EDUPSU Journal of Education, Prince of Songkla University, Pattani Campus
Vol.33 No.1: January - April, pp. 170-183

Exploring the construct measured by PSU-TEP Reading Test: a cognitive processing model of reading

Thanyapa Palanukulwong*
Ph.D. (Applied Linguistics), Associate Professor
Department of Foreign Languages, Faculty of Liberal Arts, Prince of Songkla University
*Corresponding Author: thanyapa.c@psu.ac.th

Received: August 14, 2020 **Revised:** December 24, 2020 **Accepted**: 5 January, 2021

Abstract

One of the main concerns of a test is constructed validity - whether it tests what it is designed to test. The primary aim of this study was to explore the construct underlying the PSU-TEP reading test. The stimulated recall data produced by 16 proficient participants in the four parallel forms of the test were analyzed for cognitive processes the participants relied on to complete the reading test. The findings showed that slow and careful global reading was the most predominant strategy the participants used to get correct answers. Other frequently-used strategies were choice elimination, the use of vocabulary knowledge, local-search reading based on lexical access, and slow and careful local reading, respectively. The findings were consistent in all four test forms. The results suggest that the PSU-TEP reading test measured the construct of reading abilities which all contribute to reading comprehension: careful local and global reading, and vocabulary knowledge. In other words, the participants who achieved high scores on the PSU-TEP reading test possessed the abilities they need in real-life reading, especially the ability to read slowly and carefully to comprehend sentences and overall texts, and had a good command of vocabulary, which plays an important role in reading. The fact that choice elimination emerged as one of the frequently-used strategies was because of the use of multiple-choice questions in the test. Recommendations for further research and reading test developers are provided.

Keywords: Test Construct, Reading Test, Cognitive Processes for Reading

EDUPSU วารสารศึกษาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี
ปีที่ 33 ฉบับที่ 1 เดือนมกราคม – เมษายน 2565, หน้า 170-183

การศึกษาความตรงเชิงโครงสร้างของแบบทดสอบ PSU-TEP ด้านการอ่าน: รูปแบบกระบวนการคิดด้านการอ่าน

ธัญภา พลานุกูลวงศ์*
Ph.D. (Applied Linguistics), รองศาสตราจารย์
สาขาวิชาภาษาต่างประเทศ คณะศิลปศาสตร์ มหาวิทยาลัยสงขลานครินทร์
*ผู้ประสานงาน: thanyapa.c@psu.ac.th

วันรับบทความ: 14 สิงหาคม 2563 **วันแก้ไขบทความ:** 24 ธันวาคม 2563 **วันตอบรับบทความ:** 5 มกราคม 2564

บทคัดย่อ

ข้อกังวลประการหนึ่งของแบบทดสอบคือแบบทดสอบได้วัดสิ่งที่ต้องการวัดหรือไม่ การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาว่าความตรงเชิงโครงสร้างของข้อสอบทักษะการอ่านของมหาวิทยาลัยสงขลานครินทร์ นักวิจัยใช้ข้อมูลทางวาจาที่รวบรวมจากการกระตุ้นการเรียกความจำคืน ของผู้เข้าสอบที่มีความสามารถทางภาษาระดับสูงจำนวน 16 คน ซึ่งทำข้อสอบการอ่านในการจัดสอบที่ใช้ข้อสอบคู่ขนานรวม 4 ครั้ง เพื่อศึกษากระบวนการรู้คิดและกลวิธีที่ใช้ระหว่างทำข้อสอบ ผลการวิจัยพบว่าในภาพรวม กลวิธีการอ่านที่กลุ่มตัวอย่างนำมาใช้บ่อยที่สุดในการหาคำตอบที่ถูกต้องคือ *กลวิธีการอ่านอย่างช้า ๆ และพินิจพิจารณาในระดับย่อหน้าและระดับบทอ่าน* กลวิธีที่ใช้รองลงมาคือ *การตัดตัวเลือกทิ้ง การใช้ความรู้ด้านคำศัพท์ การกวาดหาคำที่มีความหมายคล้ายคลึงกันในคำถามและบทอ่าน และการอ่านอย่างช้า ๆ และอย่างพินิจพิจารณาในระดับประโยค* กลวิธีเหล่านี้ถูกพบในการทำข้อสอบทั้ง 4 ครั้ง แสดงให้เห็นว่าข้อสอบทักษะการอ่านของมหาวิทยาลัยสงขลานครินทร์วัดทักษะซึ่งมีความสำคัญต่อความเข้าใจในการอ่าน นั่นคือความสามารถในการอ่านอย่างช้า ๆ และอย่างพินิจพิจารณาในระดับประโยคและระดับองค์รวม กล่าวอีกนัยหนึ่งคือ กลุ่มตัวอย่างที่เข้าสอบแบบทดสอบวัดทักษะการอ่านของมหาวิทยาลัยสงขลานครินทร์ที่ได้คะแนนสูง จะต้องมีความสามารถในการอ่านในชีวิตจริง โดยเฉพาะอย่างยิ่งความสามารถในการอ่านอย่างช้า ๆ และอย่างพินิจพิจารณาเพื่อเข้าใจทั้งระดับประโยคและระดับบทอ่าน และต้องมีความรู้ด้านคำศัพท์เป็นอย่างดี เพราะความรู้ด้านคำศัพท์มีความสำคัญต่อความเข้าใจในการอ่าน การพบว่าการตัดตัวเลือกทิ้งเป็นกลยุทธ์ที่กลุ่มตัวอย่างใช้บ่อยเช่นกัน เป็นเพราะว่าลักษณะข้อสอบที่ใช้ในแบบทดสอบเป็นแบบตัวเลือก การวิจัยในครั้งนี้มีข้อเสนอแนะสำหรับการศึกษาครั้งต่อไปรวมไปถึงสำหรับผู้พัฒนาข้อสอบวัดทักษะการอ่าน

คำสำคัญ: ความตรงเชิงโครงสร้างของแบบทดสอบ แบบทดสอบการอ่าน กระบวนการคิดด้านการอ่าน

Introduction

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. 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 judgement 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. A test is said to have construct validity if it measures just the ability 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).

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).

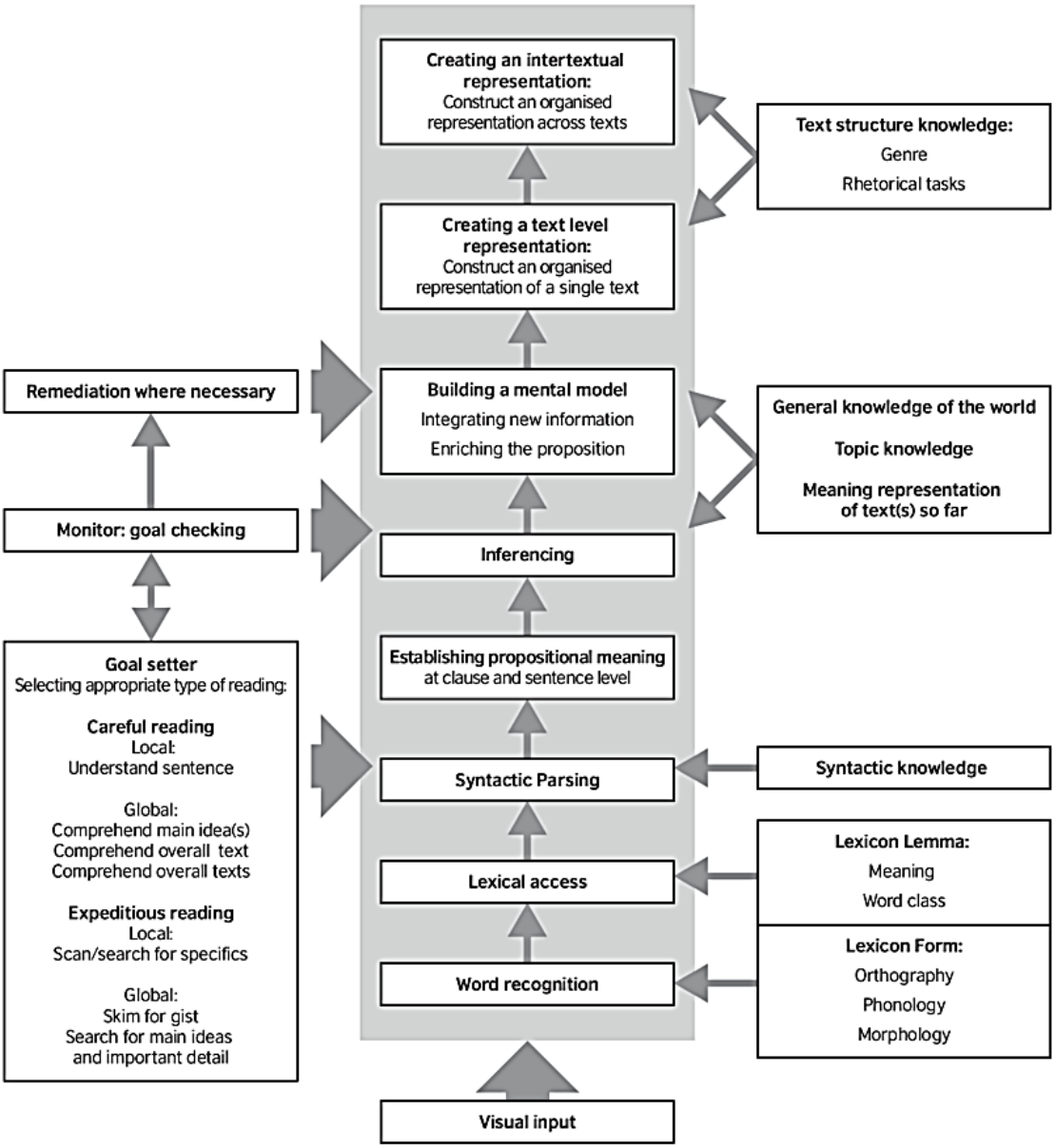
Alderson (2000) argued that "the validity of a test relates to the interpretation of the correct responses to items, so what matters is not what test constructors believe an item to be testing, but which responses are considered correct, and what process underlies them" (p. 97). The use of verbal protocol 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, Bachman, Perkins, & Cohen, 1991).

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 & Weir, 2009; Brunfaut & McCray, 2015). The model is shown in Figure 1 below.

Figure 1

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



The model 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 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, so-called lower- and higher-level processing (Urquhart & Weir, 1998, cited in Weir, Hawkey, Green & Devi, 2012a). 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* (multi-text representations).

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.*

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, Unaldi, & Devi, 2012b).

Verbal reports can be gathered either concurrently or after a language event, or retrospective reports. 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 (Green, 1998).

Stimulated recall is an introspective method of eliciting the thought processes taking place while an individual is doing a task. Stimulated recall procedures 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 in that no training is required for the participants to carry out a task (Gass & Mackay, 2017).

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. Rupp, Ferne, and Choi (2006), for example, used concurrent verbal protocols to look 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 and 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. The data indicated that the test as a whole sampled widely from the construct of reading, and the findings provided key information for test validation purposes.

Anderson et al. (1991) used think-aloud protocols to identify the strategies that test-takers employed when taking a multiple-choice reading comprehension test. No relationship between strategies used and item types was found, nor any relationship between item difficulty and test-takers' ability to understand main ideas, direct statements, and inferences.

Lim (2017), in an attempt to validate reading questions, used eye-tracking data to investigate Chinese ESL learners' reading processes in the iBT TOEFL reading section. The results showed that the participants predominantly exercised careful reading skills below paragraph level. Also, expeditious reading rarely occurred.

Purposes of the Study

Prince of Songkla University Test of English Proficiency (PSU-TEP) is a skill-based proficiency test, consisting of 4 sub-tests: Structure and Reading, Listening, Writing, and Speaking tests. The test is offered to PSU students and 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.

One major concern of any test is whether it tests what it is supposed to test. As a constructor and 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 primary aim of this study was, therefore, to explore the construct measured by the PSU-TEP Reading Test by examining the cognitive processing

of test-takers while completing reading comprehension items. Using Khalifa and Weir's (2009) model of cognitive processing in reading and stimulated recall would reveal whether the PSU-TEP Reading Test measured what it was set to measure. The following research questions were formulated in this study.

1. What cognitive processes do test-takers employ while tackling reading comprehension questions?

2. What does the PSU-TEP reading test measure?

Methodology

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 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).

Research Instruments

Four parallel forms of the PSU-TEP reading part administered in February, April, June, and December, 2016 were used as research instruments. Each form consisted of 2 passages, each followed by 10 multiple-choice questions, totaling 20 items. The questions focus on checking test-takers' understanding of the gist, main idea, purpose of the text, details, word meaning, references, inferences, and the tone of the passage.

Stimulated recall analyses

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

Table 1

The coding framework for the stimulated recall data

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)

Table 1 [Continued]

Code	Definition
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) 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 reading test administered in February, April, June, and December, 2016.

The participants were familiarized with the nature of retrospective stimulated recall procedures 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 reading task.

Each participant took the test on different occasions. Immediately after the completion of the reading comprehension 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 … ?*", etc.

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 97.2, 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 comprehension items. 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).

Research Findings

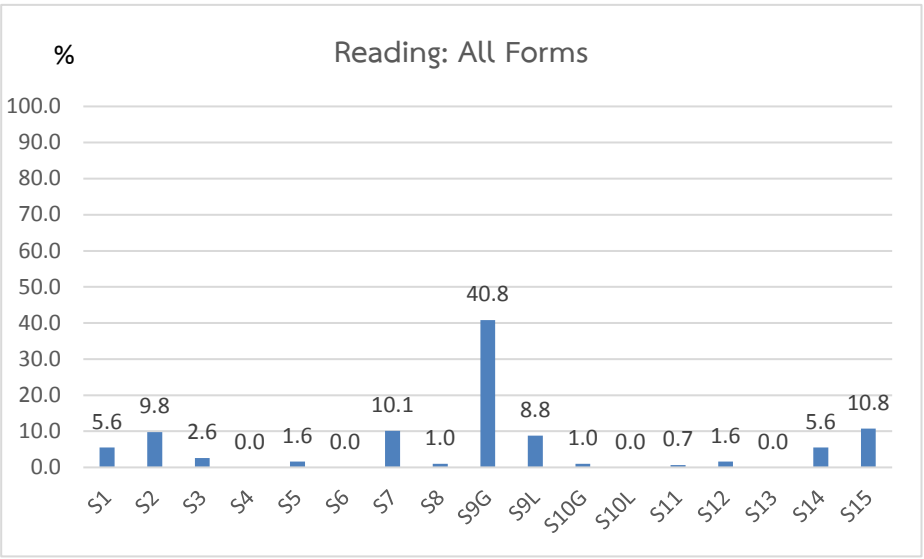
Insights were gained through the detailed analysis of the stimulated recall data produced by the 16 participants who were purposively selected to participate in stimulated recall sessions. Several cognitive processes revealed may reflect what the test really measured.

The stimulated recall data showed that before approaching the reading test, only one out of the 16 participants indicated that she previewed the reading passage to get a general idea of what it was about before attempting the questions.

Analysis showed that the reading tasks involved a wide range of strategies. The most popular strategy the participants activated to arrive at correct answers was strategy No. 9G (*Slow and careful global reading*). Of the total of 306 strategies the 16 participants used to arrive at correct answers, they heavily relied on strategy No. 9G, *reading the text or part of it most often globally* (40.8% of the total strategies used). Other four popular strategies employed at a relatively same frequency were strategy No.15 (*choice elimination*, 10.8%), No.7 (*knowledge of vocabulary,* 10.1%), No.2 (*local-search reading based on lexical access,* 9.8%), and No.9L (*slow and careful local reading,* 8.8%). Evidence of other strategies activated, No.1 and No.14, was found, but they occurred in a less frequent manner. However, there was no use of strategies Nos. 4, 6, 10L, and 13. Figure 2 below presents an overview of the strategies used.

Figure 2

Cognitive strategies employed in reading tasks across four test forms



Detailed analyses into the stimulated recall data in each test form showed a rather similar pattern of strategies used. Strategy No.9G (*slow and 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* most often. Other popular strategies across the four forms included No. 2 (*local-search reading based on lexical access)*, No. 7 (*vocabulary knowledge)*, and No.15 (*choice elimination*). Strategy No.9L (*slow and careful local reading*), and No. 2 were frequently used in three out of the four forms.

Of all 80 strategies used to get correct answers in Form 1, the top three strategies were No. 9G (*slow and careful global reading,* 22.5%), No.2 (*local-search reading based on lexical access*, 17.5%), and No.9L (*slow and careful local reading*, 15%). In Form 2, of all 87 strategies used for correct answers, the participants mainly relied on No. 9G *(Slow and careful global reading,* 48.3%), followed by No. 7 (*knowledge of vocabulary*, 11.5%), and No. 15 (*choice elimination*, 10.3%). Nearly half of 72 and 67 successful strategies that helped the participants taking Form 3 and Form 4 to answer the questions successfully were No.9G (47.2% and 46.3%, respectively) with 8.3% and 13.4% of No.9L (*slow and careful local reading*) as the second most popular strategies, respectively. The next popular strategies were No. 15 (*choice elimination*) and No.7 (*knowledge of vocabulary).*

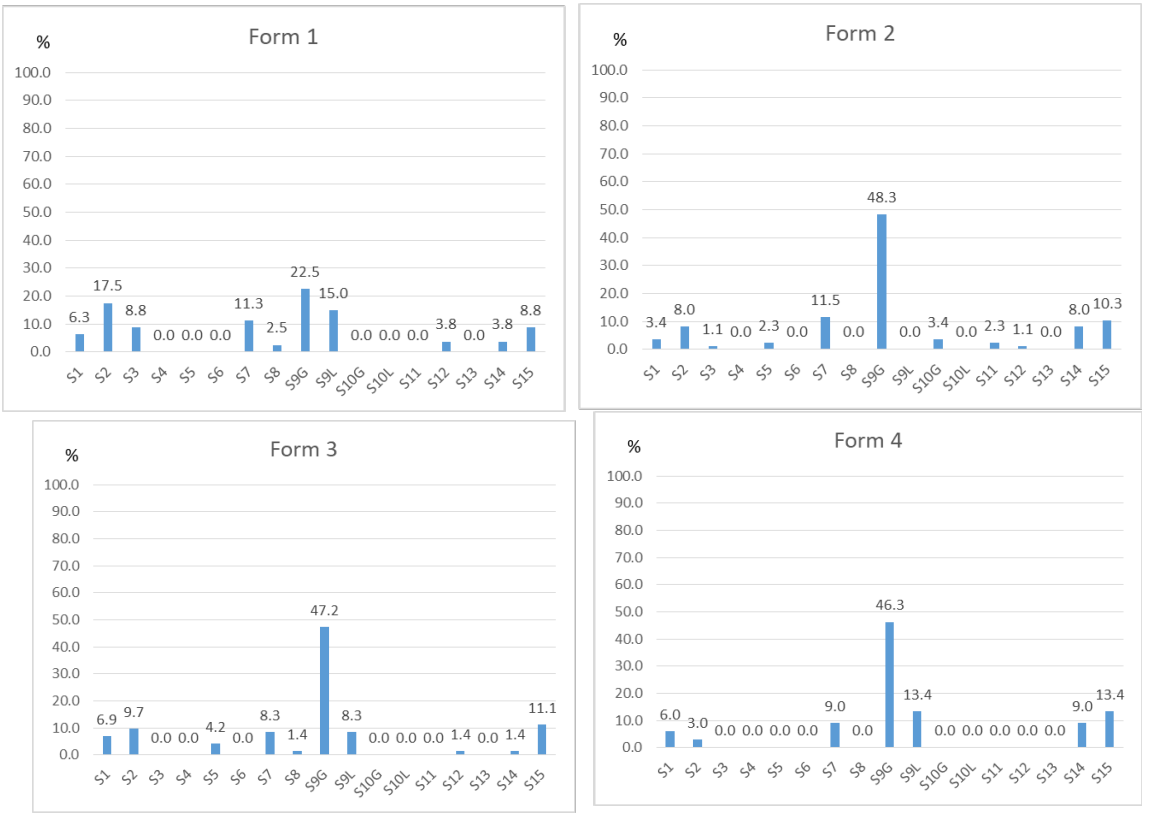
As previously discussed, a combination of strategies was used before the participants arrived at a correct answer. So, it should be noted that although some participants who explicitly stated to eliminate choices, they ended up eliminating choices to narrow down their choices after the use of other strategies.

A relatively consistent pattern of strategy use was found in the four parallel forms, with *slow and careful global reading* emerging as the most popular strategy and

slow and careful local reading, *knowledge of vocabulary*, *local-search reading based on lexical access* being among the other most popular ones, indicating that the four test forms measured relatively the same abilities, or construct. Figure 3 below displays the frequency of strategy use in all four forms of the reading tasks.

Figure 3

Cognitive strategies employed in reading tasks in each test form



Discussion

To answer the first research question, different strategies were found in helping the participants arrive at correct answers. Overall, *slow and careful reading at both global and local levels* (Nos. 9G and 9L), *knowledge* of *vocabulary* (No.7), and *local-search reading based on lexical access* (No.2) appeared to have been most significantly beneficial in helping the participants arrive at correct answers. This is confirmed by analysis of each test form. *Choice elimination* (No.15) emerged as one of the frequently-used strategies because of the use of multiple-choice questions in the test.

For the second research question, since the cognitive processes the participants employed for correct responses in the reading part were *slow and careful reading at both global and local levels* and *vocabulary knowledge*, which plays a very important role on reading tests (Alderson, 2000), it might be possible to conclude that the PSU-TEP Reading Test measured the test-takers'

abilities to *read slowly and carefully at both sentence and text levels,* and their *vocabulary knowledge*, something that readers need in real-life reading. The combination of these abilities appeared to have been most significantly beneficial in helping participants arrive at correct answers. The types of strategies as reported using by the participants were in the range of reading strategies proposed in Khalifa and Weir's (2009) model of cognitive processing in reading.

Slow and careful reading, based on Khalifa and Weir's (2009) model of cognitive processing in reading, is intended to extract the complete meaning from presented material, and it can take place at a local and global level (within or beyond the sentence up to the level of the complete text). It involves slow, careful, linear, incremental reading for comprehension. The fact that *slow and careful reading* at text level (strategy No. 9G) was found to be most frequently used by the participants and that *slow and careful local reading* (strategy No.9L) was one of the most used strategies, reveals that the participants may first processing texts to establish *local comprehension* of the text (i.e. understanding of propositions at the sentence and the clause) and then afterwards, *global comprehension* (understanding of macro- propositions including the main ideas, and the links between the micro-propositions). The combination of these two strategies, with the participants' vocabulary knowledge, helped them to arrive at correct answers.

The absence of expeditious reading such as *scanning* and *skimming* was noted in this study. *Expeditious reading* involves quick, selective, and efficient reading to access desired information in a text. Expeditious reading is difficult because it demands rapid recognition and sufficient practice in reading in the target language. In fact, many studies have found that for some readers, reading quickly, selectively, and efficiently is more difficult than reading carefully and efficiently (Khalifa & Weir, 2009). In addition, the fact that the participants were informed that they could spend as much time as they desired to complete each task before reporting their cognitive processes might have led to the absence of expeditious reading in the present study.

The findings of the present study were consistent with Lim's (2017). Lim found that the participants predominantly exercised careful reading skills below paragraph level and that expeditious reading rarely occurred.

Recommendations

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

1. Implication

1.1 The use of test-takers' cognitive processes in the process of test validation has been proved to provide valuable insights into what thinking processes test-takers were undergoing during a test, and this can help enhance the validity of a test. Test validation should therefore include the use of test-takers' cognitive processing.

1. 2 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. Areas of Further Research

The findings of this current study showed a limited range of strategies as proposed by Khalifa and Weir's (2009) model of cognitive processing in reading. The participants mainly approached the reading tasks with slow and 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) study: quick and selective search reading followed by intensive careful reading of relevant parts.

These findings, however, were subject to at least one limitation. The participants were informed that they could spend as much time as they needed while completing the test to encourage them to finish all items for stimulated recall data.

It is therefore recommended that 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 under time constraints in order to reflect the reading processes that readers would employ in real-world academic reading context. Another extensive study of the cognitive processing of test-takers of the PSU-TEP Reading Test is then called for, using a larger number of proficient test-takers.

References

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.

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,* Cambridge: Cambridge University Press.

Heaton, J. B. (1988). *Writing English language tests.* Longman: New York.

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

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

Lim, H.J. (2017). Exploring test-takers' cognition in a high-stakes reading test: An eye-tracking study. *The Journal of Asia TEFL*, *14*(3), 482-500.

Messicks, S. (1989). Meaning and values in test validation: The Science and ethics of assessment. *American Educational Research Association, 18*(5), 5-11.

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.

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

Weir, C. J., Hawkey, R., Green, A. & Devi, S. (2012a). The cognitive processes underlying the academic reading construct as measured by IELTS. In L**.** Taylor & C. J. Weir (Eds.), *Studies in Language Testing 34* (pp. 212-269). Cambridge: Cambridge University Press.

Weir, C. J., Hawkey, R., Green, A., Unaldi, A. & Devi, S. (2012b). The relationship between 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* (pp.37-119). Cambridge: Cambridge University Press.