



Proceedings of the 1st English Education International Conference (EEIC) in conjunction with the 2nd Reciprocal Graduate Research Symposium (RGRS) of the Consortium of Asia-Pacific Education Universities (CAPEU) between Sultan Idris Education University and Syiah Kuala University

November 12-13, 2016, Banda Aceh, Indonesia



LOOKING AT PBT (PAPER AND PENCIL-BASED TEST) AS A STANDARD TEST

Hendra Heriansyah

Syiah Kuala University, Banda Aceh, INDONESIA

Email: hendrisa@unsyiah.ac.id

Abstract

Generally, there are two kinds of test namely teacher made test and standard test. TOEFL is one of international standard tests that is acknowledged by many people around the world especially for those who want to pursue their further education in English speaking countries. This paper addresses and reviews a model of TOEFL that is commonly taken by test takers related to Paper and Pencil-Based Test (PBT). The structure of this paper falls into three parts. The first part is introduction that presents a general description of the TOEFL. The second part is discussions that consist of the description that becomes the underlying model of language test, the sections of test, and some things that are relevant to the test such as validity, reliability, practicality, and washback. The third is conclusion related to the discussion results and some remain that still become the problematic issues.

Keywords: TOEFL, PBT, validity, reliability, practicality, washback.

INTRODUCTION

TOEFL is a registered trademark of Educational Testing Service (ETS). The test was originally developed at the Center for Applied Linguistics led by a linguist who names Dr. Charles A. Ferguson. The TOEFL Committee of Examiners comprises 12 specialists in the field of linguistics, language testing, English as a foreign or second language, teaching or research. The main responsibility of this committee is to advise on TOEFL and make sure that the measurement of English language proficiency test is valid according to the current trends and methodologies.

The purpose of Test English as a Foreign Language (TOEFL) is to measure the ability of non-native speakers of English to use and understand North American English as it is spoken, written, heard in college and university settings (www.examenglish.com/TOEFL/index.php). It is also an admission requirement for non-native English speakers at many English-speaking colleges and universities. In addition, most people who take the TOEFL are planning to study at colleges and universities where medium of instruction is in English. It is also used by many government agencies, scholarship programs, and licensing/certification agencies that use TOEFL scores to evaluate English proficiency. The score of test is valid for two years and cannot be used again when a candidate's language proficiency has changed since the date of the test. Colleges and universities usually consider only the most recent TOEFL score.

In dealing with the model of language test, TOEFL follows Canale and Swain's model of communicative competence. Canale and Swain (1980, cited in Fulcher & Davidson, 2007) present a model of knowledge what they called 'communicative competence'. They (ibid) argue 'communicative competence refers to the interaction between grammatical competence or knowledge of the rules of grammar, and sociolinguistic or knowledge of the rules of language use' (p.

38). In terms of assessment, according to Canale and Swain (1980, cited in Fulcher & Davidson, 2007) a test needs to include both aspect of communicative competence through tasks that require communicative performance. The communicative competence itself is composed of grammatical competence, sociolinguistic competence, and strategic competence (Fulcher & Davidson, 2007). It means that the use the target language in real communication is determined or influenced by some competencies above. Therefore, in order to get success in using L2, someone has to be able to use the competency above through actual language use in daily interaction.

So far we have looked at the aim of TOEFL and the underlying model of language. This following part discusses some things that deal with test review. The kind of language test that is reviewed in this paper is TOEFL especially paper and pencil-based test (PBT). The TOEFL PBT is a test of English proficiency with an academic focus and typically it is used in a North American academic context (en.wikipedia.org/wiki/TOEFL). In order to take this test, test takers can register through online or by using the registration form provided in the Supplemental Paper TOEFL Bulletin. Tests are administrated 6 times each year and the length of test is 2 hours. There are three sections within this test namely Listening, Structure and Written Expression, and Reading Comprehension. For these each section will be discussed as follows.

Listening Comprehension

Listening comprehension is the first part given in PBT, and then it is followed by Structure and Written expression section and the last part is Reading comprehension section. The language is used in listening section of course is native language or target language (L2). Listening is one of the receptive skills. Therefore, aural or audio is the channel of presentation that is used in this part in which the test takers have to listen to the conversations or talks through their own earphone. Sometime, the conversations or talks take place too fast because the native speakers speak quickly. The test takers have 40 minutes to answer 40 questions based on the input material. This section consists of 3 parts and the topic of conversations or talks can vary in this section. The first part contains 30 questions that relate to short conversations about hobby, favourite movie, and borrowing a book from the library, for example. The second part comprises 8 questions about longer conversations such as history, science, a scientific invention, etc. The last part has 12 questions about lectures and talks. These may contain special topic, discussion between a lecturer and some students, seminar discussion, lectures presentation and so on. In listening section, the test items are multiple choices, in which test takers have to choose one of the four options provided. In order to get a correct answer the test takers have to listen carefully the content of conversation, lectures or talks that are being talked and catch some important information are relevant to the topics, conversations, or talks that are being spoken or discussed and then quickly look at the options or responses given on the test sheet. In addition, test takers have to infer or interpret the meanings explicitly or implicitly what both of the speakers have spoken including vocabularies, idiomatic expressions, or phrasal verbs that are used in that conversations. The test taker may answer the questions after they have listened to the conversations, lectures or talks. The input that is listened by test takers and the questions that arise after that will influence participant's responses or answers.

Apart from some things above that deal with listening comprehension section, Brown (2004, cited in Kamyab, 2009) mentions there are some skills that can be used for listening comprehension what he called micro and macro skills. In terms of this, I only opt some skills that I consider that they are important for test takers when taking the TOEFL PBT. In relation to micro skills, there are some skills such as (1) recognize reduced forms of words; and (2) recognize that a particular meaning may be expressed in different forms. It is very common in TOEFL that both speakers use contractions or connected sounds (e.g. I'll go or I've got) when they are doing a conversation. In this case, the test takers have to be able to detect these sounds from the words that are pronounced by the speakers. Besides, the participants should be familiar with the expressions or statements may have the same meanings but these are expressed with different ways; for example active forms are changed to be passive forms (e.g. John read a book – a book was read by John), or perhaps the ways how to invite someone (e.g. I would like to invite you...; Would you like to come...?). Relating to macro skills, I presume the following skills are important for test takers. They are (1) recognize the communicative functions of utterance, according to situations and goals; (2) detect such relations as main idea,

supporting idea, new information, generalization, and exemplification; (3) distinguish between literal and implied meanings.

Regarding the macro skills above, the test takers have to recognize some communicative functions or speech acts that are used by the speakers such as requesting, apologizing, asking and giving information, accepting and refusing of an invitation, etc. These examples of language functions which may be used in a conversation, but it depends on the purpose of talk and where the conversation setting takes place (e.g. in a canteen where someone wants to request something like drink, here the language function used is 'requesting or how to request something). In listening test, the test takers need to identify also the main ideas of conversations or talks that are being spoken including the supporting ideas that relate to the conversation. In addition, the important thing after identifying some important information is that the test takers have to be able to distinguish and infer both explicit and implicit the meanings of conversations. These are some skills that may contribute or help the test takers in listening comprehension test.

Structure and Written Expression

Structure and written expression is second part within TOEFL PBT after listening section. The test takers are given 25 minutes to answer 40 questions in this section. This section comprises two types. The first type is about incomplete sentences in which the test takers have to complete the sentences by choosing one of the options from the four alternative answers provided on the question sheet. The second type is about error analysis where the test takers have to identify which the error that appears among the four underlined words. In order to be able to answer both the types of the question, the participants have to read the whole sentences first and then try to find or detect what is the correct answer for the question. All the questions in this section are multiple choices.

Reading Comprehension

This is the last section from TOEFL PBT. The channel of presentation that is used in this part is visual where the test takers can see directly the input materials given. The input materials that are presented in this section are in the form of reading passages. There are four or five passages given in reading comprehension test with the number of words for each passage is about 250-300 words. The kinds of topic can vary such as Geography, History, Science, Sociology, etc. The test takers are given 55 minutes to answer 50 questions by reading and understanding a number of reading passages provided. The participants have to select one out of the four options that they suppose it is the correct answer. The answers that will be chosen by test takers depend on their comprehension towards the content of reading passages. Moreover, that there are some things that can help the test takers in understanding the passages such as mastering of adequate vocabulary, knowing the kinds of text types (genres), having the background knowledge about the topic (schemata), skimming and scanning skill. For instance, skimming is required when we want to look for generally main idea of the reading passage, while scanning is needed if we want to look for specific information such as the date of event, place of the event, or the name of the person in that story. Knowing a variety of text types will help test takers in reading test because they can know what kind of text is tested. Schemata can be also an important part in reading test. Perhaps, it will help the test takers in order to give a better understanding about the topic because they have previous background knowledge about the reading passages that are questioned. In relation to mastery of vocabulary, I consider that there is no guarantee for someone who has sufficient vocabulary will help him in reading test. However, mastery of adequate vocabulary is helpful us when taking a test.

Aside from several matters above there are also some other skills that may help us in reading test. Brown (2004, cited in Kamyab, 2009) lists some skills that can be used for reading comprehension. For micro skill there are two skills that contribute in this section namely (1) recognize a particular meaning may be expressed in different grammatical; and (2) recognize cohesive device in written discourse and their role in signalling the relationship between and among clause. From the reading passages we can see and find there are some expressions that have the same meanings but the author uses different way in expressing them (e.g. I will visit my grandmother next month or I am going to visit my grandmother next month). In terms of this, the test takers have to be able to recognize that the use of future tense from both sentences may the

same meaning that indicates in which an activity will do in the future. For the second skill the test takers have to be able to identify some cohesive devices that link from among the clauses or paragraph like 'however, so that, therefore, in contrast' etc. By recognizing some cohesive devices the test takers are expected to be able to detect the relationship between one sentence/paragraph and other sentences/paragraphs. In relation to macro skill, there are some skills that can be used by test takers in reading test. They are (1) infer context that is not explicit by using background knowledge; (2) detect such relations as main idea, supporting idea, new information, generalization, and exemplification; and (3) distinguish between literal and implied meanings. In reading test sometime there are some author's expressions or statements that are not implied directly. Hence, it is better for test takers try to infer them according to the context and the content of reading passages and based on the test takers' background knowledge that the test takers have. This shows that the test takers can differentiate between literal and implied meaning from a reading passage. For the second one, it is very common in reading test the main idea is questioned. In order to answer this test takers have to be able to detect the relationship between one sentence and other sentences in a paragraph for example and try to conclude what the main idea of the paragraph is.

We have talked about the kinds of test section of TOEFL PBT including some facet theories and micro and macro skills that deal with this test. Now, we will look at some characteristics of quality of a test that are relevant to validity, reliability, practicality, and test impact.

Validity

Apparently, a good language testing has some characteristics as mentioned above. One of the characteristics of a good test is validity. Hughes (1989, cited in Fulcher & Davidson, 2007) asserts that validity in testing and assessment means "to measure accurately what is intended to measure" (p. 4). From this point it can be concluded that validity means testing of what should be tested or giving a test to measure what we want to measure from our students or a particular group.

Pertaining to TOEFL PBT, it seems this test is not fully valid because what we notice that the main priority is given to grammatical competence and two receptive skills (listening and reading) rather than communicative competence or language use and productive skill. We assume that the purpose of this test has changed although Canale and Swain state that both grammatical competence and communicative competence/language use are required in this test.

Apart from there are some things that make this test has some weaknesses. Firstly, in listening section the test takers have to retain the important information in their memory based on what they hear in order to be able to answer the question. If it is a brief conversation or short talks, there is possibility for test takers to remember that information, but if it is a long conversation/talk it may be problematic for them because they have to remember all of information on their mind. In this case, it can be said that the method of testing is used in listening test is not fully to test listening comprehension because it also tests our memory. Secondly, in terms of structure section there is no clear what are the essential grammatical components that the students need in their academic work. Besides, there is clear also how a broad range of types of structure should be tested in this section and if we look at the content validity of the test we can see the test designer does not balance in giving the proportion of question which is called difficulty index in term of language testing. Thirdly, in relation to reading comprehension section, there is no clear what the kinds of genre that should be known by the test takers (e.g. is it report, description, or explanation?). It seems the text types that are tested in this section like school textbooks. In terms of vocabulary in reading test, the items generally are multiple choices but in this case the test takers get difficulty because the distracters are the words that its meanings are so far the meaning of the word is being tested. As result the test takers have to use intellectual guessing and try to understand the words in context.

Reliability

According to Richards, Platt, and Weber (1985) reliability in testing is a measure of the degree to which a test gives consistent results. They (ibid) also add "A test is said to be reliable if it gives the same results when it is given on different occasions or when it is used by different people" (p. 243). From this definition it can be said that whenever a test is given by different people in distinctive situations, so the results of the test will be the same or consistent. If the results of the test differ from the first one, perhaps, the reliability of the test can be questionable.

Aligned with this, it may be reliability of the test is not consistent especially listening section because on this part strong memorization is really needed to retain information in our mind. In addition, the speed of reading, correct interpretation and appropriate inference are some things that influence the test result while doing the test. For someone who does not usually read quickly, and has no good interpretations towards the topics of reading passages may get difficulty in inferring a text. For this reason it is possible their scores or results will be not consistent.

Practicality

There are some advantages of paper and pencil-based test. First, it is more practical because it only uses paper and pencil in which the test takers have been familiar with both these objects. It is very different from when we take TOEFL CBT (computer-based test) because perhaps not all test takers are literate in using the computer. Second, this test is objective test because we can get good result if we can answer the questions correctly. There is no subjective scoring like when we take speaking test for instance. However, there are also some problems in test administration. The first thing is if the electricity suddenly is off when we are doing listening test, it will disrupt our concentration or memory. The second thing is that it sometime happens in which the questions and responses/answers on the question sheets are blur or illiterate. These can be problematic for test takers because they have to read several times. The third thing is TOEFL PBT is an indirect test that means that the test takers are not tested based on what happens in reality. What I mean here is that a candidate's English proficiency is not tested directly. It is only based on their results or score in answering the whole questions.

Washback

McNamara (2000) states that test can also have effects particularly on the community as a whole including the school. It means that there is a test impact from each test given to students or test takers/participants. For example, the result of TOEFL score can be used as one of requirements for getting a scholarship from an educational institution or a sponsor if we have the required TOEFL score.

CONCLUSION

In conclusion, PBT is a good model of standard test; but this test does not reflect fully the actual language use because grammatical competence is more dominant. In addition, there is an assumption that test takers who obtain high score will have good English proficiency even though it may be not true because it does not show their English ability particularly in communicative competence. Apart from this, for those who can get high score that there is no guarantee that the test takers have adequate English proficiency.

REFERENCES

- Fulcher, G; & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Abingdon: Routledge.
- Kamyab, G. (2009). *Study guide: Test design (listening)*. Victoria: Deakin University.
- Kamyab, G. (2009). *Study guide: Test design (reading)*. Victoria: Deakin University.
- McNamara, T. F. (2000). *Language testing*. New York: Oxford University Press.
- Richards, J., Platt, J., & Weber, H. (1985). *Longman dictionary of applied linguistics*. Essex: Longman Group Limited.
- TOEFL. (n.d.). Retrieved from <http://en.wikipedia.org/wiki/TOEFL>
- What is TOEFL? (n.d.) Retrieved from <http://www.examenlsh.com/TOEFL/index.php>