelevant to the construct.

However, construct validation is problematic as it is a theory testing procedure; it involves a theory and the relationship of data to the theory; in construct validation, one validates a test against a theory (Palmer and Groot 1981, cited in Salehi, 2011). It is “the means by which theories are put to the test and are confirmed, modified or abandoned” (Hughes 2003: 26). Thus, validating a reading test might require a theory of reading. Therefore, the current study will be based on the reading theory that underpins the SLC reading test which defines reading as a cognitive process consisting of different sub-skills.

Construct validation also requires a proper source of data. Among other sources of data e.g test performance, test content etc. verbal reports of test takers have been seen as useful to make judgements about the construct validity of a test (Green, 1998). The following section defines verbal reports and reviews some studies that have used verbal reports for test validation.

2.4 Verbal reports and test validation

A verbal report refers to a problem solver's account of his/her mental processing (Ericsson and Simon, 1993). Thus, the main rationale for collecting verbal reports is to get access to the cognitive processes in readers' minds.

There are mainly two forms of verbal reports: think -aloud protocols (TAPs) and retrospective debriefing (RD). TAPs are the verbatim records of a problem solver's thinking aloud while completing a given task where as RDs are the problem solver's reports about how s/he performed the task (Taylor and Dionne, 2000).

However, TAPS and RDs have some limitations too. The main criticism of TAPs concerns their *veridicality* and *reactivity*; *Veridicality* concerns whether the TAPs can accurately report or represent the participants' true and complete thinking processes, while *reactivity* concerns whether the requirement to report the rating process alters the process being observed (Ericsson and Simon, 1993). Similarly, a problem with RD concerns the temporal separation between processing and reporting. Consequently, the readers may not be able to report exactly what they did during the task performance.

Despite having limitations, many reading researchers (e.g. Alderson, 1990a, 1990b; Bachman, 1990; Cohen and Upton, 2006, 2007; Song, 2008) advocate the importance of collecting verbal reports as a part of construct validation as the reports might provide a rich source of information about comprehension processes. In reading test validation, the cognitive processes involved in responding to an item are more important than what the constructors believe the item to be testing (Alderson, 2000). However, to the best of my knowledge, only three studies reviewed below have explored the construct validity of a test using verbal reports.

Alderson (1990b) collected both TAPs and RDs from his participants, who took a reading test; their verbal reports indicate that "what appears to be being tested by an item does not always match the beliefs of the test and test constructors" (p.477). The findings further indicate that the test lacks construct validity. However, there were only two participants and neither of them could report properly: one had language problems and the other had long silences during the think aloud session and could not successfully report retrospectively.

Contrary to Alderson's (1990b) findings, Nikolov (2006), who collected TAPs of 52

Hungarian children taking a Grade 7 Hungarian reading test, and Cohen and Upton (2006), who collected TAPs of 32 test takers of the *new TOEFL*, report that the reading strategies employed by the test takers were in line with the strategies intended by the test. However, both of these studies collected only the TAPs; had the test takers' RDs been collected, the RDs might offer more insights into the cognitive processes and produce more valid results.

In a similar vein, Holzkecht (2012) asked eight students to take the Austrian Matura exam and think-aloud. The results revealed that the students used all the careful reading strategies specified in the test specifications. However, the exam did not have equal focus on expeditious reading and the strategy "reading to deduce the meaning of unfamiliar words from the context" was not in use which was listed in the test specification. Thus, the findings posed a possible threat to the construct validity of the reading exam. However, the author also collected only the TAPs of the test takers.

Nevertheless, there is no final absolute measure to validate a reading test, but different kinds of evidence can be invoked in support to investigate the complex issues of reading skills in order to get more valid results (McNamara, 2009). Therefore, the current study, in addition to the verbal reports of the students, collected experts' views with regard to the skills measured by the test.

2.5 Expert judgements in test validation

Expert judgements (the views of the experts regarding the skills tested by the reading test) are considered to be important for test validation (Alderson, 1993). However, the use of expert judgements in validating a test is challenging. The enduring problem of a

research using expert judgements might be of low level of agreement among the experts (Fulcher, 1997).

Previous studies in language testing have used expert judgements for different purposes indicating different level of agreement. For example, Bejar (1983) and Flucher (1997) asked their experts to predict item difficulty and discrimination indices of the test of Standard Written English; and to sequence ten different texts in order of difficulty. Additionally, Alderson (1990a) and Alderson and Lukamani (1989) asked their experts to judge whether the test items in two reading tests were testing 'lower', 'middle' or 'higher' order abilities. All the studies indicated low agreement among the experts. However, Lumley (1993), who asked the experts to examine the place of sub-skills in English as second language (ESL) syllabus and test design, claims for a high level of agreement among the experts. But, it can be argued that the experts were forced to have agreement in Lumley's study as they had discussion until agreement was achieved (Alderson and Lumley, 1995).

Tavakoli and Barati (2011) used expert judgments to investigate the construct validity of the Cambridge English: First Paper. In the study, six university lecturers in Iran were asked to make judgments regarding the reading skills measured by the test. They also asked six test takers, the under graduate students, to match the skills assessed by each test item in the test on the reading taxonomy of reading skills, which was provided to them. The findings did not reveal substantial agreement among the experts on the skills claimed to be tested by the test, nor could any significant agreement be observed among the test takers on the skills being tested by the test. Similarly, Holzknacht (2012), besides collecting verbal reports of eight test takers, asked seven testing experts to make judgments about the reading skills tested by the Austrian Matura exam. The findings are

in line with most previous studies on expert judgment indicating little agreement among the experts on the reading skills tested by the test. The study also indicates a big gap between the strategy use and strategy judgments.

2.6. Research questions

The study aims to explore the following research questions:

1. Do test takers employ all the reading strategies specified in the SLC English curriculum to take the English reading test of the SLC examination?
2. Are the reading strategies specified in the SLC English curriculum in line with what experts think the reading test is measuring?
3. To what extent is there consistency between the strategies test takers use and the expert judgements?

Chapter 3: Methods of data collection and analysis

This chapter presents the methodological approach of the empirical investigation of the study. The chapter has three main sections: methodological framework, data collection and analysis.

3.1 Methodological framework

This study is exploring what is actually being tested by the reading test by looking at how students take the reading test and how experts see the test. It should be noted that looking at how students take the test involves exploring the complex cognitive processes the students follow to respond to the test. Therefore, it was realised that the study should be qualitative in its nature so that in-depth information on those complex issues could be collected. The students' verbal reports were thought to be the most appropriate sources as the reports were expected to provide us the first-hand evidence about the students' cognitive processes when they were engaged in taking the reading test. Additionally, in order to dig more into the issue, the views of the experts were collected. By employing verbal reports and expert judgments in a complementary fashion, the research attempted to pay heed to the complex issues related to the reading strategies.

3.1.1 Access to the field work and ethical considerations

This section describes the processes followed to get access to the field work and recruit the participants along with the careful measures used for ethical consideration.

3.1.2 Location of data collection

Kathmandu was selected as the location for the study mainly for two reasons. First, unlike most parts of Nepal, Kathmandu has a large number of schools. So, the schools could be

found within a small area which could ease the data collection. Second, my familiarity with the community of the students, their school environment, language and culture could help me to collect in-depth information for this qualitative research. Both the expectations were met while collecting the data.

3.1.3 Negotiating access for field work

Having decided on the kind of location, I needed to identify gatekeepers to help me gain access for field work. Two schools (one private and one public) in Kathmandu were purposively selected and their written permission was obtained in December, 2014 before submitting the proposal for ethics review.

After getting approval for the data collection from the ethics review committee at the Open University, I went back to Kathmandu in April, 2015. However, I could not get easy access to the schools as a devastating earthquake of 7.9 magnitude hit Nepal on April 25, taking more than 8000 lives and displacing millions. Because of this disaster, the government of Nepal decided to shut down educational institutions in 11 quake-hit regions including Kathmandu till May 14 as physical infrastructures of schools needed technical inspection. Therefore, all the schools were closed until May 16 but I had to return to Milton Keynes on May 12 (i.e the day another quake of 7.3 magnitude hit the country worsening the situation).

Consequently, I could get access only to the public school as I had gone there before the earthquake hit the country. In the case of the private school, I could not contact to any of the responsible persons in the school as the school was closed until May 8 and none of them were in Kathmandu. Therefore, I selected another private school, and had telephone conversation with the vice principal of the school. Upon my request, he visited

the school although it was closed and gave me oral permission to collect data from the students in the school.

3.1.4 Recruiting participants

At first I visited the public school which had already given me a written consent to collect the data and requested the head teacher of the school to select any four students for the data. Two boys and two girls were selected from the school. Then, the next day, their parents were invited in the school. Each of the parents along with their child was informed about the study verbally as well as through the information sheet which was written in Nepali. Then, the schedule for the data collection was set, starting on 25 April.

However, I had to face an unexpected problem on the very first day of the data collection. I had collected data from the very first participant and was waiting for the second one when we faced the deadly earthquake; I had a narrow escape there. After the earthquake, we had no network to have telephone contact for a few days and we had to sleep under the open sky, with less food and water for seven days because of the continuous aftershocks. Also, I was very much scared to go back to school. Therefore, I could not contact the participants for a week. Then, when I contacted the participants over the telephone after a week, I found that all of them had left Kathmandu. The reality was that about half a million people including students had left Kathmandu and moved to a safe place. Therefore, I contacted the head teacher again and requested him to select other students. The head teacher was so helpful that he visited the school, although it was closed.

Telephoning was the only option for us to contact the students and their parents. Therefore, the head teacher and I looked at the telephone directory of grade 10 students and selected about two numbers (two boys and two girls) randomly and the head teacher

telephoned them one after another, upon my request. However, we found that all of the students were not in Kathmandu. Then, the head teacher picked up several other numbers from the directory and continued telephoning the students until he could make sure that four students were available in Kathmandu and each of them could visit the school next day with one of their parents or a member (senior to her/him) in the family. Therefore, it was not possible to have gender balance of the students. In the case of another school (the private school) too, I had to request the vice principal to select and contact the students for the data. We followed a similar process for the student selection and we had similar problems to find the students and make gender balance. Consequently, we ended up with seven girls and one boy as our student participants from the two schools. However, it did not affect my data as the study did not aim to make any gender-based comparison.

It was more difficult to recruit the subject experts as all the seven experts, whom I had requested for the data, had left Kathmandu after the earthquake. Therefore, I had to contact another group of experts who could meet the selection criteria (see 3. 2.2) to be my participant and were available in Kathmandu. With the help of a teacher at Tribhuvan University, Kirtipur I collected the telephone directories of MahendraRatna Campus, Tanhachal and Tribhuvan University, Kirtipur and selected about 20 teachers from the institutions, thinking that I could find at least seven to eight experts living in Kathmandu. However, I could find only four of them living in Kathmandu. Fortunately, all of them happily accepted my request to become my participant and then with the help of the experts, I contacted three more experts.

3.1.5 Timing of the study

The data collection for the study took place between 15 April to 11 May, 2015. It was the first month of the new academic year in the schools in Nepal. Despite the natural disaster, I managed to secure access to enough participants and collected sufficient data to answer my research questions within the period originally scheduled for data collection.

3.1.6 Ethical considerations

Approval for the research was obtained by submitting a preliminary project proposal to the University ethics committee before commencing the research. For the protection of human rights and data, in addition to The Ethics Principles for Research involving Human Participants at the Open University, British Educational Research Association (BERA) Ethical Guidelines (2011) and European Association for Language Testing and Assessment (EALTA) Ethical Guidelines were followed.

Both the parents and students were informed about the research orally as well as through an information sheet before obtaining the consent of the parents for their permission to record, archive and publish the data related to their children, anonymously. Both the information sheet and consent form were written in Nepali language as most of the parents could not understand English (See appendix, A). The parents' consent was obtained as the children were 14 -16 years old. Additionally, the students' oral consent was obtained.

However, it should also be noted that, as mentioned in 3.1.4, the head teachers were requested to telephone the students and parents. It was my compulsion to request the head teachers in the adverse situation as I did not know any of the students and their parents, and telephoning was the only option for us to contact them. One might also argue that there might have been a kind of pressure on the students and their parents to

accept the request from the head teacher. However, due to the devastating earthquakes and the availability of my time, I had to rely on the head teachers, and to avoid pressure on the students and the parents, the head teachers were very much politely telling them that it was not a compulsion at all for them to be a participant in the research; their participation on the study was completely voluntary.

With regard to the subject experts, they were informed about the research through the information sheet prior to the written consent (See Appendix B for the information sheet and consent form).

During the research, care was taken to avoid causing mental, social or professional harm to the participants by ensuring anonymity in the report, and respecting their wishes to withdraw from the research when they wanted.

3.2 Data collection and analysis

3.2.1 Verbal reports

Following Alderson (1990b) both forms of verbal reports (TAPs and RDs) from the same students were collected in order to maximise the richness of information gathered (Taylor and Dionne, 2000). The verbal reports were collected in order to address the first research question of the study.

It is important to choose students who reflect the target population of the test as closely as possible, namely the students who have completed Grade 10. However, as the investigation was carried out at the beginning of the school year, it was impossible to find students who had just completed the Grade. Therefore, it was decided to select the students studying at Grade ten, the prospective students of the SLC exam, and eight students (seven girls and one boy) aged between 14 -16, studying at Grade 10 (four in a public school and four in a private school) in Nepal participated in this study. All of them

were Nepali native speakers. They had learned English at schools for an average of 12.5 years.

3.2.1.1 Think-aloud and RDs procedure

The data from the students were collected in their respective schools. In order to ensure that the participants were not disturbed by outside noise, a very quiet room was selected and both the sessions (think-aloud and retrospective debriefing) were audio recorded.

The following four major steps were taken to collect the data.

1. The students were told what was required of them following a standardized set of instructions introduced by Taylor and Dionne (2000: 424). The instructions were not tailored to the research questions, in other words, the students were not asked to verbalize which reading strategy they were employing, as it was important not to influence the range of cognitive processes reported. The instructions were given in the native language of the participants to guarantee complete understanding.
2. A warm up task was conducted in order to overcome the inherent challenges of thinking-aloud while engaged in complex cognitive tasks, such as reading. Warm up exercises were expected to make the students feel more confident and less anxious (Ericsson and Simon, 1993), and reduce the need for probing questions and prompting during the verbal report sessions (Afflerbach and Johnson, 1984). Therefore, using another reading test of the SLC exam, the students were given an example of how thoughts can be verbalized and they practised thinking-aloud until they felt thoroughly comfortable with the process involved in thinking-aloud.
3. Immediately after the warm up session, they were asked to take the real test and think-aloud using the language of their choice (i.e either Nepali or English) in order

to ensure that there was no hindrance caused by language use. However, all the students used English to verbalise their thoughts. No time limit was set for the session in order to allow them enough time to think aloud; the individual session lasted approximately 35 to 40 minutes. During each individual think-aloud session, the researcher positioned herself next to the students and took some notes related to their activities. Following Cohen and Upton (2006), the participants were reminded to keep thinking-aloud whenever they remained silent for about 30 seconds; they were politely prompted: "Please say aloud what you are thinking now." However, this was not necessary for most of the participants. It should be noted that the participants were individually trained and also took the test individually.

4. Immediately after they completed the test, they were interviewed about how they solved each item in the test. The researcher used the notes she collected during the think-aloud sessions for asking questions. Although all the students used English to verbalise their thoughts, Nepali was mostly used in their interview as most of them preferred Nepali to use in the interview.
5. Finally, they answered a questionnaire (See appendix c) which includes some questions related to the students' background and the test. The questions were adapted from Holzkecht (2012).

3.2.1.2 Data analysis

The verbal reports of each participant were transcribed following the convention of the previous studies (Alderson, 1990b; Nikolov, 2006) and the transcripts were organized according to the individual test item. Thus, even if a student came back to an item after some time, for example to check the answer, this part was put together with the first part

related to that item. To make the analysis systematic, the TAPs and RDs on each test item were put together. Then, those verbal reports were coded using the software called NVivo10 (Lewins and Silver, 2014). While coding the reports, both the top down and bottom up approaches were followed. Following the top down approach, 12 different codes were developed on the basis of the reading strategies specified in the SLC curriculum, prior to the start of coding the data. Five other codes emerged through the data increasing the total number of the codes to 14. Then, the frequency of each strategy was calculated on an item basis for each participant. It should be noted that the repeated use of the same strategy within a single test item by an individual was counted only once so that the number of participants using the strategy for the same item could be easily identified. Finally, their responses to the questionnaire were analysed. It should be noted that the questionnaire consists of both the open ended and close ended questions. Thus, while analysing the close ended questions the number of frequency for each choices was counted but in the open ended questions, the main points made by the students were collected.

3.2.2 Expert Judgements

In order to address the second research question, following Tavakoli and Barati (2011), it was decided to collect subject experts' views with regard to the inclusion of the reading strategies in the test.

Seven subject experts from Nepal took part in the study. In order to ensure that the right person was selected, two basic selection criteria were set: the experts should have completed a course on 'Language Assessment' at post graduate level and have been involved in English test item writing or marking for at least three years. Four of the

participants had experience in teaching and testing English for about 20 years and the rest about four years.

3.2.2.1 Procedures

First of all, a judgment instrument was developed based on the reading strategies listed in the SLC curriculum. The instrument consisted of a table, which needed to be filled in by the experts indicating which of the strategies listed in the SLC curriculum they thought were tested by each individual item. They were asked to take the test as the students do and decide the strategies assessed by each of the items in the test. They could choose more than one strategy for each item. If they believed that an item was not testing any of the strategies, they could indicate this by ticking a column labeled “other” and mentioning which strategy they believed was being tested. After the judgment, they had to answer a few open-ended questions related to the judgment process(See appendix D); the judgment instrument were adapted from Holzkecht (2012).

They were allowed to take the test home and complete the tasks required; thus there was no communication among the judges. They did the task individually and returned the instrument within four days. The frequency of reading strategies in each item was calculated before they took part in the focus group discussion (FGD). The main purpose of FGD was to discuss the items which received multiple responses i.e the items which showed discrepancies among the experts’ views.

The FGD was conducted in a room of a restaurant, which was convenient for most of the experts to visit. It was ensured that the room was very peaceful so that the experts could fully concentrate on the discussion. In the FGD, unlike Lumley (1993) who forced the experts to have an agreement, the experts were simply encouraged to give reasons for their choices. An attempt was made to have discussion about the strategies which

indicated discrepancies. English was used throughout the discussion as all the experts could speak English very well. The FGD was audio recorded.

3.2.2 2 Data analysis

In order to make the analysis systematic and clear, the data were analysed in four steps. First, the frequency of the strategy judgements was calculated on an item and task basis. Then, the category “other” was looked at in more detail with regard to the comments provided by the experts in order to identify the reading strategies not specified in the SLC curriculum. Then, the answers to the open-ended questions were analyzed. Finally, the information obtained from the FGD were analysed on item basis by collecting the views expressed by the experts.

Chapter 4: Results

This chapter presents the results of the study. For the systematic presentation of the results, they are presented in the order of the three research questions. Therefore, the reading strategies used by the students are presented at first, which are followed by the judgments made by the experts and the final section makes a comparison between the strategies used by the students and the expert judgments. The final section makes an analysis of the test material being investigated in the study.

4.1 Reading strategies used by the students

The first research question aimed to explore the reading strategies employed by the students to respond to the test. Therefore, the TAPs and RDs provided by the students were minutely analyzed using some codes. Table 1 presents the coding scheme used to analyze the data.

Table 1: Categories used in the coding scheme

*DetailUnd = Reading for detailed understanding

*GeneralUnd= Reading for general understanding

*UndThemes= Reading for understanding the underlying themes and ideas of the text

*UndArgument= Reading for understanding of an argument

*Gist= Reading for gist

*SpecificInfo= Reading for specific information

*TextOrganisation = Reading for understanding texts organization

*TextContinuation = Reading for anticipating the likely continuation of the interrupted text

*DiagrammaticInfo = Reading for interpreting the information presented in diagrammatic forms

*LexicalItems= Reading for deducing the meaning of unfamiliar lexical items

*AppreciatingTx = Reading for appreciating literary text

*DictionaryUse= Using authentic English dictionary

Wmatch= Word matching

Smatch= Sentence matching

OwnInterp= Own interpretation

GSmatching= Matching grammatical structure

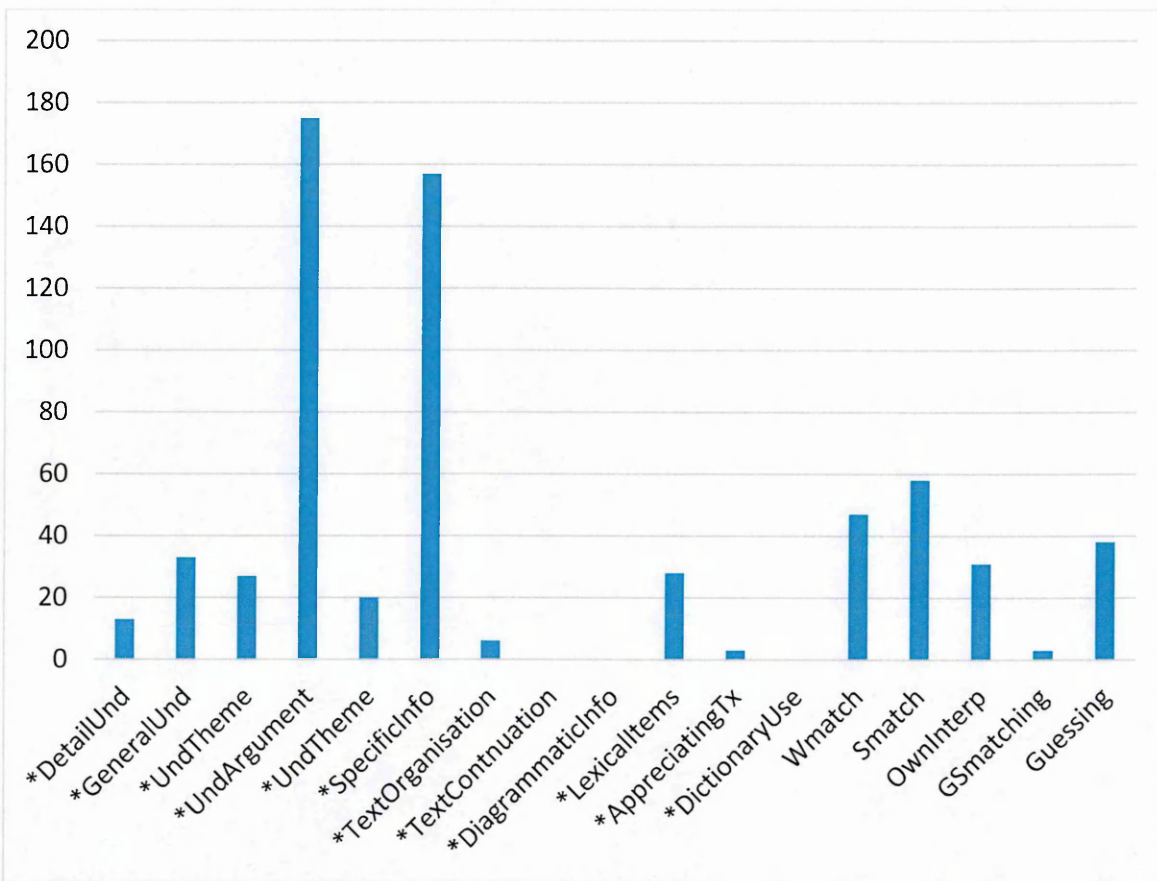
Guessing= Guessing

***= reading strategies listed in the SLC curriculum**

The analysis of the data indicates that the students employed 14 different strategies but only nine of them (DetailUnd, GeneralUnd, UndThemes, UndArgument, Gist, SpecificInfo, TextOrganisation, LexicalItems and AppreciatingTx) are specified in the SLC curriculum. No evidence was found for the employment of the three strategies: TextContinuation, DiagrammaticInfo and DictionaryUse, which are specified in the SLC curriculum. However, there were evidences for the use of five other strategies (Smatching, Wmatching, OwnInterp, GSmatching and Guessing) which are not specified in the SLC curriculum.

It should be noted that following Cohen and Upton (2006), it was felt that some effort at quantifying the verbal report data would help to lend more rigor to statements about the frequency of strategy use. Hence, simple quantitative units like numbers and percentage are used, though they play no major role in the data interpretation. There were altogether 639 occurrences of the strategies. The overall frequencies of the strategy use are presented in Diagram 1.

Diagram 1: Total Frequencies of strategy use



*** = Reading strategies listed in the SLC curriculum**

In the following section, all the strategies are presented mostly in the order of their frequency (highest to lowest). Additionally, each of the strategy is seen on the basis of the framework introduced in 2.2 which includes two levels of engagement (global and local) and two types of reading (careful and expeditious). Some examples are also drawn from

the students' verbal reports in order to illustrate the application of the reading strategies. However, it should be noted that the examples presented do not merely represent the strategy mentioned; they can be an example for the application of other strategies too as, in many cases, multiple strategies were assigned to the same chunk of the TAPs data.

4.1.1 Reading for understanding of an argument

The most common strategy used by the students was UndArgument. This strategy involves careful reading of a text at a global level to understand an argument made through the statements. The strategy was identified 175 times (making up 27.4% of all strategy use). It was employed across all the item-types, except matching item. Here is an example from the transcriptions of a participant's TAPs for the strategy use.

Participant 3 Task 2 Item 17(true/false item)

(Reads the statement) The writer thought Jack's mother was not so wise... Um here the meaning is ... Jack's mother is not wise...Um let me read... (Reads some sentences from the text) Um here it is: Jack's mom, I decided was a wise woman. It means ...the writer... um wise woman... he decided was a wise woman...um this means Jack's mother was wise. So, the sentence is false.

This excerpt indicates that the student is trying to understand the argument made by the writer, particularly through the sentence 'Jack's mom, I decided was a wise woman'.

4.1.2 Reading for finding specific information

The second most frequent strategy was SpecificInfo. This strategy involves expeditious reading (speed reading) of a text at a local level to find specific piece of information. Very similar to UndArgument, this strategy was observed across all the item-types, except the

matching item. It was observed 157 times (making up 24.6% of all strategy use). Here is an example from the transcription of the TAPs for this strategy use.

Participant 7 Task 4 Item 13 (short answer question)

What was wrong with the wipers?... Um I think I have seen the words 'wipers' somewhere in the first paragraph... (Reads the first two sentences)...No no not here should be in the second paragraph... jumps to the second paragraph...Um here it is... his wipers had not been working.

It can be seen in the excerpt that the student just focuses on a piece of information related to the 'wipers' and quickly goes through the text.

4.1.3 Reading and matching similar sentences

The third most common strategy was Smatching. This strategy involves expeditious reading of a text at a local level. However, one might argue that this strategy involves careful reading of a text. But in the study, the students were found picking up a sentence from the test items and looking for a similar sentence in the text. Therefore, it is included under expeditious reading. An example of the employment of the strategy by a student to respond to the question 'How do parents make trouble to themselves?' has been presented below.

Participant 2 Task 3 Item 21 (short answer question)

How do parents make trouble to themselves?...Um um...I think I have seen similar sentence in the text... um parents make trouble to themselves, parents make trouble to themselves (Reads 7th and 8th sentence in the text)...no no not here... (jumps to the first line)...uh it's here...they make trouble for themselves.

In the excerpt, it can be seen that the student is looking for a similar sentence in the text and matches the sentence with the test item. Smatching was applied to respond to three different item-types: true/false, gap filling and short answer questions.

4.1.4 Reading and matching lexical items

Wmatching was the fourth most common strategy. It involves expeditious reading of a text at a local level. It was employed to solve all the item-types, except ordering sentences. The following excerpt illustrates the employment of Wmatching.

Participant 3 Task 4 item 25 (gap filling)

R: Okay, let's see this item now. They can submit their essays latest by... How did you find the words to fill in this gap?

S: It was very easy. So I just quickly looked for the words 'latest by' and found the words there in the text. See here it is written 'latest by 31 December, 2013'. So I wrote 31 December, 2013.

The excerpt clearly shows that the student just focused on the words 'latest by' and quickly looked for the words in the text and matched the words.

4.1.5 Reading and making a guess

Guessing was employed particularly when the students were not able to understand either the text or the questions. However, it is very difficult to place it within the framework used in the study (i.e levels of engagement and reading types) as guessing could be made without necessarily reading the text and/or without necessarily being engaged with the text. It was employed to all the item-types, except the matching item.

The following excerpt indicates the application of Guessing to respond to the true/false item 'Domestic walls have fragmented the world'.

Participant 1 Task 1 Item2 (True/false item)

R: Okay, you think this statement is true. Why?

S: Actually, I could not understand this sentence. Therefore, I just guessed the answer.

R: How did you guess then? Was there anything like a sentence or word etc. that helped you to guess the answer?

S: No no nothing.

This excerpt indicates that the student is not able to understand the test item and just makes a wild guess without reading the text. However, she is able to respond to the item correctly.

4.1.6 Reading for general understanding

The strategy 'GeneralUnd' involves reading a text carefully at a global level to have general understanding of the text. However, one might also argue that it involves expeditious reading of a text. But, in the study, the students were found reading the texts carefully to have general understanding of the text. Therefore, this has been included under careful reading of the text. The students used this strategy to solve three different items: true/false, matching and ordering sentences. The following excerpt indicates the application of GeneralUnd to respond to the short answer question 'What is the title of the essay?'

Participant 8 Task 4 Item 29 (Short answer question)

R: In order to answer this question, I could see that you read the whole text. What were you doing?

S: I was trying to understand overall meaning of the poem as I could not find the title there.

4.1.7 Reading for detail understanding

This strategy involves careful reading of a text at a global level for understanding all the information mentioned, either explicitly or implicitly, in the text. The strategy was employed to respond to two different tasks: gap filling and ordering sentences. The following excerpt indicates the application of DetailUnd to respond to the gap filling item.

Participant 1 Task 1 Item 6 (gap filling item)

R: Okay, fearless situation makes people can get ... of all kinds. What have you written here?

S: I am a bit confused with this. I have written the word 'knowledge'.

R: I could see that you were reading the whole poem for several times before answering this question. Why?

S: I could not find the answer. So, I tried to understand each line of the poem.

4.1.8 Reading for understanding themes and ideas in the text

This strategy involves careful reading of a text at a global level. The strategy was employed to three different tasks: gap filling, true/false items and ordering sentences. The following example indicates the application of UndThemes to respond to the true/false item.

Participant 5 Task 1 Item 3 (true/false item)

(Reads the statement) 'The poet prays for his individual freedom'. Um here the meaning is ... oh yes for individual freedom means for him only. No no it should be false. Um let me read the poem... um here it is: Into the heaven of freedom, my Father; let my country awake ...So he is praying for the freedom of his country but not for his individual freedom.

This excerpt indicates that the student is trying to understand the underlying themes and ideas presented through the lines of the poem.

4.1.9 Reading for understanding the gist

This strategy involves expeditious reading of a text at a global level. In the study, it was employed to solve four different items: true/false, gap filling, ordering sentences and short answer questions. The following example indicates the application of this strategy.

Participant 4 Task 3 Item 20 (ordering sentences)

R: Can you remember how you ordered the sentences given here?

S: I read the text and understood the gist of the text. That helped me to order the sentences.

4.1.10 Reading for understanding texts organisation and appreciating literary text

These strategies involve careful reading of a text at a global level for understanding the paragraphs or the ideas in the text and appreciating the literary text. There seems to be a strong relationship between item-type and the employment of these strategies as both of these strategies were observed only for 'ordering sentences'. The test contains only two

Items (11 and 20) that require the students to order the sentences (see appendix E). The following example indicates the application of these strategies.

Participant 7 Task 2 Item 11 (ordering sentences)

R: Okay, how did you order these sentences?

S: It's a bit easy. I really enjoyed the text; I found the story very much interesting. I could easily understand the text.

R: Do you know the order of the events presented in the text.

S: Yes, I do. I can remember the sequence very well. I also enjoyed doing this exercise.

The excerpt indicates that the student appreciates the text as she finds the story interesting. Also, she tries to understand how the information is organised in the text.

4.1.11 Reading for inferring the meaning of unfamiliar lexical items and matching grammatical structure

Both of these strategies involve expeditious reading of a text at a local level. Very similar to TextOrganisation and AppreciatingTx, there seems to be a strong relationship between item-type and the employment of these strategies as they were observed only for the item 'finding the words with similar meaning'. The test contains four items of such types (Item 7 to 10). Here is an example from their TAPs which indicates the employment of these two strategies to find a word that has a similar meaning with the word 'searched'.

Participant 4 Task 2 Item 9 (finding similar words)

Searched um um searched uh...(reads from the beginning of the second paragraph and stops at 'looked for'...(Repeats the phrase 'looked for')... um...here it is said 'I

looked for my khalasi but could not find him'. Um ...he could not find means his Khalasi was not there ...I think searched means looked for... it should be the answer here...(7) um also both searched and looked are in past form.

The transcription indicates that the student first looks at the context where the phrase 'looked for' is used and tries to understand the meaning with the help of another word 'find'. He also compares the verb forms of 'search' and 'look'.

4.1.12 Reading and making own interpretation

Finally, evidences for the use of OwnInterp were found in their verbal reports. OwnInterp involves reading and interpreting the text using their own logic rather than making effort to understand the arguments made in the text. However, it is very difficult to include this strategy under the framework used in the study as it might involve either careful reading or expeditious reading, also either at local or at global level engagement. It was applied to three different item types: true/false, gap filling and short answer questions. The following excerpt exemplifies the employment of OwnInterp.

Participant 8 Task 3 Item 18 (True/False item)

R: Okay, you think that the statement 'Freedom from the parents made children turn positive' is false. Why do you think so?

S: I know that freedom from the parents is not good for children. If we are free, we do many naughty things. So the result will be negative.

R: I see. It's your logic.

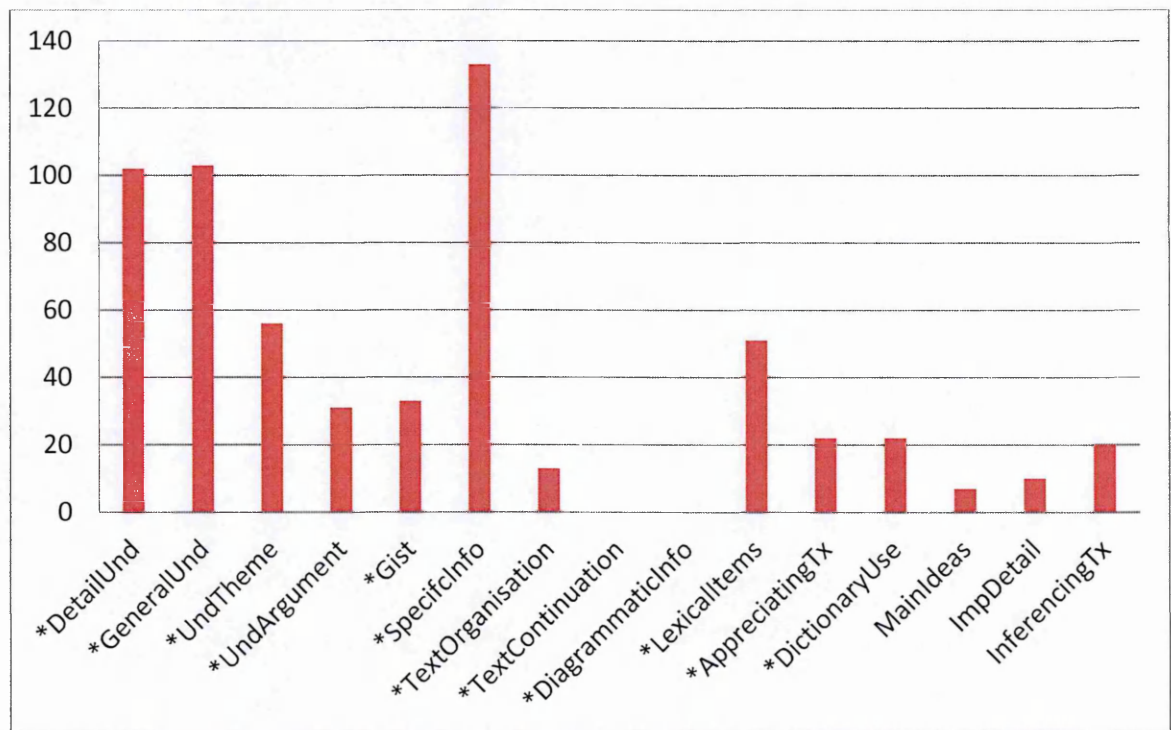
Here, the student makes his own interpretation of children's freedom without reading the text. However, the text interprets it differently indicating that freedom for children can turn positive results.

4.2 Expert judgements on the reading strategies

In order to analyze the expert judgments data, the same codes used in the verbal reports for the reading strategies specified in the SLC curriculum were used. However, the test was also judged to be testing three other strategies: careful reading for main ideas, careful reading for important details and careful reading for making inference of the text. The three strategies are coded as MainIdeas, ImpDetail and inferencingTx respectively.

It should be noted that, very similar to the verbal reports, the frequencies of each of the strategies were counted on an item basis from all the seven experts so that the level of agreement among the experts could be found. The overall strategy judgments made by the experts on the test are presented in diagram 2.

Diagram 2: Total frequencies of strategy judgements



* = Reading strategies listed in the SLC curriculum

Diagram 2 reveals that the test was judged to be testing 13 different strategies. However,

only ten of them are specified in the SLC curriculum. The test was judged not to be testing the two reading strategies: AppreciatingTx and DiagrammaticInfo, which are specified in the SLC curriculum. However, the test was judged to be testing three other strategies (MainIdeas, ImpDetail and InferencingTx), which are not specified in the SLC curriculum.

A close inspection of the judgements on an item basis reveals that, with the possible exceptions of a few items, the great majority of the items were judged to test at least five different strategies or more. The most extreme items were Item 1 and Item 3 (the true/false items) which were judged to test nine different strategies each. However, with the possible exceptions of Item 24 (matching item), the amount of agreement among the judges for the items was very low; the judges show a great degree of variation in each item. For Item 24, six out of seven judges agreed that it is testing LexicalItems and five of them go for DictionaryUse.

The most extreme disagreement concerns the judgements of GeneralUnd between Judges 4 and 5. While Judge 4 thought that none of the items was targeting GeneralUnd, Judge 5 indicated that all of the items, except Item 7 to Item 10, were testing GeneralUnd. Similar kind of disagreement can be identified throughout the data. For instance, Judge 3 did not think that any of the items was testing UndThemes. In contrast, Judge 4 thought that 14 items were doing so. Similarly, Judge 7 classified only four items to be focussing on LexicalItems, while Judge 2 assigned this strategy to 15 different items. Thus, a close observation of the judgements made by individual expert on an item basis confirms that the judges were unable to agree on which reading strategies are being tested by the individual items.

Consequently, item-type wise analysis also indicates low level of agreement among the experts with regard to the strategies tested by each item types. For instance, the true

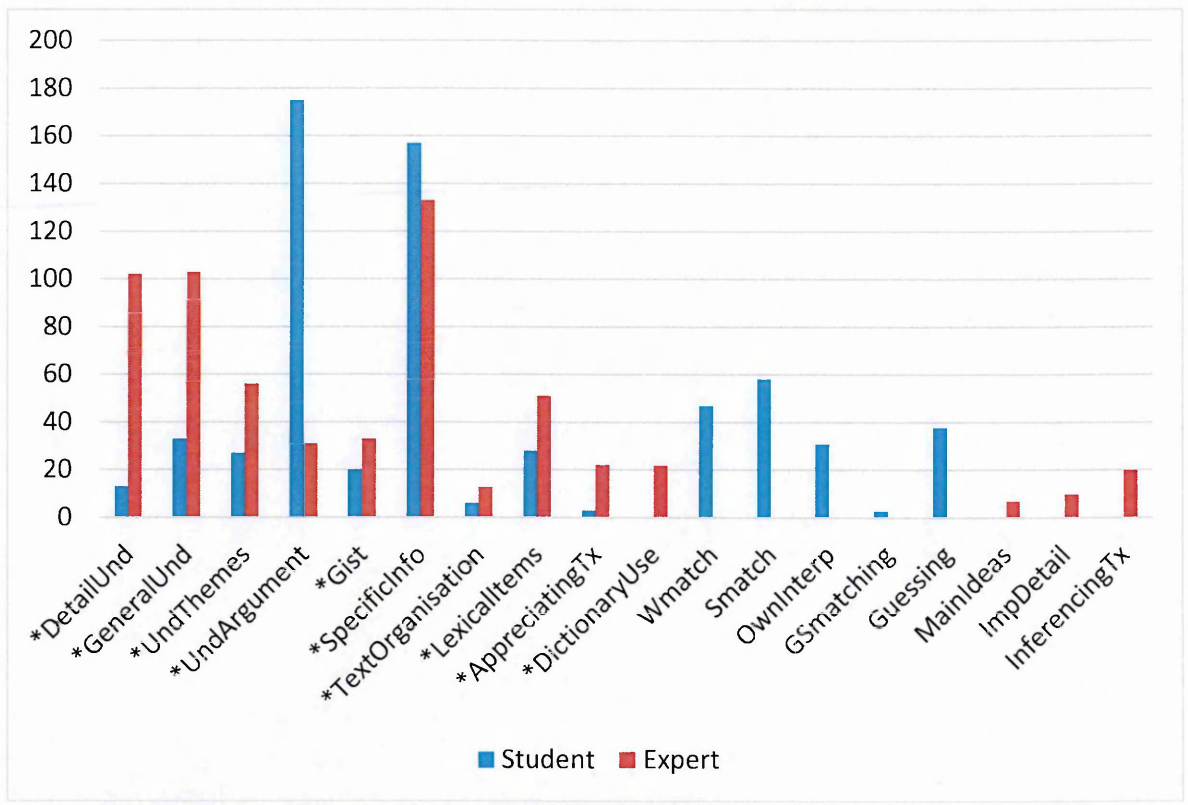
false items were judged to be testing nine different strategies: DetailUnd, GeneralUnd, UndThemes, UndArgument, Gist, SpecificInfo, LexicalItems, AppreciatingTx and InferencingTx). However, the experts had considerable agreement on just two reading strategies: SpecificInfo and GeneralUnd. Similarly, the Gap Filling items were judged to be testing six different strategies: SpecificInfo, GeneralUnderstanding, DetailUnd, UndThemes, Gist and InferencingTx, but there was considerable amount of agreement among the judges only for SpecificInfo and GeneralUnd. Very similar to the True/False items and Gap Filling items, the experts had very low level of agreement on most of the strategies assigned to the Short Answer Questions, Finding Similar Words and Ordering sentences. However, there was a high level of agreement among the experts with regard to the strategies tested by the Matching item, which was judged to be exclusively targeting LexicalItems and DictionaryUse. Among the seven experts, six of them thought that the item is testing LexicalItems and five of them also thought that the item is testing DictionaryUse.

It should also be noted that the experts could not agree as to whether the individual items were eliciting expeditious or careful reading strategies as almost all of the items include judgements of both types of reading strategies.

4.3 Consistency between the strategy use and the expert judgements

The third research question aims to explore the consistency between the strategy use and the expert judgements. Close inspection of the results indicates that due to low level of agreement among the experts on most strategies, the gap between the two is very big. The overall comparison is presented on diagram 3.

Diagram 3: Overall comparison between strategy use and expert judgements

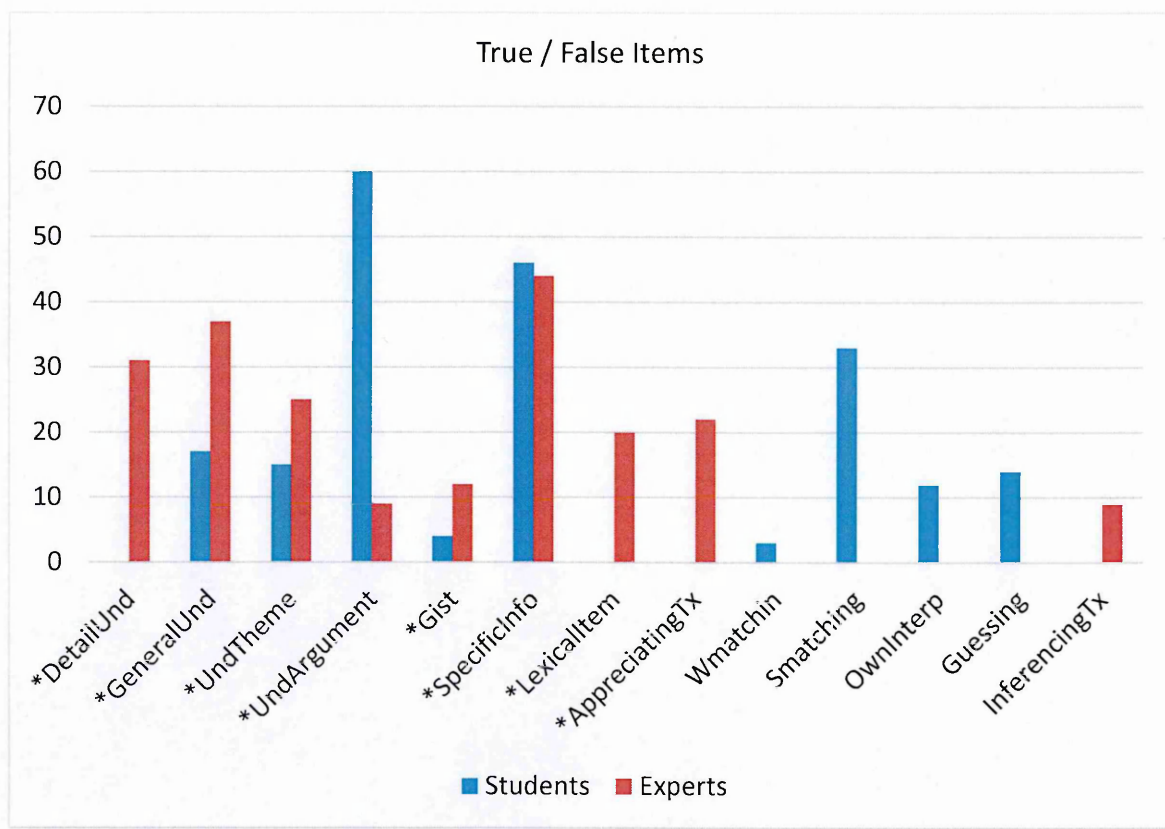


* = Reading strategies specified in the SLC curriculum

Diagram 3 reveals that among the 18 strategies, there is a considerable consistency only on SpecificInfo and TextOrganisation; there is some amount of consistency on seven other strategies (DetailUnd, GeneralUnd, UndThemes, UndArgument, Gist, LexicalItems and AppreciatingTx). However, the students seemed to be using five different strategies that are not specified in the SLC curriculum, but none of the items on the test was judged to be testing those strategies. Conversely, the experts thought the test tested three other strategies that are not specified in the SLC curriculum but no student seemed to be using those strategies. The following section presents item-type wise comparison.

4.3.1 True/false items

Diagram 4: Comparison of strategy use and strategy judgment for True/False items

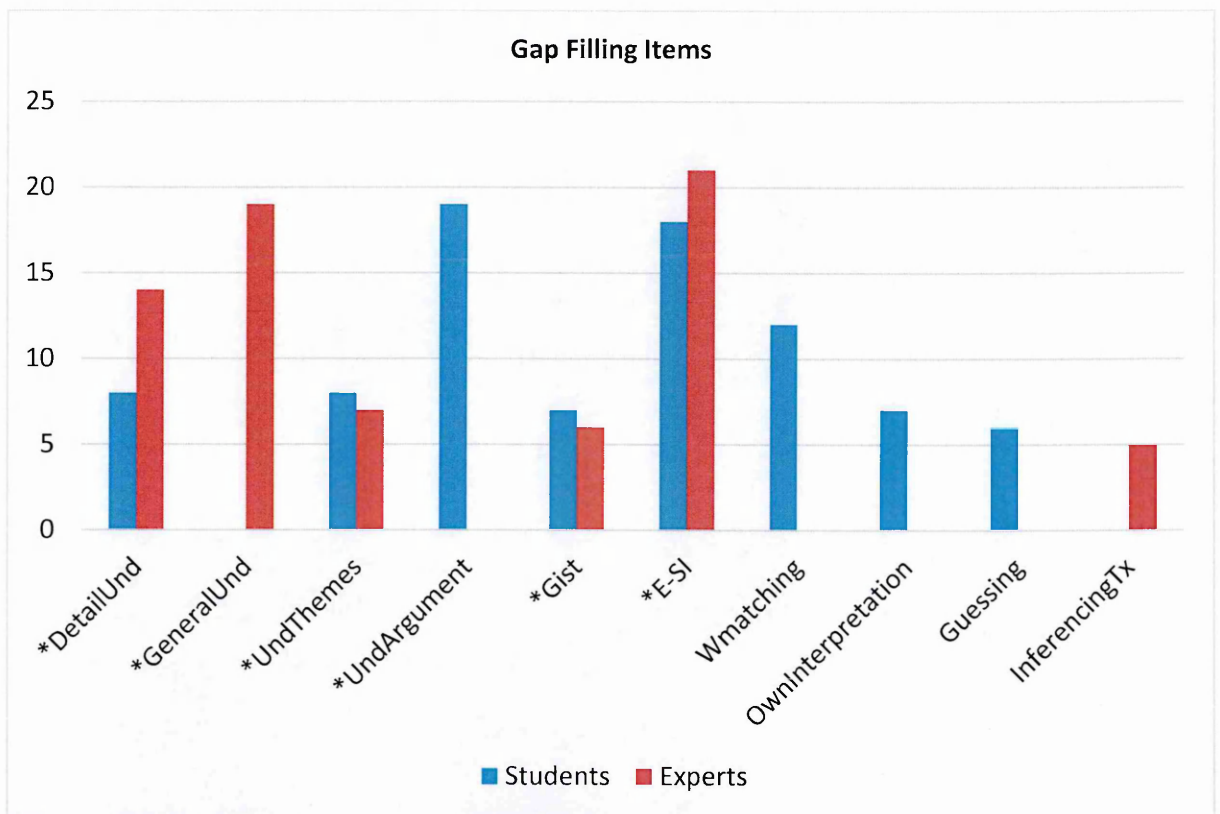


* = Reading strategies specified in the SLC curriculum

Visual inspection of the data reveals that among the 13 strategies assigned to the true/false items, only five of them (GeneralUnd, UndThemes, UndArgument, Gist and SpecificInfo) received some amount of consistency. There was no evidence for the use of DetailUnd, LexicallItems and AppreciatingTx to respond to true/false items, but the experts thought that the items tested those strategies.

4.3.2 Gap filling items

Diagram 5: Comparison of strategy use and strategy judgements for Gap Filling Items



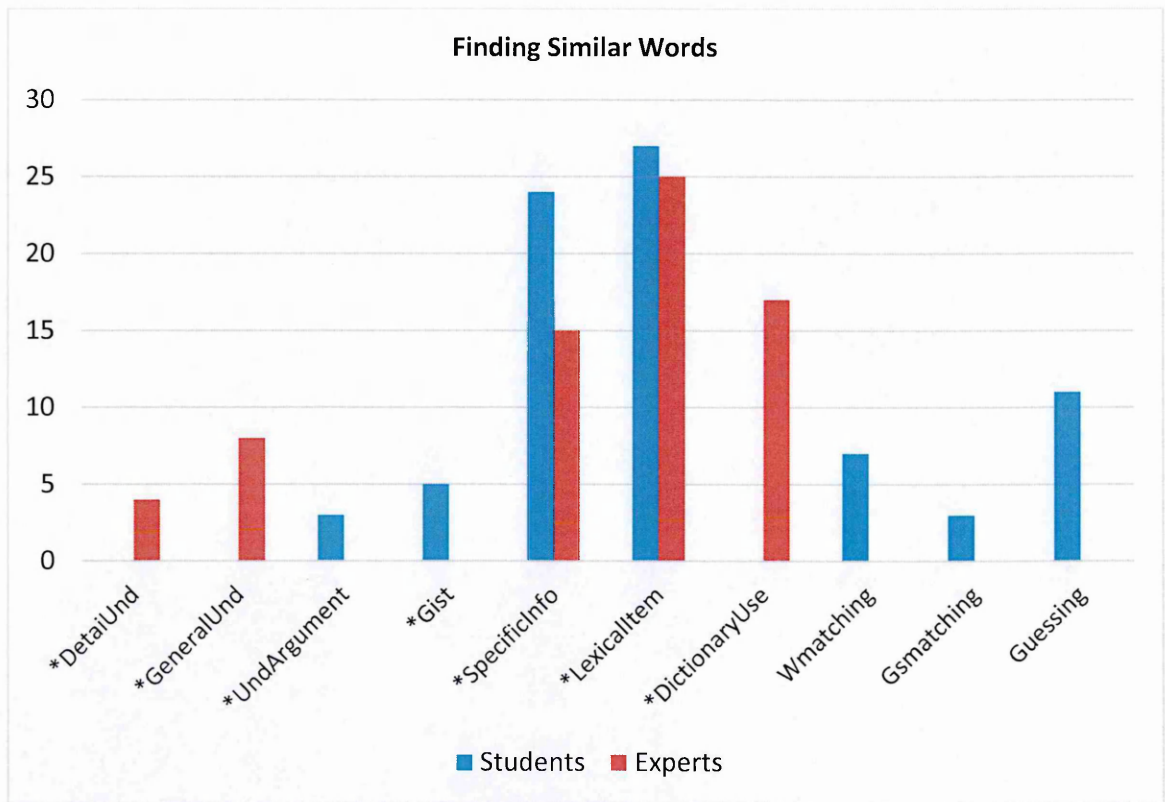
*=Reading strategies specified in the SLC curriculum

Diagram 5 indicates a considerable consistency in terms of the three strategies: SpecificInfo, UndThemes, and Gist; DetailUnd also receives some amount of consistency. However, there is a complete inconsistency in terms of the other six strategies assigned to the 'gap filling items'. For instance, the experts thought that Item 25 tested five different strategies: DetailUnd, GeneralUnd, UndThemes, Gist and SpecificInfo. However, none of the strategies, except SpecificInfo were identified in the verbal reports. The item was solved by means of UndArgument, SpecificInfo, Wmatching and Gist.

4.3.3 Finding similar words

Diagram 6: Comparison of strategy use and strategy judgements for Finding Similar

Words

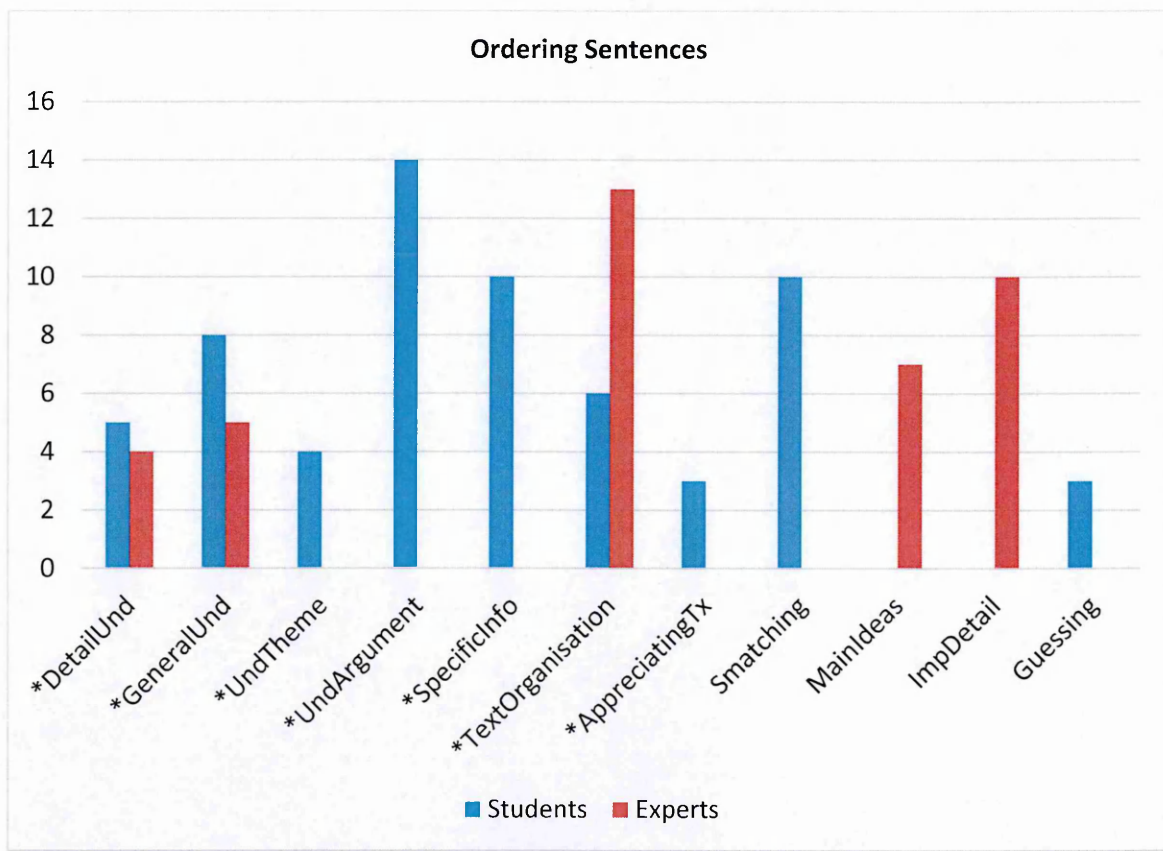


*=Reading strategies mentioned in the SLC curriculum

The inconsistency between the strategy use and strategy judgements for the items 'finding similar words' is stronger than for the true/false and gap filling items as, except SpecificInfo and LexicalItems, the expert judgements completely contradicted the findings of the verbal reports data.

4.3.4 Ordering sentences

Diagram 7: Comparison of strategy use and strategy judgements for Ordering Sentences

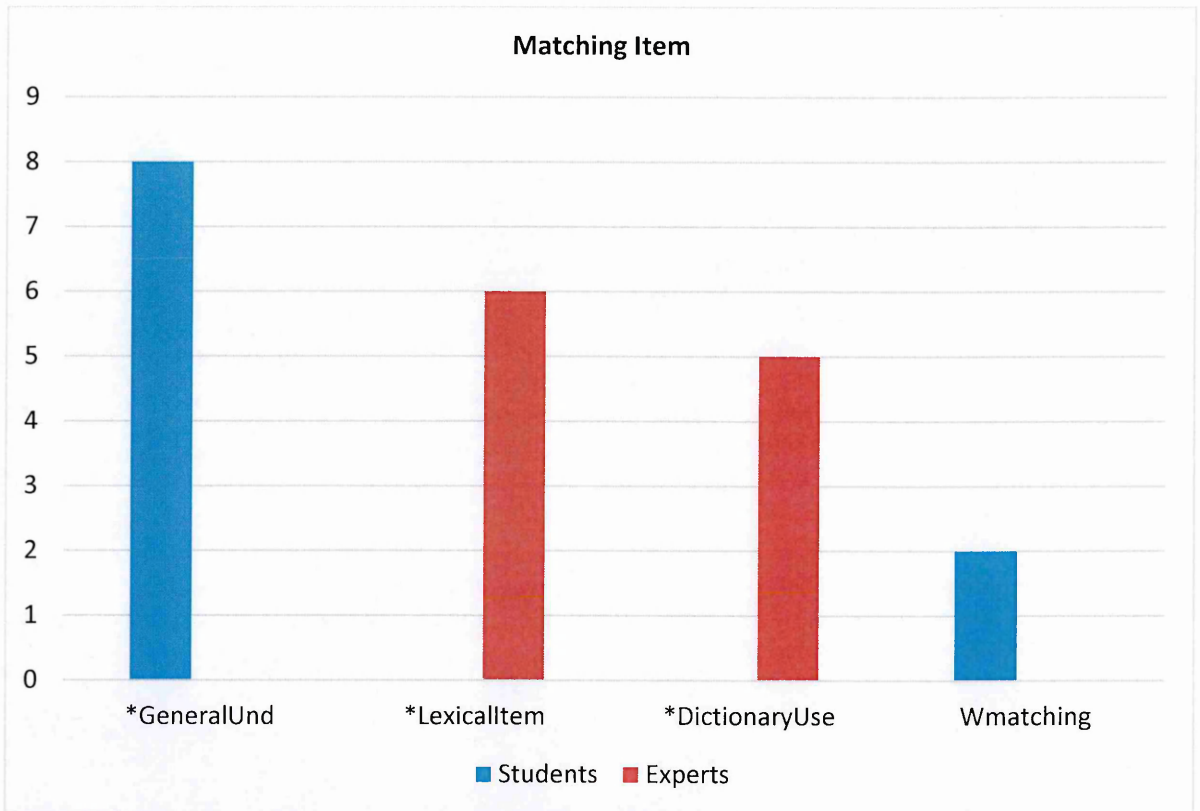


*=Reading strategies mentioned in the SLC curriculum

There is a big gap between the strategy use and strategy judgements in terms of the items 'Ordering Sentences'. Among the 11 strategies assigned to these items, some amount of overlapping can be seen only on three strategies: DetailUnd, GeneralUnd and TextOrganisation.

4.3.5 Matching item

Diagram 8: Comparison of strategy use and strategy judgements for Matching Items

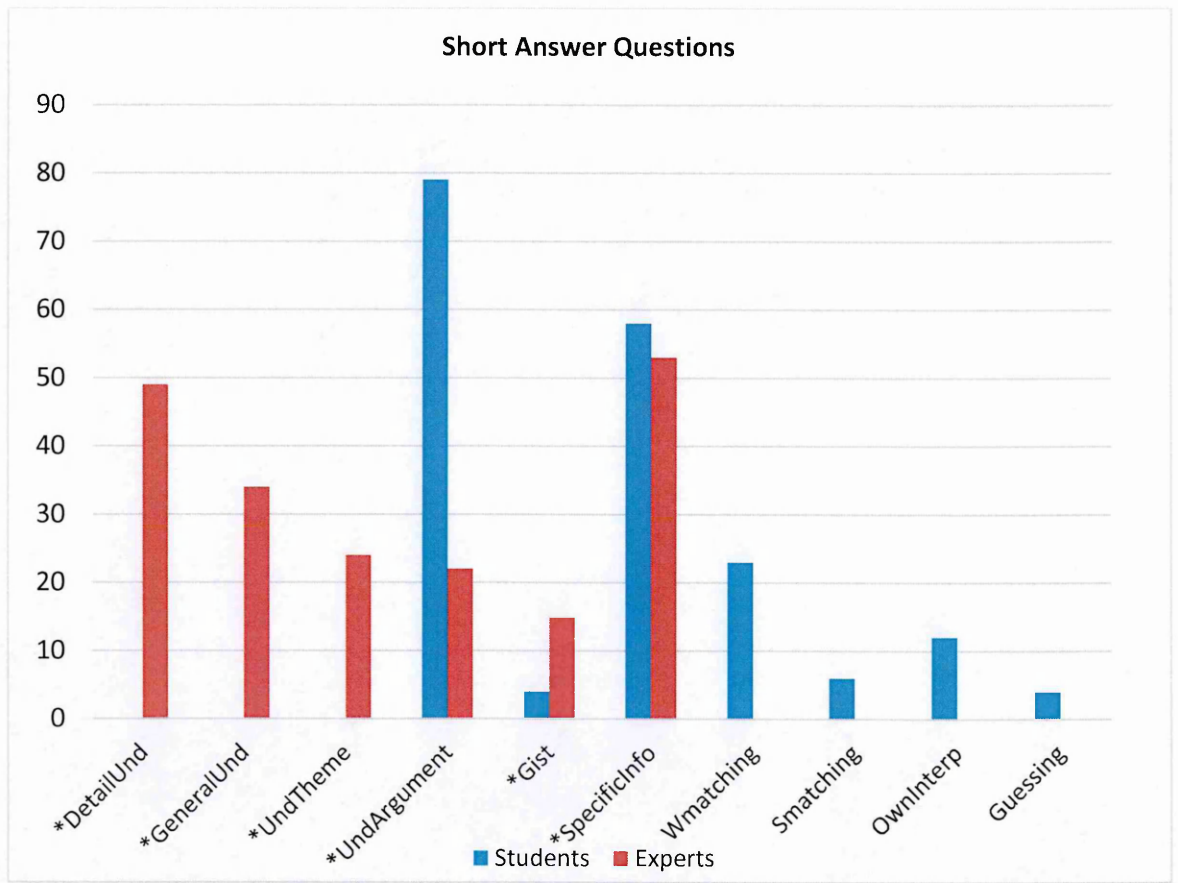


*=Reading strategies mentioned in the SLC curriculum

Diagram 8 indicates complete inconsistency between the strategy use and strategy judgements. The verbal reports show that all the eight students employed GeneralUnd to solve the Matching Item (Item24) and two of them also applied Wmatching. However, the experts thought that the item tested LexicallItems and DictionaryUse.

4.3.6 Short answer questions

Diagram 9: Comparison of strategy use and strategy judgements for Short Answer Questions



*=Reading strategies mentioned in the SLC curriculum

Very similar to the other item-types, the expert judgements contradicted the findings of the verbal report data for almost all the strategies. Among the 10 different strategies assigned to these items, only SpecificInfo received noteworthy agreement; UndArgument and Gist also indicate some amount of overlapping. Except these three strategies, the strategies which were found in the verbal reports did not occur in the expert judgements data and vice versa.

4.4 Analysis of the test material

The test being investigated is the reading section of the SLC English exam (set 5, the set for far western region) of the academic year 2013/2014 (See Appendix E). The section consists of four texts (poem, narrative text, explanatory text and advertisement) with different lengths; two of the texts are taken from their textbook, but most of the items in the textbook have been modified. Among the four texts, most of the students found the poem to be the most difficult one while the narrative text the easiest one. The test contains 29 items in total, which are divided into six different item-types: true/false (8), gap filling (3), short answer questions (11) ordering sentences (2), finding similar words (4) and matching (1).

Among the six item-types, matching item was considered to be the easiest one while sentence ordering the most difficult one by most of the students, despite the fact that all the students were familiar with all the item-types included in the test. Additionally, all the students reported that all the instructions given for each item types were clear to them. However, a close observation of the test creates some confusion. For instance, the instruction given for ordering the sentences in Item 20 is: Rewrite the sentences in an appropriate order. The instruction is lacking clarity in the sense that it does not clearly tell whether the sentences need to be ordered on the basis of the organization of the text or just in a logical order.

An analysis of the test also reveals that the test requires students to make some assumptions. No clear answer can be given to some questions. For example, the gap filling item (Item 6) 'Fearless situation makes people can get... of all kinds' can be answered differently; possible words to fill up the gaps might be perfection, freedom, thoughts, optimism etc.

Similarly, some test items are completely irrelevant to the text, based on which the test items are designed. For example, one of the true false items (Item 4), 'A stream can have good effects in the desert', has no relation to the poem. The test also contains so many grammatical errors. Additionally, in some cases, the texts do not provide any information to answer the test items. For instance, (Item 13) 'What was wrong with the wipers?' The text does not provide any information about the real problem with the wipers.

Additionally, there are lots of repetitions of the item-types and so many grammatical and typological errors.

Chapter 5: Discussion

This chapter is designed to discuss the findings of the study. The findings are discussed in the order of the research questions followed by a discussion on the construct validity of the test and the methodological issues related to the study. The final section makes critical observation of the test being investigated.

5.1 Research questions

5.1.1 Research Question 1

The first research question in this study was: Do test takers employ all the reading strategies outlined in the SLC English curriculum to take the reading test of the SLC examination? Contrary to Salehi's (2011) findings, which indicated a very good coverage of all the reading strategies intended by the high-stakes test, the verbal reports in this study indicate that only 9 (out of 12) of the reading strategies were used by the students. However, the two studies differ in terms of the methodologies as Salehi used factor analysis, a quantitative method, and this study is a qualitative one using verbal reports. But, the results are in consistent with the results of the study by Alderson (1990b), which also used verbal reports.

In order to make comparisons between the strategies in terms of their occurrence, some efforts were made to count the number of frequencies for each of the strategy use in each test item. It should be noted that none of the students was found to be inclined to use a certain strategy; all the students were found using all the strategies. Therefore, none of the strategy's frequency rate was affected by individual preference of the students.

The frequencies of the reading strategies found in this study reveal several important findings. One of the most important findings concerns the number of strategies measured by each of the test items. With the possible exceptions of the Items 5, 18, 19 and 24, the students used at least five different strategies to respond to the test-items. Therefore, as Alderson (1990a) states, "it is unlikely that any test item can be unambiguously said to be testing any one skill" (p. 436).

Item-type wise analysis also indicates that each of the item types, except matching item, seemed to measure minimum seven reading strategies. However, only two or three strategies were frequently employed by the students. Additionally, as claimed by Anderson, Bachman, Perkins and Cohen (1991) item-type is not a very reliable predictor of the patterns of strategy use; the students employed similar type of strategies to approach different items. Consequently, the strategies did not cluster on an item-type basis, with the possible exceptions of TextOrganisation and AppreciatingTx (occurring only on ordering sentences) and LexicalItems and GSmatching (occurring only on finding similar words). However, Farr, Pritchard, and Smitten (1990) argue that readers employ processes which are specifically prompted by the type of task they are asked to perform.

The strategies applied to some item-types seem to be affected by the students' pre exposure to the text. For instance, in each task, short answer questions appear only at the end preceded by several other questions and the students were found strictly following the order of the questions in each task while taking the test. Therefore, the students had a lot exposure to the text by the time they reached these items. Because of this, as Cohen and Upton (2006) argue, the items were less demanding than they probably would have been, had they appeared as the sole items accompanying the text. Consequently,

the range of what was being measured by those items was somewhat constricted due to the repeated prior exposure to the text.

It should also be noted that, as mentioned in 4.1, among the 14 different strategies, the two most frequent strategies were UndArgument and SpecificInfo. The possible reason for using UndArgument most frequently might be the case that the test items, to a vast majority, required the students to understand the argument made either implicitly or explicitly. Similarly, the possible reason for the application of SpecificInfo very frequently could be the nature of the test material. The test is designed in such a way that single text is followed by several questions. In order to answer those questions, the students might have gone back to the text for several times and looked for specific piece of information related to those questions.

The test seems to have equal focus on both types of reading (careful and expeditious) just like the IELTS exam did (Weir, Hawkey, Green and Devi, 2006). UndArgument and SpecificInfo were the first two most common strategies employed by the students. Also, the two strategies were not tested separately as UndArgument mostly occurred in combination with SpecificInfo. This finding contradicts the claim made by Weir, Huizhong and Yan (2000) that, traditionally, there has been a lack of focus on expeditious reading in the testing of reading.

With regard to the reading level, very similar to new TOFEL (Cohen and Upton, 2006), the successful completion of this test required the students to have both a local and global level understanding of the test passages. Some of the items like ordering sentences indeed challenged the respondents to understand the text as a whole understanding the arguments made either implicitly or explicitly mentioned in the text and also understand lexical, grammatical and logical links in order to determine the logical order of the

sentences, whereas some other items like matching item and finding similar words were simply focusing on word level understanding.

The study also concerns the lack of evidence for the employment of DictionaryUse, DiagrammaticInfo, and TextContinuation, which are specified in the SLC curriculum. The reason for not using DictionaryUse by the students was obvious; following the rules of the SLC exam, the students were not allowed to see any kind of dictionary while taking the test. The students did not employ DiagrammaticInfo as the test did not contain any diagrammatic information. Similarly, as the test contains no interrupted text, there was no evidence for the application of TextContinuation.

Conversely, the results indicate the use of five other reading strategies which are not specified in the SLC curriculum. Among them, Smatching, Wmatching, and Guessing were, the third to fifth respectively, the most common strategies employed by the participants. Very similar to the claim made by Yang (2006), the students were consciously matching words and sentences and also guessing the answers either on the basis of the context or just making a wild guess.

Another important finding concerns the relationship between test item difficulty level and strategy use. It was really difficult to decide the type of strategy they were using for too easy item (e.g Item 24) as the adoption of any strategy was not clearly observable to that item. This finding supports the claim made by Alderson, Clapham and Wall (1995: 177) that when responding to too easy items, readers are likely to display only “highly automated” processes as they are “less subject to conscious control” and are thus “unreportable”. This might further indicate that low difficulty level of some of the items might have had an impact on the results of the findings as certain processes might not

have been captured. Conversely, it was found that sentence ordering was the most difficult item, to which they employed the highest number of strategies.

The study also supports the claim made by Carrell (1992) that “the use of certain strategies does not always lead to successful reading comprehension” (p.168). In some cases, all the students were found to be approaching a test item in a similar way, but some ended up with correct answers while the others with incorrect answers. For instance, in order to find the words with similar meanings in Item 9 and Item 10, all the students tried to infer the meaning in context (using LexicalItems) but only three of them ended up with correct answers. This example also clarifies another finding in this study that sometimes the participants seemed to demonstrate the skills supposedly required by an item, yet, failed to answer the item correctly, however, sometimes the students also appeared to respond correctly to some items without necessarily displaying the skill(s) in question (Alderson 1990b).

It should also be noted that, RD was used to tease out points from the TAPs and both the approaches gave similar results. RD was especially helpful to make decisions about the use of the five different strategies: GeneralUnd, DetailUnd, UndThemes, TextOrganisation and AppreciatingTx as it was very difficult to make decisions regarding their use just by listening to the TAPs of the students.

5.1.2 Research question 2

The second research question of the study was: are the reading strategies specified in the SLC English curriculum in line with what experts think the test is testing? The study reveals a considerable lack of consistency among the majority of expert judges on the strategies measured by the test. The high degree of variability among the expert judges found in the study were, therefore, in line with what previous studies have claimed

i.e. experts are unable to predict which reading strategies are tested by individual test items (Alderson, 1990a; Alderson and Lukmani, 1989).

The findings of the study are against Lumley (1993). However, unlike the present study which focused on a standardized high-stakes reading test, Lumley, initiated his project with an English for Academic Purpose (EAP) test designed principally to provide diagnostic information, and then developed the reading skills to describe them. Lumley reported very substantial agreement among the expert judges on the skills measured by the individual test items.

According to the expert judgements, ten of the reading strategies specified in the SLC curriculum seem to be tested by the test. However, the expert judgements had considerable agreement only on five strategies. Additionally, the test was judged to test three more strategies, which are not specified in the SLC curriculum, despite having very low level of agreement among the experts. Conversely, the test was judged not to be testing TextContinuation and DiagrammaticInfo that are specified in the SLC curriculum. In the FGD, they expressed the obvious reason for not testing DiagrammaticInfo that the test does not contain any diagrammatic information. With regard to TextContinuation, most of them expressed their worry for not including an interrupted text.

Another interesting finding is: among the plural strategies judged to be tested by the same item, the experts seem to agree with only one or two of the strategies. For instance, Item 27 (short answer question) is judged to be measuring six different strategies but they have noteworthy agreement only on E-SI. However, all the experts argued that it is not an easy task to say exactly what strategy is measured by a particular item as the reading strategies do not have water tight definitions; there is always some amount of overlapping.

Finally, this study indicates that the experts could not agree as to whether the individual items were eliciting expeditious or careful reading strategies as each of the items, except the items 11, 20 and 24, was judged to be testing both the expeditious and careful reading strategies.

5.1.3 Research question 3

The third research question of the study was: To what extent is there consistency between the strategies test takers employ and the expert judgements? The results reveal that the majority of the expert judges and the test-takers did not show consistency on the skills to be assessed by the test. In other words, there was a big gap between the strategy use and strategy judgements. Thus, the results found in the study confirm the findings of Tavakoli and Barati (2011).

Another important finding concerns the identification of reading strategies not specified in the SLC curriculum. As mentioned in 4.3, both the verbal reports and expert judgements data indicate the test to be measuring some other strategies which are not specified in the SLC curriculum. However, those strategies are mutually exclusive, i.e. those found in verbal reports did not occur in expert judgements data and vice versa. In other words, there was complete inconsistency between the strategy use and strategy judgements in terms of those strategies.

Item-wise analysis reveals a big gap between the strategies employed by the students and the expert judgements in each item. In all the item-types, except matching item, it can be seen that among the several strategies assigned to the items, there is agreement between the strategy use and strategy judgements only on one or two of the strategies. The most extreme item-type was the matching item which indicated a complete inconsistency between the strategy use and strategy judgements. The verbal reports of

the data reveals that all the students applied GeneralUnd and two of them also applied Wmatching to respond to the item while the item was judged to be testing Lexicalltems and DictionaryUse by almost all the experts. In the FGD, six (out of seven)experts mentioned that the test seems potential to test Lexicalltems as the students may not know the meaning of the words and they may see the context to find the meaning. However, very interestingly, seven (out of eight) students through their RDs disclosed that all the words were very common to them. Similarly, with regard to DictionaryUse, five of the experts viewed that the item seems to be potential to measure this strategy. However, following the rules of the SLC, the students were not allowed to use any type of dictionary while taking the test. Therefore, there might have been complete inconsistency between the strategies used by the students and the expert judgements in terms of these two strategies.

5.2 Construct validity of the test with regard to the reading strategies

The main aim of the research is to investigate the construct validity of the test. The results of this study, contrary to the findings of some previous studies on test validation (Cohen and Upton 2006; Nikolov 2006; Salehi, 2012), reveal that the test seems to lack construct validity as it does not seem to cover all the reading strategies specified in the SLC curriculum.

The results indicate three important threats to construct validity of the test. The biggest threat to the validity of the test is: the test construct is under-represented as the test does not seem to measuring all the reading strategies specified in the SLC curriculum. The verbal reports of the students do not show any evidence for the use of DiagrammaticInfo, TextContinuation and DictionaryUse. The findings related to DiagrammaticInfo and TextContinuation are also supported by the expert judgements. It

is obvious that we encounter a lot of diagrammatic information in our day to day life situation. Therefore, it seems important for the students to develop the skills to understand diagrammatically presented information. Also, good readers are able to make predictions of what comes next in the text (Alderson, 2000). Therefore, TextContinuation and DiagrammaticInfo should be tested by an exam such as the SLC exam. Future versions of the test could focus directly on these two strategies.

The second area of possible improvement for the validity of the test could be in relation to the testing of AppreciatingTx, TextOrganisation and LexicalItem. Out of the 639 occurrences, AppreciatingTx and TextOrganisation were observed only three and six times respectively in the verbal reports. However, these two strategies are given due emphasis in the SLC curriculum. Similarly, the verbal reports indicate very few examples of the application of LexicalItem. The expert judgements data also indicate somehow similar picture. But, understanding and acquiring academic vocabulary is an important skill for students (Coxhead and Nation, 1998). Therefore, it seems important to test whether the students are able to infer the meaning of unfamiliar words from the context in an exam such as the SLC exam. Although verbal reports indicated the participants' struggle to understand some vocabulary items in the test, this was often not an issue as most of the items did not target this strategy. Therefore, future versions of the test could focus directly on these strategies.

The third threat related to the validity of the test concerns the findings of the strategies which are not specified in the SLC curriculum. This finding indicates that the test seems to measure something that is irrelevant to the test construct (Hughes, 2003). The test developers should be aware of the overrepresentation of the test construct as a valid test measures all and only the skills that are supposed to test (Hughes, 2003).

This study has opened further areas of investigation into test validation. Contrary to Phakiti's (2003) findings that the use of reading strategies in a test situation changes across language tasks, the students in the study seem to follow similar strategies to approach different tasks. Therefore, the construct validation inquiry process is to observe more tests consisting of more tasks so that the consistency of test-takers' use of the reading strategies across different tasks/test papers can be observed, and can find out whether the findings of this study will be the same or different.

5.3 Methodological considerations

A very large database was collected with the help of TAPs and RDs (an average of an hour of audio per student). All the information was analyzed rather than being selective in terms of test items as Upton and Cohen (2006) did. Enormous effort was needed for the transcription and exploratory analysis as both the tasks were extremely time-consuming. However, the data have provided new insights into what the eight students did while taking the test.

There might be, however, several methodological concerns for the interpretation of the results. Because of the intensive nature of the study, no more than a handful of participants were involved. Therefore, one needs to be tentative when extrapolating the findings of the study. Similarly, the data was coded only by the researcher, coding would be more robust when coded by more than one person (Mackey and Gass, 2005). Furthermore, although enormous effort was made to code the data minutely and consistently, it is still possible that there were some inconsistencies while coding the data (Cohen and Upton, 2006).

Similarly, the linguistic level of the students might have also played important role in the use of the reading strategies. As mentioned in 3.2.1, the students had just started Grade

10 at the time when the data were collected. This means, they still have one complete year of study before they appear in the SLC exam. Additionally, it is worthy to mention that all of the students used English to verbalise their thoughts. However, Nepali was mostly used in their interview as most of them preferred Nepali to use in the interview. One of the possible reasons for their preference might be: while thinking-aloud they were simply reading the text and talking to themselves but during the RD session, they had to communicate with the researcher, a native speaker of Nepali. Therefore, they might not have felt confident enough to use English during the interview. However, there was no effect on the data as both the TAPs and RDs gave the same results.

Another more general methodological issue is the difference between reading in a test situation and reading in a non-test situation. Alderson (1990b) claims that “reading and taking a test are not the same thing” (p. 468). In the study by Cordon and Day (1996), the participants were found using more reading strategies in a test situation than in a non-test situation. Having being concerned with the issue, the students in the study were told to treat the test like the real test and most of the SLC exam rules were followed while taking the test; only the major difference was: the students were asked to think aloud in this study. All the students took the tests seriously. Therefore, as Nikolov (2006) claims, it might be sensible to argue that the verbal reports were collected under test conditions and, thus, the use of reading strategies were similar to what the SLC test-takers would use without verbalizing their thoughts. However, the levels of anxiety would be much higher in a real exam situation. This could have had an impact on the reading strategies used. The strategy use might have also been affected by the presence of the researcher. Furthermore, the fact that the students were verbalizing as they worked through the items might have had some influence on how they responded to the task; it is impossible

to eliminate the reactive effects of verbal report on task performance (Upton and Cohen, 2006).

Another methodological issue concerns the technique used for collecting RDs. Following the notes of caution sounded by Cordon and Day (1996) that “the process of immediate retrospection may interfere with the ability under investigation” (p.288), the participants were interviewed only after they completed the whole test. It should be noted that the test contains 29 test items and the interview was taken starting from the first item. Consequently, there was a time gap between the performance and the interview on each test item. Additionally, as Yang (2006) did, the interview might have been more effective if the recordings were played (stopping after each item) and the students were asked to explain how they comprehended the text and what strategies they adopted to respond to each of the items. It should also be noted that the students undoubtedly applied other strategies as well that were not described in the verbal reports.

A further concern is that, following Alderson (1990b) a distinction was not made between the strategies used for test items that were answered correctly as opposed to those answered incorrectly. A closer look at this variable might provide more insights into the strategy use.

It is also worthy to mention that although previous research on reading claimed for the detrimental effects of thinking-aloud on reading (Cordon and Day, 1996), the majority of students through their RDs reported that thinking-aloud did not have any negative impact on their understanding. Only two of the eight students had negative comments about the effects of thinking-aloud on understanding the text. Thus, the results might support the claim made by Cohen and Upton (2006) that the reading strategy use would not be affected by the think-aloud process.

Similarly, there are also some methodological issues with regard to the expert judgments. Firstly, the experts in the FGD were not asked to discuss and make an agreement on the items which indicated discrepancies. Had it been done so, the level of agreement might have been improved. The experts in the FGD were simply asked to provide their opinion for their choices. Secondly, as shown by Lumley's (1993) study that the level of agreement might have been increased if the experts were trained on the task and were asked to discuss their findings. However, the experts were not trained as it was thought "any agreement among cloned raters would simply indicate the success of the cloning process" (Alderson, 2000: 96). The third issue concerns the expertise of the experts. As mentioned in 3.2.2, the experts did differ in terms of their working experience; four of them had about 20 years of experience while the rest had just about four years of experience in the field of test designing. Therefore, there might have been low level of agreement among the experts themselves. Additionally, although they were trained in designing language tests, they were not specially trained on reading assessment.

Another possible limitation of the expert judgements might be the time gap between the task completion and the FGD. The experts, through the FGD suggested that it would have been much better if all of them were brought together for a whole day session so that they would be able to participate in the FGD immediately after they completed the tasks which would enable them to be more critical and tell more clearly about their judgement process.

Another limitation concerns the experts' involvement in the test design. None of them were involved in designing the reading test, being investigated. Had they been involved in the task, they would have provided more insights into the test purpose. Also, it might

have been more interesting to see the gap between the intention of the test designers and the real strategies employed by the students.

Finally, another limitation of this study concerns the word limitation for the dissertation. Because of the word limit problem, item-wise analysis (for those 29 items) was not possible. Had it been possible to discuss the reading strategies tested by each of the items, as Alderson (1990b) did, the study would reveal more information on the validity issues of the test. Future research might address this limitation.

Chapter 6: Conclusion

The primary aim of this study was to determine whether the English reading test of the SLC exam is actually measuring what it claims to measure. The results of the study might degrade the construct validity of the test as it is found that the test could not cover all the reading strategies specified in the SLC curriculum, thus under-representing the test construct. The identification of five different reading strategies not specified in the SLC curriculum indicates a possible construct overlap to another part of the SLC exam. Thus, the test lacks construct validity as a valid test measures all and only the reading strategies it aims to test (Hughes, 2003). However, the test as a whole requires the students to use reading strategies to gain both a local and global understanding of the test passages as well as to use both the careful and expeditious reading strategies. But, it should be noted that, as pointed by Yang (2006), it was very difficult to compartmentalize all the activities as there was some overlapping between the reading strategies.

The study also reveals important findings regarding the usefulness of the two research methodologies. It confirms the usefulness of verbal reports to get insights into the cognitive processes in readers' mind and it advocates for the complementary use of TAPs and RDs. In many cases, it was very hard to identify which reading strategy had actually been employed, using TAPs alone. In such cases, decisions were made with further analysis of RDs.

However, the study casts doubt on the usefulness of expert judgements to investigate the reading strategies tested by the test. There were two main problems associated with the expert judgements. First, the experts did not agree to a sufficient degree on the reading strategies tested by the individual items or item-types. Second, there was a big gap between the strategy use and strategy judgements.

It is worth noting that the contradictory results from the comparison between the expert judgements and strategy use indicate the need for further research concerning the methodology. Future investigations could expand on the use of two research methods and also include a third method to triangulate the findings. Additionally, future research might also need to conduct content analysis and student performance analysis in order to shed more light into the issues.

The results of this study should be treated with caution and future investigations should try to overcome the limitations of this study. For instance, the study could be extended to a greater number of participants so that the results could be generalised. Secondly, the data could be coded by more than one researcher. Furthermore, expert judgement investigation with specifically the experts who are really involved in the test item writing might provide more insight into the issues. Similarly, training to the experts might also be desirable and conduction of the FGD immediately after the completion of the task might yield better results.

Lessons from the study are important for the improvement of the test. The study indicates a considerable lack of alignment between the test item and the reading strategies. Therefore, the test designers should make efforts to address the issues. Unfortunately for the test developer, it is apparent that the relationship between item type and response strategy may not be straightforward as item-type is not a good predictor of reading strategy use (Weir et al., 2006). More importantly, an effort should be made to cover all the strategies mentioned in the SLC curriculum.

In the case of some strategies, there seems to be a kind of contradiction between the strategies specified in the SLC curriculum and the examination practice. For instance, the curriculum clearly highlights the importance of learning and testing the strategy

'DictionaryUse'. However, the students are not allowed to see any kind of dictionary in the examination. In order to solve the problem, either it is necessary to clearly mention in the curriculum that the strategy should not be tested in the SLC exam or the students should be allowed to use a dictionary in the exam.

The strategy use also has been affected by the design of the test. For instance, as the same text is followed by several questions, the students have a lot exposure to the test by the time they reach to the final items. As a result of which their strategy use might have been affected. Additionally, there are many problems associated with the test itself like grammatical errors and repetitions of the same item-types.

Similarly, in some cases the questions are designed in such a way that the students can get right answers just by guessing. For instance, while responding to Item 20 (Ordering Sentences), five of the students could not understand the text, but still managed to put the sentences in the correct order. Similarly, in order to respond to the true/false items, the students were making just wild guesses for several times.

To conclude, it is hoped that this study has successfully illuminated some of the darker corners of the students' mental processing when they were engaged in reading and responding to the test items. It is also hoped that it has not only contributed to validating a reading test, but has also offered some possible ways to look at some theoretical and methodological perspectives for assessing reading strategies.

References

- Afflerbach, P., & Johnson, P. (1984) 'On the Use of Verbal Reports in Reading Research', *Journal of Reading Behaviour*, 16, pp. 307-322.
- Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008) Skills and Strategies: Their Differences, Their Relationships and Why It Matters, in Mokhtari K. & Sheorey, R. (Eds.), *Reading Strategies of First- and Second Language Learners - See How They read* Norwood, MA, Christopher-Gordon Publishers, Inc, pp. 11-24.
- Alderson, J. C. (1990a) 'Testing Reading Comprehension Skills (Part 1)', *Reading in a Foreign Language*, 6 (2), pp. 425-438.
- Alderson, J. C. (1990b) 'Testing Reading Comprehension Skills (part 2)', *Reading in a Foreign Language*, 7 (1), pp. 465-503.
- Alderson, J. C. (1993) Judgements in Language Testing, in Douglas, D. & Chapelle, C. (Eds.), *A new Decade of Language Testing Research*, Alexandria, VA, TESOL, pp. 46-57.
- Alderson, J. C. (2000) *Assessing Reading*, Cambridge, Cambridge University Press.
- Alderson, J. C., Clapham, C., & Wall, D. (1995) *Language Test Construction and Evaluation*. Cambridge, Cambridge University Press.
- Alderson, J. C., & Lukmani, Y. (1989) 'Cognition and Reading: Cognitive Levels as Embodied in Test Questions', *Reading in a Foreign Language*, 5(2), pp. 253-270.
- Alderson, J. C., & Lumley, T. (1995) 'Responses and Replies', *Language Testing*, 12 (1), pp. 121-130.
- Anderson, N. J., Bachman, L. F., Perkins, K., & Cohen, A. (1991) 'An Exploratory Study into the Construct Validity of a Reading Comprehension Test: Triangulation of Data Sources'. *Language Testing*, 8 (1), pp. 41-66.

- Bachman, L. F. (1990) *Fundamental Considerations in Language Testing*, Oxford, Oxford University Press.
- Bejar, I. I. (1983) 'Subject Matter Experts' Assessment of Item Statistics', *Applied Psychological Measurement*, 7, pp. 303-310.
- British Educational Research Association (BERA, 2011) *BERA Ethical Guidelines for Educational Research*, London, BERA.
- Brown J. D. & Abeywickrama, P. (2010) *Language Assessment: Principles and Classroom Practices*, New York, Pearson Education.
- Carrell, P. L. (1992) 'Awareness of Text Structure: Effects on Recall', *Language Learning*, 42 (1), pp. 1-20.
- Carver, R. (1997) 'Reading for One Second, One Minute, or One Year from the Perspective of Raiding Theory', *Scientific Studies of Reading* 1 (1), pp. 3-43.
- Cohen, A. D., & Upton, T. A. (2006) 'Strategies in Responding to the New TOEFL Reading Tasks', *ETS Research Report Series, 2006* (1), pp. i-162.
- Cohen, A. D. & Upton, T. A. (2007) 'I Want to Go Back to the Text: Response Strategies on the Reading Subtest of the New TOEFL', *Language Testing*, 24 (2), pp. 209- 250.
- Cordon, L. A., & Day, J. D. (1996) 'Strategy Use on Standardized Reading Comprehension Tests', *Journal of Educational Psychology*, 88 (2), pp. 288-295.
- Coxhead, A., & Nation, P. (1998) 'The specialized Vocabulary of English for Academic Purposes', *Wellington Occasional Papers in Applied Linguistics*, 1 (1), pp. 31-55.
- Cronbach, L. J. (1988) Construct Validation after Thirty Years, in Robert L. L. (Ed.), *Intelligence, Measurement, Theory, and Public Policy*, Urbana, Ill, University of Illinois Press.

- Davis, F. B (1968) 'Research in Comprehension in Reading', *Reading Research Quarterly* 3 (4), pp. 499-545.
- Ericsson, K. A., & Simon, H. A. (1993) *Protocol Analysis*, Cambridge, MA, MIT Press.
- Farr, R., Pritchard, R., & Smitten, B. (1990) 'A Description of What Happens When an Examinee Takes a Multiple-choice Reading Comprehension Test', *Journal of Educational Measurement*, 27 (3), pp. 209-226.
- Fulcher, G. (1997) 'Text Difficulty and Reading Formulae: Reading Formulae and Expert Judgements', *System*, 25(4), pp.497-513.
- Grabe, W., & Stoller, F. L. (2002)*Teaching and Researching Reading*, Edinburgh, Pearson Education Limited.
- Green, A. (1998) *Verbal Protocol Analysis in Language Testing Research*, Cambridge: Cambridge University Press.
- Holzknicht F. (2012). *Establishing one part of a validity argument: Reading strategies in the standardized Austrian Matura exam*. An unpublished MA Thesis, Lancaster University, England.
- Hughes, A. (2003) *Testing for Language Teachers (2nd Ed.)*, Cambridge, Cambridge University Press.
- Kamil, M. & Pearson, P. D. (1979) 'Theory and Practice in Teaching Reading', *New York University Education Quarterly*, 10 (2), pp. 10-16.
- Katalayi, G.B. & Sivasubramaniam, S. (2014) 'A Context-Based Perspective to Investigating the Construct Validity of Reading Assessments', *Mediterranean Journal of Social Sciences*, 5 (27), pp. 847-860.
- Khalifa, H., & Weir, C. J. (2009) *Examining Reading - Research and Practice in Assessing Second Language Reading*, Cambridge, Cambridge University Press.

- Khaniya, T.R. (2010) *Examination for Enhanced Learning*, Kathmandu, Students Publication.
- Lewins, A. & Silver, C. (2014) *Using Software in Qualitative Research: A Step-by- Step Guide*, London, Sage Publications Ltd.
- Lumley, T. (1993) 'Reading Comprehension Sub-skills: Teachers' Perceptions of Content in an EAP Test', *Melbourne Papers in Language Testing*, 2 (1), pp. 25-60.
- Mackey, A., & Gass, S. M. (2005) *Second Language Research: Methodology and design*, Mahwah, NJ, Lawrence Erlbaum.
- McNamara, T. (2009) 'Validity in Language Testing: The Challenge of Sam Messick's Legacy', *Language Assessment Quarterly*, 3 (1), pp. 31-51.
- Messick, S. (1988) The Once and Future Issues of Validity: Assessing the Meaning and Consequences of Measurement, in Wainer, H. & Braun, H. (Eds.), *Test Validity*, Hillsdale, NJ, Erlbaum, pp. 3-45.
- Messick, S. A. (1992) Validity of Test Interpretation and Use, *Encyclopaedia of Educational Research (6th Ed.)*, New York, Macmillan.
- Ministry of Education (2015) *Press Conference*, Kathmandu, Office of the Controller Examination.
- Moore, T., Morton, J. & Price, S. (2012) Construct Validity in the IELTS Academic Reading Test: A Comparison of Reading Requirements in IELTS test Items and in University, in Taylor, L. and Weir, C.J. (Eds.), *IELTS Collected Papers 2: Research in Reading and Listening Assessment*, Cambridge, Cambridge University Press, PP. 120-211.

- Nikolov, M. (2006) 'Test-Taking Strategies of 12- and 13- Year- Old Hungarian Learners of EFL: Why Whales Have Migraines' *Language Learning*, 56 (1), pp. 1-51.
- Olshavsky, J. E. (1977) 'Reading as Problem Solving: An Investigation of Strategies', *Reading research Quarterly*, 12 (4), pp. 654- 674.
- Phakiti, A. (2003) 'A Closer Look at the Relationship of Cognitive and Metacognitive Strategy Use to EFL Reading Achievement Test Performance', *Language Testing*, 20 (1), pp. 26-56.
- Rosenshine, B.V. (1990), Skills Hierarchies in Reading Comprehension, in Spirio, R.J., Bruce, B. C. & Brewer, W. F. (Eds.), *Theoretical Issues in Reading Comprehension*, Hillsdale, NJ, Lawrence Erlbaum, pp. 535- 59.
- Salehi, M. (2011) 'On the Construct Validity of the Reading Section of the University of Tehran English Proficiency Test', *Journal of English Language Teaching and Learning*, 2 (222), PP. 129-159.
- Sash, P. K. (2063). *Dilemmas and challenges of teaching English in Nepal*. Sanothimi, Nepal: NCED.
- Schedl, M., Gordon, A., Carey, P. A. & Tang, K. L. (1996) 'An Analysis of the Dimensionality of TOEFL Reading Comprehension Items', *TOEFL Research Report 53*, Princeton, NJ, Educational Testing Service.
- Shiotsu, T. (2010) *Components of L2 Reading: Linguistic and Processing Factors in the Reading Test Performances of Japanese EFL learners*, Cambridge, Cambridge University Press.
- Shrestha, P. (2003) *How Parallel Are Parallel Tests? A Study of the Two Versions of the School Leaving Certificate (SLC) English Examination (2002) in Nepal*, An Unpublished MA Thesis, Lancaster University, England.

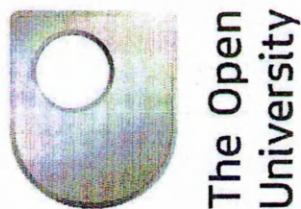
- Song, M.Y. (2008) 'Do Divisible Sub-skills Exist in Second Language (L2) Comprehension? A Structural Equation Modelling Approach', *Language Testing*, 25 (4), pp. 435-464.
- Tavakoli, E., & Barati, H. (2011) 'Investigating the Construct Validity of the FCE Reading Paper in Iranian EFL context', *Journal of Language Teaching and Research*, 2 (1), PP. 239-247.
- Taylor, K. L. & Dionne, J. (2000) 'Assessing Problem-Solving Strategy Knowledge: The Complementary Use of Concurrent Verbal Protocols and Retrospective Debriefing', *Journal of Educational Psychology*, 92 (3), pp. 413- 425.
- Urquhart, A. H., & Weir, C. J. (1998) *Reading in a Second Language: Process, Product and Practice*, New York, Addison Wesley Longman.
- Weir, C. J. (2005) *Language Testing and Validation: An Evidence-Based Approach*, Basingstoke, Palgrave Macmillan.
- Weir, C., Hawkey, R., Green, A., & Devi, S. (2012) 'The Cognitive Processes Underlying the Academic Reading Construct as Measured by IELTS', *IELTS Research Reports 9*, pp. 157 -189.
- Weir, C. J., Huizhong, Y., & Yan, J. (2000) *An Empirical Investigation of the Componentiality of L2 Reading in English for Academic Purposes*, Cambridge, Cambridge University Press.
- Weir, C. J., Yang, H. and Jin, Y. (2000) *An Empirical Investigation of the Componentiality of L2 Reading in English for Academic Purposes*, *Studies in Language testing*, 12, Cambridge: Cambridge University Press.

Williams, E., & Moran, C. (1989) 'Reading in a Foreign Language at Intermediate and Advanced Level with Particular Reference to English', *Language Teaching*, 22 (4), pp. 217-228.

Wu, R.Y (2014) *Validating Second Language Reading Examinations: Establishing the Validity of the GEPT through Alignment with the Common European Framework of Reference*, Cambridge, Cambridge University Press.

Yang, Y.F. (2006) 'Reading Strategies or Comprehension Monitoring Strategies?' *Reading Psychology*, 27, pp. 313- 343.

Appendix A: Information sheet and parents' consent form



Faculty of Education and
Language Studies
The Open University
Walton Hall
Milton Keynes
United Kingdom
MK7 6A
Tel +44(0) 1908 274 066
www.open.ac.uk

20 April, 2015

आदरणीय अभिभावक ज्यू,

म सरस्वती दवाडी हाल वेलायतको Milton Keynes मा रहेको The Open University मा मास्टर Level मा अध्ययनरत रहेको र उक्त Level को लागि थेसिस लेख्दै रहेको कुरा जानकारी गराउन चाहन्छु। मेरो रूचीको विषय Language Testing रहेको र मैले अनुसन्धान गर्न लागेको विषय एस.एल.सी. परीक्षाको अंग्रेजी विषयको Reading Test को वैद्यता भएको कुरा पनि अवगत गराउन चाहन्छु।

यसै सन्दर्भमा मैले हजुरको बच्चासँग उक्त परीक्षाको विषयमा केही सूचना लिन चाहन्छु। यहाँलाई यो कुरा पनि निश्चित बनाउन चाहन्छु कि म यस अनुसन्धानमा अंग्रेजी विषयको Test को वैद्यता हेर्नको लागि मात्र इच्छुक छु न कि हजुरको बच्चाको Performance सँग। यदी हजुरले आफ्नो बच्चालाई यस अनुसन्धानमा सहभागी हुन दिनु भएको खण्डमा हजुरको बच्चाले निम्न कार्यहरू गर्नु पर्नेछ।

- अंग्रेजी विषयको Reading Test लिने र आफ्नो सोचाइहरू व्यक्त गर्ने।
- Reading Test सम्बन्धी अन्तर्वार्ता दिने।
- एउटा प्रश्नावली भर्ने।

यहाँलाई यो कुरा पनि स्पष्ट गर्न चाहन्छु की यदी हजुरको बच्चाले मेरो अनुसन्धानमा भाग लिन नचाहेमा अथवा भाग लिइ सकेपछि पनि आफ्नो सहभागिता जनाउन नचाहेमा यहाँले सहजै निर्णय लिन सक्नुहुनेछ कुनै पनि प्रकारको करकाप गरिने छैन। मुख्य कुरा यति मात्रै हो की हजुरले वा हजुरको बच्चाले आफु सहभागी नहुने कुरा मलाई जानकारी गराउनु पर्नेछ र यदि हजुरको बच्चाले सहभागीता फिर्ता लिएमा, उनले दिएको जानकारी यस अनुसन्धानमा उल्लेख गर्ने छैन। यहाँ समक्ष यो कुरा पनि अवगत गराउन चाहन्छु कि मेरो अनुसन्धानको नतिजा हजुर लगायत अंग्रेजी शिक्षकहरू विद्यार्थीहरू र नीति नियम बनाउने निकायहरू लाई जानकारी गराउने छु।

यदि हजुरलाई मेरो अनुसन्धानको बारेमा केही जानकारी लिन इच्छा भएमा मलाई अथवा मेरो सुपरिवेक्षक डा. पृथ्वी श्रेष्ठ लाई (Prithivi.Shrestha@open.ac.uk) सम्पर्क गर्न सक्नु हुनेछ।

निवेदीका

सरस्वती दवाडी

The Open University

Milton Keynes, England

अभिभावक अनुमति पत्र

वेलायतमा रहेको The Open University अन्तर्गत मास्टर Level को एस.एल.सी परीक्षाको अंग्रेजी विषयको वैद्यता सम्बन्धी अनुसन्धानको निम्ती

अभिभावकको नाम :

अनुसन्धानकर्ताको नाम :

१. म मेरो बच्चालाई यस अनुसन्धानमा सहभागी गराउन अनुमति प्रदान गर्दछु । यस अनुसन्धानको बारेमा सम्पूर्ण जानकारी मैले सरल भाषामा प्राप्त गरेको छु र बुझेको पनि छु ।
२. मैले बुझेको छु की मेरो बच्चाले अंग्रेजी विषयको पढाई को Test लिनेछ र उसको विचार व्यक्त गर्नेछ अनि अन्तर्वार्ता दिनको साथै एउटा प्रश्नावली भर्नेछ ।
३. म यो अवगत गराउन चाहन्छु की
 - अ. मेरो बच्चाको विचार र अन्तर्वार्ताको Audio रेकर्ड गरिनेछ ।
 - आ. मेरो बच्चाले सहभागिता जनाउन नचाहेमा उसले कुनै पनि समयमा फिर्ता लिन सक्नेछ र उसले उपलब्ध गराएको जानकारी यस अनुसन्धानका लागि प्रयोग गरिने छैन ।
 - इ. यो Project अनुसन्धानको उद्देश्यले मात्र गरिएको छ ।

- ई. मलाई यो कुरा पनि थाहा छ कि मेरो बच्चाको नाम र उसले दिएको जानकारी गोप्य रहने छ र सबै कुरा नियम बमोजिम नै रहनेछन् ।
- उ. मलाई यो कुरा पनि जानकारी छ की मेरो बच्चाले दिएको जानकारी पाँच वर्ष सम्म बेलायतको Milton Keynes मा अवस्थित The Open University मा सुरक्षित रहनेछ ।
- ऊ. यस अनुसन्धानसँग सम्बन्धित कुनै पनि Publication मा मेरो मेरो बच्चाको नाम गोप्य रहनेछ ।
- ऋ. यस अनुसन्धानको सारांशको एक प्रति मलाई पनि पठाइने छ ।
- ए. मैले यस अनुमति पत्रको एक प्रति प्राप्त गर्नेछु ।

म यस अनुसन्धानको सारांश लिन चाहन्छु ।

हो

हैन

कृपया रेजा ($\sqrt{\quad}$) रगाउनुहोला

अभिभावकको हस्ताक्षर

मिति :

अनुसन्धान कर्ताको हस्ताक्षर

मिति :

साक्षीको हस्ताक्षर

मिति :

बच्चाको अनुमति उपलब्ध गराएको

उमेरको हिसाबले सानो अथवा अपरिपक्व भएकोले उसको अनुमति मैले प्रदान गराएको छु ।

अभिभावकको हस्ताक्षर

मिति :

Appendix B: Information sheet and experts' consent form



**Faculty of Education
and Language Studies**

The Open University

Walton Hall

Milton Keynes

United Kingdom

MK76AA

Tel +44 (0) 1908 274

066

www.open.ac.uk

Date: 18 January, 2015

Information Sheet

Dear Sir

I would like to inform you that I am a student at the Open University, England. I am pursuing my Masters in Research Studies at the university. I have to undertake a short research study for the partial fulfilment for the master's degree. Therefore, I am carrying out a research related to English language assessment in the School Level Certificate Exam (SLC) in Nepal.

Language assessment is an area of interest for me. The research study that I am undertaking involves investigating the construct validity of the reading test of the

SLCexam. Construct validity of a test refers to the degree to which the test measures the skills it proposes to measure.

If you participate in my research, you will need to contribute about 2 hours of your time for this study. The following will be the procedures for data collection:

- First you will be asked to make a judgement regarding the reading skill(s) each test item of the test measures. For this you need to fill out a form.
- After you make your judgement, you will be asked to participate in a Focus Group Discussion which might last for about an hour.

You are free to withdraw from the study at any time before 8 May. All you need to do is to tell me that you no longer wish to participate or that I can no longer use your data, and I will not use. At every stage your identity will be kept confidential. The data will be used for the research purpose only. The report of the research study will be shared with you, the students, teachers and the SLC Board.

If you have any query about this study, please feel free to contact myself or my supervisor who can be contacted on Prithvi.Shrestha @open.ac.uk.

Thank you for your cooperation!

Regards

SaraswatiDawadi

Open University

Milton Keynes, England

Subject Expert Consent Form

For

**A Research project under the Masters in Research Program, The Open University,
England- Reading strategies in the Nepalese School Leaving Exam: A Qualitative Study
for Establishing Construct Validity**

Name of participant:

Name of researcher:

1. I consent to participate in this project, the details of which have been explained to me, and I have been provided with a written statement in plain language to keep.
2. I understand that my participation will involve filling out the form and participating in focus group discussion and I agree that the researcher may use the results as described in the plain language statement.
3. I acknowledge that:
 - (a) I have been informed that I am free to withdraw from the project at any time without explanation or prejudice and to withdraw any unprocessed data I have provided;
 - (b) The project is for the purpose of research;
 - (c) I have been informed that the confidentiality of the information I provide will be safeguarded subject to any legal requirements;

- (d) I have been informed that with my consent the data generated will be stored at the Open University, England and will be destroyed after five years;
- (e) If necessary any data from me will be referred to by a pseudonym in any publications arising from the research;
- (f) I have been informed that a summary copy of the research findings will be forwarded to me, should I request this.
- (g) I understand that I have to write in the Expert Judgment Instrument and fill the form; and my participation at the Focus Group Discussion will be audio-tape

I wish to receive a copy of the summary project report on research findings

Yes No (please tick)

Participant's signature:

Date:

Researcher's signature

Date:

Appendix C: Questionnaire for students

Please answer the questions below which are related to the test you have just completed including a few personal questions. We need this information to be able to interpret your answers. Please tick only one box where boxes appear.

1. What is the name of your school?

Gandaki Boarding School Gyanodaya Higher Secondary School

2. Your Gender: Male Female

3. Your age (in years): _____

4. Your mother tongue: _____

5. How old were you when you started learning English? (in years) _____

6. How many years have you been learning English at school?

9 10 11 12 more

7. Is English the medium of instruction in your school?

Yes No

8. How many English classes do you have each week?

5-6 6-7 7-8 more

9. Do you read any other English text except your English textbook? If yes, what type of text like story, poem, news etc.?

10. How interesting did you find each of the text?

	very much	a bit	much	not at all
a. Poem (Rabindranath Tagor)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Narrative text about driving	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Explanatory text about family	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Advertisement for job	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

11. How easy did you find each of the text?

	very much	a bit	much	not at all
a. Poem (Rabindranath Tagor)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Narrative text about driving	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Explanatory text about family	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Advertisement for job	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

12. How familiar were you with the kinds of tasks used in the test?

	very much	a bit	much	not at all
a. True false	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Gap filling	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Matching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Finding words with similar meaning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Ordering sentences	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. Short answer questions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

13. How difficult did you find each of the tasks used in the test?

	very much	a bit	much	not at all
a. True false	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Gap filling	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Matching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Finding words with similar meaning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Ordering sentences	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. Short answer questions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

14. Please rank the following texts in order of difficulty. Write one number (1 to 4) in each of the boxes below (1 for the most difficult and 4 for the easiest one).

- a. Poem (Rabindranath Tagor)
- b. Narrative text about driving
- c. Explanatory text about family
- d. Advertisement for job

15. Please rank the following tasks in order of difficulty. Write one number (1 to 6) in each of the boxes (1 for the most difficult and 6 for the easiest one).

- e. True false
- f. Gap filling
- g. Matching
- h. Finding word meaning
- i. Ordering sentences

j. Short answer questions

16. How easy was it to understand the instructions given in each of the task?

	very much	a bit	much	not at all
a. True false	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Gap filling	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Matching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Finding words with similar meaning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Ordering sentences	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. Short answer questions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

17. Do you have any suggestions for the improvement of the test?

THANK YOU VERY MUCH FOR YOUR HELP!

Appendix D: Expert judgement instrument

Dear sir/madam

You are going to judge which types of reading strategies are elicited by each item of the SLC reading test (2070/071). Please follow the procedure outlined below.

1. Complete task 1.
2. Check your answers against the answer key (The answer key is attached herewith).
3. Go back to the first item of the task and **judge which type of reading strategy you think is elicited**. Tick the appropriate box by clicking it. You can tick **more than one box for each item**. If you tick on the 'other', please **mention which strategy you think that item is measuring**.
4. Repeat step 4 for the remaining items of task 1. Make sure to **judge each item individually**.
5. Repeat steps 2 to 5 for tasks 2, 3 and 4.
6. Please fill out the questionnaire at the end of this document.
7. Save the file with the name Expert Judgement and your first name and return to Saraswati.Dawadi@open.ac.uk.

Reading Strategies in the SLC reading test

		Item													
Task 1	A	a	Reading for detailed understanding												
		b	Reading for general understanding												
		c	Reading for understanding the underlying themes and ideas of texts												
		d	Reading for understanding of an argument												
	B	a	Reading for skimming												
		b	Reading for scanning												
	Task 2	A	a	Reading for identifying the organization of paragraphs and longer texts											
			b	Reading for anticipating the likely continuation of the interrupted text											
c			Reading for interpreting the information presented in diagrammatic forms												
d			Deduce the meaning of unfamiliar lexical items												
B			Reading for appreciating literary texts												
C		a	Using authentic English dictionary/thesaurus												
		b	Other												

		c													
		d													
Task 3	A	a													
		b													
		c													
		d													
	B														
	C	a													
		b													
		c													
	Task 4	A													
		B	a												
b															
c															
C		a													
		b													
		c													
		d													

1. Were there any reading strategies you found difficult to apply?

Yes No

2. If yes, which strategies? Please give reasons why you found it difficult to apply the strategies?

3. Were there any items you could not judge in terms of reading strategies?

Yes No

4. If yes, which items? Please give reasons why you were not able to judge the items.

5. Do you have any additional comments with regards to the application of the reading strategies to the SLC reading test items?

6. Do you have any comments for the improvement of the test with regard to the inclusion of the reading strategies listed above?

Thank you for your participation!

Appendix E: Test Paper

1. Read the poem and do the activities that follow: [5]

Where the Mind is Without Fear

Where the mind is without fear and the head is held high;

Where knowledge is free;

Where the world has not been broken up into fragments by narrow domestic walls;

Where words come out from the depth of truth;

Where tireless striving stretches its arms towards perfection;

Where the clear stream of reason has not lost its way into the dreary desert sand of dead habit;

Where the mind is held forward by thee into ever-widening thought and action....

Into that heaven of freedom, my Father

Let my country awake ——— *Rabindranath Tagor*

A. Write 'True' for true statement and 'False' for false ones: [4 X 0.5 = 2]

- Dignity is high when there is no fear in the mind.
- Domestic walls have fragmented the world.
- The poet prays to Father for his individual freedom.
- A stream can have a good effect in the desert.

B. Fill in the blanks with appropriate words from the poem: [3 X 1 = 3]

- Continuous effort leads us towards...
- Fearless situation makes people can get ... of all kinds.

2. Read the passage and do the activities that follow: [10]

The vehicles were being driven slowly because of the downpour. The visibility was poor and the wind was howling. There had been landslides in many places and driving was dangerous. Earlier the wind had been blowing forcefully but, by the time we started off, it had calmed down. The downpour had turned into a drizzle and brought thunder and lightning.

I'd been driving for an hour when the accident happened. My wipers hadn't been working, and the rain was spattering my windscreen, so I couldn't see well. I'd been stopping to clean my windscreen every few minutes. I had just started the engine again when my tyres started to slip. The truck slipped onto the side of the road, hit the hill, turned over and stopped. I felt and looked to see if I was hurt, but I wasn't. I had been driving quite slowly and luckily the bend was quite wide. It was very quiet, with just the sounds of music and falling rain; I'd been playing the cassette. I looked for my *Khalasi* but couldn't find him. Soon there was a long queue of vehicles and people were all round me asking questions about the accident. I heard them talking about two more accidents in which three people had died and ten others had been injured, Suddenly, someone shouted that there was a man lying beside the road. It was my *Khalasi*. He had been lying unconscious for half an hour. He was helped...

A. From the passage, find the words that are similar in meaning to the following: [4 X 1 = 4]

- a. Heavy shower b. State of being easy to see
c. Searched d. Senseless

B. Put the following sentences in the correct order: [4 X 0.5 = 2]

- a. The windscreen was often cleaned.
- b. The man was driving the truck slowly.
- c. The helper of the driver was hurt.
- d. The truck soon got in the road accident.

C. Answer the following questions: [4 X 1 = 4]

- a. Why was driving dangerous?
- b. What was wrong with the wipers?
- c. Where was the sound of music coming from?
- d. How many accidents happened altogether on that day?

3. Read the passage and do the activities that follow: [10]

When parents make a lot of rules about their children's behavior, they make trouble for themselves. I used to spend half my times making sure my rules are obeyed, and the other half answering questions like 'Jack can get up whenever he likes, So why can't I? or Why can't I play with Angela? Jack's mum doesn't mind who he plays with.' Or Jack can drink anything he likes. 'Why can't I drink wine too?' Jack's mum, I decided, was a wise woman. I started saying things like 'of course, dear. You can drink as much wine as you like'. And 'No, I don't mind how late you get up.' And 'Yes, dear, you can play with Angela as often as you like.' The results have been marvelous. They don't want to get up late any more, they've decided they don't want wine, and, most important, they've stopped playing with Angela. I've now realized (as Jack's mum realized a long time ago) that they wanted to do all these nasty things because they weren't allowed to.

A. Write 'True' for true statements and 'False' for the false ones: [4 X 0.5 = 2]

- a. The writer's parents made several rules.
- b. The writer thought Jack's mother wasn't so wise.
- c. Freedom from the parents made children turn positive.
- d. Children were curious to do strange things because they were not forbidden to do them.

B. Rewrite the sentences in the appropriate order: [4 X 0.5 = 2]

- a. Children do nasty things as they aren't allowed to.
- b. I used to spend half my times answering to kids' questions.
- c. Parents make trouble to kids making a lot of rules.
- d. Freedom to kids for doing things may have marvelous results.

C. Answer the following questions: [3 X 2 = 6]

- a. How do parents make trouble to themselves?
- b. Why should children be treated freely?
- c. What causes the bothering result?

4. Read the following advertisement and do the activities that follow: [15]

Announcement for SAARC Essay Competition

Theme: "SAARC – a regional hope"

(First published dated: 8 December 2013)

On the auspicious occasion of Twenty – Eight Anniversary of SAARC, the SAARC Information Center(SIC) invites to all interested students for Essay Competition.

The details/terms & conditions of Essay competition are:

1. The Essay competitions will be organized among the students up to class/Grade 12, the national of member states studying in Nepal.
2. Essay will be in English Language containing between 2000 – 2500 Words.
3. The Cash Prize of Nrs. 30000.00(Thirty Thousand), NRs 20000.00 (Twenty Thousand), Nrs 10000.00 (Ten Thousand) for first, second, third prize respectively and Nrs. 5000.00 (Five Thousand) each for Eight Consolation Prize along with Certificate provided by SIC.
4. For detail information, please visit SIC Website: www.saarc.sci.org

The Center requests interested students studying in Nepal public/private schools to submit their Essay directly or as following address latest by 31 December 2013

SAARC Information Centre(SIC)

Media Village, Tilganga

E-mail:info@saarc_sci.org

P.O. Box No. 26339, Kathmandu, Nepal.

A. Match the words in column 'A' with their meanings in column 'B' [4 X 1 = 4]

Column 'A'

Column 'B'

- | | |
|----------------------|--|
| a. Anniversary (i) | i. special event or happening |
| b. Organized (ii) | ii. Event in which people compete |
| c. Competition (iii) | iii. Celebration of the yearly return date/event |
| d. Occasion (iv) | iv. Arranged for |

B. Complete the following sentences with the correct words/phrases from the text: [3X1=3]

This is an ... for essay competition, announced on the ... of 28th Anniversary of the SAARC.

The students up to grade 12 can take part in the competition. They can submit their essay latest by ...

C. Answer the following questions: [4 X 2 = 8]

- What's the word limit of the essay?
- How can competitors submit their essays?
- Who are eligible for the essay competition?
- What's the title of the essay?