■ THE CONSTRUCT OF READING AND ITS OPERATIONALIZATION IN THE INTERNET-BASED TEST OF ENGLISH AS A FOREIGN LANGUAGE

MILAN MILANOVIĆ[1]
University of Kragujevac, Faculty of Philology and Arts,
English Department,
Kragujevac, Serbia

Ovaj rad se bavi definisanjem konstrukta i njegovom operacionalizacijom putem tekstova i zadataka odabranih kako bi se testiralo razumevanje pročitanog teksta u standardizovanom testu kao što je TOEFL iBT. Kako definisanje konstrukta predstavlja težak zadatak kako za istraživače tako i za iskusne tvorce testova, daje se predlog potencijalnih metoda koje na tom putu mogu biti od pomoći. U drugom delu rada pristupa se analizi sposobnosti koje se smatraju potrebnim kako bi kandidat pokazao da može sa efikasnošću da procesuira pročitani tekst, razume ga i odgovori na pitanja kakva se postavljaju u ovom testu. Tekstovi su iz domena akademske sredine koja ujedno predstavlja i ciljni jezički domen. Imajući u vidu da se ovaj test realizuje uz pomoć računara i interneta, daje se pregled formata ispitnih pitanja koja se koriste u testu i koja, zajedno sa tekstovima na osnovu kojih su formirana, predstavljaju način da se konstrukt operacionalizuje.

Ključne reči: konstrukt, jezičke veštine, razumevanje pročitanog teksta, format ispitnih pitanja, TOEFL iBT.

1. INTRODUCTION

Language tests are intended to measure one or more constructs, which need to be defined and operationalized in a particular test. In all large-scale high-stakes standardized language tests, such as TOEFL iBT, reading is assessed as a constituent part of the overall language proficiency. A standardized test, such as this one, may be fairer to test takers (Powers 2010), because by using multiple methods and formats, it

---

1   Kontakt podaci (Email): milanprofa@yahoo.co.uk

helps to reduce "test-method effect"[2] and allows test takers to demonstrate a wider range of their abilities.

In this paper I will discuss constructs of reading and question formats utilized to measure different aspects of the constructs in the Reading section of the TOEFL iBT. The Internet-based Test of English as a Foreign Language takes the integrated-language skills approach to assess not only receptive skills (reading and listening) but also the so-called productive language skills (speaking and writing). This integration of skills means that reading extracts are used as source materials for Speaking and Writing sections. These sections will not be analyzed here, although it must be noted that one such analysis would be more than useful due to potential impact which integration of skills may have on language instruction.

In this paper I will make use of the information available in the ETS's publications targeting prospective test takers and language teachers to reflect on the type of reading constructs the test is intended to measure as well as the question formats utilized to assess test takers' reading skills. The examples provided here are available in free online practice materials (through ETS's website available at http://www.toefl.org).

2. TOWARDS DEFINING A CONSTRUCT

Defining a construct is not an easy task because there are many factors to be considered, with all their various potential influences on test takers' performance and inferences based on that performance. Furthermore, there seems to be no consensus on what constitutes the construct(s) of reading and this, in turn, makes it harder for test developers to rely on the theory of reading in their effort to operationalize it through test tasks. However, the very fact that tests are designed and tasks created signalizes that there is some extent of agreement between the stakeholders on what skills readers/test takers need to possess. How do we define the construct of reading?

The following discussion was Buck's discussion of the constructs of listening, but it seems equally applicable here. According to Buck there seem to exist at least three possible paths: the first is competence-based, meaning that it takes into consideration competences that test takers need to possess. The second is termed "task-based" because it focuses on tasks a test taker should be able to complete (both in testing and real-world situations). The third option is to try and combine the first two (Buck 2001: 102). Buck suggests that the starting point should be to determine the purpose of assessment and specify the target language use situation. If there is no clearly distinguished target language situation, the first path, that of taking the competence-based approach seems the only logical possibility. There are several frameworks developed to facilitate the process of selecting the appropriate competence(s) to be assessed in a test (Bachman and Palmer 1996: 68; North and Schneider 1998; Alderson 2000: 120; Buck 2001: 94; Council of Europe 2001). The selection of the appropriate framework or combining competences outlined in more than one framework will be

---

2   For more about "test method effect" refer to Bachman (1990) and Chapelle (2000).

influenced by the purpose of assessment. There is another approach which is applicable in the situations where target language use is explicit. In that case, the construct may be defined in terms of the task corresponding to the target language use situation.

In defining the construct(s) of reading, it is the purpose of reading assessment that determines how the construct is conceptualized and operationalized. Constructs of reading should be based on theory of reading and models which have been developed to assess various aspects of test takers' reading ability. As Alderson states "every test is intended to measure one or more constructs" (Alderson 2000: 118) which will have to reflect the theory from which the ability to be tested is derived. In other words, if a construct is derived from a theory of communicative competence, the abilities to be measured are those of the communicative nature. Consequently, tasks designed to measure such constructs will target the test takers' communicative competence and their responses will provide the basis for making inferences of their ability to use such competence in non-test situations.

The construct is operationalized through reading passages and the tasks/items which are designed to assess test takers' reading abilities. Each of the tasks designed in the test of reading is used to measure a part of the construct. For example, if the construct of reading, in a particular test, is defined so as to include "skimming", it is operationalized by test items requiring test takers to go quickly through the test and identify specific pieces of information in the passage. The construct of reading may be rather complex to include skills such as evaluation, synthesis or categorization. With such construct, test items need to be carefully designed to measure all the abilities identified as the construct. In other words, if the operationalization of the theoretical construct fails to represent all its theoretical aspects, e.g. because it is incomplete, the construct is underrepresented. Construct-underrepresentation is not the only trap that test developers may fall into. A test may be designed so as to assess the abilities which are not specified by the construct definition, which exemplifies construct-irrelevant variance (Messick 1989). For example, if a test taker's performance in a computer-assisted test of reading is influenced by their ability to use the mouse to scroll the text, then maybe the test measures not only the test taker's reading ability but also their computer literacy.

The construct which underlies a language test is detailed in the test specifications. However, often enough test specifications, particularly in high-stakes standardized tests, are considered proprietary information and not easy to get hold of. So how are test users, including test takers, made aware of the construct being measured in a particular language test? This depends on the testing organization as well as on the test purpose. In the first part of this paper, I will provide a description of the Reading section and outline the constructs of reading as measured by this section of the TOEFL iBT, using the information provided in the test preparation materials, more specifically in the ETS's publication entitled "TOEFL iBT Tips: How to prepare for the TOEFL iBT" (ETS 2007). An attempt to define constructs of reading in this test will be followed by specification of reading question formats used to operationalize the construct of reading in TOEFL iBT.

3. DESCRIPTION OF THE READING SECTION OF THE INTERNET-BASED TEST OF ENGLISH AS A FOREIGN LANGUAGE

The Reading section of the TOEFL iBT measures the ability of non-native speakers of English to understand university level academic texts and passages. Test takers are given between 60 and 100 minutes to complete this section which consists of 3-5 texts on different topics. The difference in time allocated to the Reading section comes from the distinction between the short and the long version of the section. In the long version section, test takers are required to read an additional text and respond to additional questions, whose purpose ranges from experimental to equating (for full review of TOEFL iBT refer to Alderson 2009). To secure test takers' undivided interest for all texts within the section, test developers do not reveal which questions are experimental so that the candidates have to attempt every question within the section. Unlike the texts provided in the previous version of TOEFL, i.e. in the computer-based Test of English as a Foreign Language, in which there were four to five texts whose length did not exceed 350 words, in the TOEFL iBT the number of texts remained almost the same (3-5), but the length of the reading passages increased to 600-700 words. The TOEFL iBT uses passages from real university-level textbooks, and the questions, whose number is in the range from 12 to 14 per passage, related to them are developed so as to measure test takers' basic comprehension, their ability to make inferences or read for detail, and a new "reading to learn" tasks are added. The excerpts in the actual test undergo as little modification as possible to ensure that the test measures what it is intended to measure. In some cases, the topic covered by a passage may be totally unfamiliar to test takers, but this is of little importance as the passage contains all the information needed to answer the question. The Reading section is linear, not computer-adaptive, unlike the previous, computer-based test (for review see Banerjee 2003), and allows for a full range of abilities to be measured. This also means that test takers can go back and forth and change their answers, or they can select the Overview tool on their Toolbar (Milanović 2010: 20) to check if all questions have been attempted before submitting their response. Once they have submitted their response, the answers can no longer be retrieved and/or changed. There are words and phrases which are underscored and when clicked on, they open a glossary feature for a test taker to view a definition or explanation of the term. Note taking is allowed throughout the test, but notes are collected and destroyed at the end of each administration to ensure test item security.

4. THE CONSTRUCT OF READING IN TOEFL IBT

The Reading section is aimed at measuring test takers' ability to understand university-level texts (ETS 2007). This leads us to assume that target language situations are those typical of a university campus, in which the abilities of academic reading will be vital for a language user to use the information from the texts for a variety of communication-related purposes. Now that we know what potential target language situations are, we may assume that test tasks designed for TOEFL iBT are those corresponding to tasks students may be faced with in target language situations.

However, it may be argued that it seems that excerpts and tasks used in TOEFL iBT are not entirely authentic, in the sense that academic reading calls for much longer texts than are those used in TOEFL iBT. Another issue raised is that of the nature of comprehension – would test takers ask themselves exactly those questions that can be found in the Reading section? No matter what the answer to this question is, there is no doubt that reading is one of the skills essential for the academic success and for this reason test users expect from a testing service to provide tasks which will address the most important purposes for academic reading.

4.1 PURPOSES FOR ACADEMIC READING AND RELATED CONSTRUCTS

In the ETS's publication, there are identified three purposes for academic reading: Reading to find information, Basic comprehension, and Reading to learn. Although there is no obvious attempt for these to be labelled as the constructs of reading, there seems to be enough information to deduce that the reading construct in the Reading section of the test will address the purposes for academic reading as identified in the ETS publication (ETS 2007):

Reading to find information addresses the issues of reading fluency and rate, so the skills necessary for this purpose include skimming and scanning. The latter is of particular interest in items requiring from a test taker to identify key facts and important information.

Basic comprehension involves understanding the general topic or main idea, major points, important details and facts. It seems logical that if test takers are to grasp details and important facts, they need to understand vocabulary used within the context provided.

Reading to learn is concerned with engaging an array of skills needed to recognize the organization and purpose of a passage and to understand relationships between ideas. This reading purpose also involves making inferences about the ways in which ideas throughout the passage connect. Questions developed to address this purpose include categorizing information within a chart or making a summary by ordering answer choices into "the best" sequence.

The theory is operationalized through selection of texts (in the Reading section) and the test items/questions developed to measure comprehension of the input materials. However, what is measured on a test is not so much the ability as the interaction between the ability and the tasks developed to measure it, so that the use of appropriate tasks "can enhance the validity of [...] inferences, provided they are appropriately chosen and explicitly related to [...] constructs" (Alderson 2000: 117). So what are the appropriate texts and tasks/questions/items to help measure the construct(s)? To attempt to answer this question, the nature of the texts will be referred to and an overview of the most commonly used question formats in the Reading section of the test will be provided.

4.2 TEXTS IN THE READING SECTION OF TOEFL IBT

Reading passages are taken from university-level textbooks, meaning that they represent authentic input data and cover a multitude of subjects a prospective

student could encounter at a university. Regardless of the subject field of the passage, test takers are not required to have any background knowledge of the topic. On the contrary, the influence of background knowledge is what is to be avoided because it may compromise test validity by increasing content bias or resulting in a disadvantage for those test takers who know little or nothing about the topic. For this reason, items are not script-based, i.e. they do not call for integrating information from the text with background knowledge (Pearson and Johnson 1978).

As is the case with any language which forms input in a test, the reading samples may be categorized in many different ways, depending on the purpose of the categorization. Many frameworks of test task characteristics agree that language of the input may be described in terms of its grammatical, pragmatic and discourse features, or they can be utilized while attempting to determine the relationship between the input and expected response (for more on frameworks see Bachman 1990; Bachman and Palmer 1996, Douglas 2000; Chapelle and Douglas 2006; Milanovic 2010). However, it is not intended to provide an analysis of any given Reading section of the test, so the details relating to the above-mentioned features will not be discussed.

4.3 QUESTION FORMATS

The Reading section of the test uses a wide range of selected responses in the form of multiple-choice questions with one or more correct answers. The questions refer to the reading passage they follow and are either textually explicit or textually implicit. They either refer to the information found within one or more passages, requiring that test takers either synthesize information or make inferences about what they have read. Apart from the traditional multiple-choice (MC) question format with one correct answer, the Reading section includes multiple-choice questions with more than one correct answer. To respond to some of these tasks, test takers are required to manipulate the mouse and order the answer choices following the instructions. The same is true for the question format introduced by the latest generation of TOEFL – 'reading to learn' questions. In this question format, test takers are supposed to synthesize information from the passages and manipulate the computer equipment in order to fill in various charts. Throughout the test administration test takers are allowed to see the text, and additional help is offered in the form of the so-called 'glossary feature'. The latter provides an explanation or a definition of a term or special purpose words and phrases (ETS 2007:10).

4.3.1 TRADITIONAL "FOUR-OPTION SINGLE ANSWER" MULTIPLE CHOICE QUESTIONS

The items typical of this format are those requiring test takers to scan the text for key facts and important information, to demonstrate they can identify the main idea, major points, important facts and details, to comprehend an argument or an idea by making inferences, to display their knowledge of grammar and vocabulary in context (Example 1), etc.

Example 1[3]:
The word pose in the passage is closest in meaning to

☐ claim   ☐ model   ☐ assume   ☐ present

Click on an oval to select your answer. To choose a different answer, click on a different oval.

4.3.2 NON-TRADITIONAL "FOUR-OPTION SINGLE ANSWER" MULTIPLE CHOICE QUESTIONS

There are usually two types of questions in this format: Single-answer questions with four choices ask test takers to "insert a sentence where it fits best" while paraphrase questions prompt test takers to paraphrase a sentence from the passage.

In the case of "insert a sentence where it fits best" questions, test takers are prompted to take a look at a certain number of squares (Example 2) indicating where a sentence can be added to a passage. To place the sentence at the appropriate place in the text, test takers are asked to click on a square to add the sentence to the passage. They can change their mind and click on a different square to place the sentence somewhere else. Since the sentence fits best at only one place in the text, this is a single-answer four-choice question format.

Example 2[4]:
Look at the four squares [■] that indicate where the following sentence could be added to the passage.
This made it easy for the artists to paint and display them for the rest of the cave dwellers.
Where would the sentence best fit?
Click on a square [■] to add the sentence to the passage.

Paraphrase questions assesses the test takers' ability to select the answer choice which best expresses the essential information in the highlighted sentence in the passage. Additionally, test takers are informed that incorrect choices change the meaning or leave out essential information. Exploring the relationships between test questions and text in TOEFL 2000 Project, researchers discuss these relationships in terms of types of information requested, type of match, and plausibility of distracters (Jamieson *et al.* 2000: 30). Type of match is of interest for this paper because it refers to "the way in which examinees process text to respond correctly to a question". Specifically, this may include at least four strategies test takers use in processing the text: locating, cycling, integrating, and generating. Integrating tasks seem to be particularly demanding since they require that examinees "pull together two or more pieces of information from the text according to some type of specified relation". In practice this means that test takers

---

3   Example taken from http://www.toefl.org, free online practice.
4   Example taken from ETS (2006).

need to be able to make comparisons and/or contrasts, or, for instance, to identify cause and effect relations. (Jamieson *et al.* 2000: 22). This becomes evident in test items which call for paraphrasing information, or making inferences, but it is also evident in the question format discussed below.

4.3.3 "READING TO LEARN" QUESTIONS

New partial-credit "reading to learn" questions with more than four choices and more than one possible correct answer. These questions are aimed at testing the test taker's ability to recognize how the passage is organized and understand the ways in which facts and ideas in different parts of the passage are related to one another (ETS 2007). Test takers are prompted to either sort information and place the text options provided into a category chart (Figure 1) or to complete the summary. The former requires that test takers click on the choices and drag them over the screen to a desired box within the chart. The answers can be changed before the final response is submitted at the end of the section.
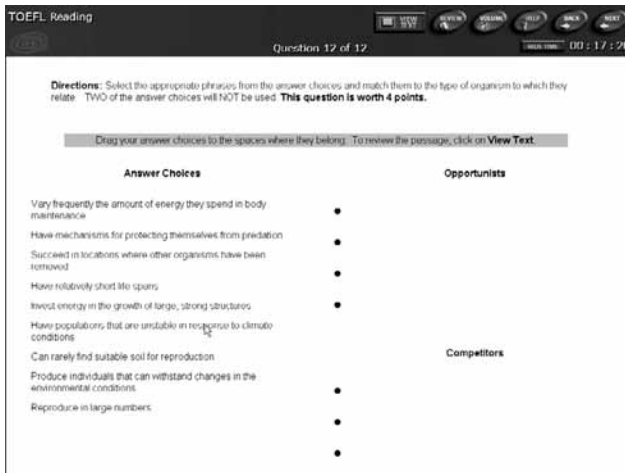

Figure 1: Category chart question example (ETS: 2007)

5. CONCLUSION

In the first part of this paper, what is discussed are difficulties that many researchers have encountered while trying to define a construct of particular degrees of ability in language. There seem to be many different ways to define constructs, and many researchers agree that purpose of assessment and target language use situations, in many respects, determine the way in which a construct will be defined. The second part of the paper is devoted to the analysis of the construct of reading in the Reading section of TOEFL iBT and operationalization of the construct through various texts and

question formats used to measure comprehension. This version of the test makes use of computer technology to create new tasks, taking into consideration issues such as construct-irrelevant variance and "test method effect". Academic reading skills are assessed by addressing the following three purposes for academic reading: basic comprehension, reading to find information, and reading to learn. Since they are aimed at testing academic reading, the reading passages selected for the Reading section cover a wide range of fields that a student would encounter at a university setting, and these passages may be classified as expository, argumentative and historical. Test takers are not required to possess any background knowledge of the topic because the questions being text-explicit or text-implicit refer to the passages. As was the case with previous versions of this test, the current version uses selected responses, although some of the question formats are made possible by computer delivery. Apart from using traditional and non-traditional multiple choice questions with single or multiple correct answers, new 'reading to learn' question format is introduced. Grammar and vocabulary are tested in all sections of the test, but in context rather than in isolation.

REFERENCES

Alderson, J. C. 2009. Test review: Test of English as a Foreign Language TM: Internet-based Test (TOEFL iBT®). *Language Testing* 26 (4), 621-631.

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

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

Bachman, L. F. and A. S. Palmer. 1996. *Language testing in practice.* Oxford: Oxford University Press.

Banerjee, J. 2003. The TOEFL CBT (Computer-based test). *Language Testing* 20 (1), 111-123.

Barron's TOEFL i-BT Test preparation with CD, 2008.

Buck, G. 2001. *Assessing listening.* Cambridge: Cambridge University Press.

Chapelle, C. A. 2001. *Computer applications in second language acquisition: Foundations for teaching, testing and research.* Cambridge: Cambridge University Press.

Chapelle, C. and D. Douglas. 2006. *Assessing language through computer technology.* Cambridge: Cambridge University Press.

Council of Europe. 2001. *Common European Framework of Reference: Learning, teaching, assessment.* Cambridge: Cambridge University Press.

Douglas, D. 2000. *Assessing languages for specific purposes.* Cambridge: Cambridge University Press.

ETS. 2006. TOEFL iBT: Sample Questions. [Internet]. Available at: http://languagetesting. info/articles/artlt.html [11.01.2011].

ETS. 2007. TOEFL iBT tips. How to prepare for the TOEFL iBT. [Internet]. Available at: http://www.ets.org/Media/Tests/TOEFL/pdf/TOEFL_Tips.pdf [11.08.2010].

Jamieson, J. *et al.* 2000. *TOEFL 2000 Framework: A Working Paper.* Princeton, NJ: Educational Testing Service.

Messick, S. 1989. Validity. In R. L. Linn (ed.) *Educational Measurement. 3rd edition*. New York: Macmillan, 13-103.

Milanović, M. 2010. Test Task Characteristics in Computer-assisted (Internet-Based) English Language tests of Reading and Listening: TOEFL i-BT. Unpublished Master's Thesis, University of Belgrade.

North, B. and G. Schneider. 1998. Scaling descriptors for language proficiency scales. *Language Testing* 15 (2), 217-262.

Pearson, P. D. and D. D. Johnson 1978. *Teaching reading comprehension*. New York, NJ: Holt, Reinhart and Winston.

Powers, D. E. 2010. The case for a comprehensive, four-skills assessment of English-language proficiency *R&D Connections* 14. Princeton, NJ: Educational Testing Service.

SUMMARY

THE CONSTRUCT OF READING AND ITS OPERATIONALIZATION IN THE INTERNET-BASED TEST OF ENGLISH AS A FOREIGN LANGUAGE

In this paper an effort is made to define the construct of reading followed by the operationalization of the construct through texts and tasks selected in order to assess reading comprehension in a standardized test exemplified by TOEFL iBT. As defining a construct is by no means an easy task for researchers or test developers, a number of potentially useful methods is suggested to help them struggle through test development process. In the second part of the paper, abilities deemed necessary for efficient text processing, text comprehension, and responding to the tasks designed to measure comprehension in this test are analyzed. TOEFL iBT makes use of academic reading as it is academic environment which is defined as a target language domain. Given that the test is administered by computer and via the Internet, there is an overview of question formats utilized in the test in order to operationalize the construct of reading.