

Test-repeaters' perceptions of difficulty on the TOEFL listening test

Anang Widodo, Novia Russilawatie & Septi Riana Dewi
Universitas Teknologi Yogyakarta, Indonesia

ABSTRACT: This case study investigates the test-repeaters' perceptions on the TOEFL Listening test. The participants were students in a private university in Indonesia. Test-takers' scores were analyzed and coded based on the name and the lowest score in each section. We collected the data of test-repeaters' perceptions from a semi-structured interview. The interview transcript was analyzed, coded, and triangulated with test-repeaters' scores. The results indicate that the participants believed that the TOEFL test was important, but it was challenging. Most participants failed on the listening sections due to a lack of preparation and problems relating to vocabulary, including pronunciation and accent used in the audio. This study emphasizes that having good planning and preparation is essential to avoid failure. This result carries crucial pedagogical implications for test-takers, TOEFL trainers, and policymakers.

Keywords: Difficulty in TOEFL test, test-repeaters, TOEFL listening test

1 INTRODUCTION

The use of the TOEFL test as a standardized English test has been applied at several universities (O'Dwyer et al. 2018; Ginther & Elder. 2014). Implementing TOEFL as the standardized test at university brings positive and negative impacts for students as presented by the following studies. Hung and Huang (2019) mentioned that students believed that campus-wide standardized tests benefit their English proficiency when they work in the future. Besides its benefits, this implementation also results in problems for students who cannot meet the minimum standard score. They, consequently, need to retake the same test. The study of test-takers who should retake the test was conducted by Cho and Blood (2020) and Knoch et al. (2020). These studies investigate the preparation that test-takers made before the test and their score changes. It indicates that repeaters should spend more time to focus on test preparation before the test day to achieve their targeted scores.

One of Indonesia's private universities implements an English standardized test for its students as one of the graduation requirements. Students should take the test at the university's education, certification, and training center. The minimum score that students should get is 450–500, depending on the minimum standard score set by each faculty. For many test-takers, the minimum standard score set is considered high. Many of them failed to meet the minimum standard score set by their college. They, therefore, had to retake the test several times until they achieved the score required. The number of repeaters is recently increasing since some faculties raise the minimum standard score for their students. During the period of the study, there were 1,690 test-takers registered and 592 test-takers or 35% of the test-takers failed to meet the minimum standard score and should retake the test several times.

A standardized English language test as a requirement for graduation is also implemented in more than 90% of universities in Taiwan (Wu and Lee 2017). This study showed that the implemented policy gets many criticisms because it is considered that the test cannot improve students' English language proficiency. However, students have positive perceptions of this English graduation benchmark policy. The study conducted by Li et al. (2012) showed that the CET (College

English Test), an English standardized test designed to measure Chinese undergraduate students' English proficiency, motivated test-takers to make a more significant effort to prepare for the test. They eagerly put more effort to learn English skills tested in CET. These two previous studies are relevant to the current study because they have similar test-takers' perspectives toward the university's standardized test. The regulation implemented impacts students as test-takers because they have to take this test. The previous studies proved that this policy positively influences students to study harder to improve their English skills to meet the standard score. However, those who should retake the test several times possibly will find it stressful and exhausting. Thus, this current study investigates perceptions of test-takers who have failed several times to meet the minimum standard test score set by faculty as a requirement for their graduation.

Although there are some researches about test-takers' perceptions (Galikyan et al. 2019; Fan 2014; Pearson 2019, Hamid et al. 2019; Wu and Lee 2017; Li et al. 2012), there is no research focusing on test-takers' perception toward difficulty on the TOEFL Listening test. This current study focuses on investigating this issue. It is essential to investigate this issue to raise test-takers' awareness about the large number of test-takers who should retake the test and their problems on the test. Besides, it hopefully gives policymakers new insights to make some regulations that address the TOEFL test repeaters. This study investigates the most challenging section in the TOEFL test and test-takers' perceptions of their TOEFL test failure.

2 LITERATURE REVIEW

Listening is one of the skills tested in a standardized test. Chang et al. (2013) mention that students face some problems such as fast speech rates, accent, connected speech, colloquial usages and slang, and transient information. Brindley and Slatyer (2002) concluded some factors affecting the difficulty of listening comprehension tasks into three categories. They are as follows, 1) the nature of the input, including vocabulary, speech rate, accent, amount of redundancy, length of the passage, syntactic complexity, propositional density, discourse structure, noise level, register; 2) the nature of the assessment task. This category includes clarity of instructions, amount of context provided, availability of question preview, and response format, and 3) the individual listener factors. This category comprises memory, knowledge, interest, motivation, and background. In terms of the Listening section in the TOEFL test, Kostin (2004) investigated the difficulty Listening section, particularly in the difficulty of TOEFL dialogue items. The result showed that test-takers find it difficult at 1) when the dialog consists of two or more negatives, 2) when the meaning is not stated indirectly and test-takers should draw an inference from the dialogue, and 3) utterances patterns in the dialogue.

The following studies focus on test-takers perceptions of the standardized test. Galikyan et al. (2019) figured out that student test-takers have positive perceptions of the test because it can show their English language abilities. Fan (2014) mentioned that VET (Versant English Test) receives positive perceptions from its test-takers since it is reflected in test-takers' proficiency in spoken English. Some test-takers' negative perceptions toward the standardized test can be seen from the following studies. Pearson (2019) highlighted that test-takers have negative perceptions of the speaking and writing assessment accuracy in IELTS. They mistrust single examiner marking. They request detailed feedback on their test performance and an explanation about their incorrect answers. In line with the previous study, Hamid et al. (2019) also found negative perceptions of test-takers towards the standardized test that they took. Although they believed that the IELTS test was fair, they assumed that it did not provide an accurate measure of test-takers' proficiency and did not test test-takers' actual language proficiency. The study conducted by Tsai and Tsou (2009) revealed that instead of focusing on enhancing students' English competency, implementing standardized ELP tests as a graduation benchmark requirement turns the teaching and learning process into test-oriented. Besides, most participants mentioned that the standardized test implemented did not assess their learning outcomes.

Test-takers' perceptions toward the test are influenced by many factors such as English proficiency, cultures, socio-economic backgrounds, and English learning conditions Pham and Bui (2019). This study showed that the Vietnamese test-takers who live in the South were more supportive of implementing the ELPTs (English Language Proficiency Tests) than test-takers in Central Vietnam because they believed that the test motivated them to learn and improve their English proficiency. They also agree that the test can enhance competitiveness in their future workplace and further education. Such perspective reflects that students from developing and urban contexts (South) acknowledge the ELPTs (English Language Proficiency Tests), the policy, and English language skills are right for their future. Although several studies about test-takers' perceptions have been conducted, there is no study about test-takers' perceptions in the Indonesian context. This study, thus, investigates test-takers' perceptions towards their failure in the TOEFL test as a requirement for graduation in one of the private universities in Indonesia.

3 RESEARCH METHOD

The data of test-takers' scores were collected from all test-takers who registered at the university's education, certification, and training center in a one-year duration. To get the data, the researchers sent a consent letter to the head of this department. After getting permission, the researchers received the data of test-takers scores and used it as one of the main data for this study. All data used in this study will remain confidential. During the study, there were 1.690 test-takers registered. However, only 1.098 test-takers or 65% of the total participants could meet the minimum standard score from their faculty, while 592 test-takers or 35% of test-takers had to retake the test twice to seven times. Of 592 repeaters, 16 test-takers should retake the test more than 5 times. For data of the interview, the researchers selected participants through the purposive random sampling from those who failed the test more than 5 times. After selecting the participants, the researchers contacted them to ask their permission and willingness to have an interview. In the beginning, the researchers contacted 9 participants, but 3 of them did not give any responses. The participants involved in a semi-structured interview were 6 test-takers aged 22–23 years old consisting of 1 female and 5 male students. After getting permission from the participants, an interview schedule was set and it was conducted online by calling each participant through WhatsApp.

L2 teaching and learning complexity require an in-depth study to understand students and their behaviors and experiences (Nassaji 2015). A Case Study design is more appropriate to figure out this complexity since it deals with a specific case on specific participants. Yin (2017, p.14) stated that "A Case Study is an empirical method that investigates a contemporary phenomenon (the "case") in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident." This study explores test-takers' perceptions of their experience after failing the TOEFL test several times and finding out why the Listening section fails them to achieve the targeted score. The data used are the test-takers' scores in a one-year duration and interview. Test-takers' scores were analyzed and coded based on the lowest score in the Listening section. To get the data of test-takers' perceptions, a semi-structured interview was conducted. The interview transcript was analyzed and given some codes and it was triangulated with the data of TOEFL test-takers' scores. A further follow-up interview was conducted with participants to confirm and clarify the data.

4 FINDINGS AND DISCUSSION

All participants have positive perceptions of the test. They believed that TOEFL is essential for their future. This finding is in line with the study conducted by Pham and Bui (2019), which showed that implementing the English standardized test motivated test-takers to improve their English proficiency. In this study, test-takers believed that it could enhance competitiveness in their future workplace and further education. However, the current study results show that participants admitted

that it is tough to get the score targeted in the TOEFL test. They found it challenging, considering that they needed 6 to 7 times to get the minimum standard score. The TOEFL test is regarded as a tough test in Indonesia and Japan (Koizumi & Nakamura 2016). This study showed many test-takers in Japanese universities need to retake the test more than once. The participants considered that the standardized test implemented was too tricky.

Based on the interview, four out of six participants agreed that listening was the most challenging section. This result is in line with a study conducted by Chang et al. (2013). This study involving low-level learners revealed that 73% of the participants perceived English language listening as difficult. In terms of difficulty faced during the test, several participants mentioned that they were not familiar with speakers' pronunciation, accent, and vocabularies used in the audio during the listening section. The data show that lack of vocabulary mastery affects test-takers' listening performance. DM realized that listening was his most difficult section since he kept failing in this section. He mentioned, *"In the Listening section, mostly I could not understand the conversation. I do not know the vocabulary used. The accent they use is very different from Indonesian people"*. As stated in the previous studies, vocabulary is one of the problems faced by students in listening (Chang et al. 2013; Brindley and Slatyer 2002) and in the TOEFL test (Kostin 2004).

The participants' other problems are the quality of the audio and listening tool used in the test and the speed of audio used. DW and DA lacked focus because the audio was not clear. It dealt with the headphones they used. DW stated, *"The speakers' voice in the Listening section is not quite clear due to the quality of the headphones used."* Fast speech rates, accent, vocabulary, length of the passage, and noise level are common problems students face in listening (Chang et al. 2013; Brindley & Slatyer 2002). NA mentioned that *"speakers speak too fast. I do not understand what they are talking about"*. NV argued, *"I do not understand TOEFL. I do not understand the listening."* Faster rate effects reduce students' comprehension in listening (Brindley & Slatyer 2002). Kostin (2004) mentioned that the dialogue's utterance patterns are some difficulties the test-takers face in the listening section.

Another challenging aspect of listening is the ability to remember information from the audio. DM stated, *"It is difficult to understand and remember all information in the Listening section, so I answer it based on my feelings and clues from a piece of information I get."* Namaziandost et al. (2018) stated EFL listening comprehension can be influenced by working memory since it depends on the storage and processing of information in the mind. Brindley and Slatyer (2002) categorized this difficulty into individual listener factors in which test-takers' memory influences their performance in the listening test.

The last issue raised by participants in this study is the lack of preparation before the test. During the period of the study, the institution did not provide test-preparation classes for test-takers. Therefore, all participants prepared it by themselves. No one of them went to an English course/training center or hired English tutors to help them prepare for the test. Chou (2015) argued that test-takers who are categorized as 'average' and 'weak' listeners tend to ignore the importance of training in other English skills, particularly in listening. Test preparation positively influences test-takers' score gain (Liu 2014; Mickan & Motteram 2008). Pearson (2019) mentioned that 40% of candidates fail the IELTS test due to lack of preparation. Pan (2016) argued that the amount of time spent on language learning before a test is significant in determining student – score improvement. Thus, it can be offered that a lack of preparation can be one of the causes of repeaters' failure in the test.

5 CONCLUSION

All participants in the study agree that TOEFL is essential for them, but it is challenging. Most participants failed in the listening section. They all mentioned that vocabulary mastery, pronunciation, working memory, dysfunctional headset, and lack of preparation are the problems they faced. Although they are aware that their English proficiency level is not good enough, they did not do much effort to prepare for the test well. Having good planning and preparation before taking

the TOEFL test is essential to avoid taking the test several times to achieve the targeted score. Consequently, they should retake the test 6 to 7 times to meet the minimum standard score set by their college. This study's results contribute to pedagogy, particularly for teachers and TOEFL instructors to explore new directions to teach TOEFL and focus on the listening section. It is essential to improve test-takers' vocabulary mastery during TOEFL training. For test-takers, good preparation before the TOEFL test will give a greater possibility to achieve the targeted score. This study also suggests that the university provides test-preparation classes for students or, particularly repeaters, to improve their test scores and avoid retaking the test several times. This study deals with the results that cannot be generalized as TOEFL test-takers' fundamental problems since the participants in this study were limited and based on specific criteria. Further studies can explore a representative and a larger number of participants involved from several universities. Besides, the study results reveal the most challenging section in the test that is still too general. Thus, further research needs to analyze and study which listening section skills that fail the TOEFL test-takers.

REFERENCES

- Bai, Y. (2020). The relationship of test-takers' learning motivation, attitudes towards the actual test use, and test performance of the College English Test in China. *Language Testing in Asia*, 10(1), 1–18.
- Barnes, M. (2016). The washback of the TOEFL iBT in Vietnam. *Australian Journal of Teacher Education*, 41(7), 158–174.
- Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, 19(4), 369–394.
- Chang, A. C., Wu, B. W. P., & Pang, J. C. (2013). Second language listening difficulties perceived by low-level learners. *Perceptual and motor skills*, 116(2), 415–434.
- Cho, Y., & Blood, I. A. (2020). An analysis of TOEFL® Primary™ repeaters: How much score change occurs? *Language Testing*, 37(4), 503–522.
- Chou, M. H. (2015). Impacts of the Test of English Listening Comprehension on students' English learning expectations in Taiwan. *Language, Culture, and Curriculum*, 28(2), 191–208.
- Fan, J. (2014). Chinese test-takers' attitudes towards the Versant English Test: a mixed-methods approach. *Language Testing in Asia*, 4(1), 6.
- Galikyan, I., Madyarov, I., & Gasparian, R. (2019). Student Test Takers' and Teachers' Perceptions of the TOEFL Junior® Standard Test. *ETS Research Report Series*, 2019(1), 1–15.
- Ginther, A., & Elder, C. (2014). A comparative investigation into understandings and uses of the TOEFL iBT® test, the International English Language Testing Service (Academic) test, and the Pearson Test of English for graduate admissions in the United States and Australia: A case study of two university contexts. *ETS Research Report Series*, 2014(2), 1–39.
- Hung, S. T. A., & Huang, H. T. D. (2019). Standardized proficiency tests in a campus-wide English curriculum: a washback study. *Language Testing in Asia*, 9(1), 1–17.
- Knoch, U., Huisman, A., Elder, C., Kong, X., & McKenna, A. (2020). Drawing on repeat test-takers to study test preparation practices and their links to score gains. *Language Testing*, 37(4), 550–572.
- Koizumi, R., & Nakamura, K. (2016). Factor structure of the Test of English for Academic Purposes (TEAP®) test in relation to the TOEFL iBT® test. *Language testing in Asia*, 6(1), 3.
- Kostin, I. (2004). Exploring item characteristics that are related to the difficulty of TOEFL dialogue items. *ETS Research Report Series*, 2004(1), i–59.
- Liu, O. L. (2014). Investigating the relationship between test preparation and TOEFL iBT® performance. *ETS Research Report Series*, 2014(2), 1–13.
- Mickan, P., & Motteram, J. (2008). An ethnographic study of classroom instruction in an IELTS preparation program. *International English Language Testing System (IELTS) Research Reports 2008: Volume 8*, 1.
- Namaziandost, E., Hafezian, M., & Shafiee, S. (2018). Exploring the association among working memory, anxiety, and Iranian EFL learners' listening comprehension. *Asian-Pacific Journal of Second and Foreign Language Education*, 3(1), 1–17.
- Nassaji, H. (2015). Qualitative and Descriptive Research: Data Type Versus Data Analysis. *Language Teaching Research*, 2015, Vol. 19(2) 129–132.
- Ockey, G. J., & Gokturk, N. (2019). Standardized Language Proficiency Tests in Higher Education. *Second Handbook of English Language Teaching*, 377–393.

- O'Dwyer, J., Kantarcioğlu, E., & Thomas, C. (2018). An Investigation of the Predictive Validity of the TOEFL iBT® Test at an English-Medium University in Turkey. *ETS Research Report Series*, 2018(1), 1–13.
- Pan, Y. C. (2016). Learners' Perspectives of Factors Influencing Gains in Standardized English Test Scores. *TEFLIN Journal*, 27 (1), January 63–81.
- Pearson, W. S. (2019). 'Remark or retake'? A study of candidate performance in IELTS and perceptions towards test failure. *Language Testing in Asia*, 9(1), 17.
- Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, 35(1), 31–65.
- Tsai, Y., & Tsou, C. H. (2009). A standardized English language proficiency test as the graduation benchmark: Student perspectives on its application in higher education. *Assessment in Education: Principles, Policy & Practice*, 16(3), 319–330.
- Wu, J., & Lee, M. C. L. (2017). The relationships between test performance and students' perceptions of learning motivation, test value, and test anxiety in the context of the English benchmark requirement for graduation in Taiwan's universities. *Language Testing in Asia*, 7(1), 9.
- Yin, R. K. (2017). *Case study research and applications: Design and methods*. Sage publications.