What Test-takers Do When Taking a Rational Cloze Test: A Cognitive Processing Model of Reading
ผู้เข้าสอบคิดอย่างไรขณะทำข้อสอบแบบเติมคำ: รูปแบบกระบวนการรู้คิดในการอ่าน

Thanyapa Palanukulwong[1]
ธัญภา พลานุกูลวงศ์[2]

Abstract

One main concern of a test is whether it tests what it is supposed to test. The primary aim of this study was to explore the construct being measured in a rational cloze test. Stimulated recall data produced by 16 proficient participants taking four parallel test forms were analyzed to determine the cognitive processes used to complete a rational cloze test. Based on Khalifa and Weir's (2009) model of cognitive processing in reading, the findings across the four test forms showed that *slow* and *careful reading at global and local level* were the most frequently-used strategies, and the two sources of knowledge used to complete the items were *lexical* and *grammatical*, suggesting that rational cloze tasks tap into the constructs of vocabulary and grammar knowledge and careful reading at both global and local level.

Keywords: Test validation, Rational cloze tests, Stimulated recall protocol, Cognitive processes for reading

บทคัดย่อ

ข้อกังวลประการหนึ่งของแบบทดสอบคือแบบทดสอบได้วัดสิ่งที่ต้องการวัดหรือไม่ การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาความตรงของข้อสอบแบบเติมคำ นักวิจัยใช้ข้อมูลทางวาจาที่เก็บจากการกระตุ้นการเรียกความจำคืนของผู้เข้าสอบที่มีความสามารถทางภาษาสูงจำนวน 16 คน ซึ่งทำข้อสอบแบบเติมคำคู่ขนาน 4 ฉบับเพื่อศึกษากระบวนการคิดที่กลุ่มตัวอย่างใช้ระหว่างทำข้อสอบและวิเคราะห์ข้อมูลตามกรอบแนวคิดของ Khalifa และ Weir's (2552) ผลการวิจัยพบว่ากลวิธีที่กลุ่มตัวอย่างใช้บ่อยที่สุดในการทำข้อสอบทั้ง 4 ฉบับ คือการอ่านอย่างพินิจพิจารณาใน*ระดับองค์รวมและระดับประโยค*โดยใช้ความรู้ด้านคำศัพท์และไวยากรณ์ ซึ่งแสดงว่าข้อสอบแบบเติมคำวัดความรู้ด้านคำศัพท์และไวยากรณ์ รวมทั้งทักษะการอ่านอย่างพินิจพิจารณาทั้งระดับองค์รวมและระดับประโยค

คำสำคัญ: การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ แบบทดสอบแบบเติมคำ การกระตุ้นการเรียกความจำคืนผ่านวาจา กระบวนการรู้คิดในการอ่าน

---

[1] Associate Professor Dr., Faculty of Liberal Arts, Prince of Songkla University, Hat Yai Campus

[2] รองศาสตราจารย์ ดร. คณะศิลปศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่

Introduction

Cloze tests, also referred to as gap-filling, comprise tasks constructed by deleting words from selected texts with test-takers having to restore the words that have been deleted. Every N^{th} word may be omitted from a selected text, and this is called a fixed-ratio cloze while tests with words deleted on a rational basis are called gap-filling tests, or rational-deletion cloze (Alderson, 2000; McCray & Brunfaut, 2018). In some cases, a pool of possible answers may be provided for test-takers to choose from while in other cases, possible answers are not provided, and the test-takers have to generate the missing word to complete the text.

The issue of what cloze tests measure has been the subject of intense debate in second language (L2) research. Some research has shown that cloze items measure the knowledge of sentence-level grammatical structure and the ability to make localized connections in the text. For example, Alderson (1979) compared cloze performance with several external measures, including a test of proficiency in English as a Foreign Language (EFL) in assessing different language skills. The results suggested that cloze in general related more to tests of grammar and vocabulary than to tests of reading comprehension. Alderson concluded that cloze tests are not suitable tests of higher-order language skills but can be a measure of lower-order core proficiency. A study by Kintsch and Yarbrough (1982, cited in Alderson, 2000) investigated the strategies required to process rhetorical structures. Kintsch and Yarbrough distinguished two levels of comprehension processes: macro-processes (global understanding) and micro-processes (local, phrase-by-phrase understanding). The questions used to test macro-processes were those asking about the topic and main ideas, while every-fifth-word cloze tests were used to investigate micro-level processes. Their findings suggested that cloze tests were not sensitive to macro-processes but related to micro-processes. On the other hand, other studies suggest that cloze tests involve discourse processing abilities and measure global-level comprehension. Among these is Bachman's (1985) study. Bachman used two cloze tests prepared from the same text, one a fixed-ratio cloze and the other with rational deletions. The findings supported the validity of cloze tests as reading comprehension tests as they were sensitive to inter-sentential or text-level constraints. Yamashita (2003) investigated whether a gap-filling (or rational deletion cloze) measured text-level processing ability. The participants completed a gap-filling test while thinking aloud by verbalizing their test-taking processes. The results showed that both skilled and less skilled participants used

text-level information more frequently than other types of information, suggesting that the gap-filling cloze could be used to test higher-order processing ability.

Given the conflicting research findings, despite the considerable interest in this test, it is still not clear exactly what is being measured by cloze tests of either type.

Test Validation

Traditionally, several types of evidence can be gathered in the process of test validation. This includes examining the outcomes of the test, i.e., test scores, the interrelation of sub-tests, and the relationship between the test and other measures of the same construct. 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 performing 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 (Bax, 2013). Alderson (2000, p. 97) argued that "the validity of a test relates to the interpretation of the correct responses to items, so what matters is not what test constructors believe an item to be testing, but which responses are considered correct, and what process underlies them." The use of verbal protocol analyses in the process of test validation makes it possible to ascertain whether a particular test actually tests what it is supposed to test (Anderson, Bachman, Perkins, & Cohen, 1991).

Verbal Protocol Analysis

Verbal protocol analysis (VPA) is a qualitative procedure used as a means of validating assessment instruments and methods. Test-takers are asked to think aloud as they work through test items, and inferences can be made directly from the data. Gathering information on test-taking processes offers insights into the process and strategies used by test-takers, which may not be available through other research methodology. Because of the intensive nature of verbal protocol research, studies typically involve no more than a handful of participants (Weir, Hawkey, Green, & Devi, 2012).

Green (1998) suggested that verbal reports can be gathered either concurrently or after a language event. Concurrent reports are generated during the process of completing the test, while retrospective reports are generated after the test-taker has finished the test task. The protocols produced are then analyzed to identify the cognitive processes involved in the completion of the

test task. Verbal protocols serve as a source of data for the researcher to infer cognitive processes and attendant information. If there is a close match between the processes that are actually employed and those that the test developers predict will be used, then the test is believed to measure what it is supposed to measure.

Based on Gass & Mackay (2017), stimulated recall is an introspective method of eliciting the thought processes taking place while an individual is doing a task. 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 the advantage over other think-aloud approaches that no training is required for the participants to carry out a task.

A number of studies adopting VPA have been conducted in L2/FL testing research to examine how test-takers respond 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. Storey 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, Ferne, & Choi (2006), using concurrent verbal protocols, looked at how 10 non-native adult readers approached a reading test with multiple-choice questions. The study showed many different representations of the construct of reading comprehension. Also, test-takers combined a variety of mental resources interactively when making a choice.

Bax (2013) investigated test-takers' cognitive processing while completing IELTS reading test items. Eye movement and stimulus recall data were collected. 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 & McCray (2015) combined the use of both eye-tracking and stimulated recall in examining 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.

Cognitive Processing Models 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 processes readers use in text comprehension when engaging in different types of real-life reading (Khalifa & Weir, 2009; Brunfaut & McCray, 2015). The model is shown in Figure 1.
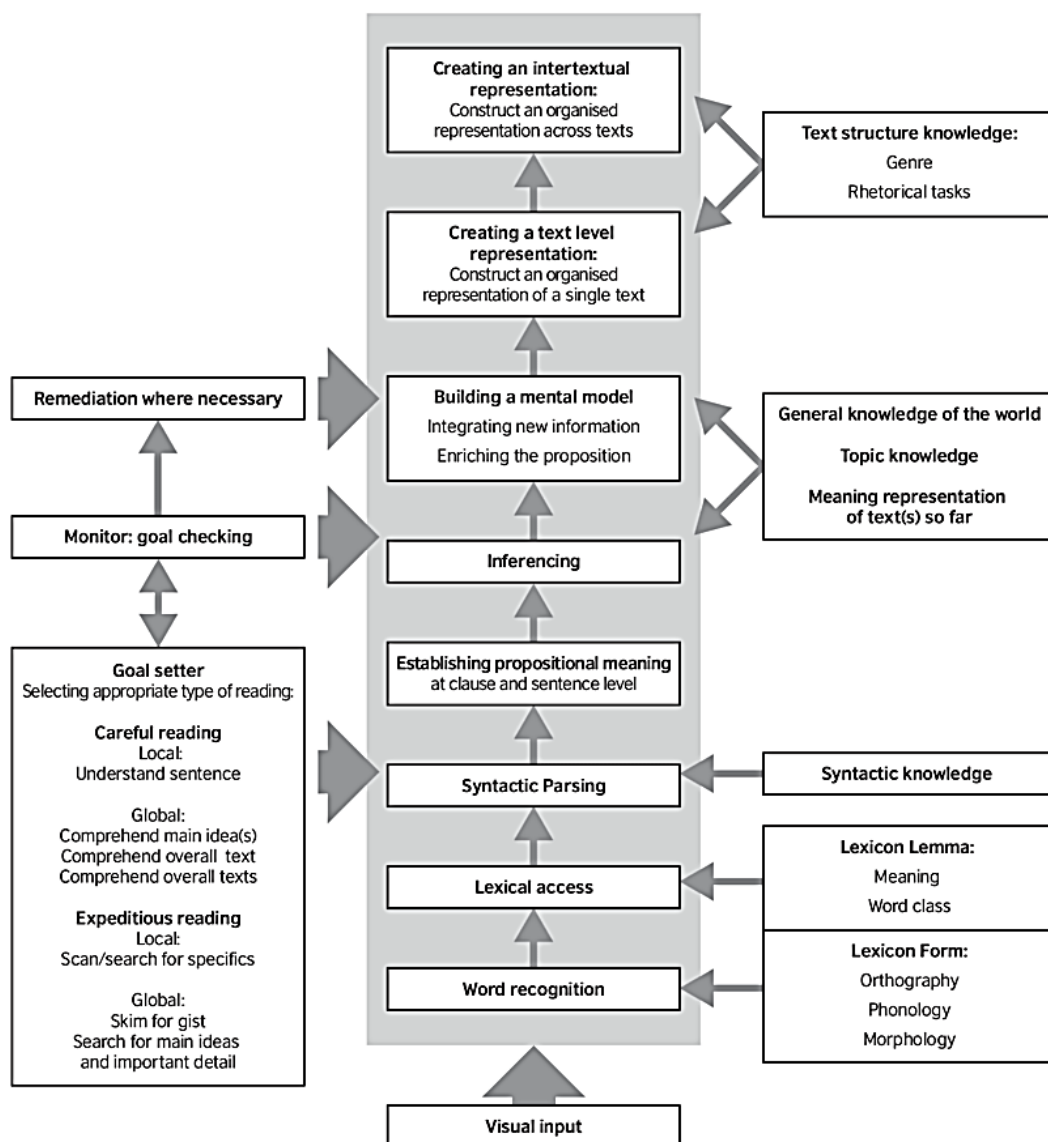

Figure 1 Cognitive model in reading (Khalifa & Weir, 2009, p. 43)

The model by Khalifa and Weir (2009) has three main components – *the metacognitive activity, the central processing core* and *the knowledge base.* In the left-hand column are *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 levels 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 on the type of reading needed to complete a specific task: *local-level reading* at the sentence and clause level, or *global-level reading* 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. *Careful reading* is intended to extract meaning from material at a local or global level, i.e., within or beyond the sentences in a text. While reading, the reader monitors the progress of their reading in line with their goals, with breakdowns triggering remediation 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, so-called lower- and higher-level processing (Urquhart and Weir, 1998, cited in Weir et al., 2012). Lower-level processes include *word recognition, lexical access, syntactic parsing,* and *establishing propositional meaning*, while higher-level processes are *inferencing, building a mental model,* and *creating a text level or intertextual representation.*

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

Purpose of the Study

There is some debate on what is being tested in a cloze test, and there has been little research investigating what a rational cloze test tests (Read, 2000). This study therefore aimed to

examine what is being measured in a rational cloze test. Using Khalifa and Weir's (2009) model of cognitive processing in reading, this study was designed to examine what test-takers did when they were completing rational cloze items. The following research questions were formulated in this study.

      1. What cognitive processes do test-takers employ while completing rational cloze items?

      2. What does the rational cloze test measure?

Research Methodology

1. Participants

      Sixteen PSU students from the Faculties of Medicine, Engineering, Science, and Liberal Arts, were selected to be the participants. They were high proficiency learners of English as identified by their English teachers based on their performance of English subjects and, thus, were purposively selected to be part of the retrospective stimulated recall protocol. Proficient participants were targeted 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).

2. Research Instruments

2.1 Rational Cloze Test of the PSU-TEP Structure and Reading Test

      The PSU-TEP is a skill-based proficiency test developed by Prince of Songkla University, consisting of 4 sub-tests: Structure and Reading, Listening, Writing, and Speaking tests. The test is offered to the public four times a year, each time with a parallel form of the test developed by the same team of test developers based specifically on the same test specification.

      The PSU-TEP Structure and Reading Test consists of three parts: error recognition, rational cloze test, and reading. Four parallel forms of the rational cloze test administered in February, April, June, and December 2016 were used as research instruments. Each form consisted of 2 passages, with 12 words and 13 words rationally removed in the first text and the second text respectively, totaling 25 items. Approximately equal numbers of content and function words were rationally deleted in each passage. Test-takers completed the passages by choosing appropriate words from the choices given.

2.2 Stimulated recall analyses

To seek information on the cognitive processes that the 16 participants engaged in to find answers to each cloze item, Khalifa and Weir's (2009) model of cognitive processing in reading, outlined in Figure 1, formed the theoretical framework on which the analyses of the stimulated recall data were based. Strategies number 1-12 representing the cognitive processes the participants engaged in to find answers to each cloze item are those used in Weir et al. (2012) with three additional codes added.

Table 1 The coding framework for the stimulated recall data on the rational cloze items.

Strategy Code	Definition
1	Match words that appear in the question with exactly the same words in the 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 the 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 (general/topic knowledge)
13	Collocation
14	Guess
15	Choice elimination

Following Brunfaut and McCray's (2015) study, two extra codes, No. 13 (*collocation)*, and No. 14 (*guess)* were added. Also, after a random sampling of about 25 % of the stimulated recall data, No. 15 *(choice elimination)* was another code added as it was found to be an activity contributing to either correct or incorrect answers.

Data Collection

To answer the two research questions, the 16 proficient students chosen to take the four test forms and participate in the retrospective stimulated recall sessions were randomly divided into groups of four; each group completed one of the parallel forms of the rational cloze test 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 was told that the researcher was interested in what they were thinking and what was in their mind while engaging in the rational cloze task.

Each participant took the test on different occasions. Immediately after the completion of the 25 rational cloze items, the participants participated in the retrospective stimulated recall sessions carried out on a one-to-one basis either by the researcher or a research assistant. They expressed their thoughts in their first language (Thai), and the conversations were audio-recorded throughout the sessions.

Questions directed to the participants during the retrospective stimulated recall interviews were, for example, "*Could you tell me how you started doing the test?"; "What were you thinking here / at this point / right then?" "Do you remember thinking anything when you…?".*

Data Analysis

The audio-recorded recall data were transcribed and coded by the researcher based on Khalifa and Weir's (2009) model of cognitive processing in reading and the three additional codes (see Table 1). To ensure the coder reliability of the stimulated recall data, an external coder was asked to code 25 % of the total recall data. The percentage of similarity between the two coders was 94.7, indicating a very high inter-coder reliability value (Green, 1998). The stimulated recall data were analyzed for frequency and percentages to examine the cognitive processes the 16 participants used while processing the cloze items.

Findings

In the analysis of stimulated recall data, the occurrence of each of the strategy codes was calculated. The analysis mainly focused on the correctly-answered items since, from a validation perspective, these would reflect the intended aspect of reading (Brunfaut & McCray, 2015).

The stimulated recall data showed that before approaching the items, no participant indicated that they previewed the passages to get a general idea of the text.

Across all the test forms, the four strategies most frequently-used were (in descending order) 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 (*read the text or part of it slowly and carefully beyond sentence level*). From a total of 415 strategies which the participants relied on to arrive at correct answers, 119 cases (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) or *choice elimination* (No. 15) were found to have been used to help the participants obtain the correct answers. This is well illustrated in Figure 2 below.
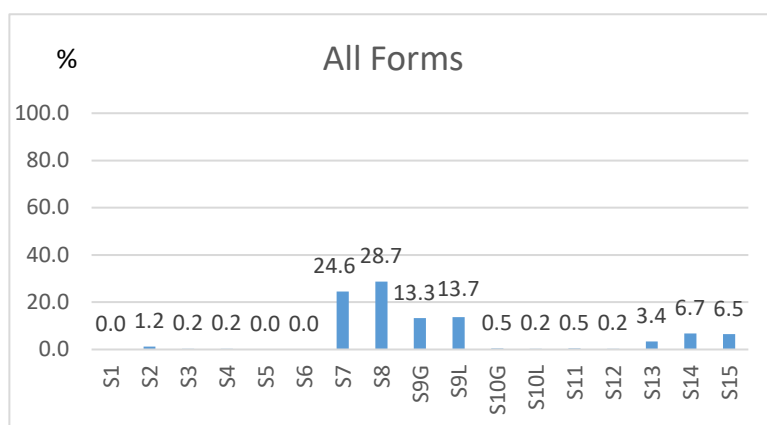

Figure 2 Cognitive strategies employed in rational cloze tasks across four test forms

The analysis of all the strategies used to obtain correct answers in each test form showed a quite similar pattern of strategy use. As shown in Figure 3, the most popular strategy in each form was strategy No. 8 (*use of grammar knowledge*), which accounted for 22.1 % of the total strategies which the participants taking Form 1 used to arrive at correct answers, 30.8 % of the total strategies for Form 2, 33 % for Form 3 and 28.3 % for Form 4. The second most popular strategy was No. 7

(*use of vocabulary knowledge*), which accounted for 25.3 % of the total effective strategies used for Form 1, 20.6 % for Form 2, 24 % for Form 3 and 28.3% for Form 4. Strategy No. 9 (*global careful reading*) and No. 9 (*local careful reading*) were the next most frequently- used strategies; participants who took Forms 2- 4 read the text more *globally* while those who took Form 1 approached the text more *locally.* A few cases of strategy No. 14 (*guessing*) and No. 15 (*choice elimination*) were found for each test form.

No evidence of the use of the 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 high scores. Moreover, this might be due to the fact that the texts used for the cloze tasks were generally short, consisting of a single paragraph or a few paragraphs; thus the need for skimming for general idea seemed unnecessary. Also, the 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.
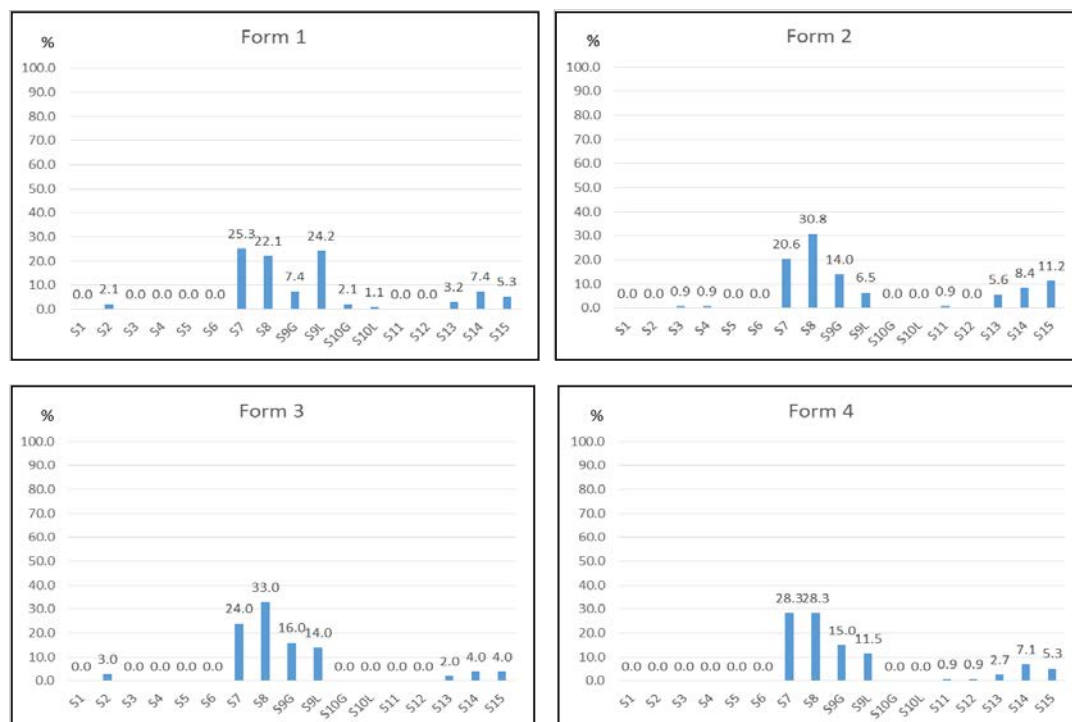

Figure 3 Cognitive strategies employed in rational cloze tasks in each test form

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

Conclusion

The study aimed to explore the cognitive processes that test-takers employed during the completion of rational cloze items in an attempt to validate what the rational cloze tests in the PSU-TEP Structure and Reading Test measured. Insights were gained through the detailed analysis of stimulated recall data produced by the 16 participants. For the first research question, the analysis showed that strategies No. 7 (*use knowledge of vocabulary*), No. 8 (*knowledge of grammar*), No. 9 (*global careful reading*), and No. 9 (*local careful reading*) were the four most frequently-used strategies, recorded for all the four test forms. The results showed that slow and careful reading at both a global and local level were the most frequently-used strategies when the participants were approaching the rational cloze items and two sources of knowledge they used to complete the items were lexical and grammatical knowledge, an aspect of so-called low-level cognitive processing.

The contribution of the participants' lexical and syntactic skills was effective in helping them to carefully read and comprehend the texts to the extent that they could successfully complete the gaps. This association between the strategies used and the sources of knowledge on which the participants most heavily relied seems to suggest that the answer to the second research question is that the rational cloze tasks under investigation tapped into both the constructs of vocabulary knowledge and grammar knowledge and careful reading at both global and local levels.

The type of reading the participants employed when faced with a text, i.e. careful reading at both global and local levels was the metacognitive activity of *a goal setter* in the left-hand column on Khalifa and Weir's (2009) cognitive model in reading in Figure 1. The decision to select this type of reading affected the levels(s) of cognitive processing to be activated in the central processing core of the model, i.e. the use of lexical and syntactic knowledge for text comprehension.

Since these cognitive processes are among the so-called lower-level processes, it might therefore be possible to conclude that the rational cloze items managed to tap mostly lower-level processing, i.e., grammar and vocabulary knowledge. This is in line with Weir (2013), who claimed that cloze tests do not reflect the reader's ability to comprehend beyond the sentence level. The

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 meaning. Rational cloze tests, therefore, appear only to reflect the processes involved in careful local reading to establish comprehension in the sentence.

The findings of the present study support previous studies that cloze tests measure a limited part of reading proficiency (Weir, 2005). For example, Alderson's (1979) suggested that cloze tests were not suitable tests of higher-order language skills. Kintsch and Yarbrough's (1982, cited in Weir, 2013) demonstrated 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.

Recommendations for Further Studies and Test Developers

The findings of the study provide valuable insights into the use of test-takers' cognitive processes in the process of test validation. The use of the stimulated recall is recommended for test validation research. Although the methodology was quite laborious, stimulated recall proved to be useful in revealing the cognitive processes employed to arrive at correct answers.

A more extensive study of the cognitive processing employed by test takers in completing rational cloze items is needed, using a larger number of proficient test-takers, with every test-taker being assigned to take the same test form, in order to examine the construct being measured in the test.

In spite of the finding that the participants in the present study predominantly arrived at the correct answers by relying on their lexical and grammatical knowledge, suggesting that the rational deletion cloze test measured low-level processing skills, many scholars agree that a rational cloze test can be a good measure for testing language skills. Alderson (2000) suggested that the rational cloze format can be a technique used to determine what being tested, and that this type of test format is much more under the control of the tester than other types of cloze. Selected content words can be deleted to test understanding of the overall meaning of the text; function words deleted to test mainly grammatical sensitivity while removing words that are essential to the main idea, or deleting words which carry the text's coherence can measure the overall understanding of the text. Weir (2013) and Bachman (1985) agreed that the selective deletion procedure makes it

possible for test developers to design the content to be measured. Other advantages of rational deletion cloze are that they can be relatively easy to construct, and texts can be selected to satisfy any appropriate contextual parameter, e.g., text type, size, and topic (Weir, 2013).

References

Alderson, J. C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13(2), 219-223.

Alderson, J.C. (2000). *Assessing reading*. 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. (1985). Performance on the cloze test with fixed-ratio and rational deletions. *TESOL Quarterly*, 19(3), 535-556.

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

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.

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. In M. Milanovic (Ed.), *Studies in Language Testing 5* (pp. 1-175). Cambridge: Cambridge University Press.

Khalifa, H. & Weir, C. J. (2009). Examining reading: Research and practice in assessing second language reading. In M. Milanovic & C.J. Weir (Eds.), *Studies in Language Testing 29: Examining reading* (pp.1-342). Cambridge: Cambridge University Press.

McCray, G., & Brunfaut, T. (2018). Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. *Language Testing*, 35(1), 51–73.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

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 cloze 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. In M. Milanovic & C.J. Weir (Eds.), *Studies in Language Testing 37: Measured constructs* (pp.103-256), Cambridge: Cambridge University Press.

Weir, C., Hawkey, R., Green, A., & Devi, S. (2012). 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. In L. Taylor & C.J. Weir (Eds.), *Studies in Language Testing 34*: *IELTS collected papers* (pp.212-269). Cambridge: Cambridge University Press.

Yamashita, J. (2003). Processes of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing*, 20(3), 267-293.