Complete Research Report Form 1 (For a single project or a sub-project)

Complete Research Report

The Construct Validity of the PSU-TEP Structure and Reading Test

การศึกษาความตรงเชิงโครงสร้างของข้อสอบ PSU-TEP ทักษะโครงสร้างภาษาและการอ่าน

Associate Professor Thanyapa Palanukulwong (Ph.D.)

This research project has been financially supported by Faculty of Liberal Arts, Prince of Songkla University

Abstract

The primary aim of this study was to explore the construct measured by the three different parts of the PSU-TEP Structure and Reading Test through (1) analyses of test scores and test items and (2) stimulated recalls to investigate test-takers' cognitive processing based on Khalifa and Weir's (2009) cognitive model in reading. Statistical analysis was done on the test scores obtained by 941 test-takers sitting for the four parallel forms administered in 2016. The four test forms were found to be reliable tests (Cronbach's Alpha, rtt = .80, .75, .84 and .66, respectively). However, the tests were quite difficult for their target population. All test forms possessed a relatively large number of items with acceptable difficulty and discrimination indexes and efficient distractors. Stimulated recall data, collected from 16 proficient participants taking the four test forms, demonstrated that the three parts of the test: error recognition, rational cloze part and the reading part measured what they were supposed to measure. However, the absence of expeditious reading (scanning and skimming) was noted in the reading part of the test. Recommendations for further research and test developers are provided.

Executive Summary

One major concern of the use of the PSU-TEP Structure and Reading Test is whether it tests what it is supposed to test. Thus, the primary aim of this study was to explore the construct measured in the different parts of the PSU-TEP Structure and Reading Test. Various means were employed for data collection: analyses of test scores and test items, factor analysis and stimulated recalls to investigate test-takers' cognitive processing based on Khalifa and Weir's (2009) cognitive model in reading. There were two groups of participants: (1) 941 test-takers who sat for the four parallel forms of PSU-TEP Structure and Reading Test administered in February, April, June, and December 2016; (2) 16 proficient PSU students selected to take the four tests and participated in the stimulated recall methodology.

Statistical analysis and factor analysis were done on the test scores obtained by 941 test-takers. The four test forms were found to be reliable tests (Cronbach's Alpha, rtt = .80, .75, .84 and .66, respectively). The tests were quite difficult for its target population, with one form being more difficult than the others (Mean = 24.14, SD = 7.77; 24.48, SD = 7.11; 27.00, SD =

9.97; and 21.22, SD = 6.45, respectively from a total score of 60). All test forms possessed a relatively large number of items with acceptable difficulty and discrimination indexes and efficient distractors. Unfortunately, no systematic and meaningful patterns of construct being measured in the four different parts of the test emerged through the use of factor analysis so the investigation of construct being measured were mainly based on stimulated recall data.

Stimulated recall data, collected from 16 proficient participants taking the four test forms, demonstrated the use of *grammar knowledge* emerged as the only significant cognitive process activated while the participants were completing error recognition part. *Knowledge of vocabulary* and *grammar* and *careful reading at global and local level* were the four most used strategies by the participants while approaching rational cloze items. S*low and careful reading* at both global and local level and *vocabulary knowledge* were the three most frequently activated strategies when the participants engaged in the reading part. The use of these types of cognitive processes seemed to suggest that the three parts of the test: error recognition assessing language knowledge, rational cloze part measuring structure and reading, and the reading part assessing reading skills, measured what they were supposed to measure. However, the absence of expeditious reading (scanning and skimming) was noted in the reading part of the test.

Further studies are needed to look into the statistics aspects of the test and the cognitive processing of test-takers while completing the test. A group of 100 test-takers with varying proficiency levels should be recruited to take all the four test forms to establish its reliability and test difficulty. Another extensive study of the cognitive processing of test-takers of the PSU-TEP Structure and Reading Test is called for, using a larger number of proficient test-takers; every test-taker should be assigned to take all the same test forms in order to examine the construct being measured in the test. Strict time constraints should also be imposed on both the rational cloze and reading parts to encourage the use of expeditious reading which is critical for reading in higher education context. It is also recommended the PSU-TEP test developers include a variety of task types that requires both expeditious and careful reading with both global and local information processing.

Acknowledgements

The present research project would not have been completed without the help and support of many people and sources along the way.

First, my utmost thanks and sincere gratitude go to the Department of Languages and Linguistics, Faculty of Liberal Arts, for providing me the access to their test items and to Faculty of Liberal Arts, Prince of Songkla University, for a research grant.

I would like to express my deep gratitude to Dr. Anchana Rukthong, Dr. Usa Intharaksa, and Lecturer Apinya Jongwattanapaiboon for their assistance in the collection of stimulated recalls.

Special thanks go to my research assistant, Ajarn Pun (Lecturer Sakesit Petchinalert) for helping me in data analyses as well as to Ms. Buaboocha Boonyapassawee for her statistical expertise.

The successful completion of the present research project would have been far from reality without the 16 participants who voluntarily agreed to take part in the study. To each and every single one of you, thank you for your participation.

Table of Contents

List of Figures

List of Tables

The Construct Validity of PSU-TEP Structure and Reading Tests

1. Background of the study

A proficiency test is a test used to measure how suitable candidates are for performing a certain task or following a specific task (Heaton, 1997). McNamara (2000) notes that a proficiency test will look to the future situation of language use without any reference to the previous process of teaching. Some proficiency tests such as the Test of English as a Foreign Language (TOEFL), the First Certificate of English (FCE), the International English Language Testing System (IELTS), and the Test of English for International Communication (TOEIC) are utilized all over the world. However, these tests are costly and are only offered at specific times.

In Thailand, there are a number of acceptable proficiency tests which have been developed to measure the English proficiency of Thai students and the Thais in general. These include Chulalongkorn University Test of English Proficiency (CU-TEP) developed and administered by Chulalongkorn University; the Thammasat University Graduate English Test (TU-GET) developed and administered by Thammasat University; and the Prince of Songkla University Test of English Proficiency (PSU-TEP) developed and administered by Prince of Songkla University.

The PSU-TEP Test is a skill-based test battery, assessing test-takers' structural knowledge, reading, listening, writing and speaking abilities. Originally, the tests were to assess PSU graduate students' general English proficiency. According to the regulations of the PSU Graduate School, students are required to achieve an acceptable level of proficiency on internationally-recognized standardized tests such as the TOEFL, the IELTS, etc. However, the PSU-TEP Test developed by the Faculty of Liberal Arts, Prince of Songkla University, is another choice of test for PSU graduate students to take.

These days, the tests are available to both PSU undergraduate and graduate students and the public. The PSU-TEP Test is offered four times a year, each time with a parallel test form based on the same test specification. The test scores are valid for two years. The PSU-TEP Test consists of 4 separate sub-tests: (1) Structure and Reading (2) Listening, (3) Writing, and (4) Speaking. The test format for the Structure and Reading and Listening sub-tests is accompanied

with multiple-choice questions. The writing sub-test measures test-takers' composition ability and the speaking sub-test requires test-takers to orally communicate their ideas/opinions. The Structure and Reading sub-test is administered in 2 hours, followed by the Listening sub-test in 1 hour, the Writing sub-test taking 1 hour and the Speaking part, 10-15 minutes for each test-taker.

Every PSU graduate student studying for a Master's degree has to take the Structure and Reading sub-test and must reach the criteria set depending on their specific field of study, while those studying for a doctoral degree are required to take the Structure and Reading, Listening, and Writing sub-tests and must pass all. The Speaking part is for those in general who wish to have their speaking proficiency assessed.

The PSU-TEP test has been offered to PSU students and the public for years. In test construction, The Department of Languages and Linguistics, Prince of Songkla University, responsible for its construction and administration, follows certain standard procedures for test construction (Brown, 2004): (1) determining the purpose of the test, (2) designing test specification, (3) constructing test items, and (4) specifying scoring procedures for subjective test formats. The other standard procedures for test construction, which are establishing test reliability and test validity, have not yet been conducted. As a provider of the PSU-TEP Test, it has been the responsibility of the Faculty of Liberal Arts, Prince of Songkla University to provide valid information for stakeholders and to demonstrate the qualities of the test.

The principles that govern test design are validity and reliability (Alderson et al, 1995). Test reliability is the degree to which a measurement tool produces stable and consistent results. A test with a little measurement error will consistently rank-order the test-takers in accordance with their comparative true abilities, and this is necessary when important decisions are being made on the basis of the test results (Henning, 1987). Reliability is of primary importance in the use of both proficiency tests and classroom tests. While reliability is necessary, it alone is not sufficient. For a test to be reliable, it also needs to be valid. (Heaton, 1988; Hughes, 1989). Test validity refers to the extent to which a test measures what it is supposed to measure but nothing else (i.e. its construct). For Messicks (1989), validity is an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment. To ensure that a test actually measures what it is intended to measure (i.e. the construct) but not

other variables, it is important that the measure is actually assessing the intended construct, rather than an extraneous factor. A test is said to have construct validity if it can be demonstrated that it measures just the ability which it is supposed to measure. According to Hughes (1989: 27), "It is through construct validity that language testing can be put on a sounder, more scientific footing". Every test must be as valid as possible (Heaton, 1988). The test must aim to provide a true measure of the particular skill which it is intended to measure. Validity relates to the uses made of test scores and the ways test scores are interpreted and is relative to test purpose (Alderson et al, 1995).

2. Test validation

Traditionally, several types of evidence can be gathered in the process of test validation. These include examining the outcomes of the test, i.e. test scores and item performance, test reliability, the interrelation of sub-tests, and the relationship between the test and other measures of the same construct, and factor analysis. However, it has been argued that traditional approaches to construct validation are inadequate in that they largely ignore the processes that test-takers are actually doing to produce answers to questions. Recent thinking in language testing has recognized the importance of gathering information on test-taking processes in the investigation of test construct. The use of verbal protocol analysis in the process of test validation makes it possible to ascertain whether a particular test actually tests what it is supposed to test (Anderson, et al., 1991; Cohen, 2006)

2.1 Statistical analysis of test score and item performance

Test scores are the evidence of test-takers' performances on the items of the test, which are related to the construct or constructs being measured. Test scores were used for the estimation of test reliability and test difficulty.

Based on Classical Item Analysis, two statistical analyses can help in analyzing each objective test item: item facility analysis and item discrimination analysis (Alderson et al, 1995, Brown, 2005). Item Facility (IF, also called Item Difficulty) measures the level of difficulty of an item in the test. It tells us how difficult a test item is by analyzing the percentage of test-takers who correctly answer a given item. The Item Facility value ranges from 0.00 to 1.00. An item with an IF close to 1.00 is a very easy item, while an item close to 0.00 is a very difficult one.

For example, an item with an IF of 0.95 would indicate that 95 percent of test-takers answered correctly, while an item with an IF of 0.25 would be a difficult item since only 25 percent of test-takers correctly answered it. Ideal items for a norm-referenced test should have an average IF of .50. Too easy or too difficult items fail to discriminate among test-takers; too difficult items will lead to guessing.

Item Discrimination (ID) is the capacity of the test to discriminate among different candidates. The discrimination index measures the extent to which the results of an individual item correlate with the results of the whole test. It tells us how each test item discriminates or how well it distinguishes between test-takers at different levels of ability, indicating the degree to which an item separates the test-takers who performed well from those who did poorly. If the item is working well, the top-scoring test-takers should know the correct answer better than the low-scoring ones. The highest discrimination index value possible is +1.00, which is achieved if all the test-takers in the top group get an answer right and none of the low group does. Often, an discrimination index of +0.40 or above is accepted. If test items are intended to have a high discrimination index value, their IF values should be close to 0.50 (Alderson et. al, 1995).

The need for discrimination is particularly important for a proficiency test. Ideal items in a proficiency test should have an average of Item Facility of 0.50 with the highest discrimination index value. These ideal items would be considered well-centered as exactly half of the test-takers (50%) could answer them correctly, while the other half could not be able to get the answer right. However, items rarely have an IF of exactly 0.50. Therefore, items within a range of 0.30 and 0.70 are acceptable (Brown, 2005). With multiple-choice test items, a low discrimination index value can often be explained by the performance of one or more of the distractors (Alderson et. al, 1995).

In order to have good multiple-choice test items to be used in a good test, a further statistical analysis of different parts of each item, i.e. Distractor Analysis, is needed. Apart from the stem, a multiple-choice test has options presented to the test-taker to choose the correct answer from. One of the options, or alternative choices, is the correct answer, whereas the others function as distractors. The use of a distractor efficiency analysis will help examine the degree to

which the distractors are attracting the test-taker who does not know the correct answer. This can be done by numerically analyzing each option chosen by the test-taker (Brown, 2005).

2.2 Factor analyses of test scores

Factor Analysis, a group of analytical and statistical techniques to represent a set of observed variables in terms of a smaller number of hypothetical variables, is one choice for establishing construct validity in language testing. For construct validation, the observed variables are test scores, and hypothetical variables are test construct (Bachman, 1990). Factor Analysis shows whether there are components that are shared in common by the tasks or whether different components underlie the variables under consideration. However, the factorial approach is criticized because it focuses on the outcome of a test rather than upon the process a reader activates when approaching a test item. Khalifa and Weir (2009: 38) caution against relying on factor analysis for test validation because it does not focus on "what is actually happening when a reader processes text under test condition".

2.3 Verbal protocol analysis

Verbal Protocol Analysis (VPA) is a qualitative procedure currently used as a means for supplementing data obtained from quantitative techniques in testing and assessment. Verbal protocols are increasingly playing an important role in the validation of assessment instruments and methods. One approach to establish what a test actually measures is to ask test-takers to think aloud as they work through the test items. The use of verbal protocol is a qualitative methodology with no statistical procedures being applied to the verbal report data. Inferences can be made directly from the data without the need to quantify the data for numerical analyses. Gathering information on test-taking processes is accepted as part of construct validation. It offers insights into the process and strategies used by test-takers, which may not be available through other research methodology. However, because of the intensive nature of verbal protocol research, studies typically involve no more than a handful of participants (Weir, et al., 2012a).

Verbal report, data collected by asking an individual to vocalize what is going on in their mind as they are performing a task, can be gathered from individuals either concurrently or after a language event. Concurrent reports are generated at the same times as a test-taker is completing the test, while retrospective reports are generated after the test-taker has finished the test task and is asked to verbalize once the test task is complete. The protocols produced are then analyzed to identify the cognitive processes involved in the completion of the test task. Verbal protocol serves as a source of data for the researcher to infer cognitive processes and attended information. If there is a close match between the processes that are actually employed and those the test developers predict will be used, then the test is believed to measure what it is supposed to measure (Green, 1998).

Stimulated recall is an introspective method used as a means for eliciting the thought processes taking place while an individual is doing a task or an activity. Stimulated recalls take place after an event, with a prompt that stimulates recall of the mental process and aids the participant in mentally reengaging with the original event. It has an advantage over other think-aloud approaches in that no training is required for the participants to carry out a task. Through stimulus, the prompt in stimulated recalls is a guide to memory verbalization (Gass and Mackay, 2017).

A number of studies adopting VPA have been conducted in L2/FL testing research to examine how test-takers responded to test items that measure language skills. Storey (1997), for example, used concurrent think-aloud and immediate retrospective recalls to investigate L2 learners' test-taking process on a cloze test. Comparing observed behaviors and inferred processes with those predicted in a model of ideal performance of the ability being tested, the researcher found that different items entailed varying degrees of construct validity. Some test-takers used theoretically expected reading processes, while others showed test-wiseness, using strategies such as selecting an option on the basis of elimination.

Rupp et al. (2006), using concurrent verbal protocols, looked at 10 non-native adult readers approaching a reading test with multiple-choice questions. The study showed multiple different representations of the construct of reading comprehension revealed through the characteristics of the test items. Also, test-takers viewed responding to the multiple-choice questions as a problem-solving task and they combined a variety of mental resources interactively when choosing a choice.

Bax (2013) investigated test-takers' cognitive processing while completing IELTS reading test items. Eye movement of a random sample of the participants was tracked. Stimulus recall data were collected to assist in interpreting eye-tracking data. The study found that successful and unsuccessful test-takers differed significantly in their ability to read expeditiously and their focus on particular aspects of test items.

Brunfaut and McCray (2015), with a combined use of both eye-tracking and stimulated recalls, examined the cognitive processing of 25 test-takers while completing Aptis reading tasks. A wide range of cognitive processes was found, including lower- and higher-level processes. The data indicated that the test as whole sampled widely from the construct of reading, and the findings provided key information for the test validation purposes.

3. Cognitive Processing Model in Reading

Khalifa and Weir's (2009) model of cognitive processing in reading integrates cognitive and metacognitive processes with reading. The principal concern is with the mental process readers use in text comprehension when engaging in different types of real-life reading (Khalifa and Weir, 2009; Brunfaut and McCray, 2015). The model is shown in Figure 1 below.

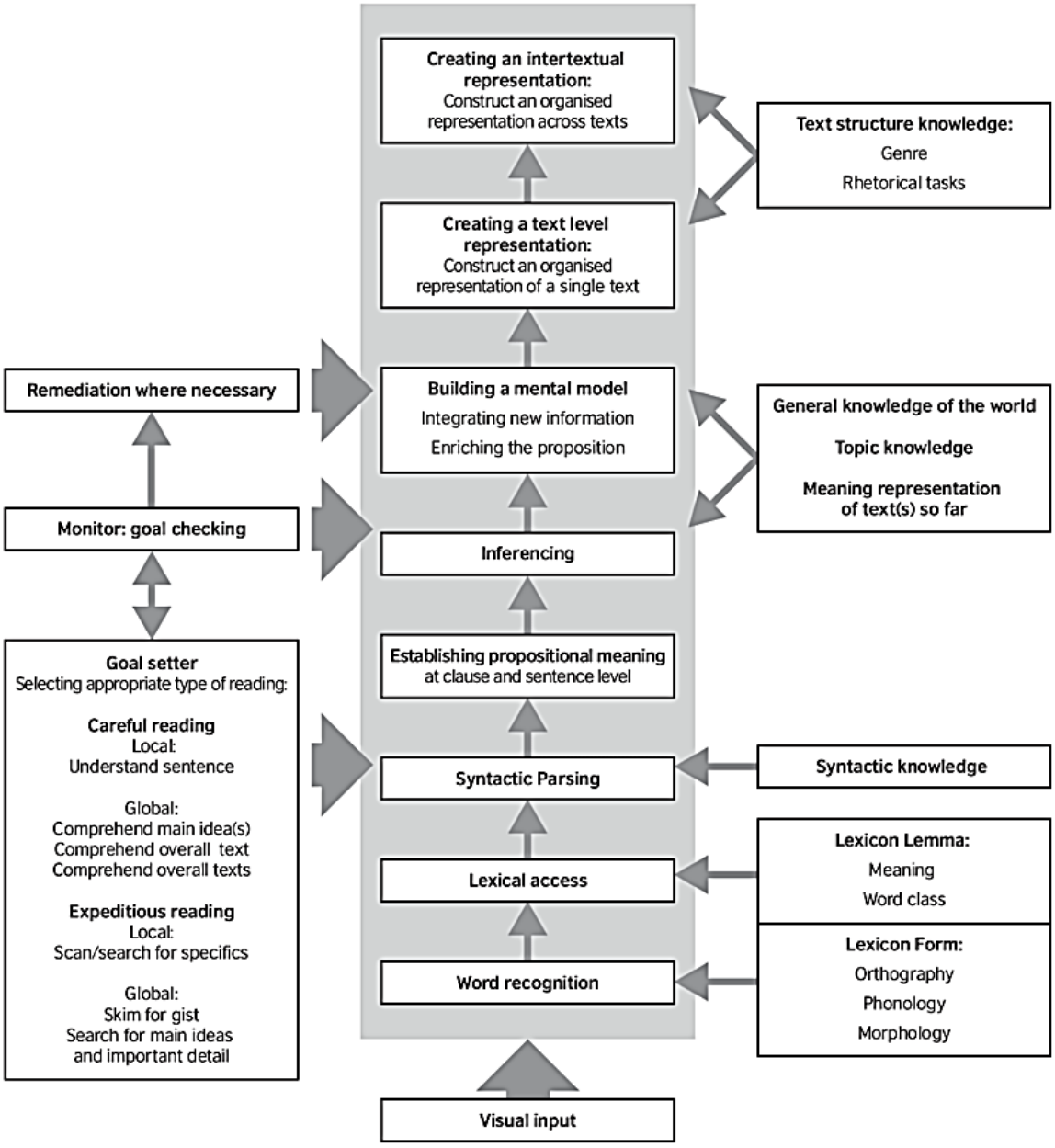


Figure 1: Khalifa and Weir's (2009, p 43) cognitive model in reading

As can be seen in Figure 1, the model of cognitive processing in reading has three main components - metacognitive activity, the central processing core and the knowledge base. In the left-hand column is the **metacognitive processes** of *a goal setter* in deciding what types of reading to use when faced with a text, which will affect ***the levels of processing*** to be activated in the central core of the framework. The *monitor* can be applied to each of the processing level

activated in response to the goal setter's instruction. The ***knowledge base*** required for comprehending texts is in the right-hand column.

The ***metacognitive activities*** involve *setting goals, monitoring, and remediating* where necessary. In *goal setting*, the reader decides the types of reading needed to complete a specific task: *local reading* at the sentence and clause level, or *global level* to understand the text beyond sentence and clause level. Readers may also decide to employ *expeditious reading* or *careful reading*. *Expeditious reading* is quick, selective and efficient reading to access desired information in a text, i.e. scanning, search reading and skimming. Skimming is reading to obtain gist and/or main idea or general impression of a text while scanning involves reading selectively to achieve specific reading goals such as specific words/numbers, events at the local word level. Search reading involves locating information, at both a local and global level, on predetermined topics without having to establish proposition for the whole test. *Careful reading* is intended to extract meaning from presented material at a local or global level, i.e. within or beyond the sentence texts. While reading, the reader monitors that their reading is progressing in line with generated goals, and breakdowns trigger remediation of reading behavior where necessary.

The ***central processing core*** in the middle column of the figure represents a hierarchical system of eight different levels of cognitive processing to be activated as a result of reading. These are divided into two levels of processing: the so-called lower- and higher-level processing (Urquhart and Weir, 1998, cited in Weir et al, 2012a). Lower-level processes include *word recognition, lexical access, syntactic parsing*, and *establishing proposition meaning*, while higher-level processes are *inferencing, building a mental model*, and *creating a text level or intertextual representation. Word recognition* involves recognizing the printed symbols, sounding out the words, and using information on expected grammatical forms. *Lexical access* concerns retrieval of information about the form and word meaning from the reader's vocabulary knowledge to establish word meaning. *Syntactic parsing* is the integration of grammar knowledge of the language (word order, word form, and structural elements) used in establishing *propositional meaning at the clause/sentence level. Inferencing* is the use of a reader's own knowledge of the world, the text topics to comprehend the text. Finally, the integration of individual propositions into overall text meaning is referred to as *building a mental model*.

Knowledge base as illustrated in the right-hand column of the figure consists of various knowledge sources readers may have to help them successfully complete the reading task while processing the texts: *lexical lemma, syntactic knowledge, world and topic knowledge and text structure knowledge*.

The framework is used as the information source to explore the cognitive processes in the present study.

4. Purposes of the study

As a provider of the PSU-TEP test, it has been the responsibility of the Faculty of Liberal Arts, Prince of Songkla University, to supply valid information for stakeholders. To demonstrate the validity of the test, this test validation research was thus conducted to:

1. examine the test scores, item performance (item difficulty and item discrimination) and reliability of four forms of the PSU-TEP Structure and Reading Test administered in 2016.
2. examine the cognitive processes the test-takers employed while completing different parts of the test.
3. explore the construct measured in the test.

5. Research questions

The following research questions were formulated in this study.

1. To what extent is the PSU-TEP Structure and Reading Test a reliable test?
2. What cognitive processes do test-takers employ while completing the test tasks?
3. What does the test measure?

6. Scope of the study

The study focused on both the product of the test, i.e. test scores, and the cognitive processes while test-takers went about processing test tasks. It aimed to provide quantitative information about the test: the reliability, the item difficulty, item discrimination and the

distractor efficiency of the PSU-TEP Structure and Reading Test administered in 2016. Results would demonstrate whether the test is right for the main population of test-takers, the majority of whom are PSU graduate students. More importantly, the use of verbal protocols (stimulated recalls) would reveal whether the PSU-TEP Structure and Reading Test measured what it was set to measure. A group of PSU students were chosen to take the four test forms and participated in the retrospective stimulated recall protocol. The data obtained were analyzed afterwards, specifically based on Khalifa and Weir's (2009) model of cognitive processing in reading, claimed to be the most tenable and productive theoretical basis for establishing the construct validity of a test (Khalifa and Weir, 2009).

7. Definition of terms

Test construct: the cognitive processes that test-takers activate on completing test items successfully (Messick, 1995)

Cognitive processes: activation of 12 response strategies proposed in Weir et al.'s study (2012a) of cognitive model of reading and three additional strategies (See Table 2 and 3, page 14 and 15, respectively)

Error recognition: task consisting of a complete sentence, with four underlined part, one of which is grammatically wrong and the test taker has to identify this grammatically wrong part

Rational cloze test: task consisting of a text from which a number of words are removed and the test taker has to complete the text by selecting a word from a number of options provided for each individual gap

8. Research methodology

8.1 Population/ Participants

The participants of the present study were of two groups:

Participants for quantitative analysis: 941 test-takers completing the four forms of the PSU-TEP Structure and Reading test were participants of the study. Their test performances on different test tasks were analyzed to examine the test's difficulty, discrimination index values, distractor efficiency, reliability, and factor analysis.

Participants for retrospective stimulated recall data: 16 PSU students were purposively and voluntarily selected to be the participants. They were not among the 941

test-takers but were students chosen from different fields of study at Prince of Songkla University (from the Faculty of Medicine, Engineering, Science, and Liberal Arts). They were high-proficient learners of English as identified by their English teachers and were purposively selected to be part of the retrospective stimulated recall protocol. Proficient participants were targeted on in order to be sure that each participant would be able to complete the test tasks and mirror their cognitive processes as described in Khalifa and Weir's (2009) cognitive model of reading, without using test-taking strategies such as using existing knowledge and/or guessing (Weir, 2005). These 16 participants were randomly assigned to take 4 parallel forms of PSU-TEP Structure and Reading Test administered in 2016.

8.2 Research Instruments

8.2.1 PSU-TEP Structure and Reading Test

Four parallel forms (Form 1, 2, 3 and 4) of PSU-TEP Structure and Reading Test administered in February, April, June, and December 2016, respectively, were used as research instruments. Every parallel form was divided into 3 parts with a total score of 60: 15 items of error recognition, 25 items of rational cloze passages, and 20 items of two reading passages (each with the length of 350-400 words). The first part was supposed to test structure knowledge while the second part, the rational cloze passages, was to measure both structure and reading and the last part assessed reading skills.

Based specifically on the same test specification, the three different parts of all PSU-TEP Structure and Reading test forms were developed by three independent teams of test developers at the Faculty of Liberal Arts, to ensure content consistency across test forms. In other words, one group of test developers focused primarily on the error recognition part, while another group was responsible for the rational cloze passages; and the reading part was taken care of by another separate group of test developers.

For each test administration, test-takers have 2 hours to complete the tasks. Answer sheets are computer cards to be machine-scored and calculated. An overview of the PSU-TEP Structure and Reading Test is shown in Table 1 below.

Table 1: The PSU-TEP Structure and Reading Test

Part	Item Types	Number of Items	Points	Ability Measured
1	Error recognition	15	15	**Structure** (subject-verb-agreement; relative clause; gerund and infinitive; present and past participial phrases; passive voice; part of speech; pronouns; word order; signal words, punctuations)
2	Two rational cloze passages (Multiple-choice)	25	25	**Structure and Reading** (equally number of function and content words deleted)
3	Two reading passages (Multiple-choice)	20	20	**Reading Skills** (understanding gist, main ideas/purpose of the text/ topic of text, title of text, tone of the passage; understanding details; word meaning; references; inferences

8.2.2 Stimulated recall analyses

To seek information on the cognitive processes that the participants engaged in to find answers to each item of the test tasks, Khalifa and Weir's (2009) model of cognitive processing in reading, outlined in Figure 1, page 8 of this paper, formed the theoretical framework on which the analyses of stimulated recall data were based.

Strategies Nos. 1-12 representing the cognitive processes the participants engaged in to find answers to each item in different test tasks are those used in Weir et al. (2012a) and three additional strategies, Nos. 13-15 were added.

Table 2: The coding framework for the stimulated recall data on the reading tasks

Strategy Code	Definition
1	Match words that appear in the question with exactly the same words in text (local – scan reading based on word recognition)
2	Quickly match words that appear in the question with similar or related words in the text (local - search reading based on lexical access)
3	Look for parts of the text that the writer indicates to be important (global, text level)
4	Read key parts of the text such as the introduction and conclusion (global, selective reading at text level)
5	Work out the meaning of a difficult word in the question (local, word recognition)
6	Work out the meaning of a difficult word in the text (local, word recognition)
7	Use knowledge of vocabulary (lexical knowledge)
8	Use knowledge of grammar (syntactic knowledge)
9	Read the text or part of it slowly and carefully (careful reading- establishing propositional meaning – global or local)
10	Read relevant part of the text again (careful reading- global or local) re-reading relevant part (local global)
11	Use knowledge of how texts like this are organized (text structural knowledge)
12	Connect information from the text with knowledge already have (general/topic knowledge)
13	Collocation
14	Guess
15	Choice elimination

Following Brunfaut and McCray's (2015), two extra codes, *collocation*, and *guess* were added to the framework. Also, after a random sample of about 25% of stimulated recall data, choice elimination was another code added as it was found to be an activity contributing to either correct or incorrect answers.

Since Khalifa and Weir's (2009) model of cognitive processing was developed in the context of reading in mind (Brunfaut and McCray, 2015), it seemed that only certain strategies were applicable to the error recognition task. So, another set of coding framework was adopted by the researcher to better suit the nature of error recognition task. This set consisted of certain

strategies in Khalifa and Weir's (2009) model of cognitive processing and some additional ones as shown in Table 3 below.

Table 3: The coding framework for the stimulated recall data on the error recognition task

Strategy Code	Definition
6	Work out the meaning of a difficult word in the text (local, word recognition)
7	Use knowledge of vocabulary (lexical knowledge)
8	Use knowledge of grammar (syntactic knowledge)
9	Read the text or part of it slowly and carefully (careful reading- establishing propositional meaning –local)
13	Collocation
14	Guess
15	Choice elimination

8.3 Data collection

Data were collected as follows:

8.3.1 Test score collection

To answer the first research question, test scores of 941 test-takers sitting for the four forms of the PSU-TEP Structure and Reading Test administered in February, April, June, and December 2016, were collected. Descriptive statistics and analyses of the item difficulty, item discrimination, distractor efficiency, and test reliability were performed.

8.3.2 Stimulated recall collection

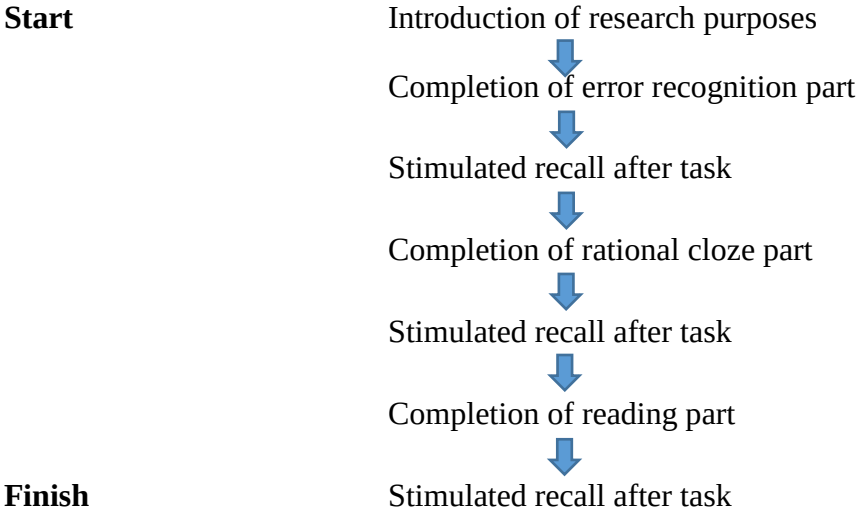
To answer the second and third research questions, 16 proficient students from Faculty of Medicine, Science, and Engineering at Prince of Songkla University, were chosen to take the four test forms and participated in the retrospective stimulated recall sessions.

Sixteen participants were informed of the research purposes and signed a consent form. They were randomly divided into groups of 4; each group was to complete a different test form administered in February, April, June, and December 2016.

The participants were familiarized with the nature of retrospective stimulated recalls before they were asked to take the tests. Each participant took the test at different occasions. Each participant was told that the researcher was interested in what they were thinking at the time, what was in their mind at that time while engaging in the test tasks. Then, they were instructed to complete each part of the test, starting with the error recognition part, rational cloze passages, and reading passages. No time constraints were imposed on test completion.

Immediately after the completion of each test part, the participants were asked to participate in the retrospective stimulated recall sessions carried out on a one-to-one basis either by the researcher or a research assistant. The sessions were conducted to investigate the cognitive processes employed by the participant while responding to each test part. The participant expressed their thought in their first language (Thai) and was simultaneously audio-recorded. Then, the participant continued doing the next test component followed by another stimulated recall interview. Data collection ended when the participant finished the last part and responded to the prompts given by the researcher/research assistant. Their stimulated recall data were recorded and analyzed afterwards to reveal cognitive processes they employed while taking each part of the four test forms.

Figure 2: Flow Chart for data collection



The following questions, some of which were based on Gass and Mackay (2000), were directed to the participants during the retrospective stimulated recall interviews to see how they had approached the tasks and items:

Could you tell me how you started doing the test?

What were you thinking here/ at this point/ right then/ while answering this question?

Do you remember thinking anything when you…?

What were you paying attention to, while completing this task?

Why did you choose to select this choice?

It took about two and a half hours for each participant to complete the required task: completing test tasks and reporting their thoughts. The retrospective stimulated recalls were audio-recorded and transcribed. The data obtained were divided into three parts: (1) the recall data based on error recognition part assessing test-takers' structural knowledge, (2) recall data from the rational cloze passages measuring structure and reading, and (3) reading passages assessing reading skills. Each data set derived from the stimulated recall sessions was coded and analyzed qualitatively to establish the cognitive processes employed by the 16 participants during test completion.

8.4 Data analysis

Data were analyzed as follows.

The first set of research data was quantitative information. This included the test scores of 941 test-takers sitting for the four forms of the PSU-TEP Structure and Reading Test administered in 2016. Data were analyzed using SPSS. Descriptive statistics (frequency, percentage, mean and standard deviation) and difficulty and discrimination indexes were used with test scores and test items. Cronbach's alpha was employed to assess the reliability of the four test forms.

The use of factor analysis was carried out to investigate 941 test-takers' underlying ability of structure and reading capacity captured by the four tests. ONE-WAY analysis of variance and Mann Whitney U were used to investigate the differences among test scores obtained by the 941 test-takers in the four test forms.

The qualitative data comprised the simulated recall data produced immediately after test completion to gain insights into how they approached the tasks**.** The audio-recorded recall data were transcribed by the research assistant. The transcripts on the error recognition part were coded by the researcher based on a specially formulated coding framework adapted from Khalifa and Weir's (2009) model while those on the rational cloze and reading parts were coded based on Khalifa and Weir's (2009) model of cognitive processing in reading and three additional codes (see Table 2 and 3 on page 14 and 15, respectively). In order to ensure the coder reliability of stimulated recall data, an external coder, who was an experienced researcher specializing in English language testing and the use of stimulated recall procedure, was asked to code 25% percent of the total recall data. The percentage of similarity between the two coders was 96.06, indicating a very high inter-coder reliability value (Green, 1998). The stimulated recall data were analyzed to examine the cognitive processes the 16 participants used while going about processing the test tasks. The frequency and percentage of strategies used by the 16 participants were also analyzed using descriptive statistics.

9. Findings

The findings of this study were from two sets of information: quantitative and qualitative data. The former is descriptive statistics on test scores and items of the four forms of the PSU-TEP Structure and Reading Test while the latter is the results of stimulated recall analyses.

9.1 Descriptive statistics

The quantitative findings are presented according to the first research question.

RQ 1: To what extent is the PSU-TEP Structure and Reading Test a reliable test?

In order to answer the first research question, whether the four parallel forms produced consistent and stable scores, data on the performance of all 941 test-takers sitting for the four different test forms of the PSU-TEP Structure and Reading Test were gathered and analyzed to establish the reliability and difficulty of each test form.

The test-takers' scores on four different occasions were analyzed and shown in Tables 4-5 below.

Table 4: Administration of the PSU-TEP Structure and Reading Test on Four Occasions

Month/Test Form	No. of Test-takers	Max	Min	Mean	SD	Reliability
February/ Form 1	256	47	6	24.14	7.77	0.80
April/ Form 2	192	46	9	24.48	7.11	0.75
June/ Form 3	184	52	9	27.00	8.97	0.84
December/ Form 4	309	51	8	21.22	6.45	0.66

Table 4 describes the test-takers' results of performance on the four different test forms. The number of test-takers in each form was different, with Test 4 having the largest number of test-takers (309). The reliability values of the 4 forms were high: 0.80, 0.75, 0.84 and 0.66, respectively. Reliability is a necessary characteristic of a good test. To be a valid test, the test must be first reliable (Heaton, 1988).

In terms of test difficulty, as suggested by the mean scores, the 4 forms of PSU-TEP Structure and Reading Test administered at different times of the year had different levels of difficulty. From a total score of 60, Form 1 and 2 seemed to have the same difficulty level ($\bar{x}$ = 24.14, SD = 7.77 and $\bar{x}$ = 24.48, SD = 7.11, respectively). Form 3 was the least difficult one ($\bar{x}$ = 27.00, SD = 8.97), while Form 4, with its lowest mean scores ($\bar{x}$ = 21.22, SD = 6.45), was the most difficult test form among the four. As can be seen from the mean and range of scores, the test-takers, in general, did not perform well. Their mean scores were less than half of the total score of 60. Such low mean scores indicated that the PSU-TEP Structure and Reading Test was quite difficult for its target test-takers.

Table 5 below describes all the detailed results of the participants taking the PSU-TEP Structure and Reading Test on four different occasions.

Table 5: Descriptive Statistics - four forms of the PSU-TEP Structure and Reading Test

Form 1 (1st Administration: February, 2016, N = 256)					
Part	**Total**	**Max**	**Min**	**Mean**	**SD**
Error Recognition	15	13	0	5.64	2.55
Rational Cloze	25	21	2	10.46	3.96
Reading Passages	20	16	1	8.03	2.85
Total	60	47	6	24.14	7.77

Form 2 (2nd Administration: April, 2016, N = 192)					
Part	**Total**	**Max**	**Min**	**Mean**	**SD**
Error Recognition	15	14	1	6.19	2.48
Rational Cloze	25	21	3	11.46	3.71
Reading Passages	20	15	2	6.82	2.57
Total	60	46	9	24.48	7.11
Form 3 (3rd Administration: June, 2016, N = 184)					
Part	**Total**	**Max**	**Min**	**Mean**	**SD**
Error Recognition	15	15	1	6.25	2.72
Rational Cloze	25	23	2	12.01	4.44
Reading Passages	20	17	1	8.73	3.07
Total	60	52	9	27.00	8.97
Form 4 (4th Administration: December, 2016, N = 309)					
Part	**Total**	**Max**	**Min**	**Mean**	**SD**
Error Recognition	15	13	0	4.73	2.43
Rational Cloze	25	22	2	9.53	3.33
Reading Passages	20	18	1	7.00	2.90
Total	60	51	8	21.22	6.45

Details of different parts of each test form provided further evidence that the four forms had varying degrees of difficulty. The performance of 309 test-takers sitting for Form 4 was striking. The highest scores of the whole test form were 51 out of 60, the second highest scores among the 4 test forms, but its lowest scores were only 8. By the same token, some test-takers on Form 4 also performed very well on different parts of the test, 13 out of 15 on error recognition, 22 out of 25 on rational cloze, and 18 out of 20 on the reading part. However, the minimum scores on different parts were as low as 0 out of 15 on error recognition, 2 out of 25 on rational cloze and only 1 out of 20 on reading passages, suggesting that the test-takers on the fourth administration of the test were of extremely varying levels of language proficiency.

In order to test whether the 4 test forms were statistically different in terms of difficulty, One-way ANOVA was conducted and the results in Table 6 below confirmed that as a whole, the four forms of the PSU-TEP Structure and Reading Test administered throughout 2016 were significantly different in terms of difficulty ($F = 24.20$, $p < .01$).

Table 6: ONE-WAY analysis of variance

Test Form	Form 1		Form 2		Form 3		Form 4		F	Sig.
	X	SD	X	SD	X	SD	X	SD		Differences
Error recognition (total scores = 15)	5.64	2.55	6.19	2.48	6.25	2.72	4.73	2.43	19.99**	0.00
Rational Cloze (total scores = 25)	10.46	3.96	11.46	3.71	12.01	4.44	9.53	3.33	19.81**	0.00
Reading (total scores = 20)	8.03	2.85	6.82	2.57	8.73	3.07	7.00	2.9	21.53**	0.00
Total scores (60)	**24.14**	**7.77**	**24.48**	**7.11**	**27.00**	**8.97**	**21.22**	**6.45**	24.20**	0.00

** significant at .01 level

The mean scores of Form 1, Form 2, Form 3 and Form 4 were 24.14 (SD = 7.77), 24.48 (SD = 7.11), 27.00 (SD = 8.97), and 21.22 (SD = 6.45), respectively. Form 3 seemed to be the easiest form of the PSU-TEP Structure and Reading Test since the test-takers scored the highest, while Form 4 was the most difficult one, with 309 test-takers' mean scores being the lowest. Form 1 and Form 2, however, seemed comparable in terms of difficulty, with average mean scores of 24.14 and 24.48, respectively.

Analyses of multiple comparisons were performed to further examine the differences among the four test forms and the findings are presented in Table 7 below.

Table 7: Multiple comparisons of the mean scores among the four test forms

Test Form		Mean Differences
Form 1 ($\bar{x}$ = 24.14, SD = 7.77)	Form 2	-.339
	Form 3	-2.845*
	Form 4	2.928*
Form 2 ($\bar{x}$ = 24.48, SD = 7.11)	Form 1	-.339
	Form 3	-2.505*
	Form 4	3.268*

Form 3 ($\bar{x}$ = 27.00, SD = 8.97)	Form 1	2.845*
	Form 2	2.505*
	Form 4	5.773*
Form 4 ($\bar{x}$ = 21.22, SD = 6.45)	Form 1	-2.928*
	Form 2	-3.268*
	Form 3	-5.773*

* significant at .05 level

The table shows that the mean scores of Form 1 were not significantly different from those of Form 2, whereas they were significantly different from the mean scores of Form 3 and Form 4.

Therefore, it can be concluded that both Form 1 and 2 are comparable in terms of difficulty, and they were significantly different from Form 3 and Form 4, indicating that only the two forms of PSU-TEP Structure and Reading Test were parallel; the other two forms were not, even though all forms were supposed to have the same level of difficulty.

To sum up, statistical analyses showed that of all four forms of PSU-TEP Structure and Reading Test administered in 2016, only Form 1 and Form 2 had the same level of test difficulty, while Form 3 and Form 4 were significantly different in terms of test difficulty. To be specific, Form 4 was the most difficult test, while Form 3 was the easiest one. Also, the low mean scores obtained by all the 941 test-takers taking the four forms of the test across different administrations in 2016 suggest that the PSU-TEP Structure and Reading Test, represented by the four forms, was quite difficult for its target test-takers.

Thus, the answer to the first research question was that all the four forms of the PSU-TEP Structure and Reading Test administered in 2016 had acceptable reliability estimates (r = .80, .75, .84 and .66, respectively). In addition, the four test forms had significantly different levels of difficulty. Form 1 and Form 2 had a comparable difficulty level ($\bar{x}$ = 24.14, SD = 7.77 and $\bar{x}$ = 24.48, SD = 7.11, respectively). Among the four test forms, Form 4 was the most difficult one, with the lowest mean score of 21.22 (SD = 6.45). Form 3, on the other hand, was the easiest one as it yielded the highest mean score of 27.00 (SD = 8.97).

Furthermore, items in each test form were analyzed in order to reveal their difficulty and discrimination indexes. Items with difficulty indexes ranging from 0.30 – 0.70 are considered good items. Discrimination indexes above 0.30 are acceptable in terms of discrimination (Brown, 2005). Results are shown in Table 8 below.

Table 8: Difficulty and Discrimination Indexes across four tests

Test Form	Item Difficulty		Item Discrimination	
	Acceptable	**Poor**	**Acceptable**	**Poor**
Form 1	40 (66.67%)	20 (33.33%)	25 (41.66%)	35 (58.33%)
Form 2	48 (80%)	12 (20%)	15 (25%)	45 (75%)
Form 3	49 (81.66%)	11 (18.33%)	42 (70%)	18 (30%)
Form 4	35 (58.33%)	25 (41.67%)	24 (40%)	36 (60%)

Each test form consisted of 60 items: 15 items on error recognition, 25 items on rational cloze, and 20 items on reading. The number of items with acceptable difficulty ranged from 58.33% to 81.66%; items with acceptable discrimination indexes were in a range of 25 to 70%. Form 3 consisted of the highest number of acceptable items in terms of difficulty (81.66%), and 70% of its items could effectively discriminate the good from the poor test-takers. Items in Form 2 were well accepted in terms of difficulty (80%) but only 25% of its items were accepted in terms of discrimination power. Two-thirds of items in Form 1 (66.67%) were of acceptable difficulty, and nearly half of the items were accepted regarding their discrimination power. Among the four test forms, Form 4 contained the least number of acceptable items regarding their difficulty (58.33%) and yielded the second lowest number of acceptable items as far as discrimination was concerned (40%). Obviously, Table 8 shows a number of items needing improvement in terms of difficulty and discrimination.

In order to investigate the efficiency of the options in each of the four test forms, analyses into how distractors managed to distract the test-takers who did not know the correct answer were done. To do this, the percentages of test-takers in the high, middle and low groups who chose each of the distractors were analyzed. Efficient distractors should attract more of the test-takers in the low group than those in the higher groups while inefficifient distractors were those

working in the opposite way, distracting more proficient test-takers than less proficient ones. Non-functioing distractors were those not chosen by any test-takers. The results of distractor analysis were shown in Table 9 below

Table 9: Distractor analyses across the four test forms

Distractors			
Test Form	**Efficient**	**Inefficient**	**Non-functioning**
Form 1	142 (78.89%)	38 (21.11%)	0
Form 2	143 (79.44%)	37 (20.56%)	
Form 3	148 (82.22%)	32 (17.78%)	
Form 4	136 (75.56%)	44 (24.44%)	

In terms of the effectiveness of the distractors used in all forms of the PSU-TEP Structure and Reading Test, by comparison, it can be seen from Table 9 that all test forms contained both effective and ineffective distractors. However, there were no non-functioning distractors.

One hundred eighty distractors in each test form (3 distractors for each item, a total of 60 items in a test) were analyzed for their efficiency in distracting poor test-takers. Table 9 showed the number of ineffective distractors ranging from 17.78 to 24.44% of the total distractors. With the least number of ineffective distractors (32 out of 180, 17.78 %), Form 3 seemed to be the most efficient test of all four forms, while Form 4 containing the greatest number of inefficient distractors (44 out of 180, 24.44%). These inefficient distractors were subject to revision.

Overall, although the four test forms were found to be relatively difficult for their target test-takers, their items performed relatively quite well in terms of item difficulty and discrimination; most distractors were quite efficient. Certain items and distractors, however, needed improvement (see Appendix A for difficulty and discrimination indexes, distractor analysis).

9.2 Factor Analysis

The test scores obtained by 941 test-takers who took the four parallel forms of the PSU-TEP Structure and Reading Test administered in 2016 were used for the analysis. Results from factor analyses of the four test forms of the PSU-TEP Structure and Reading Test are presented in the table below and also in Appendix B.

Table 10: Factor analysis: Components of the three parts of the test

Part	Form1	Form2	Form3	Form4
Error (Items1-15)	KMO=0.638, df=105, sig=0.01	KMO=0.588, df=105, sig=0.01	KMO=0.606, df=105, sig=0.01	KMO=0.534, df=91, sig=0.05
	6 components/ Test items: 1) 7, 1, 13 2) 12, 9, 10, 6 3) 2, 5 4) 8, 3 5) 11, 15 6) 4	6 components/ Test items: 1) 12, 2, 11 2) 8, 1 3) 15, 4, 5 4) 14, 3, 7 5) 10 6) 9, 13	7 components / Test items: 1) 6, 12, 13 2) 1, 9, 2 3) 8, 4 4) 15, 3 5) 7, 10 6) 11, 5 7) 14	6 components/ Test items: 1) 4, 11, 10 2) 5, 3, 8 3) 2, 12 4) 15, 13 5) 9, 14, 6 6) 7
Cloze (Items16-40)	KMO=0.722, df=300, sig=0.01	KMO=0.643, df=300, sig=0.01	KMO=0.699, df=300, sig=0.01	KMO=0.608, df=300, sig=0.01
	9 components/ Test items: 1) 31, 25, 26, 29, 32 2) 21, 33, 34, 40 3) 35, 17, 18, 23, 16 4) 37, 38, 27 5) 30, 39 6) 22 7) 36 8) 24, 28 9) 20	9 components/ Test items: 1) 38, 35, 31, 39, 34, 26, 37 2) 20, 17, 30, 32 3) 23, 28, 18 4) 33 5) 22, 19 6) 16, 29 7) 25, 21 8) 40, 24 9) 27	10 components/ Test items: 1) 18, 27, 24, 20, 29 2) 31, 35, 40, 26 3) 22, 33, 32 4) 38 5) 28, 34, 19, 37 6) 16, 25 7) 23, 30 8) 21 9) 39 10) 17	10 components/ Test items: 1) 25, 27, 24, 36 2) 40, 34, 37 3) 18, 39, 17, 19 4) 22, 33 5) 38 6) 29, 21 7) 32, 35 8) 23, 16, 30 9) 28 10) 26, 31

Table 10: Factor analysis: Components of the three parts of the test (cont'd)

Part	Form1	Form2	Form3	Form4
Reading (Items41-60)	KMO=0.711, df=190, sig=0.01	KMO=0.525, df=190, sig=0.05	KMO=0.644, df=190, sig=0.01	KMO=0.638, df=190, sig=0.01
	7 components/ Test items: 1) 48, 44, 51, 55, 52, 58, 43 2) 57, 46, 54 3) 53, 47 4) 56, 45 5) 49, 50 6) 41, 59 7) 60	10 components/ Test items: 1) 51, 49 2) 44, 56 3) 53, 47 4) 41,58 5) 54 6) 46, 42 7) 45 8) 48, 43 9) 55, 59 10) 57, 52	9 components/ Test items: 1) 43, 49, 56, 58 2) 41, 45 3) 47 4) 60, 51, 50 5) 54, 53, 46 6) 48, 44 7) 57 8) 52 9) 42	7 components/ Test items: 1) 43, 44, 45, 54 2) 50, 47, 60 3) 57, 56, 59, 58 4) 46, 51 5) 52, 41 6) 48 7) 49,

When the Eigenvalues of higher than 1 was taken into account, the absolute values contributing to each component lower than .03 were suppressed. Table 10 presents the number of components in each test part together with the number of test items in each component.

The number of components ranged from 6 to 7 in the error recognition part, 9 to 10 components in the rational cloze part and 7 - 10 in the reading part. However, a close look at each component in each test part did not yield any systematic and meaningful patterns of construct being measured in the three parts of the PSU-TEP Structure and Reading Test so that relationship among the variables could be seen (see detailed analyses and construct specified in the test specification as compared to components extracted by factor analysis in Appendix B). Thus, the investigation of the construct of the test would be done based mainly on stimulated recall data.

9.3 Quantitative Findings: Stimulated Recall Data

A second set of data to answer the second and third research questions was based on the stimulated recall data produced after the test-takers completed the test.

RQ 2: What cognitive processes do test-takers employ while completing the test tasks?

RQ 3: What does the test measure?

In order to gain insights into test-takers' cognitive processing while completing each test task, retrospective stimulated recall data collected on the items of each task type were divided into three sets: (1) recall data based on the error recognition part assessing test-takers' structural knowledge and (2) data from the rational cloze part measuring structure and reading; and (3) reading passages assessing reading skills. The three sets were analyzed according to two different coding schemes (see Tables 2 and 3 on pages 13-15).

Table 11 describes the 16 participants' performance on the four parallel forms of the PSU-TEP Structure and Reading Test.

Table 11: Sixteen participants' performance on the four test forms

Form 1 (1st Administration: February, 2016; N = 4)					
Part	**Max. Scores**	**Max**	**Min**	**Mean**	**SD**
Error Recognition	15	15	4	10.75	4.99
Rational Cloze	25	18	10	14.75	3.59
Reading Passages	20	13	9	10.75	1.71
Total	**60**	**43**	**29**	**36.25**	**7.27**
Form 2 (2nd Administration: April, 2016; N = 4)					
Part	**Max. Scores**	**Max**	**Min**	**Mean**	**SD**
Error Recognition	15	15	11	12.00	2.00
Rational Cloze	25	20	15	18.25	2.22
Reading Passages	20	17	6	12.50	4.80
Total	**60**	**50**	**32**	**42.75**	**7.89**
Form 3 (3rd Administration: June, 2016; N = 4)					
Part	**Max. Scores**	**Max**	**Min**	**Mean**	**SD**
Error Recognition	15	15	9	12.25	2.50
Rational Cloze	25	23	19	21.25	1.71
Reading Passages	20	16	12	14.25	1.71
Total	**60**	**53**	**45**	**47.75**	**3.59**
Form 4 (4th Administration: December, 2016; N = 4)					
Part	**Max. Scores**	**Max**	**Min**	**Mean**	**SD**
Error Recognition	15	13	10	11.50	1.29
Rational Cloze	25	23	17	20.50	2.65
Reading Passages	20	16	8	11.50	3.42
Total	**60**	**52**	**37**	**43.50**	**6.86**

The 16 participants performed well on the four parallel test forms; their mean scores, 36.25 (SD = 7.27), 42.75 (SD = 7.89), 47.75 (SD = 3.59), and 43.50 (SD = 6.86), respectively, were higher than half of the total possible score of 60, suggesting that they were proficient test-takers.

In each of the tables below, 12 response strategies based on Khalifa and Weir's (2009) model of cognitive processing in reading and three extra codes, as already shown in Table 2 and 3, page 14 and 15, respectively, served as the basis of the coding framework. To determine the

cognitive processes the 16 participants employed while taking the test, the occurrence of each of the codes was calculated.

A distinction was made between codes used for correct answers and incorrect answers. However, the analysis mainly focused on the correctly-answered items since, from a validation perspective, these would reflect the intended aspect of reading (Brunfaut and McCray, 2015). It should be noted that, while completing an item, more than one cognitive process could be activated by the participants

9.3.1 Analysis of the simulated recall data on the error recognition part

The table below gives an overview of the cognitive processes used by the 16 participants while completing the error recognition items.

Table 12: Frequency of occurrence of cognitive processes used during the error recognition

Error	Form 1				Form 2				Form 3				Form 4			
	Correct		Incorrect		Correct		Incorrect		Correct		Incorrect		Correct		Incorrect	
	(n=43)		(n=26)		(n=55)		(n=16)		(n=53)		(n=11)		(n=55)		(n=21)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
No.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.7	0.00	0.00	1.00	3.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.8	37.00	86.05	12.00	46.15	48.00	87.27	11.00	68.75	49.00	92.45	7.00	63.64	41.00	74.55	8.00	38.10
No.9 (L)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.82	0.00	0.00
No.13	4.00	9.30	4.00	15.38	0.00	0.00	1.00	6.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.14	1.00	2.33	7.00	26.92	6.00	10.91	2.00	12.50	4.00	7.55	4.00	36.36	7.00	12.73	9.00	42.86
Np.15	0.00	0.00	2.00	7.69	1.00	1.82	2.00	12.50	0.00	0.00	0.00	0.00	6.00	10.90	4.00	19.05
Total	42.00	97.68	26.00	100.00	55.00	100.00	16.00	100.00	53.00	100.00	11.00	100.00	55.00	100.00	21.00	100.00

G=Global L=Local

Overall, as evidenced in Table 12, the most frequently used strategy for the error recognition task across four different test forms was the use of lower-level skill, No. 8 (*using grammar knowledge*). Of all the 206 strategies employed for correctly answered items (N = 43, 55, 53 and 55 in all four forms), a total of 175 cases (85%) were reported using strategy No. 8. Figure 3 below best illustrates strategy use in the completion of error recognition items across all test forms.

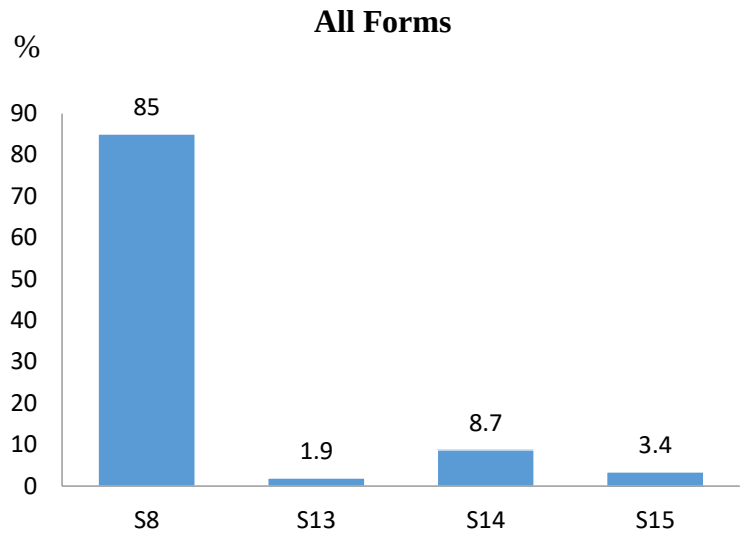


Figure 3: Cognitive strategies employed in error recognition tasks across four test forms

Obviously as shown in Figure 3 above, all the participants heavily relied on *the use of grammar knowledge* (No. 8) to arrive at correct answers, with a very few uses of *guessing* (No. 14) and *choice elimination* (No. 15).

Further analysis in four different test forms showed a relatively similar pattern of strategy use. Of all the 43 strategies used to answer the items correctly in Form 1, 86% of these strategies activated by the participants were strategy No. 8 (*use of grammar knowledge*); 87.3% in Form 2; 92.5% in Form 3; and 74.5% by participants taking Form 4. Very few cases of *guessing* (No. 14), and *choice elimination* (No. 15) were reported in each form. The graphs below best illustrate the strategy used in error recognition task in each test form.

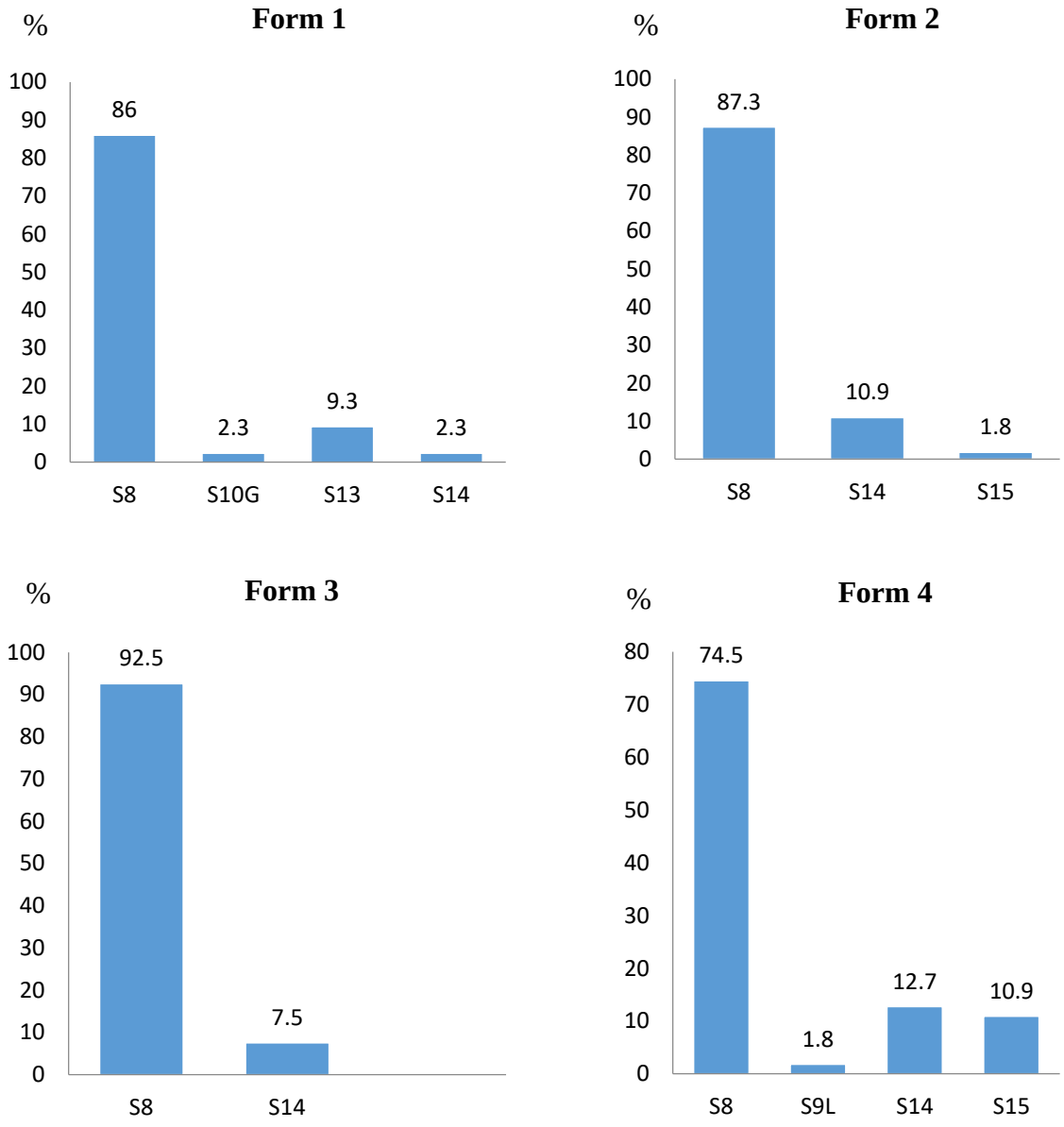


Figure 4: Cognitive strategies employed in error recognition tasks in each test form

The fact that strategy No. 8 (*using knowledge of grammar*) stood out across the four forms of the test and in each of the test forms seems to confirm that the error recognition items actually tested what they were supposed to test, i.e. the language knowledge of the test-takers. In other words, data from the stimulated recalls showed that the construct of the error recognition is the knowledge of grammar. Also, the similar pattern of strategy use for error recognition items

across the four forms of the PSU-TEP Structure and Reading Test may suggest that error recognition parts across the four test forms are parallel.

9.3.2 Analysis of simulated recall data on the rational cloze part

Before approaching the items, no participant indicated that they previewed the passages to get a general idea of what it was about. This might be due to the fact that texts used for cloze tasks were generally short, just a paragraph or a few paragraphs long; the need for skimming for general idea seemed unnecessary. Also, a number of participants may have felt that they had enough time available to carefully read the texts since they were told to spend as much time as they needed to complete the test.

An overview of the cognitive processes used by the 16 participants while completing the rational cloze tasks is presented in the Table 13.

Table 13: Frequency of occurrence of cognitive processes used during the rational cloze part

Cloze	Form 1				Form 2				Form 3				Form 4			
	Correct		Incorrect		Correct		Incorrect		Correct		Incorrect		Correct		Incorrect	
	(n=95)		(n=73)		(n=107)		(n=48)		(n=100)		(n=17)		(n=113)		(n=33)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
No.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.2	2.00	2.11	1.00	1.37	0.00	0.00	0.00	0.00	3.00	3.00	0.00	0.00	0.00	0.00	0.00	0.00
No.3	0.00	0.00	0.00	0.00	1.00	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.4	0.00	0.00	1.00	1.37	1.00	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.7	24.00	25.26	18.00	24.66	22.00	20.56	7.00	14.58	24.00	24.00	6.00	35.29	32.00	28.32	7.00	21.21
No.8	21.00	22.11	15.00	20.55	33.00	30.84	10.00	20.83	33.00	33.00	2.00	11.76	32.00	28.32	4.00	12.12
No.9 (G)	7.00	7.37	3.00	4.11	15.00	14.02	4.00	8.33	16.00	16.00	3.00	17.65	17.00	15.04	4.00	12.12
No.9 (L)	23.00	24.21	7.00	9.59	7.00	6.54	12.00	25.00	14.00	14.00	3.00	17.65	13.00	11.50	3.00	9.09
No.10 (G)	2.00	2.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	6.06
No.10 (L)	1.00	1.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	3.03
No.11	0.00	0.00	0.00	0.00	1.00	0.93	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.88	0.00	0.00
No.12	0.00	0.00	1.00	1.37	0.00	0.00	1.00	2.08	0.00	0.00	0.00	0.00	1.00	0.88	0.00	0.00
No.13	3.00	3.16	1.00	1.37	6.00	5.61	1.00	2.08	2.00	2.00	0.00	0.00	3.00	2.65	0.00	0.00
No.14	7.00	7.37	18.00	24.66	9.00	8.41	6.00	12.50	4.00	4.00	2.00	11.76	8.00	7.08	6.00	18.18
No.15	5.00	5.26	8.00	10.96	12.00	11.21	7.00	14.58	4.00	4.00	1.00	5.88	6.00	5.31	6.00	18.18
Total	95.00	100.00	73.00	100.00	107.00	100.00	48.00	100.00	100.00	100.00	17.00	100.00	113.00	100.00	33.00	100.00

G=Global L=Local

No.1	Match words that appear in the question with exactly the same words in text (local – scan reading based on word recognition	No.6	Work out the meaning of a difficult word in the text (local, word recognition)	No.11	Use knowledge of how texts like this are organized (text structural knowledge)
No.2	Quickly match words that appear in the question with similar or related words in the text (local - search reading based on lexical access)	No.7	Use knowledge of vocabulary (lexical knowledge)	No.12	Connect information from the text with knowledge already have (general/ topic knowledge)
No.3	Look for parts of the text that the writer indicates to be important (global, text level)	No.8	Use knowledge of grammar (syntactic knowledge)	No.13	Collocation
No.4	Read key parts of the text such as thee introduction and conclusion (global, selective reading at text level)	No.9	Read the text or part of it slowly and carefully (careful reading- establishing propositional meaning – global or local)	No.14	Guess
No.5	Work out the meaning of a difficult word in the question (local, word recognition)	No.10	Read relevant part of the text again (careful reading- global or local)	No.15	Choice elimination

The results in Table 12 indicate that in the completion of rational cloze tasks, a wide range of cognitive processes were used. Often, the use of more than one strategy to arrive at a correct answer was evident from the stimulated recall data.

Across all test forms, the four most popular strategies were Strategies No. 8 (*use of grammar knowledge*), No. 7 (*use of vocabulary knowledge*), No. 9 L (*read the text or part of it slowly and carefully at sentence/ clause level*) and No. 9 G (*reading the text or part of it slowly and carefully beyond sentence level*), respectively. In a total of 415 strategies participants relied on to arrive at correct answers, 119 cases of these (28.7%) were the use of strategy No. 8; 102 cases (24.6%) were strategy No. 7; 57 cases (13.73%) were of No. 9 (local) and 55 cases (13.3%) of No. 9 (global). Very few cases of *collocation* (No. 13), *guessing* (No. 14) and *choice elimination* (No. 15) were found to help the participants get the correct answers.

This is well evidenced in Figure 5 below.

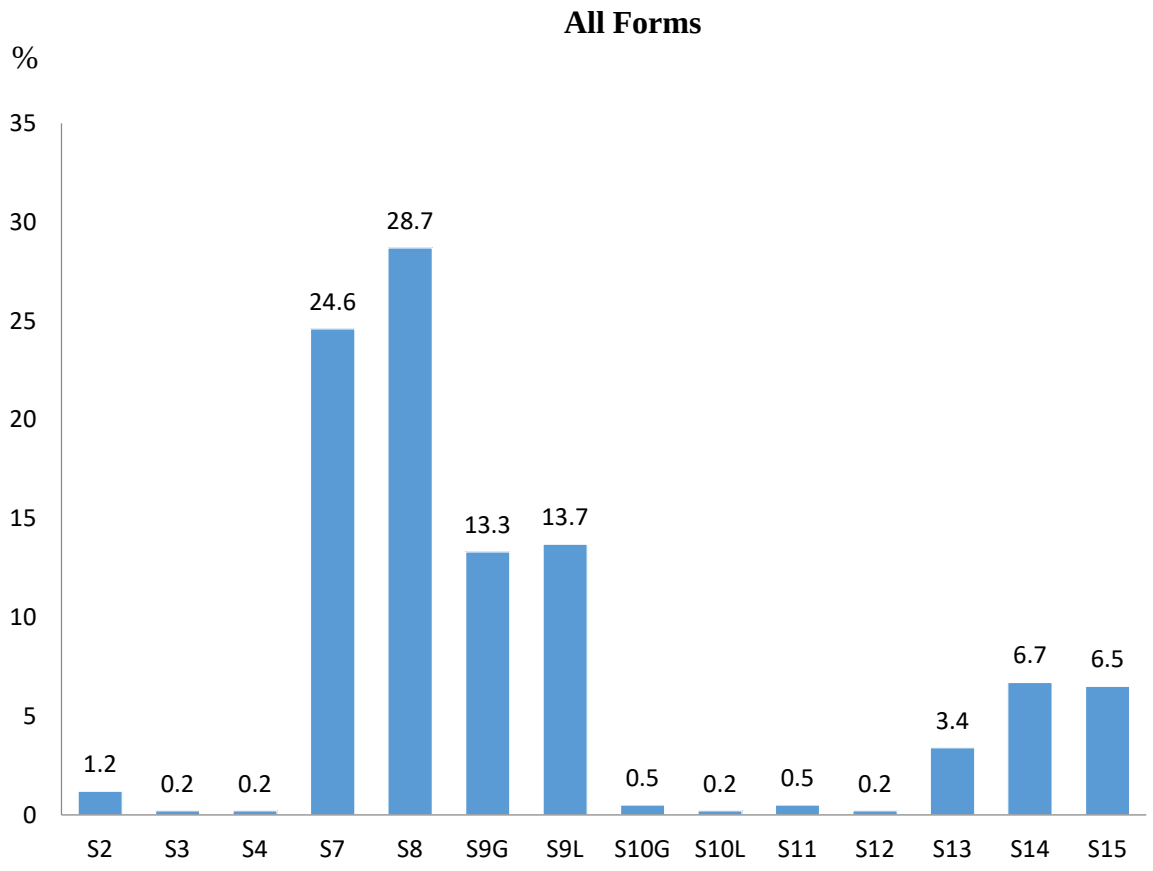
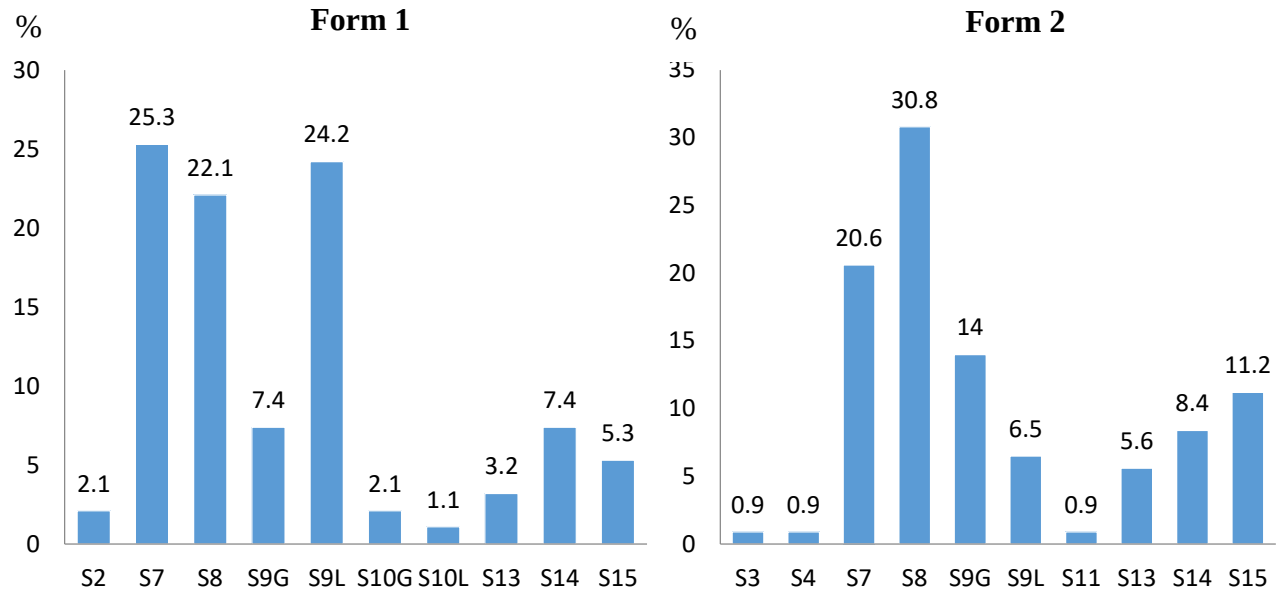


Figure 5: Cognitive strategies employed in rational cloze tasks across four test forms

Analyses of all strategies used for correct answers in each test form showed a quite similar pattern of strategy use. As shown in Figure 6, the most popular strategy in each form was strategy No. 8 (*use of grammar knowledge*): 22.1% of the total strategies participants taking Form 1 used to arrive at correct answers, 30.8% of total strategies in Form 2, 33% in Form 3 and 28.3% in Form 4. The second most frequently used strategy was strategy No. 7 (*use of vocabulary knowledge*), which accounted for 25.3% of total effective strategies used in Form 1, 20.6% in Form 2 , 24% in Form 3 and 28.3% in Form 4. Strategy No. 9, global and No. 9 local were the next most frequently used ones; participants in Forms 2- 4 *read the text more globally* while those in Form 1 approached the text more *locally.* A few cases of strategy No. 13 (*collocation*), No. 14 (*guessing*) and No. 15 (*choice elimination*) were found in each test form.

No evidence of the use of expeditious reading (*scanning and skimming*) strategy was noticeable in the data. This was not surprising due to the high-pressure situation of the test, which caused the participants to carefully read the texts to secure scores. Moreover, in the process of test completion and the stimulated recalls, the 16 participants were informed that they could spend as much time as they needed so there was no need for them to skim and scan the texts.

Figure 6 below illustrates the strategies the participants relied on while completing cloze task in each test form.



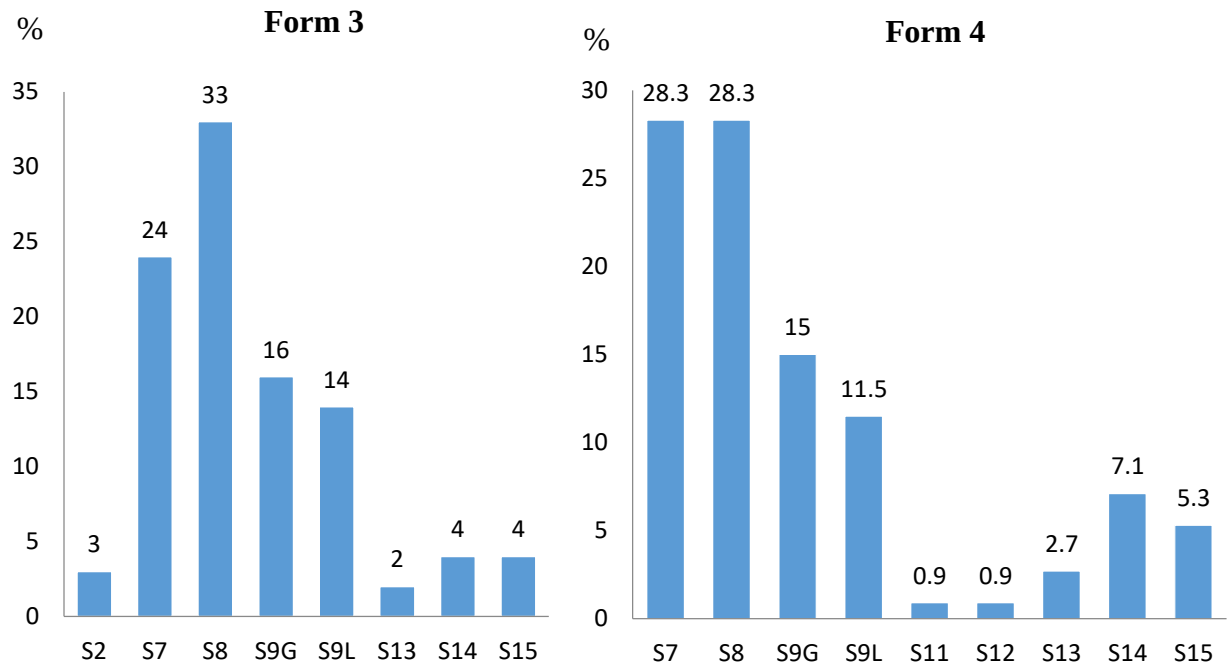


Figure 6: Cognitive strategies employed in rational cloze tasks in each test form

Again, rather similar patterns of cognitive processing found in each test form seemed to suggest that each rational cloze part tested the same construct: knowledge of grammar and vocabulary and careful reading at both global and local levels.

9.3.3 Analysis of stimulated recall data on the reading part

Table 14 below shows that the reading tasks involved a wider range of strategies.

Table 14: Frequency of occurrence of cognitive processes used during the reading part

Reading	Form 1				Form 2				Form 3				Form 4			
	Correct (n=80)		Incorrect (n=80)		Correct (n=87)		Incorrect (n=49)		Correct (n=72)		Incorrect (n=25)		Correct (n=67)		Incorrect (n=50)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
No.1	5.00	6.30	9.00	11.30	3.00	3.40	0.00	0.00	5.00	6.90	1.00	4.00	4.00	6.00	1.00	2.00
No.2	14.00	17.50	12.00	15.00	7.00	8.00	2.00	4.10	7.00	9.70	0.00	0.00	2.00	3.00	1.00	2.00
No.3	7.00	8.80	2.00	2.50	1.00	1.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.5	0.00	0.00	1.00	1.30	2.00	2.30	1.00	2.00	3.00	4.20	0.00	0.00	0.00	0.00	0.00	0.00
No.6	0.00	0.00	3.00	3.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.7	9.00	11.30	6.00	7.50	10.00	11.50	8.00	16.30	6.00	8.30	2.00	8.00	6.00	9.00	6.00	12.00
No.8	2.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.40	0.00	0.00	0.00	0.00	1.00	2.00
No.9 (G)	18.00	22.50	12.00	15.00	42.00	48.30	18.00	36.70	34.00	47.20	9.00	36.00	31.00	46.30	25.00	50.00
No.9 (L)	12.00	15.00	6.00	7.50	0.00	0.00	1.00	2.00	6.00	8.30	3.00	12.00	9.00	13.40	0.00	0.00
No.10 (G)	0.00	0.00	0.00	0.00	3.00	3.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.10 (L)	0.00	0.00	1.00	1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.11	0.00	0.00	1.00	1.30	2.00	2.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.12	3.00	3.80	7.00	8.80	1.00	1.10	1.00	2.00	1.00	1.40	0.00	0.00	0.00	0.00	0.00	0.00
No.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No.14	3.00	3.80	9.00	11.30	7.00	8.00	11.00	22.40	1.00	1.40	4.00	16.00	6.00	9.00	6.00	12.00
No.15	7.00	8.80	11.00	13.80	9.00	10.30	7.00	14.30	8.00	11.10	6.00	24.00	9.00	13.40	10.00	20.00
Total	80.00	100.00	80.00	100.00	87.00	100.00	49.00	100.00	72.00	100.00	25.00	100.00	67.00	100.00	50.00	100.00

G=Global L=Local

Only one out of the 16 participants indicated that she previewed the passages to get a general idea of what it was about before attempting the questions. Overall, the five most popular strategies activated included strategy No. 9 G (*slowly and careful global reading*), with No. 15 (*choice elimination*) being second, followed by strategies No. 7 (*knowledge of vocabulary*) and No. 2 (*local - search reading based on lexical access*), with No. 9 L (*slowly and careful local reading)* coming last.

A combination of a variety of strategies helped the participants to answer the questions successfully. Of the total of 306 strategies the 16 participants used to arrive at correct answers, the participants heavily relied on strategy No. 9 global, *reading the text or part of it most often globally* (42.48% of the total strategies used). They also used their *choice elimination* (No. 15, 10.75%) and their *knowledge of vocabulary* (No. 7, 10.13%) to understand the text. Additionally, they *matched words appearing in the questions with similar or related words in the text* (No. 2, 9.78%) before deciding on the correct answers.

Evidence of other strategies activated, No. 1 and No. 14, was found but relatively less frequent. However, there was no use of strategies No. 4, 6, 10L and 13. Figures 7 below presents an overview of strategies used.

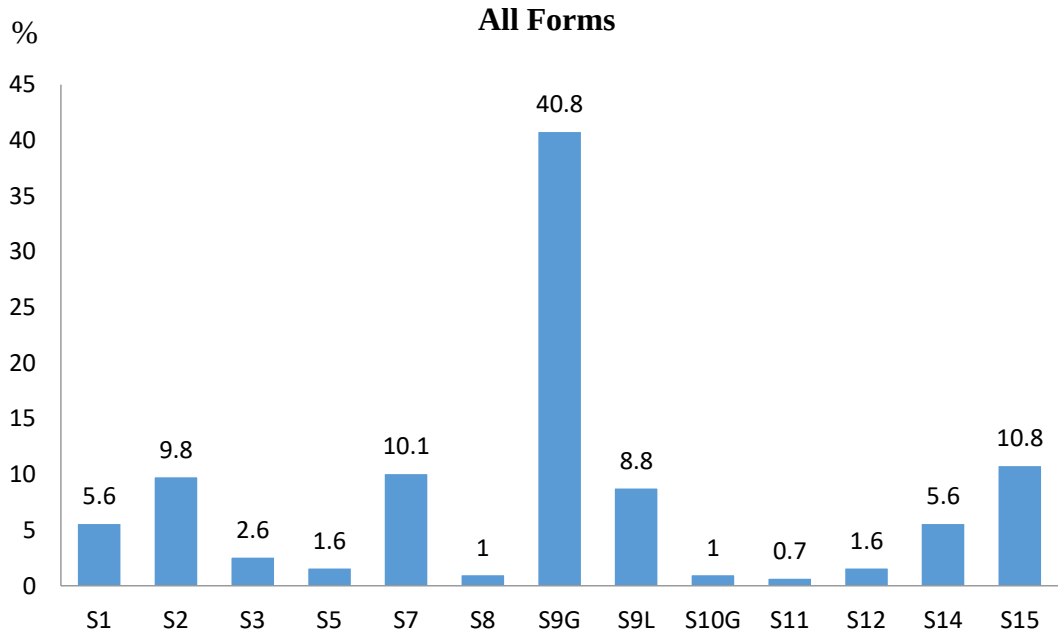


Figure 7: Cognitive strategies employed in reading tasks across four test forms

Detailed analyses into stimulated recall data in each test form showed a wider variety of strategies used. Form 4 possessed the least types of strategies employed by the participants to correctly answer the questions. Strategy No. 9G (*careful global reading*) was the most popular one; participants in each test form reported *reading the text or part of it slowly and carefully beyond sentence level*. Of all 80 strategies used to get correct answers in Form 1, the top three strategies were No. 9 (*global*, 22.5%), No. 2 (*match words in question with related words in the text*, 17.50% of) and No. 9 (*careful local reading*, 15%). In Form 2, of all 87 strategies used for correct answers the participants mainly relied on No. 9 (*global*, 48.3%), No. 7 (*vocabulary knowledge*, 11.5%) and No. 15 (*choice elimination*, 10.30%) for correct answers. Nearly half of 72 and 67 successful strategies participants in Form 3 and Form 4 used were No. 9: *global* (47.2% and 46.3%, respectively) with 8.3% and 13.4% of *local reading*, as the second most popular strategies, respectively.

Other popular strategies across the four forms included No.2 (*matching words appear in questions with similar or related words in the text*), No. 14 (*guess*) and No. 15 (*choice elimination*). Certain strategies such as strategies No 4, No. 6, No.10L and No.13 were not activated at all across the four forms with very few uses of strategies Nos. 3, 8, 10G, 11, 12, and 14.

As previously discussed, a combination of strategies was used before the participants arrived at a correct answer. So it should be noted that although some participants who explicitly stated to guess to complete the item, they ended up guessing after the use of other strategies to narrow down their choices.

A broadly similar pattern of strategy use across the four different test forms might suggest that reading passages in the four different forms tested the same abilities. Figure 8 below displays the frequency of strategy usage in all four forms.

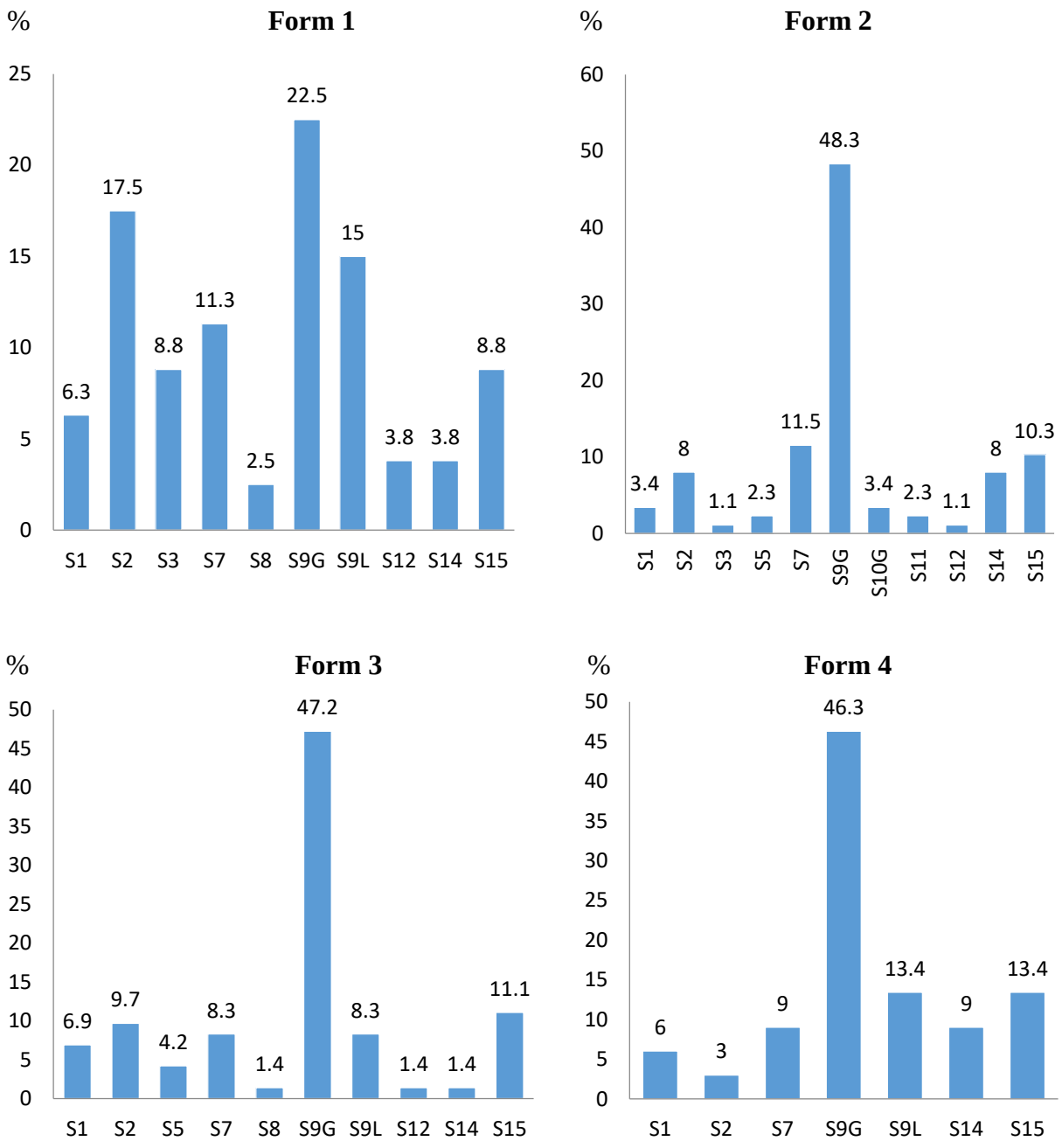


Figure 8: Cognitive strategies employed in reading tasks in each test form

10. Conclusion

One major concern of the use of the PSU-TEP Structure and Reading Test is whether it tests what it is supposed to test. Thus, the primary aim of this study was to explore the construct measured by the different parts of the PSU-TEP Structure and Reading Test through various means: analyses of test scores and test items, factor analysis and the test-takers' cognitive processing while responding to the test tasks. The findings of this study were as follows.

1. Statistical analyses performed on the test scores obtained by 941 test-takers sitting for the four test forms administered in February, April, June and December, 2016 suggested that the four test forms were reliable tests (Cronbach's Alpha, rtt = .80, .75, .84 and .66, respectively). This was a preliminary indication that the PSU-TEP Structure and Reading Test was a valid test because reliability is a necessary characteristic of a valid test (Heaton, 1988).

In terms of test difficulty, it was found that Form 4 was more difficult than the other three forms. The mean scores of Forms 1- 4 were 24.14 (SD = 7.77), 24.48 (SD = 7.11), 27.00 (SD = 8.97) and 21.22 (SD = 6.45), respectively from a total score of 60. Analyses of test items showed that each form consisted of a relatively large number of items with acceptable difficulty and discrimination indexes. Through distractor analyses, 79.03% of the distractors in Forms 1, 2, 3 and 4, were found to be efficient in distracting the less proficient test-takers.

2. Insights were gained through the detailed analysis of stimulated recall data produced by the 16 participants who were purposively selected to participate in stimulated recall protocol. Several cognitive processes revealed may reflect what each part of the test was supposed to really measure.

In error recognition completion, the use of *grammar knowledge* emerged as the only significant strategy used to help participants correctly answer the questions across the four forms of the test. The finding seemed to confirm that the error recognition items actually tested what they were supposed to test, i.e. the language knowledge of the test-taker. In other words, data from the stimulated recalls showed that the construct of the error recognition is the knowledge of grammar. Also, the similar pattern of strategy use for error recognition items across the four test forms may suggest that error recognition parts in the four tests are parallel.

The analyses of strategies used in the rational cloze items showed strategies No. 7 (*use knowledge of vocabulary*), No. 8 (*knowledge of grammar*), No. 9 global, and No. 9 local were the four most frequently-used strategies. These strategies are specified in Khalifa and Weir's (2009) model of cognitive processing in reading, appearing in all the four test forms. This association seemed to suggest that the rational cloze tasks tapped into the construct of vocabulary knowledge, grammar knowledge and careful reading at both global and local levels. The participants' lexical and syntactic contributory skills were efficient in helping them to carefully read and comprehend the texts to the extent that they could successfully complete the gaps. There is growing evidence that lexical and syntactic knowledge is an efficient indicators of careful reading ability (Weir, 2005).

A broadly consistent pattern of strategy use across four test forms suggests that, to some extent, the rational cloze tasks in each test form measured relatively the same construct.

In the reading part of the PSU-TEP Structure and Reading Test, different strategies were found in helping the participants arrive at correct answers, depending on types of information necessary for correct items. Obviously, *slow and careful reading at both global and local levels* (strategy No. 9*)* and *vocabulary knowledge* (strategy No.7) appeared to have been most significantly beneficial in helping participants arrive at correct answers. It might be possible to conclude that the construct being measured in the reading part of the PSU-TEP Structure and Reading Test were vocabulary knowledge and the ability to read carefully to comprehend sentences, main ideas and overall texts.

A relatively consistent pattern of strategy use was found in the four test forms, with *slow and careful global reading* emerging as the most popular strategy and *vocabulary knowledge, search reading based on lexical access* being among the most popular ones, indicating that the four test forms measured relatively the same construct.

The types of strategies as reported using by the participants, except for *guessing* and *choice elimination*, were in the range of reading strategies proposed in Khalifa and Weir's (2009) model of cognitive processing in reading. The absence of expeditious reading such as *scanning and skimming* was noted in this study. This might be explained by the fact that, while completing

the test and afterwards participating in the stimulated recalls, the participants were informed that they could spend as much time as they desired to complete each task before reporting their cognitive processes they used in task completion. This might have led to the absence of expeditious reading in the present study.

In conclusion, to answer the research questions dealing with the cognitive processes the participants employed while completing the test tasks and as to whether the PSU-TEP Structure and Reading Test measured what it was supposed to measure, it was found that the error recognition part did. The same cases were true for the rationalcloze part and reading parts; the two parts were demonstrated to measure what they were supposed to measure, to a great extent. However, the fact that the cognitive processes the participants employed for correct responses in the rational cloze part and the reading part were the so-called lower-level processing ones, it might be therefore possible to conclude that these two parts of PSU-TEP Reading and Structure Test managed to tap mostly lower-level processing, i.e. grammar and vocabulary knowledge and careful local reading and only one higher-level processing (careful global reading). The missing of expeditious reading (scanning and skimming) was noted.

It is arguable what is being tested in a cloze test. It might be a test of general proficiency rather than a test of reading. Evidence has been found that the test format measured a limited part of reading proficiency (Weir, 2005). For example, Kintsch and Yarbrough (1982, cited in Weir, 2013) suggested that cloze tests were not sensitive to macro processes but related to micro processes. Markham's (1985, cited in Weir, 2013) also showed that cloze tests did not require inter-sentential comprehension and that the tests do not assess global comprehension. In Kobayashi's (1995, cited in Weir 2013), cloze tests were found to measure local comprehension. This is backed up by Weir (2013: 160), who claimed that cloze tests do not reflect the reader's ability to comprehend beyond the sentence. Cloze tests appear to measure only a limited part of reading proficiency in terms of the cognitive processes presented in Khalifa and Weir's (2009) model: lexical access, syntactic parsing skills and establishing propositional level meaning. The tests reflect the processes involved in careful local reading to establish comprehension in the sentence level.

11. Recommendation for further studies

A few recommendations are suggested for researchers and test-developers at the faculty of Liberal Arts, Prince of Songkla University.

1. The use of stimulated recall methodology is recommended for test validation research. It was proved to be useful in this study in revealing the cognitive processes the participants employed to arrive at correct answers although the methodology was quite laborious. It could help test developers develop test items that reflect overall test construct and avoid testing irrelevant elements.

2. The limited range of strategies as proposed by Khalifa and Weir's (2009) model of cognitive processing in reading was found in this current study because the participants mainly approached the reading tasks, both the rational cloze items and reading passages with careful global and local reading, with the absence of expeditious reading (scanning and skimming). This is not consistent with the general approach to academic reading construct as reported in Weir et al.'s (2012b): *quick and selective search reading* followed by *intensive careful reading of relevant parts*. It is therefore recommended that the PSU-TEP test developers should include a variety of task types that requires both expeditious and careful reading with both global and local information processing under time limit.

3. Further studies are needed to look into the statistics aspects of the test and the cognitive processing of test-takers while completing the test. A group of 100 test-takers with varying proficiency levels should be recruited to take all the four test forms to establish its reliability and test difficulty. Another extensive study of the cognitive processing of test-takers of PSU-TEP Structure and Reading Test is called for, using a larger number of proficient test-takers.

List of References

Alderson, J. C. , Claphan, C. & Wall, D. (1995). *Language test construction and evaluation.* Cambridge: Cambridge University Press.

Anderson, N. J., Bachman, L., 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.

Bachman, L. F. (1990*). Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, *30*(4), 441-465.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.

Brown, J. D. (2005). *Testing in Language Programs: A Comprehensive Guide to English Language Assessment*. New York, NY: McGraw Hill

Brunfaut, T. & McCray, G. (2015). *Looking into test-takers' cognitive processing while completing reading tasks: A mixed-method eye-tracking and stimulated recall study*. British Council English Language Assessment Research Group.

Cohen, A. D. (2006). The Coming of Age of Research on Test-Taking Strategies, *Language Assessment Quarterly*, *3*(4) 307-331.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. M. & Mackey, A. (2017). *Stimulated recall methodology in Applied linguistic and L2 research*. New York: Routledge.

Green, A. (1998). *Verbal protocol analysis in language testing research: a handbook*. Studies in Language Testing volume 5, Cambridge: Cambridge University Press

Heaton, J. B. (1988). *Writing English Language Tests.* Longman: New York.

Heaton, J. B. (1997). *Classroom testing.* London: Longman.

Henning, G. (1987). *A guide to language testing*. Los Angeles, C.A.: Newbury House.

Hughes, A. (1989). *Testing for language teachers*. Cambridge, England: Cambridge University Press.

Khalifa, H. & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading*, Studies in Language Testing volume 29, Cambridge: UCLES/Cambridge University Press.

McNamara, T. F. (2000). *Language testing*. Oxford: Oxford University Press.

Messick, S. (1989). Meanig and Values in Test Validation: The Science and Ethics of Assessment. *American Educational Research Association, 18*(5), 5-11

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective. *Language testing, 23*(4), 441-474.

Storey, P. (1997). Examining the test-taking process: a cognitive perspective on the discourse close test. *Language Testing, 14*(2), 214-231.

Weir, C. J. (2005). *Language testing and validation*: *An evidence-based approach*, Basingstoke: Palgrave Macmillan

Weir, C. J. (2013). *The measurement of reading ability 1913-2012*, Studies in Language Testing volume 37, Cambridge University Press

Weir, C. J., Hawkey, R., Green, A. & Devi, S. (2012 a). *The cognitive processes underlying the academic reading as measured by IELTS the academic reading construct as measured by IELTS and the reading experiences of students in their first year of study at a British university*, Studies in Language Testing volume 34, Cambridge: UCLES/Cambridge University Press

Weir, C. J., Hawkey, R., Green, A., Unaldi, A. & Devi, S. (2012 b). *The relationship between the academic reading construct as measured by IELTS*, Studies in Language Testing volume 34, Cambridge: UCLES/Cambridge University Press

Appendix A

Item facility, item discrimination, distractor analysis of the four test forms

Form 1							
Item No.	IF	ID	Group	Options			
				A	B	C	D
1	0.620	0.35	High	0.06	**0.85**	0.01	0.08
			Middle	0.04	**0.60**	0.07	0.29
			Low	0.13	**0.41**	0.18	0.28
2	0.540	0.25	High	0.00	0.15	**0.77**	0.08
			Middle	0.04	0.31	**0.46**	0.20
			Low	0.10	0.25	**0.42**	0.23
3	0.230	0.02	High	**0.26**	0.04	0.42	0.28
			Middle	**0.21**	0.07	0.47	0.24
			Low	**0.23**	0.11	0.35	0.31
4	0.450	0.26	High	0.27	**0.60**	0.03	0.10
			Middle	0.18	**0.50**	0.08	0.24
			Low	0.15	**0.23**	0.24	0.38
5	0.320	0.33	High	**0.55**	0.05	0.17	0.23
			Middle	**0.27**	0.17	0.22	0.34
			Low	**0.15**	0.23	0.24	0.38
6	0.400	0.18	High	0.09	0.10	**0.54**	0.27
			Middle	0.13	0.21	**0.38**	0.28
			Low	0.17	0.25	**0.28**	0.30
7	0.430	0.29	High	0.08	0.01	**0.63**	0.28
			Middle	0.09	0.11	**0.44**	0.36
			Low	0.23	0.14	**0.21**	0.42
8	0.500	0.28	High	**0.73**	0.10	0.12	0.05
			Middle	**0.44**	0.20	0.21	0.15
			Low	**0.32**	0.25	0.30	0.13
9	0.310	0.29	High	**0.54**	0.14	0.23	0.09
			Middle	**0.21**	0.25	0.27	0.26
			Low	**0.20**	0.35	0.23	0.23
10	0.440	0.30	High	0.17	0.06	**0.65**	0.12
			Middle	0.24	0.19	**0.43**	0.14
			Low	0.18	0.32	**0.23**	0.27
11	0.440	0.12	High	**0.55**	0.41	0.03	0.01
			Middle	**0.41**	0.40	0.11	0.07
			Low	**0.35**	0.27	0.13	0.25
12	0.110	0.10	High	0.24	0.31	0.27	**0.18**
			Middle	0.24	0.24	0.46	**0.06**
			Low	0.41	0.21	0.27	**0.11**

Form 1							
Item No.	IF	ID	Group	Options			
				A	B	C	D
13	0.320	0.32	High	0.32	**0.54**	0.13	0.01
			Middle	0.47	**0.26**	0.20	0.07
			Low	0.41	**0.18**	0.24	0.17
14	0.250	0.11	High	0.27	**0.36**	0.17	0.21
			Middle	0.27	**0.22**	0.38	0.12
			Low	0.27	**0.18**	0.35	0.20
15	0.270	0.17	High	**0.36**	0.09	0.23	0.32
			Middle	**0.27**	0.01	0.35	0.37
			Low	**0.17**	0.14	0.32	0.37
16	0.390	0.35	High	0.15	0.03	**0.64**	0.18
			Middle	0.26	0.07	**0.36**	0.29
			Low	0.37	0.15	**0.15**	0.32
17	0.600	0.34	High	0.10	**0.82**	0.03	0.05
			Middle	0.13	**0.57**	0.19	0.10
			Low	0.23	**0.39**	0.25	0.13
18	0.540	0.17	High	**0.68**	0.31	0.00	0.01
			Middle	**0.57**	0.35	0.07	0.01
			Low	**0.35**	0.30	0.23	0.13
19	0.230	-0.03	High	**0.23**	0.35	0.35	0.08
			Middle	**0.21**	0.34	0.27	0.18
			Low	**0.27**	0.23	0.31	0.20
20	0.410	0.09	High	0.15	0.21	0.13	**0.51**
			Middle	0.15	0.36	0.08	**0.41**
			Low	0.15	0.44	0.10	**0.31**
21	0.410	0.21	High	0.10	**0.62**	0.05	0.23
			Middle	0.21	**0.34**	0.21	0.24
			Low	0.23	**0.31**	0.31	0.15
22	0.340	0.06	High	0.22	0.15	**0.36**	0.27
			Middle	0.09	0.24	**0.39**	0.26
			Low	0.21	0.24	**0.23**	0.32
23	0.600	0.37	High	0.01	0.14	**0.83**	0.01
			Middle	0.03	0.28	**0.62**	0.07
			Low	0.00	0.44	**0.32**	0.24
24	0.130	-0.01	High	0.58	0.23	0.04	**0.15**
			Middle	0.43	0.35	0.13	**0.09**
			Low	0.44	0.20	0.23	**0.14**

Form 1							
Item No.	IF	ID	Group	Options			
				A	B	C	D
25	0.460	0.32	High	0.10	**0.68**	0.04	0.18
			Middle	0.27	**0.43**	0.09	0.21
			Low	0.28	**0.25**	0.13	0.34
26	0.180	0.46	High	0.01	**0.44**	0.35	0.21
			Middle	0.10	**0.09**	0.66	0.13
			Low	0.21	**0.01**	0.54	0.24
27	0.830	0.27	High	**0.95**	0.00	0.05	0.00
			Middle	**0.84**	0.03	0.10	0.03
			Low	**0.68**	0.11	0.17	0.04
28	0.160	0.01	High	0.06	0.08	**0.17**	0.69
			Middle	0.13	0.18	**0.19**	0.50
			Low	0.21	0.21	**0.10**	0.48
29	0.550	0.37	High	0.00	0.08	**0.87**	0.05
			Middle	0.05	0.29	**0.46**	0.21
			Low	0.17	0.34	**0.35**	0.14
30	0.190	0.08	High	0.21	0.14	0.38	**0.27**
			Middle	0.32	0.13	0.36	**0.20**
			Low	0.34	0.10	0.46	**0.10**
31	0.290	0.30	High	**0.50**	0.29	0.05	0.15
			Middle	**0.23**	0.30	0.07	0.38
			Low	**0.13**	0.24	0.10	0.54
32	0.490	0.35	High	0.03	**0.78**	0.17	0.03
			Middle	0.15	**0.43**	0.31	0.11
			Low	0.13	**0.25**	0.31	0.31
33	0.650	0.45	High	0.01	0.03	0.03	**0.94**
			Middle	0.13	0.06	0.16	**0.65**
			Low	0.31	0.13	0.23	**0.34**
34	0.270	0.46	High	0.04	0.21	**0.59**	0.17
			Middle	0.10	0.58	**0.19**	0.13
			Low	0.15	0.62	**0.04**	0.18
35	0.320	0.18	High	0.00	**0.49**	0.08	0.44
			Middle	0.02	**0.22**	0.07	0.68
			Low	0.10	**0.27**	0.11	0.52
36	0.240	0.04	High	**0.28**	0.06	0.26	0.40
			Middle	**0.25**	0.18	0.28	0.29
			Low	**0.18**	0.21	0.28	0.32

Form 1							
Item No.	IF	ID	Group	Options			
				A	B	C	D
37	0.520	0.27	High	0.18	0.06	0.04	**0.72**
			Middle	0.35	0.12	0.02	**0.51**
			Low	0.46	0.18	0.04	**0.31**
38	0.630	0.31	High	0.10	0.01	**0.82**	0.06
			Middle	0.19	0.01	**0.67**	0.13
			Low	0.27	0.10	**0.37**	0.27
39	0.390	0.36	High	**0.65**	0.03	0.00	0.32
			Middle	**0.35**	0.07	0.18	0.40
			Low	**0.17**	0.14	0.38	0.31
40	0.640	0.38	High	0.08	**0.88**	0.01	0.03
			Middle	0.22	**0.64**	0.04	0.09
			Low	0.37	**0.38**	0.04	0.21
41	0.290	0.14	High	**0.40**	0.22	0.06	0.32
			Middle	**0.26**	0.20	0.07	0.47
			Low	**0.20**	0.24	0.14	0.42
42	0.580	0.14	High	0.29	**0.68**	0.01	0.01
			Middle	0.30	**0.61**	0.02	0.07
			Low	0.15	**0.42**	0.21	0.21
43	0.210	0.27	High	**0.36**	0.01	0.00	0.63
			Middle	**0.22**	0.03	0.00	0.75
			Low	**0.03**	0.07	0.03	0.87
44	0.870	0.38	High	**0.97**	0.01	0.01	0.00
			Middle	**0.95**	0.01	0.01	0.03
			Low	**0.62**	0.14	0.04	0.20
45	0.150	-0.06	High	0.03	0.79	0.03	**0.15**
			Middle	0.19	0.57	0.07	**0.18**
			Low	0.37	0.41	0.13	**0.10**
46	0.410	0.36	High	0.17	**0.64**	0.06	0.13
			Middle	0.14	**0.40**	0.21	0.25
			Low	0.17	**0.17**	0.35	0.31
47	0.280	0.03	High	0.01	0.44	**0.33**	0.22
			Middle	0.16	0.39	**0.23**	0.21
			Low	0.24	0.15	**0.28**	0.32
48	0.720	0.47	High	0.01	0.01	0.00	**0.97**
			Middle	0.19	0.03	0.01	**0.78**
			Low	0.45	0.13	0.06	**0.37**

Form 1							
Item No.	IF	ID	Group	Options			
				A	B	C	D
49	0.110	-0.06	High	**0.08**	0.27	0.58	0.08
			Middle	**0.13**	0.19	0.55	0.13
			Low	**0.11**	0.17	0.39	0.32
50	0.450	0.38	High	0.05	0.27	**0.68**	0.00
			Middle	0.10	0.38	**0.49**	0.03
			Low	0.25	0.37	**0.15**	0.23
51	0.440	0.30	High	0.13	0.00	**0.67**	0.21
			Middle	0.20	0.10	**0.41**	0.28
			Low	0.20	0.18	**0.23**	0.39
52	0.590	0.42	High	0.01	**0.87**	0.04	0.08
			Middle	0.10	**0.56**	0.12	0.21
			Low	0.13	**0.32**	0.25	0.30
53	0.220	-0.06	High	0.40	**0.21**	0.22	0.18
			Middle	0.22	**0.23**	0.22	0.32
			Low	0.18	**0.23**	0.35	0.24
54	0.260	0.16	High	0.23	0.14	**0.36**	0.27
			Middle	0.21	0.30	**0.21**	0.28
			Low	0.08	0.46	**0.24**	0.21
55	0.580	0.36	High	0.17	**0.81**	0.01	0.01
			Middle	0.31	**0.60**	0.05	0.05
			Low	0.39	**0.30**	0.10	0.21
56	0.200	0.10	High	0.06	0.45	0.23	**0.26**
			Middle	0.30	0.29	0.19	**0.22**
			Low	0.58	0.18	0.13	**0.11**
57	0.410	0.19	High	0.10	0.13	**0.55**	0.22
			Middle	0.18	0.15	**0.41**	0.25
			Low	0.21	0.28	**0.25**	0.25
58	0.500	0.47	High	0.03	0.17	**0.81**	0.00
			Middle	0.12	0.31	**0.50**	0.06
			Low	0.17	0.49	**0.17**	0.17
59	0.380	-0.20	High	0.51	0.01	0.15	**0.32**
			Middle	0.21	0.05	0.35	**0.38**
			Low	0.08	0.06	0.41	**0.45**
60	0.380	-0.04	High	**0.42**	0.09	0.32	0.17
			Middle	**0.33**	0.12	0.32	0.22
			Low	**0.42**	0.14	0.31	0.13

Form 2							
Item No.	IF	ID	Group	Options			
				A	B	C	D
1	0.406	0.10	High	0.12	**0.58**	0.12	0.19
			Middle	0.17	**0.32**	0.31	0.20
			Low	0.23	**0.38**	0.27	0.12
2	0.438	0.28	High	0.08	0.13	0.10	**0.69**
			Middle	0.32	0.20	0.08	**0.40**
			Low	0.35	0.23	0.17	**0.25**
3	0.422	0.33	High	0.08	0.21	0.02	**0.69**
			Middle	0.13	0.40	0.10	**0.38**
			Low	0.15	0.50	0.12	**0.23**
4	0.641	0.23	High	**0.81**	0.06	0.12	0.02
			Middle	**0.64**	0.14	0.16	0.07
			Low	**0.48**	0.23	0.23	0.06
5	0.594	0.33	High	0.08	0.10	0.00	**0.83**
			Middle	0.20	0.14	0.03	**0.63**
			Low	0.21	0.29	0.19	**0.31**
6	0.125	-0.14	High	0.21	**0.06**	0.10	0.63
			Middle	0.17	**0.15**	0.10	0.58
			Low	0.19	**0.15**	0.19	0.46
7	0.448	0.22	High	0.06	0.08	**0.69**	0.17
			Middle	0.23	0.05	**0.39**	0.34
			Low	0.33	0.19	**0.31**	0.17
8	0.313	0.23	High	0.10	0.04	**0.52**	0.35
			Middle	0.03	0.17	**0.25**	0.55
			Low	0.19	0.12	**0.21**	0.48
9	0.474	0.13	High	0.21	0.15	**0.63**	0.00
			Middle	0.40	0.13	**0.43**	0.05
			Low	0.31	0.23	**0.38**	0.08
10	0.380	0.07	High	0.04	0.29	0.12	**0.56**
			Middle	0.15	0.20	0.38	**0.27**
			Low	0.12	0.13	0.37	**0.38**
11	0.229	0.18	High	0.25	0.13	**0.38**	0.23
			Middle	0.38	0.19	**0.18**	0.25
			Low	0.37	0.29	**0.15**	0.19
12	0.323	0.33	High	**0.56**	0.02	0.40	0.02

Form 2							
Item No.	IF	ID	Group	Options			
				A	B	C	D
			Middle	**0.32**	0.16	0.51	0.01
			Low	**0.10**	0.31	0.50	0.10
13	0.344	0.22	High	**0.54**	0.21	0.15	0.10
			Middle	**0.30**	0.31	0.15	0.25
			Low	**0.23**	0.29	0.23	0.25
14	0.557	0.24	High	**0.73**	0.02	0.12	0.13
			Middle	**0.57**	0.07	0.10	0.25
			Low	**0.37**	0.06	0.27	0.31
15	0.495	0.19	High	0.04	0.21	0.13	**0.62**
			Middle	0.03	0.25	0.19	**0.52**
			Low	0.13	0.44	0.10	**0.33**
16	0.162	0.05	High	0.27	0.44	**0.19**	0.10
			Middle	0.24	0.48	**0.13**	0.16
			Low	0.21	0.33	**0.19**	0.27
17	0.813	0.32	High	0.02	0.00	0.00	**0.98**
			Middle	0.02	0.09	0.02	**0.86**
			Low	0.19	0.13	0.12	**0.56**
18	0.646	0.29	High	**0.79**	0.17	0.04	0.00
			Middle	**0.73**	0.16	0.05	0.07
			Low	**0.37**	0.31	0.13	0.19
19	0.052	0.00	High	0.62	**0.08**	0.06	0.25
			Middle	0.48	**0.03**	0.28	0.20
			Low	0.37	**0.06**	0.44	0.13
20	0.500	0.19	High	0.08	0.15	0.10	**0.67**
			Middle	0.11	0.11	0.31	**0.47**
			Low	0.23	0.29	0.10	**0.38**
21	0.307	0.24	High	0.31	0.15	**0.54**	0.00
			Middle	0.47	0.30	**0.24**	0.00
			Low	0.40	0.29	**0.19**	0.12
22	0.516	0.21	High	0.25	**0.67**	0.02	0.06
			Middle	0.22	**0.49**	0.14	0.16
			Low	0.27	**0.40**	0.12	0.21
23	0.516	0.23	High	**0.79**	0.02	0.10	0.10
			Middle	**0.39**	0.11	0.33	0.17
			Low	**0.46**	0.06	0.27	0.21

Form 2							
Item No.	IF	ID	Group	Options			
				A	B	C	D
24	0.396	0.24	High	0.02	0.12	0.27	**0.60**
			Middle	0.11	0.08	0.44	**0.36**
			Low	0.25	0.13	0.37	**0.25**
25	0.495	-0.13	High	0.12	0.42	**0.40**	0.06
			Middle	0.16	0.57	**0.23**	0.05
			Low	0.06	0.44	**0.37**	0.12
26	0.443	0.26	High	**0.67**	0.10	0.10	0.13
			Middle	**0.43**	0.09	0.18	0.30
			Low	**0.23**	0.29	0.27	0.21
27	0.177	0.00	High	0.38	0.42	**0.13**	0.06
			Middle	0.41	0.30	**0.24**	0.06
			Low	0.44	0.33	**0.12**	0.12
28	0.344	0.18	High	0.08	0.06	0.40	**0.46**
			Middle	0.14	0.27	0.20	**0.39**
			Low	0.38	0.27	0.19	**0.15**
29	0.505	0.11	High	**0.60**	0.19	0.08	0.13
			Middle	**0.49**	0.26	0.08	0.17
			Low	**0.44**	0.29	0.25	0.02
30	0.427	0.33	High	0.08	**0.77**	0.08	0.08
			Middle	0.16	**0.34**	0.17	0.33
			Low	0.13	**0.23**	0.33	0.31
31	0.568	0.35	High	0.02	0.06	0.13	**0.79**
			Middle	0.13	0.09	0.22	**0.57**
			Low	0.12	0.33	0.21	**0.35**
32	0.406	0.14	High	0.06	0.10	**0.56**	0.29
			Middle	0.20	0.19	**0.36**	0.24
			Low	0.19	0.19	**0.33**	0.29
33	0.344	-0.08	High	**0.29**	0.40	0.21	0.10
			Middle	**0.39**	0.32	0.25	0.05
			Low	**0.33**	0.23	0.31	0.13
34	0.672	0.39	High	0.02	**0.90**	0.00	0.08
			Middle	0.08	**0.72**	0.02	0.17
			Low	0.17	**0.37**	0.40	0.06
35	0.677	0.35	High	0.02	**0.92**	0.06	0.00
			Middle	0.08	**0.68**	0.14	0.10

Form 2							
Item No.	IF	ID	Group	Options			
				A	B	C	D
			Low	0.17	**0.42**	0.31	0.10
36	0.505	0.26	High	0.06	0.10	0.17	**0.67**
			Middle	0.14	0.09	0.24	**0.53**
			Low	0.21	0.17	0.33	**0.29**
37	0.375	0.20	High	0.12	0.29	**0.60**	0.00
			Middle	0.22	0.45	**0.32**	0.01
			Low	0.35	0.25	**0.25**	0.15
38	0.557	0.44	High	0.02	0.08	0.04	**0.87**
			Middle	0.18	0.17	0.09	**0.56**
			Low	0.23	0.31	0.21	**0.25**
39	0.464	0.42	High	**0.77**	0.10	0.00	0.13
			Middle	**0.49**	0.14	0.10	0.27
			Low	**0.12**	0.31	0.19	0.38
40	0.583	0.31	High	0.02	**0.81**	0.06	0.12
			Middle	0.19	**0.58**	0.08	0.15
			Low	0.27	**0.37**	0.10	0.27
41	0.287	0.21	High	0.13	0.23	0.17	**0.46**
			Middle	0.22	0.38	0.16	**0.25**
			Low	0.31	0.33	0.19	**0.17**
42	0.370	0.14	High	**0.46**	0.08	0.31	0.15
			Middle	**0.39**	0.07	0.36	0.18
			Low	**0.25**	0.17	0.40	0.17
43	0.396	0.45	High	**0.75**	0.08	0.08	0.10
			Middle	**0.35**	0.22	0.26	0.17
			Low	**0.12**	0.17	0.62	0.10
44	0.391	0.31	High	0.15	0.23	0.02	**0.60**
			Middle	0.17	0.32	0.16	**0.35**
			Low	0.13	0.33	0.29	**0.25**
45	0.193	0.17	High	**0.29**	0.23	0.13	0.35
			Middle	**0.18**	0.33	0.33	0.16
			Low	**0.12**	0.44	0.31	0.13
46	0.307	-0.06	High	0.25	0.38	0.10	**0.27**
			Middle	0.20	0.28	0.17	**0.34**
			Low	0.25	0.19	0.27	**0.29**
47	0.266	0.18	High	0.08	**0.44**	0.08	0.40

Form 2							
Item No.	IF	ID	Group	Options			
				A	B	C	D
			Middle	0.23	**0.22**	0.11	0.44
			Low	0.23	**0.17**	0.13	0.46
48	0.198	0.06	High	0.25	0.46	**0.25**	0.04
			Middle	0.22	0.47	**0.22**	0.10
			Low	0.35	0.37	**0.12**	0.17
49	0.458	0.39	High	0.12	0.08	**0.77**	0.04
			Middle	0.23	0.22	**0.41**	0.15
			Low	0.25	0.44	**0.23**	0.08
50	0.516	-0.05	High	**0.50**	0.06	0.21	0.23
			Middle	**0.56**	0.10	0.18	0.16
			Low	**0.46**	0.13	0.15	0.25
51	0.255	0.32	High	0.10	**0.50**	0.19	0.21
			Middle	0.20	**0.20**	0.27	0.32
			Low	0.17	**0.10**	0.52	0.21
52	0.302	0.02	High	0.33	0.29	**0.35**	0.04
			Middle	0.35	0.27	**0.28**	0.09
			Low	0.38	0.27	**0.29**	0.06
53	0.479	0.06	High	0.21	0.04	**0.60**	0.15
			Middle	0.33	0.02	**0.45**	0.19
			Low	0.31	0.10	**0.40**	0.19
54	0.188	0.11	High	**0.29**	0.08	0.10	0.54
			Middle	**0.18**	0.10	0.28	0.43
			Low	**0.10**	0.19	0.21	0.50
55	0.089	0.14	High	**0.19**	0.38	0.27	0.15
			Middle	**0.02**	0.45	0.32	0.20
			Low	**0.10**	0.33	0.35	0.23
56	0.568	0.19	High	0.04	0.04	**0.79**	0.13
			Middle	0.24	0.10	**0.50**	0.16
			Low	0.29	0.15	**0.46**	0.10
57	0.318	0.07	High	0.17	**0.37**	0.31	0.15
			Middle	0.23	**0.33**	0.30	0.15
			Low	0.27	**0.25**	0.33	0.15
58	0.380	0.26	High	**0.62**	0.21	0.13	0.04
			Middle	**0.34**	0.20	0.19	0.26
			Low	**0.21**	0.27	0.12	0.40

Form 2							
Item No.	IF	ID	Group	Options			
				A	B	C	D
59	0.484	0.05	High	0.02	0.35	0.12	**0.52**
			Middle	0.10	0.28	0.13	**0.49**
			Low	0.15	0.27	0.13	**0.44**
60	0.375	-0.05	High	0.08	0.19	0.38	**0.35**
			Middle	0.03	0.13	0.39	**0.45**
			Low	0.17	0.15	0.38	**0.27**

Form 3							
Item No.	IF	ID	Group	Options			
				A	B	C	D
1	0.450	0.46	High	**0.74**	0.08	0.16	0.02
			Middle	**0.38**	0.21	0.25	0.15
			Low	**0.28**	0.22	0.42	0.08
2	0.310	0.42	High	**0.56**	0.20	0.20	0.04
			Middle	**0.26**	0.21	0.24	0.29
			Low	**0.14**	0.14	0.70	0.28
3	0.540	0.48	High	0.12	0.08	**0.74**	0.06
			Middle	0.17	0.11	**0.60**	0.13
			Low	0.16	0.24	**0.26**	0.34
4	0.430	0.34	High	**0.70**	0.26	0.02	0.02
			Middle	**0.33**	0.51	0.10	0.06
			Low	**0.34**	0.38	0.12	0.16
5	0.140	0.18	High	0.28	0.20	0.26	**0.24**
			Middle	0.26	0.24	0.38	**0.12**
			Low	0.34	0.32	0.28	**0.04**
6	0.360	0.32	High	0.30	**0.56**	0.06	0.06
			Middle	0.44	**0.29**	0.18	0.11
			Low	0.34	**0.26**	0.26	0.14
7	0.360	0.22	High	0.26	**0.52**	0.16	0.06
			Middle	0.18	**0.31**	0.32	0.18
			Low	0.18	**0.28**	0.32	0.24
8	0.320	0.34	High	**0.50**	0.04	0.20	0.32
			Middle	**0.30**	0.04	0.26	0.40
			Low	**0.16**	0.10	0.30	0.44

Form 3							
Item No.	IF	ID	Group	Options			
				A	B	C	D
9	0.670	0.52	High	0.04	**0.94**	0.00	0.02
			Middle	0.12	**0.63**	0.04	0.21
			Low	0.12	**0.42**	0.06	0.64
10	0.590	0.44	High	0.04	0.08	**0.82**	0.06
			Middle	0.01	0.24	**0.55**	0.20
			Low	0.30	0.36	**0.40**	0.38
11	0.310	0.10	High	**0.38**	0.06	0.24	0.32
			Middle	**0.29**	0.10	0.36	0.26
			Low	**0.28**	0.14	0.40	0.18
12	0.430	0.50	High	0.14	**0.76**	0.00	0.10
			Middle	0.26	**0.32**	0.11	0.30
			Low	0.32	**0.28**	0.18	0.22
13	0.370	0.42	High	**0.64**	0.08	0.14	0.14
			Middle	**0.30**	0.04	0.29	0.38
			Low	**0.22**	0.08	0.18	0.52
14	0.460	0.22	High	0.06	0.08	**0.64**	0.22
			Middle	0.15	0.18	**0.42**	0.25
			Low	0.12	0.24	**0.40**	0.24
15	0.520	0.40	High	0.08	0.00	**0.74**	0.18
			Middle	0.14	0.12	**0.51**	0.23
			Low	0.32	0.10	**0.34**	0.24
16	0.510	0.36	High	0.20	0.08	0.02	**0.70**
			Middle	0.21	0.27	0.01	**0.50**
			Low	0.30	0.26	0.08	**0.36**
17	0.350	0.40	High	0.18	0.24	0.08	**0.52**
			Middle	0.27	0.20	0.20	**0.32**
			Low	0.18	0.32	0.32	**0.18**
18	0.770	0.50	High	0.06	**0.92**	0.02	0.02
			Middle	0.11	**0.87**	0.00	0.02
			Low	0.40	**0.44**	0.06	0.10
19	0.650	0.52	High	0.14	0.04	**0.82**	0.02
			Middle	0.14	0.13	**0.69**	0.04
			Low	0.14	0.28	**0.36**	0.22
20	0.440	0.44	High	0.14	0.12	**0.64**	0.08
			Middle	0.20	0.10	**0.42**	0.29
			Low	0.28	0.18	**0.24**	0.30
21	0.430	0.36	High	**0.66**	0.28	0.02	0.06

Form 3							
Item No.	IF	ID	Group	Options			
				A	B	C	D
			Middle	**0.38**	0.35	0.12	0.15
			Low	**0.30**	0.14	0.24	0.32
22	0.250	0.38	High	**0.50**	0.16	0.18	0.16
			Middle	**0.20**	0.39	0.11	0.30
			Low	**0.10**	0.38	0.12	0.40
23	0.320	0.06	High	0.24	**0.30**	0.24	0.22
			Middle	0.31	**0.35**	0.08	0.26
			Low	0.36	**0.24**	0.06	0.34
24	0.510	0.38	High	0.30	0.04	**0.64**	0.02
			Middle	0.20	0.10	**0.55**	0.14
			Low	0.20	0.28	**0.26**	0.26
25	0.440	0.44	High	0.14	**0.62**	0.00	0.24
			Middle	0.30	**0.48**	0.05	0.19
			Low	0.52	**0.16**	0.08	0.22
26	0.670	0.58	High	0.02	0.04	0.02	**0.90**
			Middle	0.19	0.04	0.07	**0.70**
			Low	0.48	0.02	0.16	**0.32**
27	0.800	0.46	High	0.02	**0.96**	0.02	0.02
			Middle	0.05	**0.96**	0.08	0.01
			Low	0.14	**0.56**	0.20	0.10
28	0.520	0.42	High	0.12	0.10	**0.76**	0.06
			Middle	0.17	0.19	**0.48**	0.18
			Low	0.12	0.24	**0.34**	0.28
29	0.540	0.82	High	**0.94**	0.04	0.02	0.00
			Middle	**0.51**	0.15	0.30	0.04
			Low	**0.12**	0.44	0.16	0.26
30	0.520	0.60	High	0.08	**0.86**	0.00	0.06
			Middle	0.15	**0.45**	0.24	0.15
			Low	0.16	**0.26**	0.36	0.20
31	0.520	0.58	High	0.04	0.04	0.16	**0.76**
			Middle	0.06	0.15	0.27	**0.58**
			Low	0.34	0.06	0.44	**0.16**
32	0.650	0.22	High	0.12	0.06	0.10	**0.76**
			Middle	0.17	0.14	0.11	**0.60**
			Low	0.26	0.08	0.10	**0.56**
33	0.450	0.32	High	0.22	0.10	**0.68**	0.00
			Middle	0.17	0.32	**0.36**	0.15

Form 3							
Item No.	IF	ID	Group	Options			
				A	B	C	D
			Low	0.24	0.16	**0.36**	0.24
34	0.320	0.40	High	0.44	**0.54**	0.00	0.02
			Middle	0.61	**0.27**	0.01	0.11
			Low	0.52	**0.12**	0.18	0.18
35	0.860	0.36	High	**0.98**	0.00	0.00	0.02
			Middle	**0.93**	0.02	0.01	0.04
			Low	**0.64**	0.10	0.12	0.14
36	0.230	0.30	High	0.12	0.04	**0.42**	0.42
			Middle	0.20	0.17	**0.18**	0.49
			Low	0.10	0.32	**0.12**	0.72
37	0.210	0.52	High	**0.54**	0.24	0.18	0.04
			Middle	**0.12**	0.44	0.29	0.15
			Low	**0.02**	0.28	0.38	0.32
38	0.410	0.40	High	0.06	0.16	0.16	**0.62**
			Middle	0.27	0.24	0.10	**0.39**
			Low	0.36	0.18	0.24	**0.22**
39	0.320	0.20	High	0.40	**0.44**	0.04	0.12
			Middle	0.26	**0.30**	0.20	0.24
			Low	0.18	**0.24**	0.30	0.28
40	0.330	0.48	High	0.30	0.16	**0.50**	0.04
			Middle	0.32	0.17	**0.39**	0.12
			Low	0.46	0.12	**0.06**	0.62
41	0.490	0.26	High	**0.62**	0.02	0.14	0.22
			Middle	**0.50**	0.10	0.14	0.26
			Low	**0.34**	0.10	0.24	0.32
42	0.280	0.28	High	**0.44**	0.12	0.20	0.22
			Middle	**0.26**	0.31	0.13	0.30
			Low	**0.18**	0.34	0.22	0.26
43	0.330	0.34	High	0.24	**0.52**	0.02	0.22
			Middle	0.23	**0.31**	0.19	0.27
			Low	0.24	**0.18**	0.44	0.14
44	0.690	0.56	High	**0.94**	0.02	0.02	0.02
			Middle	**0.71**	0.10	0.14	0.05
			Low	**0.42**	0.20	0.22	0.16
45	0.550	0.62	High	0.00	0.02	0.08	**0.90**
			Middle	0.07	0.13	0.33	**0.46**
			Low	0.14	0.26	0.28	**0.32**

Form 3							
Item No.	IF	ID	Group	Options			
				A	B	C	D
46	0.700	0.42	High	0.10	0.00	**0.88**	0.02
			Middle	0.23	0.05	**0.71**	0.01
			Low	0.24	0.16	**0.52**	0.08
47	0.280	0.40	High	0.16	**0.48**	0.20	0.16
			Middle	0.17	**0.27**	0.30	0.25
			Low	0.12	**0.08**	0.42	0.38
48	0.290	0.10	High	0.04	**0.32**	0.32	0.32
			Middle	0.26	**0.32**	0.23	0.19
			Low	0.20	**0.22**	0.36	0.22
49	0.540	0.62	High	0.00	0.08	**0.88**	0.04
			Middle	0.05	0.15	**0.50**	0.30
			Low	0.30	0.16	**0.28**	0.26
50	0.240	0.22	High	0.14	0.38	0.08	**0.40**
			Middle	0.17	0.44	0.20	**0.18**
			Low	0.20	0.36	0.24	**0.20**
51	0.780	0.28	High	0.00	0.00	0.12	**0.88**
			Middle	0.11	0.02	0.06	**0.82**
			Low	0.48	0.06	0.06	**0.60**
52	0.330	0.20	High	**0.40**	0.32	0.16	0.14
			Middle	**0.36**	0.31	0.25	0.08
			Low	**0.20**	0.24	0.52	0.04
53	0.390	0.42	High	0.00	0.04	**0.62**	0.36
			Middle	0.05	0.04	**0.35**	0.57
			Low	0.12	0.14	**0.22**	0.52
54	0.570	0.40	High	**0.76**	0.10	0.12	0.00
			Middle	**0.55**	0.11	0.27	0.18
			Low	**0.38**	0.10	0.36	0.16
55	0.280	-0.10	High	0.58	**0.16**	0.14	0.12
			Middle	0.24	**0.37**	0.07	0.32
			Low	0.22	**0.26**	0.14	0.36
56	0.590	0.68	High	0.02	0.06	0.04	**0.88**
			Middle	0.02	0.25	0.10	**0.63**
			Low	0.14	0.54	0.10	**0.20**
57	0.280	0.26	High	0.08	0.32	0.20	**0.40**
			Middle	0.10	0.27	0.35	**0.29**
			Low	0.22	0.28	0.34	**0.18**
58	0.400	0.26	High	0.18	**0.48**	0.10	0.24

Form 3							
Item No.	IF	ID	Group	Options			
				A	B	C	D
			Middle	0.15	**0.45**	0.10	0.29
			Low	0.18	**0.24**	0.24	0.34
59	0.380	-0.14	High	**0.30**	0.32	0.10	0.28
			Middle	**0.40**	0.31	0.10	0.19
			Low	**0.42**	0.40	0.04	0.12
60	0.330	0.22	High	0.16	0.08	**0.48**	0.24
			Middle	0.26	0.10	**0.31**	0.33
			Low	0.30	0.12	**0.22**	0.34

Form 4							
Item No.	IF	ID	Group	Options			
				A	B	C	D
1	0.204	-0.09	High	0.49	**0.19**	0.28	0.04
			Middle	0.45	**0.16**	0.24	0.15
			Low	0.33	**0.29**	0.25	0.13
2	0.304	0.16	High	**0.35**	0.19	0.27	0.19
			Middle	**0.34**	0.24	0.28	0.13
			Low	**0.19**	0.31	0.36	0.14
3	0.239	0.22	High	0.24	**0.35**	0.28	0.13
			Middle	0.24	**0.24**	0.33	0.20
			Low	0.35	**0.13**	0.35	0.17
4	0.239	0.09	High	0.36	0.20	0.20	**0.23**
			Middle	0.24	0.22	0.23	**0.31**
			Low	0.25	0.21	0.39	**0.14**
5	0.544	0.46	High	0.14	**0.78**	0.08	0.13
			Middle	0.11	**0.54**	0.20	0.14
			Low	0.19	**0.32**	0.26	0.38
6	0.476	0.28	High	0.23	**0.65**	0.05	0.08
			Middle	0.20	**0.43**	0.08	0.28
			Low	0.14	**0.37**	0.18	0.31
7	0.395	-0.08	High	0.41	**0.30**	0.10	0.18
			Middle	0.23	**0.45**	0.07	0.24
			Low	0.36	**0.38**	0.02	0.24

Form 4							
Item No.	IF	ID	Group	Options			
				A	B	C	D
8	0.246	0.37	High	**0.48**	0.13	0.16	0.23
			Middle	**0.19**	0.29	0.20	0.31
			Low	**0.11**	0.49	0.18	0.23
9	0.291	0.12	High	0.13	**0.41**	0.23	0.23
			Middle	0.13	**0.22**	0.28	0.37
			Low	0.13	**0.29**	0.39	0.19
10	0.285	0.15	High	0.13	**0.35**	0.31	0.22
			Middle	0.20	**0.30**	0.40	0.12
			Low	0.19	**0.20**	0.56	0.05
11	0.343	0.00	High	0.10	0.24	**0.31**	0.34
			Middle	0.20	0.25	**0.29**	0.26
			Low	0.27	0.21	**0.31**	0.20
12	0.395	0.40	High	0.14	0.18	**0.57**	0.12
			Middle	0.09	0.35	**0.43**	0.22
			Low	0.30	0.29	**0.17**	0.25
13	0.162	0.13	High	0.31	0.25	0.17	**0.25**
			Middle	0.39	0.33	0.15	**0.13**
			Low	0.46	0.31	0.11	**0.12**
14	0.382	0.36	High	0.07	**0.58**	0.28	0.07
			Middle	0.23	**0.28**	0.40	0.09
			Low	0.20	**0.21**	0.40	0.18
15	0.220	0.17	High	0.02	0.12	**0.34**	0.53
			Middle	0.08	0.14	**0.19**	0.59
			Low	0.17	0.12	**0.17**	0.55
16	0.709	0.13	High	**0.82**	0.07	0.06	0.06
			Middle	**0.65**	0.15	0.10	0.10
			Low	**0.69**	0.14	0.01	0.15
17	0.304	0.11	High	0.11	0.33	**0.35**	0.22
			Middle	0.18	0.27	**0.32**	0.22
			Low	0.15	0.30	**0.24**	0.30
18	0.388	0.28	High	0.16	0.17	0.16	**0.51**
			Middle	0.10	0.27	0.22	**0.41**
			Low	0.29	0.24	0.25	**0.23**
19	0.430	0.17	High	**0.51**	0.20	0.28	0.01
			Middle	**0.44**	0.15	0.35	0.07

Form 4							
Item No.	IF	ID	Group	Options			
				A	B	C	D
			Low	**0.33**	0.21	0.38	0.07
20	0.223	-0.03	High	0.22	0.31	**0.22**	0.25
			Middle	0.24	0.43	**0.21**	0.10
			Low	0.29	0.31	**0.25**	0.15
21	0.207	0.13	High	0.16	**0.28**	0.23	0.34
			Middle	0.12	**0.20**	0.34	0.34
			Low	0.14	**0.14**	0.32	0.39
22	0.262	0.27	High	**0.42**	0.20	0.36	0.01
			Middle	**0.23**	0.28	0.37	0.11
			Low	**0.15**	0.23	0.37	0.25
23	0.385	-0.03	High	0.22	**0.35**	0.29	0.14
			Middle	0.24	**0.41**	0.20	0.15
			Low	0.27	**0.38**	0.21	0.13
24	0.495	0.36	High	0.16	0.12	0.10	**0.63**
			Middle	0.10	0.18	0.17	**0.55**
			Low	0.23	0.29	0.23	**0.26**
25	0.291	0.28	High	**0.45**	0.07	0.06	0.40
			Middle	**0.27**	0.23	0.10	0.39
			Low	**0.17**	0.30	0.18	0.36
26	0.162	0.22	High	**0.31**	0.33	0.16	0.23
			Middle	**0.11**	0.32	0.21	0.35
			Low	**0.10**	0.43	0.20	0.26
27	0.227	0.55	High	0.01	**0.57**	0.11	0.31
			Middle	0.03	**0.16**	0.24	0.57
			Low	0.13	**0.01**	0.27	0.58
28	0.595	0.08	High	0.10	**0.71**	0.13	0.05
			Middle	0.18	**0.51**	0.24	0.08
			Low	0.15	**0.63**	0.14	0.07
29	0.385	0.16	High	0.25	0.01	**0.48**	0.27
			Middle	0.20	0.02	**0.36**	0.41
			Low	0.27	0.05	**0.32**	0.36
30	0.463	0.29	High	0.14	0.11	0.16	**0.58**
			Middle	0.22	0.20	0.08	**0.50**
			Low	0.38	0.26	0.07	**0.29**
31	0.337	0.14	High	0.36	0.11	**0.41**	0.13

Form 4							
Item No.	IF	ID	Group	Options			
				A	B	C	D
			Middle	0.38	0.18	**0.33**	0.10
			Low	0.30	0.26	**0.27**	0.17
32	0.275	0.42	High	**0.52**	0.19	0.36	0.05
			Middle	**0.24**	0.29	0.41	0.06
			Low	**0.10**	0.30	0.58	0.02
33	0.469	0.29	High	0.23	0.12	**0.59**	0.06
			Middle	0.18	0.21	**0.50**	0.11
			Low	0.17	0.30	**0.30**	0.23
34	0.408	0.46	High	**0.66**	0.22	0.10	0.04
			Middle	**0.38**	0.32	0.15	0.16
			Low	**0.20**	0.46	0.26	0.08
35	0.463	0.39	High	0.16	**0.70**	0.14	0.00
			Middle	0.27	**0.41**	0.16	0.15
			Low	0.37	**0.31**	0.15	0.17
36	0.249	0.35	High	0.04	0.35	0.19	**0.42**
			Middle	0.11	0.37	0.26	**0.26**
			Low	0.07	0.55	0.31	**0.07**
37	0.382	0.30	High	0.10	0.30	**0.59**	0.01
			Middle	0.27	0.32	**0.32**	0.10
			Low	0.31	0.20	**0.29**	0.20
38	0.654	0.38	High	0.05	**0.84**	0.06	0.05
			Middle	0.05	**0.66**	0.04	0.25
			Low	0.10	**0.46**	0.08	0.36
39	0.437	0.24	High	0.24	0.04	0.17	**0.57**
			Middle	0.36	0.14	0.07	**0.43**
			Low	0.35	0.27	0.07	**0.32**
40	0.330	0.26	High	0.11	0.01	0.42	**0.46**
			Middle	0.20	0.09	0.38	**0.33**
			Low	0.24	0.23	0.33	**0.20**
41	0.249	0.28	High	0.05	0.36	**0.40**	0.20
			Middle	0.08	0.27	**0.24**	0.40
			Low	0.12	0.26	**0.12**	0.50
42	0.243	0.01	High	0.63	**0.23**	0.24	0.02
			Middle	0.33	**0.27**	0.20	0.20
			Low	0.10	**0.21**	0.39	0.30

Form 4							
Item No.	IF	ID	Group	Options			
				A	B	C	D
43	0.385	0.40	High	0.07	0.05	0.23	**0.64**
			Middle	0.15	0.15	0.38	**0.33**
			Low	0.08	0.25	0.43	**0.24**
44	0.469	0.39	High	0.13	0.04	**0.71**	0.13
			Middle	0.36	0.10	**0.42**	0.10
			Low	0.33	0.21	**0.32**	0.13
45	0.233	0.39	High	0.01	0.22	0.27	**0.48**
			Middle	0.08	0.26	0.49	**0.17**
			Low	0.20	0.21	0.49	**0.10**
46	0.353	0.18	High	0.07	0.16	**0.43**	0.37
			Middle	0.04	0.27	**0.36**	0.33
			Low	0.08	0.30	**0.25**	0.37
47	0.638	0.56	High	**0.90**	0.06	0.01	0.02
			Middle	**0.66**	0.04	0.20	0.10
			Low	**0.35**	0.04	0.49	0.13
48	0.485	0.27	High	**0.60**	0.16	0.12	0.11
			Middle	**0.51**	0.15	0.16	0.18
			Low	**0.33**	0.17	0.21	0.29
49	0.201	0.10	High	0.04	**0.28**	0.49	0.20
			Middle	0.07	**0.17**	0.59	0.17
			Low	0.19	**0.18**	0.48	0.15
50	0.421	0.44	High	0.18	0.11	0.05	**0.64**
			Middle	0.26	0.22	0.10	**0.43**
			Low	0.33	0.35	0.12	**0.20**
51	0.311	0.05	High	**0.33**	0.06	0.37	0.24
			Middle	**0.32**	0.06	0.31	0.30
			Low	**0.27**	0.12	0.36	0.25
52	0.469	0.39	High	0.13	0.05	**0.69**	0.13
			Middle	0.17	0.08	**0.44**	0.31
			Low	0.17	0.26	**0.30**	0.27
53	0.476	0.31	High	0.18	0.10	0.05	**0.67**
			Middle	0.27	0.18	0.13	**0.42**
			Low	0.33	0.11	0.18	**0.37**
54	0.223	0.30	High	0.17	**0.43**	0.19	0.20
			Middle	0.14	**0.16**	0.52	0.18

Form 4							
Item No.	IF	ID	Group	Options			
				A	B	C	D
			Low	0.17	**0.13**	0.44	0.25
55	0.256	0.04	High	0.31	0.16	**0.28**	0.25
			Middle	0.34	0.16	**0.26**	0.24
			Low	0.25	0.17	**0.24**	0.33
56	0.278	0.35	High	**0.46**	0.11	0.31	0.12
			Middle	**0.27**	0.09	0.26	0.38
			Low	**0.11**	0.06	0.39	0.44
57	0.107	0.07	High	0.31	0.31	0.22	**0.16**
			Middle	0.42	0.30	0.19	**0.09**
			Low	0.39	0.20	0.32	**0.08**
58	0.317	0.44	High	**0.61**	0.04	0.30	0.04
			Middle	**0.22**	0.10	0.54	0.14
			Low	**0.18**	0.17	0.50	0.15
59	0.294	0.16	High	0.13	**0.40**	0.40	0.07
			Middle	0.17	**0.27**	0.40	0.16
			Low	0.24	**0.24**	0.30	0.23
60	0.550	0.33	High	0.05	0.12	**0.71**	0.12
			Middle	0.10	0.18	**0.55**	0.17
			Low	0.25	0.23	**0.38**	0.13

Appendix B

Components extracted from factor analysis and constructed as specified in the test specification

Error: Form 1 (6 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	7	Preposition	Collocation Grammar knowledge	13	Preposition	Grammar knowledge
	1	S-V agreement	Grammar knowledge	14	Relative pron.	Grammar knowledge
2	12	Word order	Grammar knowledge Collocation Guess	10	Infinitive to	Grammar knowledge No
	9	Gerund	Grammar knowledge No	6	Passive	Grammar knowledge
3	2	Relative pronoun	Grammar knowledge	5	Adjective	Grammar knowledge
4	8	Pronoun	Grammar knowledge Collocation Guess	3	Passive	Grammar knowledge
5	11	Tense	Grammar knowledge Re-read relevant part	15	Adjective	Grammar knowledge
6	4	Tense	Grammar knowledge			

Error: Form 2 (6 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	12	Tense/If Clause	Grammar knowledge	2	Pronoun	Grammar knowledge
	11	Infinitive	Grammar knowledge Guess			
2	8	Tense	Grammar knowledge	1	S-V agreement	Grammar knowledge
3	15	Signal Words	Grammar knowledge	4	Part of Speech	Grammar knowledge
	5	Relative Clause	Grammar knowledge Test-taking Strategy			
4	14	Parallel Structure	Grammar knowledge	3	Passive	Grammar knowledge
	7	Preposition	Grammar knowledge			
5	10	Infinitive	Grammar knowledge			
6	9	Signal Words	Grammar knowledge Guess	13	Adjective Order	Grammar knowledge Guess

Error: Form 3 (7 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	6	Passive	Grammar knowledge	12	Word Order	No Explanation
	13	Preposition	Grammar knowledge			
2	1	S-V agreement	Grammar knowledge	9	Gerund	Grammar knowledge
	2	Relative Clause	Grammar knowledge			
3	8	Pronoun	Grammar knowledge	4	Tense	Grammar knowledge Guess
4	15	Part of Speech	Grammar knowledge	3	Passive	Grammar knowledge
5	7	Preposition	Grammar knowledge	10	Infinitive	Grammar knowledge
6	11	Tense	Grammar knowledge	15	Part of Speech	Grammar knowledge
7	14	Punctuation	Grammar knowledge			

Error: Form 4 (6 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	4	If clause/Tense	Grammar knowledge	11	Plural Noun	Grammar knowledge Guess
	10	Part of Speech	Grammar knowledge Test-taking strategy			
2	5	Infinitive	Grammar knowledge Guess Test-taking strategy	8	Preposition	Grammar knowledge Guess Careful reading (local)
3	3	S-V agreement	All Wrong	12	Preposition	Grammar knowledge
4	15	Tense	No Explanation	13	Passive	Grammar knowledge Test-taking strategy
5	9	Punctuation	Grammar knowledge	14	Tense	Grammar knowledge
	6	Adverb	Grammar knowledge Guess Test-taking strategy			
6	7	Relative Pronoun	Grammar knowledge			

Cloze: Form 1 (9 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	31	Vocab/reading	Use knowledge of vocabulary Collocation Guess Test-taking strategies Careful reading (local)	25	Verb form/structure	Careful reading (local) Use knowledge of grammar Careful reading (global) Guess Test-taking strategies
	26	Vocab/reading	Careful reading (local)	29	Preposition/ structure	Use knowledge of vocabulary Test-taking strategies Careful reading (global)
	32	Relative pronoun/ reading	Use knowledge of vocabulary Careful reading (local)			
2	21	Adverb/ structure and reading	Use knowledge of grammar Careful reading (local)	33	Vocab/reading	Use knowledge of vocabulary Careful reading (local)
	34	Vocab/reading	Use knowledge of vocabulary Careful reading (local)	40	Signal word/ reading	Careful reading (local) Careful reading (global)
3	35	Pronoun/ structure	Use knowledge of grammar Careful reading (local)	17	Part of speech/structure	Use knowledge of grammar
	18	Vocab/ reading	Use knowledge of vocabulary Collocation Guess	23	Determiner/ reading	Use knowledge of vocabulary Read relevant part (global)
	16	Vocab/reading	Use knowledge of vocabulary Guess			
4	37	Vocab/reading	Match words appearing in question	38	Vocab/reading	Look for important parts of the text

Component	Item	Construct	Strategy	Item	Construct	Strategy
			Careful reading (global) Test-taking strategies Read key parts of the text			Careful reading (local)
	27	Adjective/ reading	Use knowledge of vocabulary Use knowledge of grammar Careful reading (local)			
5	30	Determiner/ structure	Use knowledge of grammar Test-taking strategies Careful reading (local)	39	Adverb/ reading	Use knowledge of grammar
6	22	Vocab/ reading	Use knowledge of vocabulary Careful reading (local)			
7	36	Adverb/ structure	Use knowledge of grammar			
8	24	Vocab/reading	Use knowledge of vocabulary Use knowledge of grammar Guess	28	Adverb/ structure	No Explanation
9	20	Connector/ Reading	Use knowledge of vocabulary Use knowledge of grammar Guess Test-taking strategies			

Cloze: Form 2 (9 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	38	Vocab/reading	Use knowledge of vocabulary Test-taking strategies	35	Passive voice/structure	Use knowledge of grammar
	31	Vocab/reading	Use knowledge of vocabulary Careful reading (global)	39	Vocab/ structure	Use knowledge of vocabulary Test-taking strategies
	34	Relative pronoun/ structure	Collocation	26	Participial phrase/ structure	Use knowledge of grammar Guess Test-taking strategies
	37	Part of speech/structure	Use knowledge of grammar Guess			
2	20	Verb form/structure	Use knowledge of grammar Collocation	17	Determiner/ structure	Use knowledge of vocabulary Use knowledge of grammar
	30	Part of speech/structure	Use knowledge of vocabulary Use knowledge of grammar Guess Test-taking strategies	32	Determiner/ structure	Use knowledge of grammar Guess Careful reading (global)
3	23	Vocab/ reading	Use knowledge of vocabulary Test-taking strategies Careful reading (global)	28	Vocab/reading	Use knowledge of vocabulary Text structural knowledge Careful reading (local)
	18	Preposition/ structure	Test-taking strategies Careful reading (local)			

Component	Item	Construct	Strategy	Item	Construct	Strategy
4	33	Preposition/Structure	Guess Careful reading (global)			
5	22	Part of speech/structure	Use knowledge of vocabulary Use knowledge of grammar Test-taking strategies Careful reading (global)	19	Relative pronoun/ reading	No Explanation
6	16	Vocab/ reading	No Explanation	29	Signal word/ reading	Use knowledge of vocabulary Use knowledge of grammar Test-taking strategies Careful reading (global)
7	25	Quantitative adjective/structure	Careful reading (global)	21	Verb form/structure	Use knowledge of grammar Guess Test-taking strategies Careful reading (local)
8	40	Vocab/reading	Use knowledge of vocabulary Use knowledge of grammar Guess Test-taking strategies	24	Signal words/ reading	Test-taking strategies Careful reading (local) Careful reading (global)
9	27	Preposition/ structure	Use knowledge of grammar			

Cloze: Form 3 (10 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	18	Pronoun/reading	Use knowledge of grammar	27	Comparative/reading	Use knowledge of grammar
	24	Vocab/reading	Use knowledge of vocabulary Use knowledge of grammar Careful reading (local)	20	Vocab/reading	Match words appearing in question Use knowledge of grammar Careful reading (global)
	29	Signal words/reading	Use knowledge of vocabulary Use knowledge of grammar Careful reading (local)			
2	31	Vocabulary /reading	Use knowledge of vocabulary	35	Vocab/ reading	Use knowledge of vocabulary
	40	Vocab/ reading	Use knowledge of vocabulary Guess Test-taking strategies Careful reading (global)	26	Relative pronoun/reading	Use knowledge of grammar
3	22	Signal word./reading	Use knowledge of grammar	33	Pronoun/ reading	Use knowledge of grammar
	32	Vocab/ reading	Use knowledge of vocabulary Careful reading (global)			
4	38	Signal word/reading	Use knowledge of vocabulary Guess Test-taking strategies Careful reading (global)			

Component	Item	Construct	Strategy	Item	Construct	Strategy
5	28	Vocab/reading	Use knowledge of vocabulary Test-taking strategies Careful reading (global)	34	Vocab/reading	Careful reading (global)
	19	Relative pronoun/reading	Use knowledge of grammar	37	Adverb/ reading	Guess Careful reading (global)
6	16	Vocab/reading	Match words appearing in question Collocation	25	Vocab/reading	Use knowledge of grammar Careful reading (local)
7	23	Vocab/reading	No Explanation	30	Pronoun/ reading	Use knowledge of grammar
8	21	Signal word/reading	Test-taking strategies Careful reading (global)			
9	39	Adjective form/structure	Use knowledge of grammar			
10	17	Vocab/reading	Careful reading (local)			

Cloze: Form 4 (10 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	25	Vocab/reading	Careful reading (global)	27	Vocab/reading	Use knowledge of vocabulary Careful reading (local)
	24	Infinitive + to/structure	Use knowledge of grammar Collocation	36	S-V-A + Vocab/ structure	Use knowledge of vocabulary Use knowledge of grammar
2	40	Vocabulary/reading	Use knowledge of vocabulary	34	Adverb/reading	Use knowledge of vocabulary Guess Careful reading (global)
	37	Gerund/structure	Use knowledge of grammar			
3	18	Vocab/ reading	Careful reading (global) Use knowledge of vocabulary	39	Determiner/ reading	Text structural knowledge Careful reading (global)
	17	Tense/structure	Use knowledge of grammar Careful reading (global)	19	Vocab/reading	Use knowledge of vocabulary Guess Test-taking strategies Careful reading (local)
4	22	Passive voice/structure	Use knowledge of grammar	33	Part of speech/structure	Use knowledge of vocabulary Use knowledge of grammar
5	38	Adjective/structure	Use knowledge of grammar Guess Test-taking strategies			
6	29	Signal word/reading	Guess Careful reading (local)	21	Sentence structure	Use knowledge of grammar Careful reading (global)
7	32	Vocab/ collocation	Use knowledge of vocabulary Collocation	35	Vocab/reading	Use knowledge of vocabulary Use background knowledge Test-taking strategies Careful reading (local)

Component	Item	Construct	Strategy	Item	Construct	Strategy
8	23	Relative pronoun/ reading	Careful reading (global)	16	Adverb/ reading	Use knowledge of grammar Test-taking strategies Careful reading (local)
	30	Vocab/reading	Use knowledge of vocabulary			
9	28	Preposition/structure	Use knowledge of vocabulary			
10	26	Vocab/tense/ structure	Use knowledge of vocabulary Use knowledge of grammar Guess Test-taking strategies	31	Pronouns/reading	Use knowledge of vocabulary Use knowledge of grammar Test-taking strategies

Reading: Form 1 (7 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	48	Reference	Match search similar words	44	Detail (local)	Match search similar words
			Careful reading (global)			Careful reading (local)
			Careful reading (local)			
			Vocabulary knowledge			
			Grammar knowledge			
			Test-taking strategy			
	51	Main idea	Match scan words	55	Reference	
			Look for important parts			Look for important parts
			Vocabulary knowledge			Careful reading (local)
			Careful reading (local)			Grammar knowledge
			Background knowledge			
	58	Detail (local)	Match search similar words	43	Detail (local)	Match search similar words
			Careful reading (global)			Vocabulary knowledge
						Careful reading (local)
	52	Detail (global)	Careful reading (local)			
			Careful reading (global)			
2	57	Detail (global)	Careful reading (global)	46	Word in context	Match search similar words
			Match search similar words			Careful reading (global)
			Background knowledge			Vocabulary knowledge
			Test-taking strategy			Careful reading (local)
						Guess

Component	Item	Construct	Strategy	Item	Construct	Strategy
	54	Word in context	Careful reading (global)			
3	53	Word in context	Match scan words	47	Word in context	Vocabulary knowledge
			Vocabulary knowledge			Guess
			Guess			
			Test-taking strategy			
4	56	Global understanding (highlighted)	No explanation	45	Detail (global)	Match search similar words
						Careful reading (global)
						Background knowledge
						Test-taking strategy
5	49	Inference	All Wrong	50	Global reading (tone)	Careful reading (global)
6	41	Global understanding (topic)	Match scan words	59	Inference	All Wrong
			Careful reading (global)			
			Test-taking strategy			
7	60	Global understanding (tone)	Careful reading (global)			

Reading: Form 2 (10 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	51	Purpose (global)	Careful reading (global)	49	Reference (local)	Careful reading (global)
			Test-taking strategy			
2	44	Detail global	Look for important parts	56	Word in context (global)	Careful reading (global)
			Match search similar words			Guess
			Guess			Test-taking strategy
			Careful reading (global)			Vocabulary knowledge
			Match scan words			
3	53	Highlighted (global)	Careful reading (global)	47	Word in context	Careful reading (global)
			Test-taking strategy			Vocabulary knowledge
4	41	Main idea	Careful reading (global)	58	Reference	Careful reading (global)
			Re-read relevant part			Re-read relevant part
			Guess			
			Test-taking strategy			
			Text structural knowledge			
5	54	Detail global	Careful reading (global)			
			Vocabulary knowledge			
6	46	Detail global	Match search similar words	42	Inference	Background knowledge
			word recognition			Guess
			Careful reading (global)			Test-taking strategy
			Test-taking strategy			Match search similar words

Component	Item	Construct	Strategy	Item	Construct	Strategy
						Careful reading (global)
7	45	Global understanding (highlighted)	Match search similar words			
			Careful reading (global)			
8	48	Word in context	All Wrong	43	Detail global	Careful reading (global)
						Match search similar words
						Test-taking strategy
9	55	Paragraph main idea	Match scan words	59	Global understanding (inference)	Careful reading (global)
			Word recognition			Vocabulary knowledge
			Careful reading (global)			
			Guess			
10	57	Word in context	Vocabulary knowledge	52	Inference	Careful reading (global)
			Careful reading (global)			
			Guess			

Reading: Form 3 (9 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	43	Detail global	Careful reading (global)	49	Reference	Careful reading (local)
			Match search similar words			
	56	Detail global	Match scan words	58	Word in context	Vocabulary knowledge
			Careful reading (global)			Careful reading (global)
2	41	Purpose	Careful reading (global)	45	Detail global	Careful reading (local)
						Careful reading (global)
3	47	Word in context (local)	Vocabulary knowledge			
			Word recognition			
			Background knowledge			
4	60	Tone (global)	All Wrong	51	Topic (global)	Careful reading (global)
	50	Global understanding Tone Main idea	Careful reading (global)			
5	54	Global understanding	Test-taking strategy	53	Global (highlighted)	Test-taking strategy
			Careful reading (global)			
	46	Global understanding	Match search similar words			
6	48	Word in context	Guess	44	Global highlighted	Careful reading (global)
						Test-taking strategy
7	57	Word in context	Vocabulary knowledge			

Component	Item	Construct	Strategy	Item	Construct	Strategy
			Careful reading (local)			
			Careful reading (global)			
			Test-taking strategy			
8	52	Inference	Test-taking strategy			
			Match search similar words			
			Careful reading (global)			
9	42	Global (main idea)	Careful reading (global)			

Reading: Form 4 (7 components)

Component	Item	Construct	Strategy	Item	Construct	Strategy
1	43	Global understanding	Careful reading (global)	44	Detail global	Careful reading (global)
			Test-taking strategy			Match scan words
			Vocabulary knowledge			Test-taking strategy
			Guess			
	45	Word in context	Careful reading (local)	54	Detail (local)	Careful reading (global)
			Careful reading (global)			Test-taking strategy
			Vocabulary knowledge			
			Guess			
			Test-taking strategy			
2	50	Global (attitude)	Careful reading (global)	47	Reference (local)	Careful reading (local)
	60	Tone (global understanding)	Careful reading (global)			
			Guess			
			Test-taking strategy			
3	57	Word in context (local)	All wrong	56	Word in context (local)	Vocabulary knowledge
						Careful reading (global)
						Match search similar words
						Guess
	59	Reference (local)	Careful reading (local)	58	Reference (global)	Careful reading (global)
			Careful reading (global)			
4	46	Word in context	Vocabulary knowledge	51		All wrong

Component	Item	Construct	Strategy	Item	Construct	Strategy
			Match search similar words		Main idea (global)	
			Guess			
5	52	Inference	Careful reading (global)	41	Purpose (global)	Careful reading (global)
			Test-taking strategy			
			Guess			
6	48	Reference (local)	Careful reading (local)			
			Test-taking strategy			
7	49	Inference	Careful reading (global)			
			Test-taking strategy			

Appendix C

Inter-coder reliability of the researcher and expert = 96.09%

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
1	Test-taker1	8		8	
	Test-taker2	8		8	
	Test-taker3		14		14
	Test-taker4	8		8	
2	Test-taker1	8		8	
	Test-taker2	8		8	
	Test-taker3	8		8	
	Test-taker4	8		8	
3	Test-taker1		8		8
			15		15
	Test-taker2	8		8	
	Test-taker3		0		0
	Test-taker4	8		8	
4	Test-taker1		8		8
	Test-taker2	8		8	
	Test-taker3		14		14
	Test-taker4	8		8	
5	Test-taker1	8		8	
	Test-taker2	8		8	
	Test-taker3		8		8
	Test-taker4	8		8	
6	Test-taker1	8		8	
	Test-taker2	8		8	
	Test-taker3	8		8	
	Test-taker4	8		8	

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
7	Test-taker1	13		13	
	Test-taker2	13		13	
	Test-taker3	13		13	
	Test-taker4		8		8
			13		13
			14		14
8	Test-taker1	8		8	
	Test-taker2	8		8	
	Test-taker3		8		8
			13		13
			14		14
	Test-taker4	8		8	
9	Test-taker1		14		14
	Test-taker2	0		0	
	Test-taker3		7		7
	Test-taker4	8		8	
10	Test-taker1	8		8	
	Test-taker2	0		0	
	Test-taker3		8		8
			13		13
	Test-taker4	8		8	
11	Test-taker1	8		8	
	Test-taker2	8		8	
		10		10	
				11	
	Test-taker3		8		8
			13		13

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
	Test-taker4	8		8	
12	Test-taker1		14		14
	Test-taker2	8		8	
	Test-taker3		8		8
			14		14
	Test-taker4	8		8	
		13		13	
		14		14	
13	Test-taker1		8		8
	Test-taker2	8		8	
	Test-taker3		8		8
	Test-taker4	8		8	
14	Test-taker1	8		8	
	Test-taker2	8		8	
	Test-taker3		8		8
	Test-taker4	8		8	
15	Test-taker1	8		8	
	Test-taker2	8		8	
	Test-taker3		8		8
	Test-taker4	8		8	
		37	12	37	12
		4	4	4	4
		1	7	1	7
		0	1	0	1
		0	0	0	0
16	Test-taker1		7		7

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
16			14		14
	Test-taker2		14		14
	Test-taker3	7		7	
		14		14	
	Test-taker4		14		14
			15		15
17	Test-taker1		8		8
	Test-taker2	8		8	
	Test-taker3	8		8	
	Test-taker4	8		8	
18	Test-taker1	13		13	
	Test-taker2		13		13
	Test-taker3		7		7
	Test-taker4	7		7	
		13		13	
		14		14	
19	Test-taker1		14		14
	Test-taker2		7		7
	Test-taker3	10		10	
		14		14	
	Test-taker4		7		7
			14		14
20	Test-taker1		7		7
			9		
			14		14
	Test-taker2		8		7
					9

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
20	Test-taker3	7		7	
		14		14	
		15		15	
	Test-taker4	8		8	
21	Test-taker1	8		8	
	Test-taker2	8		8	
		9		9	
	Test-taker3		7		7
			8		8
			9		9
			10		
			14		14
			15		15
	Test-taker4	8		8	
		9		9	
22	Test-taker1		7		7
			14		14
	Test-taker2	7		7	
		9		9	
	Test-taker3		7		7
			9		9
			14		14
	Test-taker4		7		7
			14		14
			15		15
23	Test-taker1			7	
		10		10	

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
23		13			
	Test-taker2	7		7	
		9		9	
		13			
	Test-taker3	7		7	
	Test-taker4	7		7	
24	Test-taker1			7	
		8		8	
		14		14	
	Test-taker2		7		7
			9		9
	Test-taker3		7		7
			9		9
	Test-taker4		7		7
			14		14
25	Test-taker1		8		8
			14		14
	Test-taker2	8		8	
		9		9	
	Test-taker3		7		7
					8
			14		14
	Test-taker4	7			
		8		8	
		9		9	
		14		14	
		15		15	

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
26	Test-taker1		7		7
			15		15
	Test-taker2	9		9	
	Test-taker3		7		7
			9		9
			15		15
	Test-taker4	9		9	
27	Test-taker1		15		15
	Test-taker2	7		7	
		9		9	
	Test-taker3	7		7	
		8			
		9		9	
	Test-taker4	8		8	
		9		9	
28	Test-taker1		8		8
	Test-taker2		8		8
	Test-taker3		8		8
	Test-taker4		8		8
			10		
29	Test-taker1	7		7	
		15		14	
	Test-taker2	7		7	
		9		9	
	Test-taker3	7		7	
		9		9	
	Test-taker4	7		7	

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
29	Test-taker4	9		9	
30	Test-taker1	8		8	
	Test-taker2	8		8	
		15		15	
	Test-taker3	8		8	
		9		9	
	Test-taker4		8		8
			9		9
			14		14
			15		15
31	Test-taker1		8		
			14		14
	Test-taker2	7		7	
		13		13	
		14			
	Test-taker3	7		7	
		9		9	
		15		15	
	Test-taker4	7		7	
		9		9	
32	Test-taker1		7		7
			9		9
	Test-taker2	9		9	
	Test-taker3	7		7	
		9		9	
	Test-taker4	7		7	

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
33	Test-taker1	7		7	
		9		9	
	Test-taker2	7		7	
	Test-taker3	7		7	
	Test-taker4	7		7	
34	Test-taker1		9		0
	Test-taker2	7		7	
		9		9	
	Test-taker3		7		7
			14		14
			15		15
	Test-taker4	7		7	
		9		9	
35	Test-taker1	8		8	
		9		9	
	Test-taker2	8		8	
		9		9	
	Test-taker3	8		8	
		9		9	
	Test-taker4	8		8	
36	Test-taker1		8		8
	Test-taker2		8		8
	Test-taker3		8		8
	Test-taker4	8		8	
37	Test-taker1		2		
			4		4

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
37	Test-taker2	2			
		9		9	
	Test-taker3	2			
		9		9	
		15		15	
	Test-taker4	4		4	
38	Test-taker1	3		3	
	Test-taker2	9		9	
	Test-taker3	9		9	
	Test-taker4	9		9	
39	Test-taker1	8		8	
	Test-taker2		14		14
	Test-taker3		12		12
	Test-taker4		10		10
			14		14
40	Test-taker1		8		8
			9		9
	Test-taker2	9		9	
	Test-taker3		8		8
	Test-taker4	9		9	
41	Test-taker1		1		1
			9.1		9.1
	Test-taker2		1		1
	Test-taker3	9.1		9.1	
		15		15	
	Test-taker4		9		9
			14		14

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
42	Test-taker1	1		1	
		15		15	
	Test-taker2	3		3	
	Test-taker3	2		2	
		3		3	
	Test-taker4	2		2	
		3		3	
43	Test-taker1		2		2
	Test-taker2	2		2	
		7		7	
		9		9	
	Test-taker3	2		2	
		7		7	
		9		9	
	Test-taker4		1		1
44	Test-taker1	2		2	
		9 local		9 local	
	Test-taker2	2		2	
		9 local		9 local	
	Test-taker3	2		2	
		9 local		9 local	
	Test-taker4	2		2	
		9			
45	Test-taker1		1		1
			15		15
	Test-taker2		1		1
			2		2

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
45	Test-taker2		14		14
			15		15
	Test-taker3	2		2	
		9		9	
		12		12	
		15		15	
	Test-taker4		1		1
			2		2
			9		9
			15		
46	Test-taker1		7		
			6		6
			14		14
	Test-taker2	2		2	
		7		7	
		9		9	
	Test-taker3	7		7	
		9		9	
		14			
	Test-taker4	7		7	
47	Test-taker1		6		6
			7		7
	Test-taker2		2		2
			9		9
	Test-taker3		12		12
	Test-taker4	7		7	
		14		14	

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
48	Test-taker1	7		7	
		8		8	
		9		9	
	Test-taker2	2		2	
		9		9	
	Test-taker3	2		2	
		15		15	
	Test-taker4	9		9	
49	Test-taker1		2		2
			9		9
			10		10
			15		15
	Test-taker2				2
			9		9
			15		15
	Test-taker3		2		2
			15		15
	Test-taker4		2		2
			15		15
50	Test-taker1	9.1		9.1	
	Test-taker2		14		14
	Test-taker3		7		7
			11		11
			12		12
	Test-taker4		9.1		9.1
51	Test-taker1	1		1	
		3		3	

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
51	Test-taker2	0		0	
	Test-taker3	7		7	
		9.1		9.1	
		12		12	
	Test-taker4		7		7
			9		
			12		12
52	Test-taker1	9 local		9 local	
	Test-taker2	9 global		9	
	Test-taker3	9		9	
	Test-taker4	9		9	
		15		15	
53	Test-taker1		14		14
	Test-taker2		2		2
			5		5
	Test-taker3	1		1	
		7			
		14		14	
		15		15	
	Test-taker4		7		7
54	Test-taker1		7		7
			9		9
	Test-taker2	9		9	
	Test-taker3		2		2
			6		6
			9		9

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
54	Test-taker4	9		9	
55	Test-taker1	3		3	
		8		8	
		9		9	
	Test-taker2	3		3	
		9		9	
	Test-taker3	3		3	
		9		9	
	Test-taker4	9		9	
56	Test-taker1		1		1
			3		3
			14		14
	Test-taker2		2		2
			9		9
	Test-taker3		2		2
	Test-taker4	1		1	
57	Test-taker1		1		1
			14		14
	Test-taker2	2		2	
		9		9	
	Test-taker3	9		9	
		12		12	
		15		15	
	Test-taker4	1		1	
58	Test-taker1	9		9	
	Test-taker2		3		3
			9		9

Item No.	Test-taker	Researcher		Expert	
		Strategies used		Strategies used	
		Correct	Incorrect	Correct	Incorrect
58	Test-taker3	2		2	
		9		9	
	Test-taker4	0		0	
59	Test-taker1		1		1
	Test-taker2		9		9
			9.1		9.1
			10		
			12		12
	Test-taker3		9		9
			12		12
			15		15
	Test-taker4		1		1
			9		9
			15		15
60	Test-taker1	9.1		9.1	
	Test-taker2		2		2
			12		12
			15		15
	Test-taker3		12		12
			14		14
			15		15
	Test-taker4		14		14

Appendix D

Performances of the 16 participants in each part of the test

	Error	Cloze	Reading	Total
Test-taker 1	15	22	16	53
Test-taker 2	12	23	16	51
Test-taker 3	15	20	15	50
Test-taker 4	9	24	15	48
Test-taker 5	13	22	11	46
Test-taker 6	13	19	14	46
Test-taker 7	13	20	12	45
Test-taker 8	14	18	11	43
Test-taker 9	12	19	12	43
Test-taker 10	15	17	10	42
Test-taker 11	11	17	14	42
Test-taker 12	10	16	10	36
Test-taker 13	9	19	8	36
Test-taker 14	10	15	6	31
Test-taker 15	3	14	13	30
Test-taker 16	10	10	9	29