

The Impacts of the PBT TOEFL-based Progressive and Summative Learning Measurement in a Higher Educational Institution

Corry Caromawati, S.S., M.A.
Institut Teknologi Nasional
(corrycaromawati@gmail.com)

Abstract

It is commonly found that testing policy affects learning and teaching process occurring in educational institutions. This case study investigated the impacts of employing the PBT TOEFL test as the progressive and summative tests in a higher educational institution in one of the biggest towns in Indonesia. The data was collected through observation, field notes and personal interviews with six English language lecturers to get in-depth information based on their experience. The data from the interview was later transcribed. The whole data was analyzed qualitatively. The results showed that on the macro level, the test indeed affected the policy makers' decisions that demanded the lecturers to teach to the test and expected the students to study language outside the classroom. Their decisions inevitably also affected the individuals in the micro level. It is shown that even though the testing policy contributed positive impacts, it dominantly affected language and teaching process negatively. The perceived positive impacts included the practicality of administering the test and language explicit knowledge input to the students. The perceived negative impacts were categorized into two groups: (1) the negative attitude of the lecturers toward language teaching and the students' negative attitude toward learning; (2) the distorted focus of language teaching and learning. These results advocate for an evaluation towards the policy for a better language learning and teaching.

Keywords: second language test, testing impacts, washback, testing policy

It is natural that people get affected by what become their interests. This reality also happens in the language testing practice. It is basically caused by their desire to make the tests successful in order to get their expectation. In terms of testing, this phenomenon is called *impacts* (Bachman & Palmer, 1996). When the test impacts related to the aspect of the learning and teaching process, it is commonly known as *washback* (Hunges, 1989, p. 1). The impacts in testing can go within two levels: macro level (the society or educational system), and micro level (the individuals) (Bachman & Palmer, 1996, pp. 29-30).

Individuals in both levels will adjust their behavior to meet the demands of a test (Green, 2013, p. 40). In *macro* level, institutions, for example, would demand teachers to conduct effective teaching and learning process to get some qualification for being credible universities by having qualified students. In *micro* level, a test can make learners put a lot of demand to their teachers to teach the test to help them get high scores to graduate from schools. Additionally, teachers might change their teaching styles to meet the demand of their learners, etc. Their behaviors can be beneficial or harmful (Buck, 1988, p. 17). Therefore, this aspect should be considered when developing or applying a test within any specific systems (Palmer & Bachman, 1996, p. 35).

Nonetheless, many institutions, particularly in Indonesia, would put aside the idealism and consider what the market/industry demands. As a result, they prepare their students to fulfill the demands. Since the demand of many prestigious companies and graduate schools

in the global era requires the applicants to provide a language proficiency proof such as TOEFL, TOEIC, or IELTS certificate, it is inevitable that a huge number of institutions prepare their students to take them. It does not mean that these high stake tests are not appropriate, but their application should be considered thoughtfully.

Amongst the high stake tests, the PBT is more popular than the other ones. It is caused by the fact that this test is relatively affordable and there have been many commercial preparation books and mock tests sold. Moreover, the test is still acceptable by most companies and graduate schools. As a consequence, many universities apply a policy to employ the test as a measurement of students' language proficiency. Furthermore, some universities even integrate the test into their curriculum and use it for progressive and summative measurements. Although it is very well known, I believe that it is important to describe the test in this paper. The information on this test was gathered from ETS' official website.

Relevant to its name, the PBT TOEFL is a paper based test that test takers get the questions on paper and put their response on a paper answer sheet. The PBT is actually the level 1 version of the ITP TOEFL which has been around for more than forty decades. It is used to measure the test takers' English ability to use and understand the language in an academic setting in higher educational systems. It targets test takers' proficiency level groups of high beginning to intermediate students, considered to be in Level 1.

Even though the PBT TOEFL is also administered by a very credible testing company, it has both strengths and weaknesses. The strengths of the test are the highly reliable of the questions and the high practicality. In addition to that, the test is also considerably affordable compared to other high stake tests. Considering these strengths, it is not surprising that this test has been around for almost half century. Among the strengths mentioned about this test, it also has some weaknesses. The only one type of question, multiple choice, is commonly considered as the weakness of this test. This type of question is believed to have many disadvantages such as testing literacy than a meaningful assessment of learning (McAllister & Guidice, 2012). In their website, ETS itself admits that the PBT TOEFL provides limited opportunity to the students to demonstrate certain learning objectives such as creativity, speaking, writing, and social skill. This claim is supported by previous studies such as Shaftel and Shaftel (2007) who believed that students' cognition would not go to the higher level by this type of test. Additionally, this selected response type of the test may lead to students' recognition of *getting the right answer* as the main aim of education and become *guessing the answers* activity for students with low cognitive skills (McAllister & Guidice, 2012).

Regardless of the strengths and weaknesses discussed, the test is considerably popular and used widely both in ESL and EFL settings. The fact that it is a high-stake test makes it not surprising if it affects educational system in Indonesia reaching both macro and micro level. In the macro level, it is observed that the demand of this test makes many universities offer courses in TOEFL preparation in their language center or language laboratory. However, a considerable number of universities also integrate this type of test in their curriculum requiring the students to take the test as progressive and summative learning measurement. This policy inevitably affects the micro level (other stakeholders in the institutions), i.e. teachers and students.

In the literature, the studies of impacts or washback effect of the PBT TOEFL (or ITP) in developed countries had been conducted since approximately a decade ago. Alderson and Hamp-Lyons (1996), for example, investigated the washback of the ITP TOEFL in the USA. While Flaman and Furuya (1997) observed and reported the washback of the same test in a university in Japan. In developing countries, however, this topic can be considered relatively current. For instance, Kawachi-Furlan, Amorim, and Finardi studied the impacts of the TOEFL ITP preparation class on the performance of students in Southeast Brazil in 2017. Additionally, Mahmud investigated problems of teaching TOEFL to the language learners in Indonesia in 2014.

Many methods have been used to investigate impacts or washback of the test including statistical data analysis, literature review, interview, and observation. Kawachi-Furlan, Amorim, and Finardi (2017), for instance, employed statistical data analysis to measure the correlation between the students' test results before the "treatment" (the TOEFL class) and after it. Flaman and Furuya (1997), on the other hand, reported their 10 years observation and a survey involving students in a university in Japan. The data that they had gathered was used to evaluate the reliability and validity of the test. Alderson and Hamp-Lyons (1996) used observation and interview involving the students and teachers to investigate the washback of the test. Finally, Mahmud (2014) applied classroom observation and interview with the students.

The results from the previous studies showed that teaching PBT test contribute positive or negative impacts. The positive impact from the test is that it helped students improve their TOEFL score (Kawachi-Furlan, Amorim, & Finardi, 2017). On the contrary, it also gives negative impacts including not improving students' English performance (Flaman and Furuya, 1997), giving frustration considering students had basic English skills and even demotivating them (Mahmud, 2014).

Furthermore, language teachers believe that it is important for students to get adequate information about the test that they are going to take in advance. Therefore, they try their best to get the students familiarized with the TOEFL test. This is in line with what Madaus (1988) claimed that "*they foster their efforts by teaching to the content and tradition of the test*" (Madaus, 1988). Thus, what they teach mostly is the strategies to deal with questions in every section in the test. They practice some tips from the commercial TOEFL preparation book, explain grammatical rules involving metalanguage, and ask their students to memorize lists of vocabulary to get as many right answers as possible. This lead to a typical TOEFL preparation class atmosphere that has been built in the teachers' mind: teaching boringly. It is mainly caused by the use of very little communicative approach applied in the class. The dynamic of the class was found lower in the TOEFL preparation class than in the non-TOEFL preparation class, and students considered the class boring (Alderson & Hamp-Lyons, 1996). Wurdinger (2012, p. vi) also supports the claim that the result of teachers teaching to the test is a boring learning process of rote memorization. On the same book, he also adds that teachers will be stuck using approaches that do not motivate or inspire students to learn.

Teachers' attitude toward their teaching is also affected. There are some positive and negative perceptions on the practice of teaching TOEFL found. Some admitted that they liked teaching TOEFL because they did not have a lot of work for preparing the lessons, and not necessarily marked any homework (Alderson, 2000, p. 37). This fact is somehow concerning since teachers might forget how to improve students' performance on their language skills (Chapman & Snyder, 2000). Additionally, some teachers think that they are happier to teach TOEFL because they feel like helping their students to be successful in taking the test (Mahmud, 2014). Some other teachers, on the other hand, are being skeptical with this system. They state that it is difficult to make a TOEFL preparation class interesting, and structuring the teaching into a coherent framework so that students can perceive some logic to what they are learning (Alderson & Hamp-Lyons, 1996, p. 294). Some of them are fully aware that teaching TOEFL was inauthentic and non-communicative (Alderson & Hamp-Lyons, 1996, P. 285). This awareness made them feel frustrated when coming into class. This burden was getting heavier when knowing that they are going to deal with a big class (40 – 50 students) with mixed level ability, where most of them are in low English proficiency level.

Learners are, of course, also affected by the testing policy. Students who are from remote areas where they do not have enough access and opportunity to learn English will have very limited English language proficiency. Mahmud (2014) reported that most of the students involved in her study hardly understood nor spoke English and possessed not enough motivation to learn English, especially TOEFL. Students from higher economic status are able to afford going to study English in informal English schools to foster their English knowledge and skills as reported in Alderson and Hamp-Lyons (1996).

Unfortunately, students from lower status, can only expect a lot from the universities where they are studying. Based on my four year observation, most of them hold the belief that learning TOEFL strategy will help them get high TOEFL score, and this score will ultimately make them English proficient users.

The myth encourages them to memorize lots of vocabulary without understanding context, grammar rules without using it to communicate, and apply the test strategies without comprehending the content of the test as suggested by the book they use (Hamp-Lyons, 1998). The way they study does not help them to foster their English language ability to help them gain communicative competence expected by all stakeholders. The fact that there is a mismatch between their actual English ability and the level of the test leads to students' frustration in learning the real English (Mahmud, 2014).

Despite the negative *washback* mentioned, this teaching to test strategy gives the opposite effect to students with higher proficiency level. They become more comfortable with the test as they are getting more familiar with the type of questions, and think that they acquire a lot of vocabulary as supported by a study conducted by Amini and Gonzales (2012). Teachers think that they also build their critical thinking from the reading exercise they practice as proved by the study conducted by Fahim, et al. (2010).

While previous studies looked at TOEFL preparation classes, none examined the impacts of the test when it is incorporated into the curriculum. This paper briefly looks at the impactor *washback* effect of the implication of using the PBT TOEFL both as a progress and as an achievement test in a higher educational institution in Bandung.

Method

According to Gillham (2010), a case study is a study investigating a particular individual from a particular community (p. 1). Thus, it was applied in this study which investigates some parties (the policymakers, the lecturers, and the students) in a specific setting. The setting of this study took place in one of private universities in Bandung, Indonesia. The university integrated the PBT TOEFL into the curriculum so that they used the same type of test called TOEFL-like test to measure students' progressive and summative learning. The policy had been applied for two semesters when the study was conducted.

There were two classes English classes observed to get information on how the teaching and learning process occurred in the classroom affected by the test. The observation was also conducted to observe the behaviors of policy makers in response to the testing policy. One head of study program involved in the curriculum team and the head of English language course coordinator were observed for the purpose of this study to get their insight from the macro level and to understand the impact of the test from the macro level. The important data from the observation was compiled in as the field notes. To get the more in-depth information from the lecturers, six English language lecturers were interviewed in order to share their experience and insight regarding the impacts of the test. For privacy purposes, the names and any personal information of the participants are not disclosed in this paper.

The data from the interview were transcribed. The overall collected data including the field notes from the observation and interview were analyzed qualitatively. In addition to that, the interpreted data was later framed to better understand the impacts of the PBT TOEFL to macro and micro levels in the institution. The framework was discussed with the second coder to increase the reliability of the results. Finally the results of the study were found and are discussed in the following section.

Result and Discussion

Based on the data, the impacts of the testing policy employing the PBT TOEFL-like tests to measure the students' learning progress within the macro and micro levels are summarized in Table 1. The impacts in the macro level are considered to be neither positive nor negative since they only reflected the policy makers' behaviors. Whether or not their

behaviors are positive is not judged considering the limitation of the data. The impacts in the micro level, however, are categorized into positive or negative based on the teachers' perspectives.

The impact in the macro level, According to the field notes from the observation, the integration of the PBT TOEFL into the curriculum compelled the policy makers to require the teachers teach to the test. It happened since a number of students complained about how different the content taught in the classroom and what was tested. As a consequence, the policy makers chose a commercial TOEFL preparation book as the main teaching material, and teachers were asked to supplement it with their own created materials. The book itself contained strategies to answer the PBT TOEFL questions accurately. It had two books for two different levels: lower (introductory) and higher (complete). These two books were used for two different level classes as well in the institution.

Investigating the insight from policy makers personally, the field notes from the observation provided evidence that both agreed that lecturers should incorporate content (e.g. industrial engineering, architecture, etc.) into the learning process. However, they still required English lecturers put the content in the TOEFL-like questions. It is known from the data that policy makers expected the lecturers to teach the students to the test. Additionally, they also believed that the students would develop both their language skills as well as their ability to get a high score on TOEFL test by teaching them the test strategies with an assumption that they would be encouraged to have autonomous learning. As their further action, the policy makers encouraged the language lecturers to adapt lessons from commercial TOEFL preparation book published by Longman. They also implicitly (without explicit announcement to the students) encouraged the students to take additional language classes outside the offered course such as, for example, taking an informal English course. This expectation, however, was considered to be implausible considering many students come from low income family background while informal English courses in Bandung are relatively costly.

The impacts of the testing policy from the micro level insight, based on the observation and interview data, can be fallen into positive and negative categories. There are two points brought up on the positive side, i.e. practicality and explicit knowledge input. On the negative side, the lecturers focused on two aspects: (1) the negative attitude of the lecturers and learners toward the language teaching and learning; (2) the distorted focus of teaching and learning.

One of the positive impact brought up was *practicality*. Practicality mentioned by the lecturers referred to the ease they felt when they prepared the lesson and delivered it, as well as when they graded the students' tests. They thought that teaching to the test did not require a lot of effort to prepare it as much as teaching skills since they would do the same thing for different classes regardless of the students' characteristics and background. Their statement was also supported by the data from the observation in which it was observed that the lecturers barely designed a lesson plan even though they had to teach different classes, and delivered exactly the same lesson to each class. The contents that the teachers taught were all based on the Longman book, test strategies. In addition to teaching, practicality was also found in grading. It was also observed that it took less than an hour for the lecturers to grade the students' work using an answer key that they had created.

Explicit knowledge input was another positive aspect that the lecturers brought up. They claimed that teaching TOEFL strategies covered so many metalinguistics (e.g. tenses, determiners, etc.) which, they believed, would eventually develop the students' language implicit knowledge. They also claimed that at the end, students would be able to produce English language accurately. From the observation including the field notes, however, there

was no data found to back up this claim. Thus, further research is needed to investigate the claim.

From the negative side, one of the aspects that the lecturers brought up was their own negative attitude toward the teaching. The lecturers' feelings expressed in the interview were also supported by the data from the observation. It was shown that the lecturers did not seem to enjoy the lessons. They delivered the lesson in a plain and one-directed way. There were not many interactive discussion nor engaging activities. As a matter of fact, they claimed to feel some frustration teaching them.

The lecturers believed that their negative attitude toward teaching was mainly caused by the fact that the students had low level of English proficiency while the class demanded them to acknowledge more than basic knowledge of English. Moreover, they believed that the students had cognitive overload since the lesson did not seem authentic to them and too many unfamiliar meta-linguistics that they were supposed to comprehend and use. As a consequence, the students indicated low motivation to engage in the classroom and catch up with the lesson. In return, the lecturers felt disrespected and frustrated at the same time as their students seemed to be lost in the lesson.

The data from the observation strongly supported the statements from the lecturers. During the observation, the students barely showed any enthusiasm nor enjoyment in learning. It seemed like they hit the lowest level of motivation in learning English. It was indicated from the way they lie their heads on the table while the lecturers were giving their lectures, some students yawned and some even were asleep in the classroom. Additionally, the students unwillingly did the tasks given by their lecturers, frequently uncompleted the tasks and when the lecturer asked them to share their answer, they only guessed and picked one of the options without being able to justify it.

In addition to the attitudes towards teaching and learning explained above, the lecturers thought that the practice of teaching to the test made the students have a testing culture instead of learning culture. It could be indicated from the way the students treated the language similarly as how they answered the questions on the test.

A student mentioned in the interview that he had learned the test strategy and considered it as the way to treat the language. The teachers thought that it was saddening since they actually wished the students to learn the language while learning the test strategies at the same time. Nonetheless, they admitted that it was unavoidable since the focus of teaching and learning was also distorted due to the test preparation.

From the discussion above, we can conclude that the lecturers indeed focused on test strategies instead of language skills since they were asked to do so. In return, the students' learning focus was also distorted to testing than language skills. This fact was also found in the observation data. During the observation, the students focused on what type of questions were being tested (practiced in class) and what shortcut to answer them. Beyond that, they did not pay any attention to the topic discussed in each question.

Table 1.

The impacts of the PBT TOEFL on the micro and macro levels

Level	Impacts	
	Positive	Negative

Macro	<ul style="list-style-type: none"> - Demanding the teachers to teach test strategies - Expecting students to learn the language outside the classroom
Micro	<ul style="list-style-type: none"> - Practicality - Explicit knowledge exposure - Attitudes toward teaching and learning - The shifted focus of the language learning and teaching

Conclusion

This study investigated the impacts of a testing policy integrating the PBT TOEFL in the curriculum and used as the progressive and summative learning measurement in a higher university in Bandung. The impacts investigated included both the *macro* and *micro* levels. For the purpose of investigation, different data collection methods were applied and the data were interpreted qualitatively.

The results showed that the testing policy affected both levels. On the macro level, the policy makers wished the teachers to teach to the test in order to prepare the students taking the test and get a high score of the PBT TOEFL and learn the English language and skills outside the classroom. This expectation, willingly or unwillingly, inevitable affected the micro level, especially the teachers and the students in regards to the attitudes and focus toward language teaching and learning. The teachers taught more metalanguage and the test answering questions strategies instead of fostering students' English skills while they actually needed as basic users of English language. The way they taught was also considered frustrating since the students did not seem engaged and showed minimum motivation toward the lessons that they delivered. The students, on the other hand, expected a lot from the course to help them get higher score on the test. They thought that by practicing the strategies from the book, and lesson from their teachers would help them reach their expectation. Unfortunately, their lack of English proficiency made this mission hard to accomplish. They lost their sense of what learning a language was. Their focus was about getting as many items correct as possible. Additionally, they also treated the language similarly to how they treated the test. Based on the results of this study, the institutions should feel the needs to evaluate their testing policy.

Even though careful thought was made when designing this study, it is far beyond perfection. There are some limitations that should be admitted. First, the fact that this is a case study makes it situational. The results might have been varied in different settings. Thus, more studies are needed. Second, this study only investigated the behaviors of the stakeholders of the institution, but it did not investigate the impact of the testing policy on the improvement of the students' language proficiency. Therefore, it is suggested that further study will also investigate the change of the students' TOEFL score before and after the course.

Regardless of the limitation, this study contributes information for educational institutions integrated the PBT TOEFL into the curriculum. The results must be different if the class is offered separately. Thus, the information from this study is strongly suggested to become consideration when applying the same policy as that of the setting of this study in order to successfully achieve their expected learning outcomes. More broadly and

importantly, a suitable testing policy should help the institutions to achieve their common missions: graduating students with competences for the global era which is closely related to the aims of language classes.

References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3): 280-297.
- Alderson, J.C. (2000). Testing in EAP: Progress? Achievement? Proficiency?. In Blue, G.M., et al. (Ed). (2000). *Assessing English for Academic Purposes*. Peter Lang.P.: 21-47.
- Amini, M. & Gonzalez, N.I. (2012). The Washback Effect of Cloze and Multiple Choice Tests on Vocabulary Acquisition. *Language in India*, 12: 71-91.
- Bachman, L.F. & Palmer, A.S. *Language Testing in Practice*. New York: Oxford University Press.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language testing*, 13(3): 257-279.
- Buck, G. (1988). Testing Listening Comprehension in Japanese University Entrance Examination. *JALT Journal*, 10: 15-42.
- Chapman, D. W., & Snyder, C. W. (2000). Can high stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development*, 20(6): 457-474.
- English Testing Center. (2013). TOEFL ITP. Retrieved from http://www.ets.org/s/toefl_itp/pdf/toefl_itp_score.pdf
- English Testing Center, (2003). *Linking Classroom Assessment with Students Learning*. Retrieved from https://www.ets.org/Media/Tests/TOEFL_Institutional_Testing_Program/ELLM2002.pdf
- English Testing Center, (2003). *CEFR Mapping Study*. Retrieved from https://www.ets.org/toefl_itp/research
- Flaman, W. & Furuya, C. (1997). < Surveys and reports> Evaluating the role of the test of English as a foreign language: institutional testing program (TOEFL ITP) at a Japanese National University of Technology.
- Fahim, et al. (2010). The Relationship between Test Takers' Critical Thinking Ability and Their Performance on the Reading Section of TOEFL. *Journal of Language Teaching and Research*, 1(6): 830 -837.
- Gillham, B. (2000). *Case Study Research Methods*. London: Continuum International Publishing.
- Green, A. (2013). Washback in Language Assessment. *International Journal of English Studies*, 13(2): 39-51.
- Hamp-Lyons, L. (1998). Ethical Test Preparation Practice: The Case of the TOEFL. *TESOL Quarterly*, 32(2): 329-337.
- Hunges, A. (1989). *Testing for Language Teachers*. New York: Cambridge University Press.

- Kawachi-Furlan, C. J., Amorim, G. B., & Finardi, K. R. (2017). The Interface between the TOEFL ITP and Internationalization and Language Assessment in Brazil. *Studies in English Language Teaching*, 5(2): 213.
- Mahmud, M. (2014). The EFL Students' Problems in Answering the Test of English as a Foreign Language (TOEFL): A Study in Indonesian Context. *Theory and Practice in Language Studies*, 4(12): 2581-2587.
- Madaus, G. E. (1998). The influence of testing on the curriculum. *Yearbook-National Society for the Study of Education*, 2, 71-112. Philips, D. (2001). Longman Introductory Course for the TOEFL Test (2nd ed). New York: Pearson Education, Inc.
- Shaftel, J., & Shaftel, T. L. (2007). Educational Assessment and the AACSB. *Issues in Accounting Education*, 22(2): 215-232.
- Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal*, 76(4): 513-521.
- Wurdinger, S. D. (2012). *Time for action: stop teaching to the test and start teaching skills*. R&L Education.