
Evaluating TOEFL ITP Test: A Critical Review

Ade Tika¹⁾, Ahmad Amin Dalimunte²⁾, Muhammad Dalimunte³⁾, Ayumi Pratiwy Tanjung⁴⁾, Irma Suryani⁵⁾

¹Universitas Islam Negeri Sumatera Utara
Ade.tika@uinsu.ac.id

²Universitas Islam Negeri Sumatera Utara
amindalimunte@uinsu.ac.id

³Universitas Islam Negeri Sumatera Utara
muhammaddalimunte@uinsu.ac.id

⁴Universitas Islam Negeri Sumatera Utara
ayumipratiwy.tanjung@uinsu.ac.id

⁵Universitas Islam Negeri Sumatera Utara
Irma.suryani@uinsu.ac.id

Received: 14th September 2022 | Revised: 28th January 2023 | Accepted: 15th June 2023
Email Correspondence: amindalimunte@uinsu.ac.id

Abstract

The critical objective of this present study is to conduct substantial evaluation of TOEFL as a language assessment tool. Using a qualitative method, the authors of the present study evaluated various specific issues based on a set of predetermined criteria such as practicality, test reliability, validity, authenticity, and washback impact. The analysis indicates that TOEFL ITP is both feasible and dependable. The TOEFL's dependability is based on these three components. However, unless the ITP test is altered from the previously used TOEFL, the TOEFL ITP is as reliable as the TOEFL PBT. In theory, its trustworthiness is uncertain, and more investigation is needed. It is debatable whether the TOEFL ITP test items can accurately represent this task when utilized for the placement test. TOEFL ITP is a reliable English test. Knowing the TOEFL significance encourages test takers to enhance their English skills as part of the exam. More students want to take lessons to enhance their grades.

Keywords: Assessment, TOEFL, Practicality, Test Reliability, Validity

1. INTRODUCTION

To date, more than 200 universities in Indonesia have used TOEFL ITP scores and certificates as graduation requirements and as certificates accompanying diplomas, as well as government and corporate institutions for recruitment and job promotion, scholarship selection requirements, and other purposes. It also has several interests, such as being very important when applying to various fields or institutions, companies and some universities.

Educational Testing Service (henceforth ETS), a US-based corporation, administers TOEFL (Test of English as a Foreign Language). ETS measures non-native English speakers' ability to use and understand North American English in college and university contexts. In over 80 countries, the test is used by over 4200 academic institutions, government agencies, scholarship programs, and

licensing/certification bodies (Educational Testing Service, 2010). As stated by Taufiq et al., (2017) TOEFL has been taught to students prior to entering college, and one of the intended applications of the TOEFL examinations is to monitor students' development, and schools give the tests to evaluate improvement in students' English ability. In practice, most graduate programs require a higher TOEFL score for admission than undergraduate programs at the same university. Because graduate school textbooks are often written at a higher grade level than undergraduate texts, this is required.

The TOEFL is a valid and credible test, according to ETS, and it has just revealed intentions to develop a new and better TOEFL. The new exam will be more closely aligned with the skills needed for long-term success in speaking, listening, reading, and writing (Educational Testing Service, 2010). In simple terms, the TOEFL is an English language test intended to assess a test taker's ability to communicate in North American English in a wide university or college setting. TOEFL is separated into three types: internet-based tests (iBT), computer-based tests (CBT), and paper-based tests (PBT). CBT, on the other hand, is no longer used and has been replaced by iBT. The examination is divided into four sections: (1) reading comprehension, (2) listening comprehension, (3) structure and written expression, and (4) written English test. It has a speaking test, especially in iBT.

ETS, as a test provider institution, also developed a particular format for the Asian region called ITP (Institutional Testing Program), which is the focus of this present study. ITP is a Paper-based TOEFL with no writing section. TOEFL ITP does not test the participants' ability in writing essays. Every question is multiple-choice with four response choices. TOEFL PBT measures English language skills comprehensively. In addition, in terms of the tightness of the exam, the TOEFL PBT is directly administered by ETS whereas the TOEFL ITP is administered to the respective institutions that have been trusted, such as universities or courses. TOEFL has long been trusted as one of the English proficiency tests demanded to be taken by non-native English speakers to obtain a scholarship or continue education further. To date, well-known scholarship providers including Full-bright, Australian Development Scholarships, Japanese Government Monbukagakusho Foundation accept TOEFL ITP certificate as one of the scholarship requirements.

1. TOEFL ITP Descriptions

Listening Comprehension, Structure and Writing Expression, and Reading Comprehension are the three components of ITP. Sections A, B, and C comprise the listening segment. Section A is a brief conversation that consists of two statements (typically made by a man and a woman) followed by a question. Short dialogue questions fall into four major categories: questions requiring factual information restatement, context awareness, idiom knowledge, and logical inference. Parts B and C contain items of

inference and restatement identical to Part A; however, other sorts of questions such as the major topic, identity of the speaker or audience, tone and point of view, the purpose of the speech, and location of the dialogue can also be asked. Approximately half of the material and background of longer conversations and longer lectures is intellectual in nature.

The component on the structure and written expression was created to examine the ability to select an appropriate language for standard written English (TOEFL ITP assessment series, 2013). This section is divided into two parts: Part A, Structure, and Part B, Written Expression. There are 15 incomplete sentences, with four words or phrases labeled (A), (B), (C), and (D) under each sentence (D). Test takers must select the word or phrase that best completes the statement. Look for the number of questions on the answer sheet and fill in the blanks that correspond to the letter of the answer they choose (Pyle, M. A., & Page, 2002). The underlined sentence is divided into four parts: (A), (B), (C), and (D) (D). In order for the sentence to be correct, the test taker must modify an underlined word or phrase. This segment can last up to 25 minutes.

Reading Comprehension has some potential overlap across categories; for example, factual questions involving interpreting paraphrased information necessarily require inferences and vocabulary; such questions can be classed as vocabulary or inference questions in some respects (Hilke, R., & Wadden, 1997).

3. TOEFL For Evaluation

As part of their admission, many universities, particularly in Asia, administer the TOEFL ITP as an English as a Foreign Language exam. Mahidol University International College, Salaya, in Thailand, is one of them. For students to apply to the college and meet other conditions, specific TOEFL ITP scores must be attained (Perez-Amurao, 2011). In the context of Indonesia, a Malang institution administers the TOEFL ITP twice a year for various objectives. The first exam is administered as a component of the placement examination for student who will be assigned to a specific level in the university's English language program. Before students receive their bachelor's or master's degree, a second test is given to determine their English competence for the final time.

The TOEFL ITP was chosen by these formal institutions for a number of reasons. Some of this can be attributed to the TOEFL's reputation for validity and reliability as well as the prices' reasonableness. Given the numerous assessment criteria that must be created, particularly its usage as a placement test, the researchers believe that a critical analysis of the TOEFL ITP test is essential.

To conduct the evaluation of the TOEFL test, we formulated three key questions: 1) Is it feasible to use the TOEFL ITP test as a placement test? 2) How accurate is the TOEFL ITP test for placement

exams? 3) Is the test reliable for determining English proficiency? And, 4) Does the test offer a useful return to the test taker?

2. RESEARCH METHOD

To analyze and measure the test, we followed the five main principles of evaluating a test: practicality, test reliability, validity, authenticity, and the washback effect (Brown, 2004; Templer, 2004). However, in order to respond to the given questions, the present study focused only on three key aspects; namely, practicality, reliability, and validity of the TOEFL ITP.

1. Practicality

A test is considered practical if meet these following criteria: a) it stays within budget limits, b) it can be completed by test takers within appropriate time constraints, c) the test must also have a clear distinction for administration, the test must be using available human resources appropriately, testing does not exceed available material resources and also consider the time and effort involved for design and assessment.

2. Reliability

Reliability refers to as to the consistency of measurement, the extent to which scores are similar on various forms of the same instrument from data collection opportunities (MacMillan, J.H. & Schumacher, 2001). This means that the test results are considered consistent and reliable if students of the same skill level take the assessment, they must receive the same grade or result. In addition, if students wish to repeat the assessment, their score must be the same as the previous one, assuming that the students do not learn more after taking the first assessment. Franzen (2000) argued that test reliability refers to the extent to which the test measures without error. This is closely related to the validity of the test. Test reliability can be thought of as precision; the extent to which the measurement occurs without error. Reliability is not a constant property of a test and is better thought of as a different type of reliability for different populations at different levels of the construct being measured.

3. Validity

On learning achievement tests and competency tests, validity is an indispensable requirement in test development. According to Lissitz & Samuelsen (2007), test validation used in education should involve analysis of test content and empirical analysis of test scores and item response data. The content analysis of the test is related to content validity, which furthermore requires empirical analysis to determine the construct validity. Both of these analyzes are intended so that tests in the world of education meet the standard test requirements. In addition, validity is important in determining test quality. Linn & Gronlund (1995) explained that Validity refers to the adequacy and appropriateness of

interpretations made from assessment, with respect to specific uses. This opinion is supported by (Messick, 1995) that validity is an integrated evaluative policy about the extent to which empirical facts and theoretical reasons support the adequacy and suitability of inference and action based on test scores. Sources of validity facts can be grouped into test content, response process, internal structure, relationships with other variables, and consequences of implementation tests (Cizek et al., 2008). Moreover, the existence of the validity of a test device can be known through analysis of test content and empirical analysis of item response data test scores (Lissitz & Samuelsen, 2007). Based on our evaluation, several questions arise regarding TOEFL ITP test.

3. FINDINGS AND DISCUSSION

The result of the present study is organized based on the answer of the four given key inquiries as follows.

1) Research question 1: Is it feasible to use the TOEFL ITP test as a placement test?

For a long period, ETS has been conducting the TOEFL test and confirming its usefulness. Overall, the TOEFL ITP is convenient to take for any reason because it was essentially adapted from the original TOEFL. Similar to other reasons, TOEFL ITP has been specifically defined those administrative aspects, the practicality of administering exams without procedural issues, although all equipment and materials, have been determined in advance. The three portions are divided and take a specific amount of time in the TOEFL ITP, and most students can finish every section within the permitted time. In addition, the test fee is the key factor in its selection because it is within the parameters of the budget. A test evaluation system and a predetermined manner of reporting results are also possible with computerized scoring.

2) Research question 2: How accurate is the TOEFL ITP test for placement exams?

According to Wainer & Lukhele (1997), TOEFL score was 100 percent trustworthy in their 1997 report. Even though the dependability of the TOEFL score indicated was obtained using an older paper-based TOEFL, its format is identical to the most recent TOEFL ITP. These three TOEFL components: Written Structure and Expression, Listening, and Reading Comprehension are reliable. Given that the paper-based TOEFL is source of the TOEFL ITP, it is still feasible that the dependability is present. However, the location or organization administering the test is still debating further critical evaluation. The TOEFL ITP is given on-site by each partner academic institution, unlike the TOEFL iBT or TOEFL PBT, and is then sent back to ETS for the evaluation alone.

It is possible that TOEFL ITP organizer provides a quite different conditions and amenities. Consider the number of participants as one example. Given the fact that the TOEFL ITP participant

capacity is unrestricted, the number of test takers may be low and occupy a little amount of space, or they may be high, like when almost 100 people are taking the placement test at once. The sound quality in the listening portion may vary for test takers from various universities. In reality, ETS has recommended using standard rooms in addition to other supporting facilities, and as long as the test-administering institution complies, it is thought that the TOEFL ITP will continue to be reliable.

3) Research question 3: Is the test reliable for determining English proficiency?

A test must meet several criteria to be considered valid, including face validity and content validity. In relation to the use of the TOEFL ITP for placement examinations, it is debatable whether the test items accurately represent this quality. All types of TOEFL test are designed to assess English language proficiency rather than to determine whether a person passes or fails. The TOEFL is solely used to indicate a test taker's degree of competency in English as a second language for academic reasons with scores ranging from 310 to 420, or 525 to 677. In their research portion, Faucet et al. believe that the TOEFL ITP is genuine (Kerans et al., 2014). The validity of the TOEFL ITP is, nevertheless, impacted by a few facts. This test's first flaw is in how well it measures "English language aptitude." It is well known that the TOEFL ITP only included three areas for testing: structure, listening, and reading comprehension. The TOEFL iBT is the only version that offers the speaking and writing tests. The Speaking test is used to evaluate potential teaching assistants and gauge the development of general speaking abilities (Farnsworth, 2013). The writing component, on the other hand, is likewise set up for automated scoring, which makes it simple for language teachers to evaluate in a fair and suitable way.

Test takers can improve their writing skills by performing computer exercises (Huu & Ngoc, 2016). As a result, the TOEFL ITP tested English skills only on passive English. A person who excels at speaking or writing may be unable to demonstrate his abilities. Furthermore, Templer (2004) claims that the TOEFL ITP only applies to institutions that administer the test. This argument is mostly based on the fact that the tests are conducted locally. Every institution is allowed to administer the TOEFL ITP Test, with the items tested drawn primarily from the previous test. In terms of the placement test, it is possible to conclude that the TOEFL ITP is applicable to institutions that administer the test. However, it only assesses some English skills. Furthermore, large institutions that use TOEFL ITP as proof of English language requirements, such as international scholarship providers, must conduct their own tests to clarify the score results.

4) Research question 4: Does the test offer a useful return to the test taker?

Test takers must keep an eye on their TOEFL scores in order to get admitted to a university or pass a certain level of English. TOEFL has a huge impact on their washback in general. Test takers are

encouraged to learn more English in order to comprehend the significance of the TOEFL, specifically the TOEFL ITP. Test takers are encouraged to pay attention to their TOEFL score in order to pass a specified level of English by studying more or taking a special course for it. TOEFL in general has a significant impact on their washback. The presence of private and school courses, supermarkets of real testing materials, and numerous programs aimed to train teachers or instructors has been discovered as a result of TOEFL as a standardized test. This predicament emerges as a result of the requirement for test takers to obtain a higher TOEFL score. The type of TOEFL is used to create courses, resources, and tutors (Templer, 2004) for example, there is a specific treatment generated from the TOEFL ITP.

The development of courses and materials has also been fueled by the emergence of English as a second language industry or foreign languages with an increasing number of teachers being trained to specialize in TOEFL ITP preparation. Recently, it was revealed that numerous TOEFL training institutes in many urban areas of the English learning globe are recruiting university lecturers for weekend teaching, and an increasing number of secondary schools and universities are offering specialized TOEFL preparation courses.

4. CONCLUSION

Based on the Method in conducting the evaluation of the test, the researchers have several particular questions, specifically those of some categories stated by Brown, namely practicality, test reliability, validity, authenticity, and the washback effect (Templer, 2004). However, the researchers focus only on the three of it. Practically of TOEFL ITP when it stays within budget limits, it can be completed by test takers within appropriate time constraints, the test must also have a clear distinction for administration, while reliability.

TOEFL ITP is used in an academic setting to help educators understand their students' English language skills and where they are progressing, whereas the urgency is important for scholarship programs, applying for jobs with well-known companies, and even as a condition for admission to domestic universities. The TOEFL ITP test contains multiple-choice questions that assess your English language skills in three areas: 1) Listening comprehension assesses your ability to understand spoken English; 2) Structure and written expression assesses your ability to construct structured and grammatically correct English sentences; and 3) Reading comprehension assesses your ability to read and comprehend reading material written in English. TOEFL ITP also issues certificates in three levels: bronze (460-542), silver (543-626), and gold (627-677).

Suggestions for the TOEFL ITP test in the future include to make the TOEFL ITP test available online, rather than only through campuses or institutions, and preparing a TOEFL ITP test scheme with

the Remote Proctoring model or TOEFL ITP online. The types of questions and the number of questions remain the same in the test format, but the implementation is online and participants work from their respective homes. Given that the paper-based TOEFL is the source of the TOEFL ITP, it is possible that dependability exists. The location or organization administering the test, on the other hand, is still debating further critical evaluation.

REFERENCES

- Brown, H. D. (2004). *Language Assessment Principles And Classroom Practices*. San Francisco: State University. (2nd ed.). Pearson Longman.
- Cizek, G. J., Rosenberg, S. L., & Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement*, 68(3), 397–412. <https://doi.org/https://psycnet.apa.org/doi/10.1177/0013164407310130>
- Educational Testing Service. (2010). *TOEFL internet-based and paper-based tests*. December.
- Farnsworth, T. L. (2013). An investigation into the validity of the TOEFL iBT speaking test for international teaching assistant certification. *Language Assessment Quarterly*, 10(3), 274–291. <https://doi.org/10.1080/15434303.2013.769548>
- Franzen, M. D. (2000). *Reliability and Validity in Neuropsychological Assessment Critical Issues in Neuropsychology* (2nd ed.). Springer Science & Business Media.
- Hilke, R., & Wadden, P. (1997). The Toefl and Its Imitators: Analyzing the Toefl and Evaluating Toefl-Prep Texts. *RELC Journal*, 28(1), 28–53. <https://doi.org/https://doi.org/10.1177/003368829702800103>
- Huu, D., & Ngoc, T. (2016). Assessment Model for Language Learners' Writing Practice (in Preparing for TOEFL iBT) Based on Comparing Structure, Vocabulary, and Identifying Discrepant Essays. *International Journal of Advanced Computer Science and Applications*, 7(5), 148–153. <https://doi.org/10.14569/ijacsa.2016.070523>
- Kerans, M. E., Cid-aranda, E., Kennedy-scanlon, M., & Maria, R. (2014). *to successful performance. 2014*.
- Linn, R.L. & Gronlund, N. E. (1995). *Measurement and assessment in teaching* (7th ed.). Englewood cliffs, NJ: Prentice-Hall.
- Lissitz, R. W., & Samuelsen, K. (2007). Further Clarification Regarding Validity and Education. *Educational Researcher*, 36(8), 482–484. <https://doi.org/10.3102/0013189x07311612>
- MacMillan, J.H. & Schumacher, S. (2001). *Research in Education. A Conceptual Introduction* (5th ed.). Longman.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American*

Psychologist, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>

- Perez-Amurao, A. (2011). The Types and Nature of Questions vis-à-vis Students' Test-Taking Skills as Significant Indicators of Second Language Examinees? Performance on the TOEFL-ITP Reading Comprehension Sub-Test. *Online Submission*, 6(2), 39–56.
- Pyle, M. A., & Page, M. E. N. (2002). *TOEFL Preparation Guide (Test Of English As A Foreign Language)*. New Delhi : Wiley Dreamtech.
- Taufiq, Wahyu., Santoso, D.R., & Fediyanto, N. (2017). Critical Analysis on TOEFL ITP as A Language Assessment. *1st International Conference on Intellectuals' Global Responsibility (ICIGR 2017)*, 125.
- Templer, B. (2004). High-Stakes Testing at High Fees : Notes and Queries on the International English Proficiency Assessment Market. *Journal for Critical Education Policy Studies*, 2(1), 1–24.
- Wainer, H., & Lukhele, R. (1997). How reliable are TOEFL scores? *Educational and Psychological Measurement*, 57(5), 741–758.
<https://doi.org/https://psycnet.apa.org/doi/10.1177/0013164497057005002>

