Open Research Online

Construct validity of the Nepalese school leaving english reading test

Journal Item

For guidance on citations see FAQs.

oro.open.ac.uk
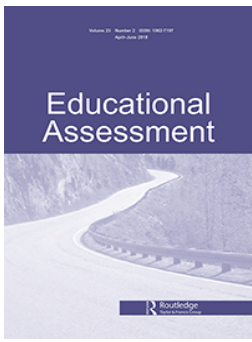
Construct validity of the Nepalese school leaving english reading test

Saraswati Dawadi & Prithvi N Shrestha

Routledge
Taylor & Francis Group

Check for updates

Construct validity of the Nepalese school leaving english reading test

Saraswati Dawadi and Prithvi N Shrestha

The Open University

ABSTRACT
There has been a steady interest in investigating the validity of language tests in the last decades. Despite numerous studies on construct validity in language testing, there are not many studies examining the construct validity of a reading test. This paper reports on a study that explored the construct validity of the English reading test in the Nepalese school leaving examination. Eight students were asked to take the test and think-aloud, followed by retrospective interviews. Additionally, seven experts were asked to make judgments regarding the skills tested by the test. The findings provide grounded insights into students' response behaviors prompted by the reading tasks, and indicate some threats to the construct validity of the test. Additionally, the study reports a low level of agreement among the experts, and a big gap between the skills used by the students and the skills that the experts thought were being examined by the test.

Validation is an important enterprise for any kind of high-stakes test. Messick (1988) points out that if the validity of a high-stakes test is not known, it might have undesirable consequences for the society at large. A high-stakes test without validation research can be "like a police force without a court system, unfair and dangerous" (McNamara, 2007, p.280).

The research reported in this paper investigated the construct validity of the school leaving certificate (SLC) examination in Nepal. The SLC examination serves several functions such as being a measure of students' language skills, a gateway to higher education, and a basic license for official employment. Additionally, the SLC results are also a measure of "what strengths and weaknesses exist in the education system at a given point in time, and how the education system is performing over the years" (Mathema & Bista, 2006, p.4). Therefore, a number of studies have been conducted to investigate the examination. For instance, Khaniya (1990) explored washback effects of the SLC English test; Shrestha (2003) compared the extent to which the two versions of the English test in the SLC examination are parallel; Bhatta (2005) investigated the determinants of student performance in the SLC examination; Mathema and Bista (2006) studied student performance in the SLC examination; Budhathoki et al. (2014) looked at the status of the SLC drop outs; and Bhattrai (2014) explored the negative consequences of the SLC examination. However, no research has been carried out to investigate the validity of the whole SLC examination in general, and the validity of its English reading test in particular. Therefore, the rationale for choosing to conduct this study starts with the need to address the validity issue of the English reading test of the SLC examination.

The study also considers 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 in general and test validation research in particular seems to be dominated by quantitative methods. But, as Song (2008) points out, quantitative approaches may be problematic as they do not capture how test takers respond to an item; a qualitative approach is necessary to reveal which skills test

CONTACT Saraswati Dawadi ✉ Saraswati.Dawadi@open.ac.uk 🌐 Faculty of Education and Language Studies, The Open University, Milton Keynes, United Kingdom.

takers really engage in when solving reading test problems. Considering this point, the study collected verbal reports from test takers and also the views of testing experts with regard to the reading skills assessed by the test.

Literature on reading test validation indicates that the studies using verbal reports and expert judgments together to validate a reading test are rare. Thus, this study seeks to make a methodological contribution to the existing research as it explores the extent to which verbal reports of test takers and expert judgments can offer a new understanding of the efficacy of the test measuring reading skills. A second potential contribution is to add more evidence of the effectiveness of verbal reports in providing insights into the complex cognitive processes in readers' minds. In terms of application, the research reported here can contribute to designing a more effective reading test and better inform language teachers and English as a second language (ESL) learners seeking to prepare for the reading test.

ESL reading and test taking strategies

When readers are engaged with a text, they activate various cognitive processes, such as parsing sentences through an analysis of the grammatical constituents, interpreting clauses and making inferences, synthesizing information as well as critically evaluating the text and monitoring comprehension. Therefore, reading has long been recognized as a complex process of text comprehension. The complex process involved in reading is aptly captured by Carrell and Grabe's (2002) definition of reading:

> When reading, a reader engages in processing at the phonological, morphological, syntactic, semantic and discourse levels, as well as engages in goal setting, texts summary building, interpretive elaborating from knowledge resources, monitoring and assessment of goal achievement, making various adjustments to enhance comprehension, and making repairs to comprehension processing as needed (p. 234).

Different taxonomies have been introduced to develop an understanding of reading processes. Among them, the most useful for the current study was 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 process or level of engagement. A distinction is made between the reading process focused on a text at a local level (i.e., the understanding of propositions at the level of micro structure) and those operating at a more global level (i.e., the understanding of propositions beyond the level of micro structure) (Khalifa & Weir, 2009, pp. 45–46). For reading types, two basic types of reading, expeditious (speed reading) and careful reading, have been identified.

The brief discussion above indicates that reading might be a difficult task even for native speakers, let alone second language (L2) learners, as reading does not mean only recognizing words but making sense of texts (Tierney, 2005). It requires involvement of a great deal of cognitive capacity available for comprehension (Pressley, 2002). Hence, it is plausible to argue that reading in an L2 can be a difficult task. L2 reading might be more difficult in a test situation because of the pressure that test takers receive to complete the test on time with a good level of success. Therefore, L2 test takers may employ different reading and test taking strategies to overcome the challenges.

a. Reading strategies: Much of the reading process is automatic in nature, which is generally known as skill. This automaticity might indicate that the reading process may be beyond the control of readers. However, readers sometimes have a significant level of active control over their reading process, which is called strategy. Afflerbach, Pearson, and Paris (2008) argue that 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" (p.15). Thus, reading strategies are the conscious mental activities that readers use to overcome the challenges they face when responding to test items. Reading strategies comprise planning, monitoring and evaluation strategies (Phakiti, 2003). Planning strategies are readers' actions of previewing or overviewing tasks and making

decisions about how the task should be done and the steps that need to be taken to accomplish the task. Monitoring strategies refer to the deliberate actions that readers use to monitor their own task performance in order to ensure that the tasks are properly done, and evaluation strategies are those strategies that test takers use to check or evaluate how well they have completed the reading test (Phakiti, 2003, 2008). Thus, reading processes are general, subconscious and more automatic, whereas reading strategies are more conscious and intentional.

The importance of gaining a better understanding of the type of strategies test takers use to respond to test items on a reading test has been widely recognize

d as a part of the process of construct validation (Cohen & Upton, 2007) "a measure of how closely a test reflects the model of reading underlying the test" (Moore, Morton, & Price, 2012, p.121). As argued by Alderson (2000) "validity of a test relates to the interpretation of the correct responses to items, so what matters is not what the test constructors believe an item to be testing, but which responses are considered correct, and what process underlies them" (p. 97). Understanding the trait being measured requires an insight into the cognitive processes test takers follow when engaged in reading tasks (Bax, 2013). Such processes are typically observed through verbal reports (see below).

b. Test-taking strategies: Test-taking strategies refer to "test-taking processes which the respondents have selected and which they are conscious of, at least to some degree" (Cohen & Upton, 2007, p.211). As the reading tasks in a test situation are time-bounded, most test takers may employ different test management or test-wiseness strategies to ensure that they can successfully accomplish the tasks within the limited time. Thus, test takers may sometimes opt out of certain processes and may not proceed through the whole text. This means test takers may use some strategies which constitute short-cuts to arriving at answers. For instance, in order to respond to a gap-filling item, test takers may simply scan the text looking for the specific information, without necessarily understanding the text. However, it should be noted that test-taking strategies do not lead to opting out or to the use of short cuts in the majority of testing situations (Cohen & Upton, 2007). Nonetheless, these strategies do not seem to appear in a non-test situation.

Verbal reports in test validation

Verbal reports are being increasingly implemented as a method of gaining insights into the comprehension processes or reading strategies that readers employ to comprehend a written text (Bax, 2013). However, very few studies have collected verbal reports as a part of test validation. It is assumed that verbal reports provide access to test takers' reading processes (Cohen & Upton, 2007). Cohen (2013) discusses two kinds of self-observation (verbal) reports: introspective (concurrent) and retrospective. Introspective reports are the verbatim records of a problem solver's thinking aloud while completing a given task whereas retrospective reports are the problem solver's reports about how s/he performed the task (Taylor & Dionne, 2000). Verbal reports have been employed in reading test research for almost 30 years. For example, Alderson (1990b) collected introspections from one participant and retrospections from another participant, who were asked to take a reading test. Their verbal reports indicated that "what appears to be being tested by an item does not always match the beliefs of the test and test constructors" (p.477). 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), having collected think-aloud protocols (TAPs) of 52 Hungarian children taking a Grade 7 Hungarian reading test, reported that the reading strategies employed by the test takers were in line with the strategies intended by the test. However, the study collected only the TAPs; had the test takers' retrospections been collected, the retrospections might have offered more insights into the cognitive processes and produced more valid results.

Cohen and Upton (2006) also collected TAPs of 32 test takers of the new Test of English as a Foreign Language (TOEFL) to explore the construct validity of the TOEFL test. The study had the investigators interacting with the test takers in order to obtain rigorous data regarding each and every alternative

choice on the test. The findings were consistent with Nikolov (2006) that the reading strategies employed by the test takers were in line with the strategies intended by the TOEFL test.

In the current study, besides collecting verbal reports from the test takers, it was also decided to collect the views of testing experts with regard to the reading strategies that they felt were being tested. The following section reviews some test validation studies that used expert judgments.

Expert judgments in test validation

Expert judgments are considered to be important for test validation (Alderson, 1993). However, literature on language testing indicates that studies that use expert judgements to validate a test are rare. To the best of our knowledge, only two studies (i.e., Holzknecht, 2012; Tavakoli & Barati, 2011) have used expert judgements to validate a language test, though a number of studies used expert judgements for different purposes: for instance, to predict item difficulty and discrimination indices of a test (Bejar, 1983; Fulcher, 1997), to judge whether the test items in a reading tests were testing 'lower,' 'middle' or 'higher' order abilities (Alderson, 1990a; Alderson & Lukmani, 1989), and to examine the place of sub-skills in ESL syllabus and test design (Lumley, 1993).

Tavakoli and Barati (2011) used expert judgments to investigate the construct validity of the Cambridge English: First (FCE) Test. In the study, six university lecturers in Iran were asked to make judgments regarding the reading skills measured by the test. Additionally, six undergraduate student test takers were also asked to match the skills assessed by each test item in the test to a reading taxonomy of reading skills, which was provided to them. The findings did not reveal a substantial agreement among the experts on the skills assessed by the test, nor could any significant agreement be observed among the test takers.

The study by Holzknecht (2012) is closely related to the current study. In the study, eight students were asked to take the Austrian Matura exam and think-aloud, and seven experts were asked to make judgements regarding the skills being assessed by the test. The results revealed that the students used all the careful reading strategies specified in the test specifications. However, the exam did not have an equal focus on expeditious reading, and the strategy 'reading to deduce the meaning of unfamiliar words from the context' was not prompted by the items although listed in the test specifications. Thus, the findings posed a possible threat to the construct validity of the reading test. The findings also indicated little agreement among the experts on the reading strategies being assessed by the test. Additionally, the study indicates a big gap between strategy use and strategy judgments.

However, to the best of our knowledge, no other studies have validated a reading test by looking at how the test takers take a test and how the testing experts see the test. Additionally, no research, to date, has been carried out to explore the construct validity of the SLC English reading test in Nepal. Therefore, the present study was designed to fill the gap.

The following research questions guided our study:

(1) What strategies do test takers employ to respond to the SLC English reading test?
(2) To what extent do testing experts think that the test assesses the strategies it claims to test?
(3) Is there consistency between strategy use and expert judgments?

Methodology

Test material

The test being investigated was the reading section of the SLC English examination paper of the academic year 2013/2014. The section consisted of four different types of texts (each with a single text): a poem, a narrative text, an explanatory text and an advertisement, with each having a different text length (see Appendix A). The test contained 29 items in total, which were 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).

Participants

Having considered the importance of selecting the students who reflected the target population of the test as closely as possible, we decided to randomly select participants from the students studying at Grade 10. Accordingly, eight students (aged 14–16 years) studying at Grade 10 in a government funded school in Nepal consented to participate in this study. All of them were Nepali native speakers. They had studied English at schools for an average of 12.5 years.

For the selection of the testing experts, two basic selection criteria were set in order to ensure that the right person was selected. The experts should have completed a course on 'Language Assessment' at a post graduate level and have been involved in English test item writing or marking for at least three years. Seven testing experts from a university in Nepal who met the criteria took part in the study. Four of the experts had experience in teaching and testing English for about 20 years and the remaining three experts had about four years' experience each. Thus, there were two different sets of testing experts, one set of highly seasoned experts and one set of relative newcomers. Such a combination was likely to capture a diversity of opinions regarding the construct validity of the test.

Data collection procedures

The study collected two forms of self-observations: introspective (concurrent) and retrospective. Accordingly, the participants were first briefed on the data collection process; they were instructed in their native language (i.e., Nepali) to ensure their understanding of the process. Then, using another reading test from the SLC exam, a warm-up task was conducted in order to assist the students to overcome the inherent challenges of thinking-aloud while engaged in reading, a complex cognitive task. The students practised thinking-aloud until they felt comfortable with the process involved.

Immediately after the warm-up session, they were asked to take the real test and think-aloud. They were allowed to use either English or their native language Nepali in order to ensure that there was no hindrance caused by the language use. 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, one of the researchers (the first author) positioned herself next to the students and took some notes related to their activities, reminding the participants to keep thinking-aloud whenever they remained silent for about 30 seconds; in these instances, they were politely prompted: "Please say aloud what you are thinking now."

Immediately after they had completed the test, they were interviewed about how they had solved each item in the test. The researcher (first author) used the notes she had collected during the think-aloud sessions as the main tool for asking the students questions. Nepali was mostly used in their interview as most of them preferred the language. Both think-aloud and retrospective interview sessions were audio-recorded.

Finally, the students answered a questionnaire which included questions related to their background information (such as their age, their mother tongue, duration of their English study, etc.) and the test (such as the clarity of the instructions, their familiarity with the test items, the difficulty of the test items for them, etc.).

Unlike students, following Holzknecht (2012), the experts were asked to take the test and decide the strategies assessed by each of the items in the test. They were provided with an instrument which consisted of a table with the list of test items and the reading strategies specified in the test specifications, and were asked to fill out the tables in the instrument to indicate which of the strategies listed in the instrument they thought were tested by each individual test item. They could choose more than one strategy for each item. If they believed that an item was not testing any of the

strategies mentioned in the instrument, they could indicate this by ticking a column labeled "other" and mentioning which strategy they believed was being tested.

After making their judgments, they had to answer a few open-ended questions related to their judgment process (e.g. how did you judge the test in terms of the reading strategies tested? How difficult did you find it to make judgments about the reading strategies tested by an item?). Then, all the experts were invited to take part in a focus group discussion (FGD) whose main purpose was to discuss the items which received contrasting responses from the experts. It should be noted that they were asked to complete the tasks independently so that there was no communication among the judges, and the frequency of reading strategies in each item being judged by the experts was calculated before they took part in the FGD. The FGD was audio recorded.

Analysis of the data

The verbal reports of each student were transcribed verbatim, the transcripts were organized according to the individual test item and the reports were coded using the software NVivo10 (Lewins & Silver, 2014). The frequency of each strategy was then calculated on an item-by-item basis for each participant. The repeated use of the same strategy within a single test item by an individual was counted only once so that the tallies for the number of the participants using the strategy for a given item would not be inflated by multiple uses of that strategy.

The expert judgments data were analyzed in three steps. First, the frequency of the strategy judgments was calculated on an item basis. The comments provided by the experts on the category "other" were then analysed. Finally, the responses to the questionnaires and the information obtained from the FGD were analyzed. The codes used in the study are summarized in Table 1.

Finally, in order to examine the consistency between the observed frequency of strategy use and strategy judgements, a chi-square analysis was run on SPSS 20. In addition, in order to control for the increased Type I rate that arises when making multiple comparisons, the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) was used, because this achieves greater statistical power than the more widely used Bonferroni procedure (see also Thissen, Steinberg, & Kuang, 2002).

Table 1. Reading strategies and coding rubric.

Strategy	Description
*DetailUnd	Reading for detailed understanding
*GeneralUnd	Reading for general understanding
*UndThemes	Reading for understanding the underlying themes and ideas
*UndArgument	Reading for understanding of an argument
*UndGist	Reading for understanding the gist of the text
*FindSpecificInfo	Reading for specific information
*UndTextOrganisation	Reading for understanding text organization
*UndTextContinuation	Reading for anticipating the likely continuation of the interrupted text
*UndDiagrammaticInfo	Reading for interpreting the diagrammatic information
*InferingWordMeaning	Reading for deducing the meaning of unfamiliar lexical items
*AppreciatingText	Reading for appreciating literary text
*DictionaryUse	Using authentic English dictionary
WordMatch	Word matching
SentenceMatch	Sentence matching
OwnInterpretaion	Reading and making their own interpretation
GrammarMatch	Matching grammatical structure
Guessing	Guessing
UndMainideas	Reading for understanding main ideas
UndImpDetail	Reading for understanding important detail
InferencingText	Reading for inferencing longer texts

* = Reading strategies specified in the SLC curriculum (Curriculum Development Center, 2009)

Results

As mentioned above, it should be noted that, following Cohen and Upton (2006, 2007), 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 and judgments. Hence, simple quantitative units like numbers and percentages were also used while analyzing both the verbal reports and experts' judgments, though they played no major role in the data interpretation.

Reading strategies used by the students

The students seemed to employ 14 different reading strategies to respond to the test items. Among them, only nine of these strategies are specified in the SLC curriculum as shown in Table 2.

There was no evidence found for the employment of the three of the 12 strategies specified in the SLC curriculum, whilst there was evidence of the use of five additional strategies which are not specified in the SLC curriculum. It should be noted that there does not seem to be a practice of developing test specifications before designing a test in Nepal. Therefore, the construct validity of the test is judged against the reading strategies in the SLC curriculum.

The overall results indicate that, among the strategies used by the students, careful reading in order to understand the argument and expeditious reading for specific information were the two most common strategies. The two strategies made up 27% and 25% respectively of all the strategies used. Both strategies were employed across all the item-types, except for the four items calling for matching (See appendix A). Here is an example to indicate how both strategies were used by a student to respond to a single test item, that is, to determine whether the statement 'The writer thought Jack's mother was not wise' was true or false.

Participant 6 task 2 item 17(true/false item)

The writer thought Jack's mother was not wise… Um here the meaning is … Jack's mother is not wise…Um let me read the text again, um I have read about Jack's mother in the text… (Reads some sentences from the last paragraph and jumps to the second paragraph and reads each line) 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.

Table 2. Reading strategies used by students.

S.N	Reading strategies specified in the SLC curriculum	Reading strategies used by students
1.	Reading for detailed understanding	Reading for detailed understanding
2.	Reading for general understanding	Reading for general understanding
3.	Reading for understanding themes	Reading for understanding themes
4.	Reading for understanding argument	Reading for understanding argument
5.	Reading for understanding gist	Reading for understanding gist
6.	Reading for finding specific information	Reading for finding specific information
7.	Reading for understanding text organization	Reading for understanding text organization
8.	Reading for understanding text continuation	
9.	Understanding diagrammatic information	
10.	Inferring word meaning	Inferring word meaning
11.	Appreciating literary text	Appreciating literary text
12.	Dictionary use	
13.		Matching word meaning
14.		Sentence matching between text and question
15.		Making own interpretation
16.		Grammar matching
17.		Guessing

However, it does not mean that the two strategies were employed together in all the cases. It should also be noted that multiple strategies were applied to respond to almost all the test items. Therefore, multiple strategies were assigned to the same TAP in many cases.

The third most common strategy was Sentence Match; that is, matching the test items with sentences in the text. The following excerpt indicates the employment of the strategy to respond to the question 'How do parents make trouble for themselves?'

Participant 2 task 3 item 21 (short answer question)

How do parents make trouble to themselves?... (Reads the first sentence in the text) Um when when parents make a lot of rules about their children's behavior, they make trouble to themselves...um oh here is the sentence very similar so I copy this answer...this sentence should be the answer.

In the excerpt, it can be seen that the student finds a sentence which is very similar to the question and copies the answer.

The strategy WordMatch was the fourth most common strategy. It was employed to solve all the item-types, except ordering sentences. The following excerpt illustrates the employment of this strategy.

Participant 3 task 4 item 25 (gap filling)

Researcher: 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?

Student: It was very easy. 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.

There was also evidence for the application of the GeneralUnd, DetailUnd, UndThemes UndGist and OwnInterpretation. Additionally, there seemed to be a strong relationship between item-type and the employment of UndTextorganisation and AppreciatingText; both of these strategies were observed only for 'ordering sentences'. Similarly, the two strategies InferingWordMeaning and GrammarMatch seemed to be used only for the item-type 'finding the words with similar meaning'.

Expert judgments on the reading strategies

As with the student verbal reports, the frequencies of each of the strategies were counted on an item-by-item basis from all the seven experts so that the level of agreement among the experts could be found.

The test was judged to be testing 13 different strategies. However, only ten of them are specified in the SLC curriculum (see Table 3).

The test was judged not to be testing the two reading strategies, UndTextContinuation and UndDiagrammaticInfo, which are specified in the SLC curriculum. However, the test was judged to be testing three other strategies (UndMainIdeas, UndImpDetail and InferencingText), which are not specified in the SLC curriculum.

A close inspection of the expert judgments on an item-by-item basis revealed that, with the possible exceptions of a few items, the great majority of the items were judged to test at least five different strategies (see Figure 1). The most extreme items were Item 1 and Item 3 (both true/false items: Dignity is high when there is no fear in mind; Domestic walls have fragmented the world, item 1 and item 3 respectively), 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 each of the items was very low.

Table 3. Reading strategies judged by experts.

S.N	Reading strategies specified in the SLC curriculum	Reading strategies judged by experts
1	Reading for detailed understanding	Reading for detailed understanding
2	Reading for general understanding	Reading for general understanding
3	Reading for understanding themes	Reading for understanding themes
4	Reading for understanding argument	Reading for understanding argument
5	Reading for understanding gist	Reading for understanding gist
6	Reading for finding specific information	Reading for finding specific information
7	Reading for understanding text organization	Reading for understanding text organization
8	Reading for understanding text continuation	
9	Understanding diagrammatic information	
10	Inferring word meaning	Inferring word meaning
11	Appreciating literary text	Appreciating literary text
12	Dictionary use	Dictionary use
13		Reading for understanding main ideas
14		Reading for understanding important details
15		Inferencing text

The most extreme disagreement concerned the judgments 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. A similar kind of disagreement could 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
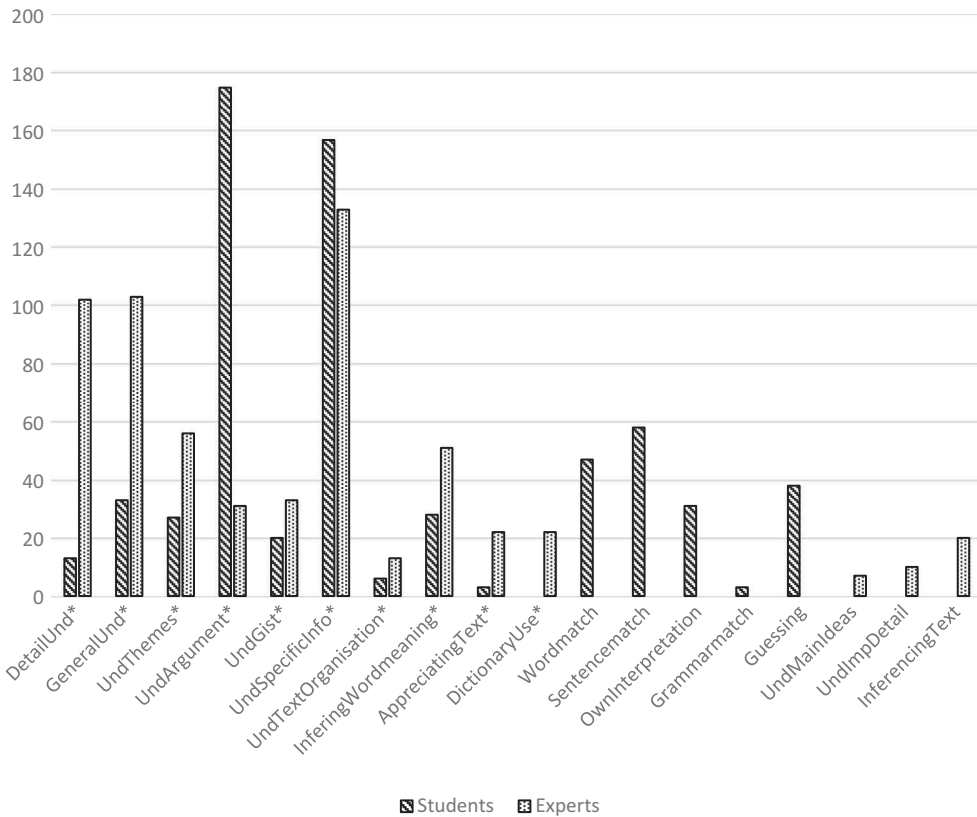

Figure 1. Overall comparison between strategy use and expert judgement.

* = Reading strategies specified in the SLC curriculum

were doing so. Thus, a close observation of the judgments made by individual experts on an item basis confirms that the judges showed a great variation in their perceptions of the strategies tested by the individual items in the test.

Consistency between strategy use and expert judgements

A close inspection of the results indicates a big gap between the strategy use by test takers and strategy judgments. The overall comparison is presented in the figure below.

Figure 1 reveals that, among the 18 strategies, there is some sort of consistency on nine strategies. In order to examine whether the consistency between strategy use and strategy judgement on those overlapping strategies can be statistically proved, a chi-square analysis was run on SPSS 20. The results are presented in the following table:

It is revealed through Table 4 that, among the nine strategies, there is a considerable consistency only on four of the strategies: UndGist, UndSpecificInfo, UndTextOrganisation and AppreciatingText (p > .05). However, there is a big gap between strategy use and strategy judgement in the rest of the strategies (p < .05). To be more specific, those strategies indicate high frequency in strategy use but low frequency in expert judgement, and vice versa. It should also be noted that the use of the Benjamini-Hochberg procedure did not materially affect the results: those effects that were statistically significant without the procedure (p < .05) remained significant when the procedure was employed.

In addition, there is a complete inconsistency in terms of some other strategies. For instance, 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 by the experts to be testing those strategies. Conversely, the experts thought the test examined three other strategies that are not specified in the SLC curriculum but no student seemed to have used those strategies.

An item-type wise comparison indicates a clear gap between the two data. The most extreme item-type was the matching item which indicates a complete inconsistency between strategy use and strategy judgements (see Figure 2).

The above figure indicates that all the eight students employed GeneralUnd to solve the Matching Item and two of them also applied WordMatch. However, almost all the experts thought that the item tested InferingWordmeaning and DictionaryUse.

In the case of other item-types, among many strategies assigned to each of the item-types, some amount of consistency could be observed only with respect to a few strategies. For instance, in the case of the item-type 'Finding similar words', ten different strategies were assigned to the item but there was agreement only on two strategies as shown in Figure 3.

Figure 3 indicates that the expert judgements contradicted the findings of the verbal reports data regarding most of the strategies. However, there is a considerable consistency between strategy use and expert judgements in terms of UndSpecificInfo (χ^2 (2) = 3.05, p = .21) and InferingWordMeaning (χ^2 (2) = .77, p = .96).

Table 4. Consistency between reading strategy use and strategy judgement.

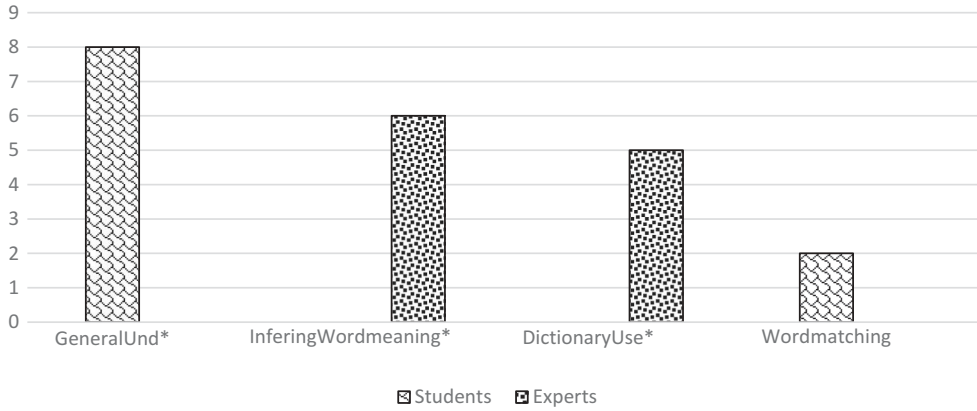| Reading strategy	Chi-square (χ^2)	Degreeof freedom	P-value	Benjamini-Hochberg P-value
DetailUnd	15.00	3	.002	.006
GeneralUnd	12.99	4	.01	.022
UndThemes	11.98	3	.00	00
UndArgument	15.00	3	.00	00
UndGist	3.88	2	.14	.18
UndSpecificInfo	2.67	3	.44	.44
UndTextOrganisation	3.60	3	.30	.33
InferingWordMeaning	9.64	3	.02	.036
AppreciatingText	2.66	1	.10	.15

Figure 2. Comparison of strategy use and strategy judgements for matching items.
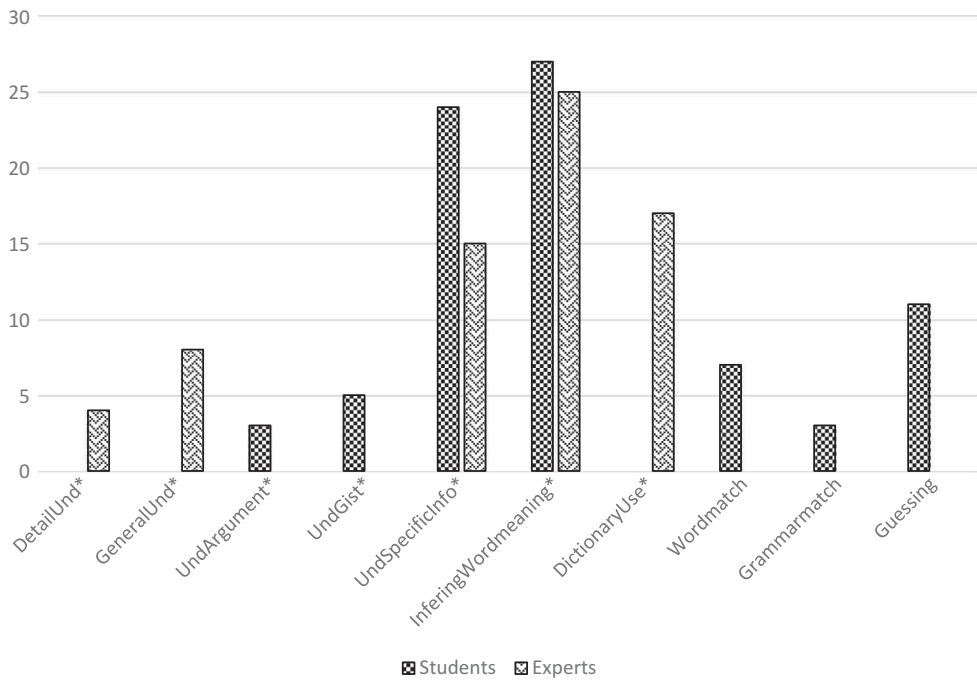
* = Reading strategies listed in the SLC curriculum

Figure 3. Comparison of strategy use and strategy judgements for finding similar words.

* = Reading strategies listed in the SLC curriculum

Summary of findings

The underlying goal of the study was to explore the reading strategies measured by the test as a process of test construct validation. The major focus was on the reading strategies used by the students while taking the test and the views of the testing experts with regard to the reading strategies examined by the test. Accordingly, both introspective and retrospective reports were collected from eight test takers along with the views of seven subject experts.

The verbal reports in this study indicated that only nine (out of 12) of the reading strategies specified in the SLC curriculum were used by the students; however, the students seemed to use five other strategies, not specified in the SLC curriculum. Among the strategies used, UndArgument and UndSpecificInfo were the most common strategies. The students seemed to use multiple strategies, at least five different strategies, to respond to each test-item, with the possible exceptions of Items 5, 18, 19 and 24.

The experts also thought that the test did not seem to test two of the reading strategies specified in the SLC curriculum (UndTextContinuation and UndDiagrammaticInfo) but that it did test three other strategies (UndImpDetail, UndMainIdeas and InferencingText) not specified in the curriculum. However, none of those three strategies were used by the students. It should, nevertheless, be noted that the experts had a low level of agreement on most of the strategies. Another interesting finding was that each of the items was judged to test a minimum of four strategies although the experts seemed to agree over only one or two of the strategies.

The results also revealed a big gap between strategy use and strategy judgements, confirming the results found in the study by Holzknecht (2012). In some cases, there was a complete inconsistency between strategy use and strategy judgements. For instance, both the verbal reports and expert judgements data indicated the test to be measuring other strategies which are not specified in the SLC curriculum. However, those strategies are mutually exclusive; that is, those found in verbal reports did not occur in expert judgements data and vice versa.

Interpretation

The verbal report data did not provide any evidence for the employment of DictionaryUse, UndDiagrammaticInfo, and UndTextContinuation, which are specified in the SLC curriculum. It can be argued that the lack of these three strategies was a result of the nature or design of the test rather than the way students were taught to read English. For instance, the reason for not using DictionaryUse by the students was obvious; following the rules of the SLC exam, the students were not allowed to consult any kind of dictionary while taking the test. The students did not employ UndDiagrammaticInfo as the test did not contain any diagrammatic information (see Appendix A). Similarly, as the test contains no interrupted text, there was no evidence for the application of UndTextContinuation.

The results indicated the use of five other reading strategies which are not specified in the SLC curriculum: WordMatch, SentenceMatch, OwnInterpretation, GrammarMatch and Guessing. However, these strategies seemed to be test taking strategies rather than reading strategies. Therefore, "during the pilot phase of test development it is crucial for test constructors to find out what their tests are actually measuring" (Cohen & Upton, 2007, p. 212).

Both verbal reports and expert judgements indicate that, with some possible exceptions, each of the test items seemed to measure multiple strategies. Therefore, as Alderson (1990a) states, "it is unlikely that any test item can be unambiguously said to be testing any one skill" (p. 436).

The strategy UndArgument was most frequently used by the test takers. A possible reason for the frequent use of UndArgument might be that the majority of 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 UndSpecificInfo, the second most frequent strategy, could be the nature of the test material. The test is designed in such a way that a single text is followed by several questions. In order to answer those questions, the students might have gone back to the text several times either to look for a specific piece of information related to those questions or to verify their answers. It should also be noted that the two strategies were not often tested separately; UndArgument mostly occurred in combination with SpecificInfo. This might further indicate that the test has an equal focus on careful and expeditious reading.

The item-type wise analysis also indicated that each of the item types, except the matching item, seemed to measure a minimum of seven reading strategies. However, in each item type, only two or

three strategies were frequently employed by the students. Additionally, as claimed by Anderson, Bachman, Perkins, and Cohen (1991), item-type was not a very reliable predictor of the patterns of strategy use as the students employed similar types of strategies to approach different items. Consequently, the strategies did not cluster on an item-type basis, with the possible exceptions of UndTextOrganisation and AppreciatingText (occurring only on ordering sentences) and InferingWordMeaning and GrammarMatch (occurring only on finding similar words). Thus, the findings were not in line with the claim made by Farr, Pritchard, and Smitten (1990) 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 seemed to be affected by the students' pre-exposure to the text. For instance, in each task, short-answer questions appeared only at the end preceded by several other questions and the students were found to be strictly following the order of the questions in each task while taking the test. Therefore, the students had a lot of 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.

With regard to the reading level, as with the new TOEFL study by 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, such as ordering sentences, indeed challenged the test-takers to understand the text as a whole, understanding arguments made either implicitly or explicitly in the text and also understanding lexical, grammatical and logical links in order to determine the logical order of the sentences, whereas other items like matching items and finding similar words simply focused on word level understanding.

The results did not indicate a clear relationship between test item difficulty level and strategy use. It was really difficult to clearly observe the strategy adopted to respond to items that were too easy for the test-takers (e.g. Item 24). This finding supports the claim made by Alderson, Clapham, and Wall (1995) 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" (p. 177). This might further indicate that the low difficulty level of some items might have had an impact on the results as certain processes might not have been captured. Conversely, it was found that sentence ordering was the most difficult item, to which the test-takers employed the highest number of strategies.

It should also be noted that it was significantly difficult to differentiate between the use of five different strategies, GeneralUnd, DetailUnd, UndThemes, UndTextOrganisation and AppreciatingText, solely by listening to their TAPs. Therefore, decisions about the application of these strategies were made with support from their retrospective reports.

Another interesting finding was that each of the test items was judged to test a minimum of four strategies. However, the experts seemed to agree with only one or two of the strategies. For instance, Item 27 (short answer question) was judged to be measuring six different strategies but the judges had noteworthy agreement only on FindSpecificInfo. The high degree of variability among the experts in the study supports the claim made by Alderson (1990a) and Alderson and Lukmani (1989) that there is a low level of agreement among testing experts. However, the experts in the FGD were not asked to discuss and reach an agreement on the items which indicated discrepancies. Had this been done, the level of agreement might have improved. The experts in the FGD were simply asked to provide the reason for their choices. Secondly, as shown by Lumley's (1993) study, the level of agreement might have increased if the experts were trained on the task and were asked to discuss their findings. But, the experts were not trained as it was thought the training might affect their judgment. It should also be noted that there might be a measurement error as it involves human subject-rating. Additionally, it should be mentioned that all the experts argued that it is not easy 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 indicated that the experts could not agree as to whether the individual items were eliciting expeditious or careful reading strategies. Each of the items, except Items 11, 20 and 24, was judged to be testing both expeditious and careful reading strategies.

Item-type analysis also revealed a big gap between the strategies employed by the students and the expert judgments. In all the item-types, among the several strategies assigned to the items, there was consistency between strategy use and strategy judgments in respect to a few strategies. This finding might support the claim made by Alderson (1990a) and Alderson and Lukmani (1989) that subject experts are unable to predict which reading strategies are tested by individual test items.

However, some practicalities related to the test should be considered with regard to the gap between strategy use and strategy judgements. For instance, five of the experts viewed that the matching item seemed to potentially measure DictionaryUse. However, following the rules of the SLC, the students were not allowed to use any type of dictionary while taking the test.

Limitations of the study

It should be noted that despite enormous efforts to code the data minutely and consistently, it is still possible that there were some inconsistencies while coding the data (Cohen & Upton, 2006). We also recognise that the students may have applied other strategies as well that were not described in the verbal reports. 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 to use more reading strategies in a test situation than in a non-test situation. In response to the issue, the students in this study were told to treat the test like the real test and most of the SLC exam rules were followed while taking the test; the only major difference was that the students were asked to think aloud in this study and take as much time as they needed to respond to each test item. 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 those which 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.

Strategy use might also have 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 and whether they articulated all their cognitive processes. It is, however, impossible to eliminate the reactive effects of verbal report on task performance (Cohen & Upton, 2006).

Another methodological issue concerns the technique used for collecting retrospective reports. 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 had completed the whole test. It should be noted that the test contains 29 test items and the interview questions started with the first item. Consequently, there was a time gap between the performance and the interview on each test item.

Similarly, two major methodological constraints with regard to the expert judgments in the study should be considered. Firstly, the experts were simply asked to make judgements without any guidelines or training on how to make judgements as it was thought that "any agreement among cloned raters would simply indicate the success of the cloning process" (Alderson, 2000, p. 96). Secondly, there was a time gap between the task completion and the FGD. If the experts had participated in the FGD immediately after they had completed the tasks, they may have been more critical and talked more clearly about their judgement processes.

Thus, it might be plausible to argue that there may be some limitations of both verbal reports and expert judgements to measure construct validity of a high-stakes instrument such as the SLC English test.

Suggestions for further research

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 seemed to follow similar strategies to approach different tasks. Therefore, further studies on construct validation need 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. In addition, this study lacks information about the effects of strategy use on test performance and whether successful and unsuccessful test takers differ in the way they process information. Such analysis might provide a better picture regarding the construct validity of the test.

The study advocates the complementary use of introspective and retrospective reports to gain insights into the cognitive processes in readers' mind. However, it casts doubt on the usefulness of expert judgments to predict which reading strategies the test items are testing as there was a low level of agreement among the experts and there was a large gap between strategy use and strategy judgments. Therefore, a need for further research concerning the methodologies used in the study has been suggested. Additionally, it is recommended that future research minimizes the time gap between the strategy judgement task completion and the FGD.

Implications of the study

Findings from the study have implications for the design of the reading test itself. 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 these 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 does not seem to be a good predictor of reading strategy use. More importantly, an effort should be made to cover all the strategies mentioned in the SLC test specification. Additionally, strategy use seemed to be affected by the design of the test. For instance, because the same text was followed by several questions, the students received a great deal of exposure to the text by the time they reached the final item which may have affected their strategy use. Therefore, shorter texts with fewer test items are recommended for the improvement of the reading test.

Conclusion

The main purpose of the study was to examine the construct validity of the SLC English reading test in Nepal. Accordingly, TAPs and retrospective reports were collected from eight test takers and the views of seven subject experts were collected regarding the strategies tested by the test. Both verbal reports and expert judgements indicated that the test lacks construct validity. For instance, neither verbal reports nor expert judgements indicated that the test tested UndTextContinuation and UndDiagrammaticInfo, which were specified in the SLC curriculum. On the one hand, the test construct was under-represented as the test did not test all the reading strategies specified in the SLC curriculum; on the other hand, the test seemed to be testing some other construct irrelevant skills; that is, testing strategies which were not specified in the curriculum. However, the successful completion of this test seemed to require the students to have both a local and global level understanding of the test passages as well as to use both careful and expeditious reading strategies. Nevertheless, there was a large gap between strategy use and expert judgements and the experts had a low level of agreement regarding the reading skills tested by the test.

Acknowledgments

References

Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Skills and strategies: Their differences, their relationships and why it matters. In K. Mokhtari & R. Sheorey (Eds.), *Reading strategies of first- and second language learners -See how they read* (pp. 11–24). Norwood, MA: Christopher-Gordon Publishers, Inc.

Alderson, J. C. (1990a). Testing reading comprehension skills (Part 1). *Reading in a Foreign Language*, 6(2), 425–438.

Alderson, J. C. (1990b). Testing reading comprehension skills (part 2). *Reading in a Foreign Language*, 7(1), 465–503.

Alderson, J. C. (1993). Judgments in language testing. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 46–57). Alexandria, VA: TESOL.

Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge, UK: 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), 253–270.

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), 41–66.

Bax, S. (2013) Readers' cognitive processes during IELTS reading tests: Evidence from eye tracking. *ELT Research papers* 13–16, London, UK: British Council.

Bejar, I. I. (1983). Subject matter experts' assessment of item statistics. *Applied Psychological Measurement*, 7(1), 303–310.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.

Bhatta, S. D. (2005). *Determinants of student performance in the SLC examination*. Kathmandu, Nepal: Ministry of Education and Sports.

Bhattrai, Y. B. (2014). *The School Leaving Certificate (SLC) examination of Nepal: Exploring negative consequences*. An unpublished MA dissertation. Ontario, Canada: Carleton University.

Budhathoki, C. B., Khatri, H. B., Shrestha, D., Rana, B. K., Sigdel, T. P., Panta, K., & Thapa, S. W. (2014). *Status of SLC dropouts and identifying ways to engaging students in co-curricular activities: Final report*. Kathmandu, Nepal: Ministry of Education and Sports.

Carrell, P. L. & Grabe, W. (2002). Reading. In Schmitt, N., editor, *An introduction to applied linguistics* (pp. 233–250). London, UK: Arnold.

Cohen, A. D. (2013). Verbal report. In C. A. Chapelle (General Ed.), *The encyclopedia of applied linguistics*. Oxford, UK: Wiley-Blackwell. doi: 10.1002/9781405198431.wbeal1261

Cohen, A. D., & Upton, T. A. (2006). Strategies in responding to the new TOEFL reading tasks. (Monograph No. 33). Princeton, NJ: ETS. Retrieved from http://www.ets.org/Media/Research/pdf/RR-06-06.pdf

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), 209–250.

Cordon, L. A., & Day, J. D. (1996). Strategy use on standardized reading comprehension tests. *Journal of Educational Psychology*, 88(2), 288–295.

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), 209–226.

Fulcher, G. (1997). Text difficulty and reading formulae: Reading formulae and expert judgments. *System*, 25(4), 497–513.

Holzknecht, F. (2012). *Establishing one part of a validity argument: Reading strategies in the standardized Austrian Matura exam*. An unpublished MA Thesis. England, UK: Lancaster University.

Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge, UK: Cambridge University Press.

Khaniya, T. R. (1990). Examinations as instruments for educational change: Investigating the washback effect of the Nepalese english exams. An Unpublished PhD thesis. Edinburgh, UK: Edinburgh University.

Lewins, A., & Silver, C. (2014). *Using software in qualitative research: A step-by- step guide*. London, UK: 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), 25–60.

Mathema, K. B., & Bista, M. B. (2006). *Student performance in SLC: Main report*. Kathmandu, Nepal: Ministry of Education and Sports.

McNamara, T. F. (2007). Language assessment in foreign language education: The struggle over constructs. *Modern Language Journal*, 91(2), 280–282.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3–45). Hillsdale, NJ: Erbaum.

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 L. Taylor & J. W. Cyril (Eds.), *IELTS collected papers 2: Research in reading and listening assessment* (pp. 120–211). Cambridge, UK: Cambridge University Press.

Nikolov, M. (2006). Test-taking strategies of 12- and 13- year- old Hungarian learners of EFL: Why whales have migraines. *Language Learning*, 56(1), 1–51.

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), 26–56.

Phakiti, A. (2008). Construct validation of Bachman and Palmer's(1996) strategic competence model over time in EFL reading tests. *Language Testing*, 25(2), 237–272.

Pressley, M. (2002). Metacognition and self-regulated comprehensionIn A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 219–309). Newark, DE: International Reading Association.

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, UK: Lancaster University.

Song, M. Y. (2008). Do divisible sub-skills exist in second language (L2) comprehension? A structural equation modelling approach. *Language Testing*, 25(4), 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), 239–247.

Taylor, K. L., & Dionne, J. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*, 92(3), 413–425.

Tierney, J. E. (Ed.). (2005). *Reading strategies and practices* (6th ed.).Boston, MA: Pearson Education, Inc.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, 27(1), 77–83.

Urquhart, A. H., & Weir, C. J. (1998). *Reading in a second language: Process, product and practice*. New York, NY: Addison Wesley Longman.

Appendix A: Test

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

a. Dignity is high when there is no fear in the mind.
b. Domestic walls have fragmented the world.
c. The poet prays to Father for his individual freedom.
d. 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]

a. Continuous effort leads us towards ………………………………..
b. Fearless situation makes people can get ……………………….. of all kinds.

2. Read the passage an 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 will. 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.

A. From the passage, find the words that are similar in meaning to the following: [4X 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 sounds 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 result 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 hen to 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 Eassay 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]

Coulmn 'A' Coulmn 'B'
a. Aniversary i. special event or happening
b. Organized ii. Event in which people compete
c. Competition iii. Celebration of the yearly return date/event
d. Occasion 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]

a. What's the word limit of the essay?
b. How can competitors submit their essays?
c. Who are eligible for the essay competition?
d. What's the title of the essay?

(Source: Office of the Controller of Examinations (2014). School Leaving Certificate Examinations: English test. Sanothimi, Bhaktapur: Department of Education.)