

Washback of TOEFL Preparation Courses on Students' Attitudes and Score Improvement

Yangting Wang (Corresponding author)

Cultural, Literacy, and Language program

The University of Texas at San Antonio

Texas, United States

E-mail: qwk936@my.utsa.edu

Becky H Huang

Department of Bicultural-Bilingual Studies

The University of Texas at San Antonio

Texas, United States

Received: April 30, 2020

Accepted: May 15, 2020

Published: May 27, 2020

doi:10.5296/ijl.v12i3.16940

URL: <https://doi.org/10.5296/ijl.v12i3.16940>

Abstract

Despite the prevalence of the TOEFL test preparation courses, research on the impact of TOEFL test preparation courses on language learning and teaching is quite limited (Green, 2013; Huang, 2018). The current study investigated the washback effects of TOEFL preparation courses on students' attitudes and TOEFL scores improvement. The study implemented a partially mixed equal status design (Leech & Onwuegbuzie, 2017). Forty-one students and four teachers in a tertiary-level Intensive English Program (IEP) participated in the study. Quantitative data consisted of students' responses to Likert-scale items and students' TOEFL scores. They were analyzed by independent sample t-tests, one-way repeated measure ANOVA, and post hoc tests. Qualitative data, which included open-ended survey items, student interviews, and teacher interviews, were analyzed using the inductive approach (Miles & Huberman, 1994) and Saldaña's (2015) coding techniques. Survey results demonstrated that students had generally positive attitudes toward the class. Additionally, a statistically significant score improvement was found for some students. However, student

and teacher interviews also revealed several issues that prevented student learning. For example, students studied TOEFL only to improve test scores; teachers rushed to cover all textbook materials and gave little homework; and there were limited classroom speaking activities. Implications regarding the teaching and administering of TOEFL classes at IEP were discussed.

Keywords: TOEFL preparation courses, TOEFL ITP, Student attitudes, Score improvement, Washback, Intensive English program

1. Introduction

Test of English as a Foreign Language (TOEFL) is an English proficiency test required for international students to enter a university in many English-speaking countries, such as the United States and Canada. More than 35 million people throughout the world have taken TOEFL (ETS, 2018). Because TOEFL is deemed as a gatekeeping test for university entry, international students seek guidance in preparing for TOEFL. Under this circumstance, over 173 tertiary-level English Language Programs (ELPs) at community college, universities, and off-campus English language centers offer TOEFL test preparation courses in the United States (TOEFL and Test Preparation, 2019). Despite a large number of TOEFL preparation classes (TOEFL class thereafter) offered in the United States, the effects of TOEFL classes on learning and teaching are underexplored. This study thus aims to investigate the impact of TOEFL on students' attitudes toward the TOEFL classes as well as its impact on score improvements.

The current study focused on the paper-based TOEFL, also known as TOEFL ITP, which is used at the research site. The TOEFL ITP evaluates non-native speakers' English language proficiency through academic topics (ETS, 2018). TOEFL ITP scores are typically only accepted by institutions internally (i.e. ELP) to make placement entrance or exit decisions. For example, in the current Intensive English Program (IEP), if students obtain over 550 points in TOEFL ITP, their admission hold can be removed and they can exit the program and enter the university. TOEFL ITP can be administered as many times as the institution wishes and are typically not administered outside of the institution (Golubovich, Tolentino, & Papageorgiou, 2018).

2. Literature Review

Washback is the effects of tests or test preparation courses on language learning and teaching (Muñoz & Álvarez, 2010). Washback can be either negative or positive (Taylor, 2005). Positive washback brings beneficial teaching practices, whereas negative washback is when the test content and format cover a narrow range of language ability and, thus, restrict the teaching situation (Gebril & Eid, 2017). According to Hughes (2003), a test may influence participants' perceptions and attitudes toward learning and affect their learning behaviors which will, in turn, impact their learning outcomes. Therefore, it is important to investigate the effects of tests and test preparation courses on students' attitudes and learning.

According to Messick (1996), the introduction of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning (p.

241). Muñoz and Álvarez (2010) indicated that most studies on the washback effects of TOEFL courses centered around teachers and their classroom practices (i.e. Barnes, 2016; Chou, 2017; Wall & Horák, 2008; 2011; Wang, 2019). Although some scholars have started to investigate the effects of tests on students' attitudes (Shih, 2007; Xie & Andrew, 2013), washback studies that focus on learners are sparse (Green, 2013). In the following sections, we discuss two aspects in the literature that are pertinent to the current study: the washback of test preparation courses on students' attitudes as well as on student learning outcomes.

2.1 Washback of Test Preparation Courses on Students' Attitudes

Because of the limited research in TOEFL preparation courses, I present research that investigated test washback on students' attitudes in general. Shih (2007) examined the impact of the General English Proficiency Test (GEPT) on students' English learning attitudes in Taiwan. He found out that students who are not interested in learning English did not care about the test. Additionally, students demonstrated resistance in learning specifically for the test. In an Educational Testing Service (ETS) research report, Golubovich, Tolentino, & Papageorgiou (2018) surveyed 249 test-takers' opinions about the TOEFL ITP (paper-based test) test in Japan, Mexico, and Indonesia and found that the participants had a positive attitude towards the TOEFL ITP. They reported that the TOEFL ITP is relatively inexpensive and is practical in assessing listening and reading skills. Zhi and Wang (2020) conducted a study to examine the impact of the National College Entrance English Exam (NCEEE) on student perceptions in a rural Chinese high school. One hundred and thirty-nine senior students participated in the study. The study indicates that students had a perceived negative attitude toward NCEEE due to its overemphasis in grammar and lack of focus on oral language. Results demonstrated the importance of test authenticity in a high-stakes exam.

2.2 Washback of Test Preparation Courses on Student Learning

TOEFL and IELTS are two major standardized English proficiency tests that influence millions of international students' academic pursuits in English speaking countries every year. Compared to TOEFL, there are more studies on the washback effects of IELTS preparation courses (Allen, 2016; Green, 2006; 2007; Hawkey 2006; Khodabakhshzadeh, Zardkanloo, & Alipoor, 2017; Mickan & Motteram 2008, 2009; Rashidi & Javanmardi, 2011; Saville & Hawkey, 2004; Gan, 2009). Findings of the effects of IELTS preparation courses on students' learning were mixed. For example, Allen (2016) investigated the effect of the IELTS test on ESL students' test preparation strategies and score gains. He found that the IELTS test yielded positive washback on students' English productive skills based on surveys and test scores. Gan (2009) examined the effects of an IELTS preparation course on one hundred and forty-six students in a Hong Kong University and found no significant difference in students' IELTS test scores between the ones who took the IELTS class and those who didn't take the class. This finding aligned with Green (2007) that there were no significant score gains for students who were in the IELTS preparation courses, compared to those who were in General English classes or IELTS-General English combined classes.

Only a few studies examined the washback effects of TOEFL courses on student learning outcomes (Choi & Papageorgiou, 2014; Yu et al., 2017). Choi and Papageorgiou (2014)

conducted pre- and post-tests to students studying in two university programs in Turkey and Thailand before and after students received English instructions. The pre- and post-score differences from the Turkey group ($N = 106$) and the Thailand group ($N = 76$) were both statistically significant. The effect sizes of the score differences for the Turkey group was consistently larger than the Thailand group, and this might be due to the more intensive English instruction that Turkish students received compared to Thai students. Choi and Papageorgiou (2014) indicated that the effects of English instructions on TOEFL ITP score gains needed to be appropriately investigated in future research.

Among the few studies on the washback of TOEFL courses, the majority investigated the Internet-based TOEFL (IBT), and the paper-based TOEFL ITP received little attention from independent researchers outside of the Educational Testing Services (ETS). Therefore, the study fills the gap by examining the effect of TOEFL ITP preparation courses by independent researchers. The study contributes to the literature in three ways: 1. Since research is particularly lacking in the washback effect on student learning outcomes (Green, 2013), the current paper contributes to the field by investigating the effects of TOEFL test preparation courses on students' score improvements. 2. This paper is one of the limited washback studies on test preparation courses that implement a mixed-methods research design. Integrating both quantitative and qualitative analysis offers a robust data interpretation (Creswell & Garrett, 2008) and contributes to a more comprehensive understanding of the test washback issues. 3. Previous research only investigated students' learning as measured by TOEFL test score gains (Gan, 2009; Green, 2007), the current study not only examines score improvement, but also provides information on how students perceive the influence of the TOEFL class on their TOEFL scores and language learning. Additionally, results from the study also offer pedagogical implications for instructors of TOEFL courses and provide curriculum development suggestions for the ELPs regarding the integration of TOEFL classes. Two research questions guided the current study:

1. What are students' attitudes toward TOEFL classes?
2. Do students' TOEFL scores improve after taking TOEFL classes?

3. Method

3.1 Study Context

Data collection was conducted at an IEP in a public university in the Southwestern United States in 2018. The required TOEFL score at the research site is 79 on TOEFL IBT (Internet-Based Test) or 550 on TOEFL ITP (Institutional Testing Program). Students at the site who do not meet the cut score are conditionally admitted and enrolled in the IEP to improve their test scores and English skills before they can take regular academic courses. There are five English levels from beginning to advanced level. Students at each level are required to take four classes each week to maintain their visa status in the United States. The four classes include three ESL regular classes (writing & grammar, reading & vocabulary, and oral communication) and one non-ESL regular class. The non-ESL regular class is either the TOEFL test preparation class or the American culture class depending on students'

English levels. Since TOEFL lessons were more advanced and were considered not suitable for beginning level students by the IEP, students in levels 1 and 2 take the American culture class and students in levels 3, 4, 5 take the TOEFL class. Based on the number of student enrollment, there is only one TOEFL class per level.

The three ESL regular classes are each 120 minutes long and are offered three times a week. The TOEFL class is shorter, which lasts 80 minutes long and twice a week. The class size is 15 students maximum. Instructors at the IEP all hold a Master's degree in Teaching English to Speakers of Other Languages (TESOL).

3.2 Research Design

The current study follows Leech and Onwuegbuzie's (2017) partially mixed equal status design where quantitative and qualitative data and analyses are given equal weight. The study is partially mixed design in that the integration of data happened after both qualitative and quantitative data were analyzed independently. Additionally, not all research questions used both quantitative and qualitative data. The second research question was solely answered by quantitative data analyses. Qualitative data included students' interviews, responses to open-ended survey questions, and teacher interviews. Quantitative data consisted of students' and teachers' responses to Likert-scale survey items and students' TOEFL ITP scores.

3.3 Participant

We conducted a pilot study in spring 2017, where we observed one level 5 TOEFL class eight times and interviewed one student in level 5. Based on the pilot study interview, some open-ended questions were added to gain a comprehensive understanding of participants' perspectives. In the current study, participants consist of one male teacher (Mr. Z) and his 26 students in spring 2018 and three female teachers (Ms. T, Ms. E, Ms. C) and their 25 students in fall 2018. All four instructors are native English speakers.

In spring 2018, the 26 student participants are from 18 to 40 years old with a mean of 24 years old ($SD = 5$). Among them, 12 were male and 14 were female. These students are from Mexico ($N = 5$), Japan ($N = 3$), China ($N = 4$), Korea ($N = 2$), Vietnam ($N = 5$), Central Asia ($N = 1$), Mideast ($N = 3$), and South America ($N = 3$).

In fall 2018, the 25 student participants are between the ages of 18 to 53 years old ($M = 26$, $SD = 10$). Student demographics were similar to that of the spring cohort. Students were informed that their participation is voluntary, will not influence their class grades, and that their information will be kept confidential.

4. Data Collection

4.1 TOEFL ITP Assessment

The TOEFL ITP assessment that the IEP uses targets the intermediate to advanced level students. This test includes three sections: listening comprehension, structure and written expression, and reading comprehension. Students enrolled in TOEFL classes at the IEP will take the TOEFL ITP test twice per semester: the first time in the middle of the semester as

their TOEFL class midterm grade, and the second time by the end of the semester as their final exam grade. Students' TOEFL class scores were based solely on the two times' TOEFL ITP scores. Table 1 shows the number of TOEFL ITP scores collected. For the 2018 spring group, 26 students' two times test scores (2018 spring mid and end-of-semester) were collected; for the 2018 fall group, 25 students' (2018 fall mid and end-of-the-semester) test scores were collected. Additionally, to compare and track students' score improvement over time, four times' TOEFL test scores were collected for the 2018 spring group (N = 13) and fall group (N = 7). The four times' TOEFL scores include the mid- and end-of-the-semester scores from the previous semester and the current semesters.

4.2 Students' End-of-Semester Survey

By the end of spring and fall 2018 semesters, students in all TOEFL classes were invited to participate in an anonymous survey regarding their attitudes toward their current TOEFL class. The survey consists of 14 Likert-scale items adapted from Huang, Zhi, & Wang (2018) and four open-ended questions that ask students to write about their attitudes toward homework, classroom activities, classroom materials, and suggestions. Students were assigned ID numbers after the data collection was complete. The Cronbach alpha reliability estimate for the 14 survey items is .94 (n = 23) for spring 2018 and .87 (n = 25) for fall 2018. The total reliability estimate is .93 (n = 48) for all students.

4.3 Students' End-of-Semester Interview

After students completed the end-of-semester survey, participants interested in the follow-up interview were invited for a face-to-face individual interview. The first six students who volunteered to participate in spring 2018 and fall 2018 were invited for the interview (See Table 1). Table 1 indicates the number of student interviewees each semester.

Table 2 presents the background information of the 12 student interviewees. The interview consists of 15 open-ended questions (See Appendix 1). We used the interview questions as a guide to ask about participants' opinions. To protect each student interviewee's identity, a pseudonym of their choice was used. Each interview lasted 20-40 minutes.

Table 1. Data collection in spring and fall 2018

	N (2018 Spring)	N (2018 Fall)
Teacher observed	1 (Mr. Z)	3 (Ms. E, Ms. C, Ms. T.)
Student end-of-semester survey	23	25
Student end-of-semester interview	6 (3 for each level)	6 (2 for each level)
Four times' TOEFL ITP scores	13 (2017 fall to 2018 spring)	7 (2018 spring to 2018 fall)
Two times' TOEFL ITP scores	26 (mid- and end-of-semester score)	25 (mid- and end-of-semester score)

Table 2. Information about student interviewees

Semester	Pseudonym	Age	Country of origin	Time studied at the IEP (months)	Levels at the IEP	Semester (teacher)
2018 Spring	Chris	26	China	36	5	Mr. Z
	Eric	27	China	36	5	
	Maria	38	Columbia	24	5	
	Hannah	21	Japan	7	4	
	Monika	21	Japan	7	4	
	Sonny	30	Korea	18	4	
2018 Fall	Lynda	24	Taiwan	4	5	Ms. T
	Asset	19	Kazakhstan	7	5	
	Von	21	China	4	4	Ms. E
	Moka	21	Japan	4	4	
	Marie	47	Mexico	7	3	Ms. C
	Anna	30	Mexico	10	3	

5. Data Analyses

To explore patterns and themes from the data itself, we followed Miles and Huberman's (1994) inductive approach for qualitative analysis. Additionally, descriptive and focused coding was used to analyze the qualitative data (Saldaña, 2015). In regards to the first research question about student attitudes toward TOEFL classes and English learning, students' attitudes were first examined by merging spring and fall 2018 student data to evaluate their general attitudes based on the Likert-scale items. To further understand students' attitudes, quantitative data (Likert-scale items) were triangulated with the qualitative data (four open-ended items, student interviews). Likert-scale items were investigated through descriptive statistics. To understand specifically how students perceived the TOEFL class, four open-ended responses were analyzed through established survey categories: homework, classroom activities, classroom materials, and other suggestions. These established categories were used as focused codes. Comments given under each category were analyzed as descriptive codes (Saldaña, 2015).

For the interview data, I followed the Rapid Identification of Themes from Audio Recordings (RITA), a technique proposed to analytically examine the qualitative data using the audio file itself instead of the transcripts. According to Neal, Neal, VanDyke, and Kornbluh (2015), the RITA technique both ensures the quality and expedites the qualitative data analysis process. One researcher listened to the recordings and jotted down all descriptive codes for each interview such as "lack opportunities to speak," "positive influence" and "do exercises." I re-listened to the recordings and identified themes as focused codes such as "like about the class," "dislike about the class," "influence of TOEFL class on TOEFL scores," "influence of TOEFL class on English learning," "attitudes about the class," "necessities of the class," "suggestions about the class."

To answer the second research question, quantitative data (Likert-scale data and TOEFL scores) were used. One-way repeated measure ANOVA and post hoc tests were performed twice to examine the effects of the TOEFL class on students' TOEFL scores over two semesters. Post hoc G*power analyses were conducted to estimate the statistical power achieved from the quantitative results.

6. Results

6.1 What Are Students' Attitudes Toward TOEFL Classes?

The Likert-scale survey items indicate that both spring and fall semester students' attitudes toward TOEFL classes were positive with a mean range of 4.37 to 5.17 agreeing with the 14 statements (N = 49) (See Table 3). Students believed that their English language skills improved (M = 4.71, SD = .94) and their TOEFL ITP scores increased (M = 4.52, SD = 1.07) after taking the class. They believed that the test strategies were helpful (M = 4.81, SD = 1.09) and the TOEFL class greatly helped the students prepare for the exam (M = 4.76, SD = 1.09). They also agreed that the amount of homework and the exercises were adequate and they felt more confident in taking the exam (M = 4.49, SD = 1.12). Additionally, the students believed that the TOEFL class is necessary and important for English language programs (M = 4.96, SD = 1.04).

Table 3. Descriptive statistics of student attitudes toward TOEFL classes

Survey items	N	Min	Max	M	S.D.
1. The TOEFL preparation class is appropriate for my level of English proficiency.	49	2	6	4.78	.99
2. My English language skills have improved after taking the class	49	2	6	4.71	.94
3. My TOEFL score has improved because of the class	48	2	6	4.52	1.1
4. I can understand the instruction in class most of the time.	48	2	6	5.17	.88
5. The TOEFL class is informative. I learn a lot from the class.	49	2	6	4.59	1.01
6. The TOEFL class is necessary and important in an English language program.	49	2	6	4.96	1.04
7. I feel more confident in taking TOEFL.	49	2	6	4.49	1.12
8. Besides preparing the TOEFL test, I also learned practical skills such as critical thinking and problem-solving skills in TOEFL class.	49	3	6	4.67	.97
9. The test strategies I learned in the class are helpful for taking the TOEFL exam.	48	2	6	4.81	1.09
10. The TOEFL class is interesting	48	1	6	4.67	1.24
11. The amount of practice/exercises we do in class is adequate (just about the right amount, not too much practice, not too little practice) for preparing for the TOEFL exam	48	1	6	4.37	1.30
12. The amount of homework we do outside of class is adequate (just about right, not too much homework, not too little homework) for preparing for the TOEFL exam	49	2	6	4.47	1.08

13. The TOEFL class greatly helps me prepare for the TOEFL exam.	49	1	6	4.76	1.09
14. The homework/exercises I did in class are helpful for learning English.	49	1	6	4.65	1.13
Valid N (listwise)	22				

Note. 1 = strongly disagree, 2 = agree, 3 = somewhat disagree, 4 = somewhat agree, 5 = agree, 6 = strongly agree

Qualitative data provided additional information that complemented and extended the quantitative results. It also shed light on potential issues students had with the TOEFL class. Three main themes emerged from student interviews and their open-ended responses. They are the need for having a TOEFL class; the effects of the TOEFL Class on students' TOEFL scores and language learning; and students' attitudes toward homework, classroom activities, and classroom materials.

6.1.1 The Need for Having a TOEFL Class

In terms of whether or not having a TOEFL class is necessary, the interviews provided rich information that further enriched the survey results. Sonny, Lynda, and Moka commented that the TOEFL class is "very necessary" in an IEP for international students who want to apply for universities. Lynda said, "other classes students may not come, but TOEFL class, everybody comes." This shows that the TOEFL class has more accountability in terms of attendance. Students saw the TOEFL class as important and necessary. The IEP can help students obtain the needed TOEFL scores set by the university.

However, the TOEFL class can be a waste of time for those who come to the program wishing to improve their English language skills. Lynda said "TOEFL test is necessary and helpful for IEP students. It's important to students to apply for universities. But for some students who did not need to pass the TOEFL...some students are immigrants and attend IEP because they want to improve their English. It's waste of their time." Sonny is one of the students who does not need to take TOEFL but is required to enroll in the class in order to maintain her visa status at the IEP. Sonny said, "I don't need the TOEFL score, I want to develop my speaking English but it's just required course in IEP. So, I just take it." This quote shows that Sonny is frustrated that she had to take the TOEFL class due to the institutional and international student policy. Excerpt 1 further indicated her reluctance to take the TOEFL class.

Excerpt 1 (4/20/2018)

R: Do you think the TOEFL class is necessary at an IEP?

S: I don't think so. I think the TOEFL and IEP have different purposes. TOEFL class is for the test, they just teach the tips, but not language. And other courses is for other skills, language skills. Communication class is not related to TOEFL class, reading class is not related to TOEFL class. Only the TOEFL class does not relate to the other language classes

S: Some students don't need the TOEFL strategies...

R: *Like you*

S: *yeah, like me. TOEFL class just teach the strategies. Other class are more beneficial for me because they are helpful for my English*

Note. R = Researcher, S = Sonny

This excerpt shows that the student perceived TOEFL class as “just teach(ing) the strategies” to cope with the TOEFL ITP. Sonny did not need to learn that as she already met the TOEFL cut score. Some students enroll in an IEP to improve their everyday English skills rather than to be admitted to the university. To these students, the TOEFL class may not be necessary.

Additionally, excerpt 1 also demonstrates a difference between the TOEFL class and the three regular ESL classes. This view was reflected by other students as well (Marie, Anna from Ms. C’s class). For instance, Anna mentioned, “I think TOEFL class is something else. Oral communication, we need activities, but in TOEFL, we just do exercise...TOEFL class is just for the exam. Oral communication other classes they help me in real life.” According to these students, compared to the other ESL classes, the TOEFL class is “something else.” It focuses less on language skills and seemed to be less helpful for students’ language improvement.

6.1.2 The Effects of the TOEFL Class on Students’ TOEFL Scores and Language Learning

When it comes to what students think about the influence of the TOEFL class on their TOEFL scores, the interviews expanded the survey findings. To one student interviewee, the TOEFL class had a little (Moka) effect, to some, it had some effects (Chris, Sonny, Eric, Lynda, Von). To others (Asset, Hannah, Monika), it had a large effect on their score improvement. However, the majority of student interviewees believed that 70%-80% of the score improvement was up to their own study and practice. For instance, when asked about “How many percentages would you say the influence of the TOEFL class is on your score?” Maria said, “A high percentage like an 80% you have to study by yourself but I think the class is very helpful.”

In regards to English language learning, the majority of the student interviewees believed that the TOEFL class helped them improve listening, grammar, and vocabulary skills. The following were quotes recorded from student interviews.

Hannah: “I can improve my listening skill in the TOEFL class, because he (the instructor) always speaks English of course.”

Monika: “I can learn formal things or business things from TOEFL, that is what I don’t learn from normal life. Because in my life, I don’t use the TOEFL words, or the scientific words. It’s kind of good chance to learn specific words.”

Maria: “yes, for example. Grammar skills. They focus on different things, for example, the difference between ‘make’ and ‘do,’ the different propositions we have to use.”

Hannah and Monika are two Japanese exchange undergraduate students and they both believed that the class was helpful for their language learning. Hannah believed that because

the TOEFL class was taught by a native speaker, she can improve her listening skills naturally. Her friend Monika reported that she can learn the academic content and vocabulary from the TOEFL class. Maria also indicated the benefits of the TOEFL class in developing various receptive skills.

TOEFL class had a positive impact on students' learning of receptive skills; however, there were no speaking or writing skills taught in the class. The following quotes are some examples.

Marie: "We never practiced speaking and writing."

Marie: "We take TOEFL ITP. We don't need to speak or write, but many students need to speak and write more. I remember all the grammar skills but I don't speaking well. And maybe I think we need to prepare for speaking well, because all the time we practices more grammar grammar grammar."

Asset: "Honestly, the classes are mostly helpful for my TOEFL scores. I train my English skills by talking with others, watching TV shows, and going to the language courses."

The quotes demonstrate that Marie was not happy with only doing grammar drills every day for class. She wanted to learn more practical skills and wanted to improve her speaking. Asset's view was more extreme, he believed that the class was mainly for improving the scores, not for language learning. Asset alluded that he used outside resources to help him improve his English productive skills. Talking with friends, watching shows, and taking language courses were more helpful for him to practice his English skills than the TOEFL class.

6.1.3 Students' Attitudes Toward Homework, Classroom Activities, and Classroom Materials

21 out of 49 students filled out the open-ended survey responses. Approximately half of the students felt that the homework was not enough and wanted more exercise. For example, students made comments, such as "We need more homework!" "We need more practice!" on the survey. Student interviews and teacher interviews further supported the finding that there was a lack of homework in TOEFL classes. Because students' TOEFL class grades were solely based on their two times' TOEFL ITP scores during the semester, the teachers only gave optional homework to students. The homework was never checked because it does not constitute their class grade. During the interviews, Lynda, Moka, and Von said they wanted the teachers to assign homework. Von said, "We must have assignments and the teacher needs to check them."

In terms of classroom activities, nine students (43%) indicated that they liked the class and the activities. However, during the interview, Asset said, "Sometimes the class is boring. (The) activities (we do are) like test, doing exercises. It's like routine, you came to TOEFL class and you do the same thing every time." This indicated that the TOEFL class activities were simply exercises. They were "good" only for test purposes. Additionally, two students from the survey felt that classroom activities could be more interesting. Additionally, three students wrote that they wish to practice more mock tests in class, not only the exercises from

the textbook. During the interviews, Marie, Anna, and Chris said the same thing about wanting to have more mock tests. Anna said, “In class, we only practice sections of the test. We need to practice real test.”

For classroom handouts and the TOEFL textbook, eleven students (52%) commented that the textbook was helpful. ID 2 mentioned that he liked that the level 5 teacher supplemented the class with TOEFL IBT materials. Three other students shared that they wished to learn more about TOEFL IBT. The following quotes illustrate the different opinions students had regarding having TOEFL IBT or ITP as course materials.

ID3: “TOEFL IBT materials are very useful. We also can try speaking practices for TOEFL IBT.”

ID 30: “Nowadays most of other university don't accept the TOEFL ITP, so our university also have to change to IBT TOEFL. It's just my opinion, our university needs to follow the trend.”

Von: “It's very helpful. I came to this program because I don't have enough TOEFL IBT scores. Now if I pass ITP, I can go directly to university.”

Anna: “No I don't want IBT because it's more difficult. If other schools they require IBT, I think our university should prepare us for IBT.”

ID 3 talked about his hope to practice speaking using the TOEFL IBT materials. He indicated that TOEFL IBT may be another option to consider for a TOEFL course, as it offers the speaking component that the paper-based test was lacking. ID 30 suggested changing the ITP course to IBT because IBT scores were more widely accepted by universities and were thus more practical. However, not all students preferred to switch to the TOEFL IBT class. Some students indicated that they did not want to take IBT as it is a much harder exam. To Von, TOEFL ITP had a lower English standard and provides him an opportunity to gain entry to the university. When it comes to having IBT or IPT preparation classes, Anna expressed her dilemma. On one hand, she wanted to take the “less difficult” ITP test. But on the other hand, she believed that it is necessary to “follow the trend” and switch to the IBT class if the IBT scores are more useful.

Additionally, because the TOEFL class was shorter than regular ESL classes, three students surveyed indicated that the class was too short and suggested to make the TOEFL class longer. For instance, ID 29 wrote, “The TOEFL class is very important for international students, we need more TOEFL class.” From the interviews, Sonny, Marie, Anna, and Moka also indicated that they wanted to change the class schedule from twice a week to three times a week like the other three ESL classes.

6.2 RQ2: Do Students' TOEFL Scores Improve After Taking TOEFL Classes?

Eleven students' four times TOEFL ITP scores were available for the spring group (From 2017 fall to 2018 spring) and seven students' four times scores were available for the fall group (From 2018 spring to 2018 fall). The spring and fall 2018 student groups' four times TOEFL ITP scores were first analyzed using the one-way repeated measure ANOVA to

detect their improvement over time. Because students' two times TOEFL scores for the spring ($N = 26$) and the fall ($N = 25$) semester were larger than their four times' TOEFL scores, paired-sample t-tests were conducted to further investigate whether a significant result can be detected within one semester period. As assessed by Shapiro-Wilk's test ($p > .05$), test scores were normally distributed at each time point. The assumptions of Mauchly's test of sphericity was met for the spring 2018 group (four tests taken from fall 2017 to spring 2018), $\chi^2(5) = 5.35$, $p = .37$. As shown in Table 4, there were significant differences over two semesters' TOEFL scores for the spring 2018 student group. $F(3, 36) = 5.83$, $p < .05$, partial $\eta^2 = .33$. Post hoc pairwise comparisons were performed. As shown in Table 5 and Figure 1, for spring 2018, significant differences between scores were only detected between the first and the third time students took the TOEFL. Students had a significant increase of 20 points (95% CI, 5.20 to 34.80) from the first to the third time taking the TOEFL test ($p = .01$). No significant difference was found between the first and the second time, between the third and the fourth time, or between the second and the fourth time. The post hoc G*Power analyses indicated that for a sample of 13 students, partial $\eta^2 = .33$, alpha value .05, a power of .99 was achieved for the repeated measure ANOVA result for spring 2018.

Table 4. Tests of within-subjects effects (Spring 2018 group)

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	4023.90	3	1341.30	5.83	.002	.33
Error (Time)	Sphericity Assumed	8279.35	36	229.98			

Table 5. Comparing the differences among students' four different times' TOEFL scores

Post hoc test with Bonferroni adjustment (Spring 2018 group)

Time (I)	Time (J)	Mean Difference (I-J)	SE	Sig.	95% Confidence Interval for Difference	Lower Bound	Upper Bound
Mid-2017Fall	End-2017Fall	-4.69	4.03	1.00	-17.38	8.00	
	Mid-2018Spring	-20.00*	4.70	0.01*	-34.80	-5.20	
	End-2018Spring	-19.23	6.48	0.07	-39.67	1.21	
End-2018Fall	Mid-2017Fall	4.69	4.03	1.00	-8.00	17.38	
	Mid-2018Spring	-15.31	5.60	0.11	-32.97	2.36	
	End-2018Spring	-14.54	7.26	0.41	-37.41	8.34	
Mid-2018Spring	Mid-2017Fall	20.00*	4.70	0.01*	5.20	34.80	
	End-2017Fall	15.31	5.60	0.11	-2.36	32.97	
	End-2018Spring	0.77	6.93	1.00	-21.07	22.61	

End-2018Spring	Mid-2017Fall	19.23	6.48	0.07	-1.21	39.67
	End-2017Fall	14.54	7.26	0.41	-8.34	37.41
	Mid-2018Spring	-0.77	6.93	1.00	-22.61	21.07

Note. *The mean difference is significant at the .05 level.

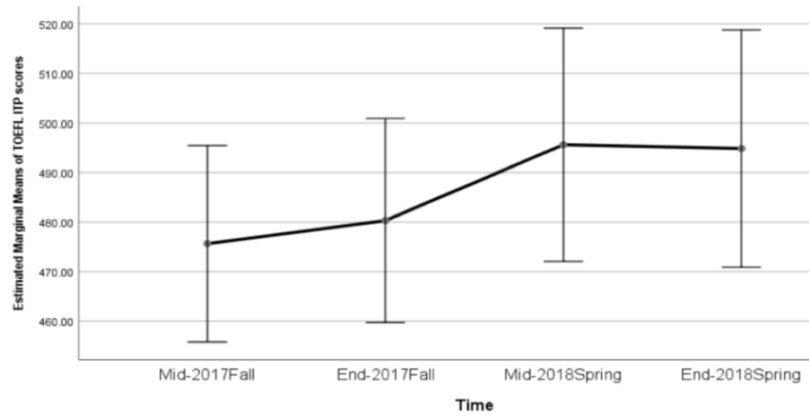


Figure 1. Estimated marginal means of TOEFL ITP scores across four different times for the spring 2018 group ($N= 11$)

Error bar shows 90% CI.

For the fall 2018 group, test scores were normally distributed at each time point as assessed by Shapiro-Wilk’s test ($p > .05$). The assumptions of Mauchly’s test of sphericity was met for fall $\chi^2(5) = 5.85, p = .33$. The estimated marginal means plot in Figure 2 indicated that students on average improved their TOEFL scores over two semesters. As indicated in Table 6, the difference in scores among the four different testing times were marginally significant: $F(3, 18) = 3.28, p = .05$, partial $\eta^2 = .35$. Finally, the post hoc G*Power analyses indicated that for a sample of seven students, partial $\eta^2 = .35$, alpha value .05, a power of .99 was achieved for the repeated measure ANOVA result for the fall 2018 group.

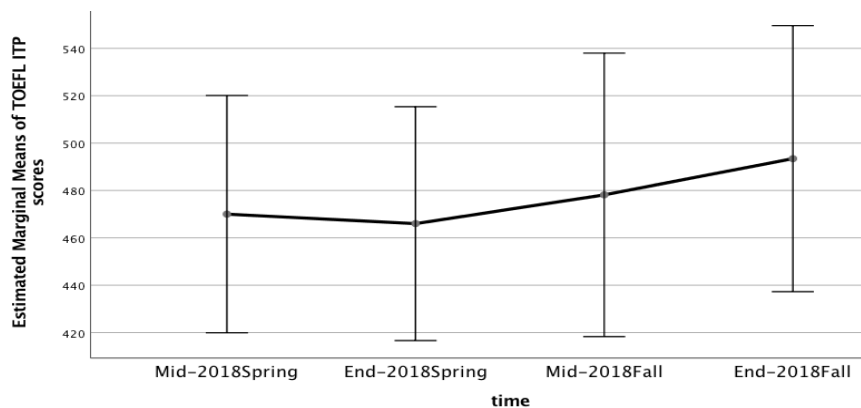


Figure 2. Estimated marginal means of TOEFL ITP scores across four different times for the 2018 fall group ($N = 7$)

Error bar shows 90% CI.

Paired-sample t-tests were performed to further examine whether there was a significant result within one semester period (two-time points: mid and end-of-the-semester in spring and fall) and whether the result was consistent with the one-way repeated measure ANOVA using data collected from the four times' test scores. Results further confirmed that students' TOEFL scores taken in the mid- and end-of-the-semester were not statistically significant with $p = .68$ for 2018 fall students and $p = .07$ for 2018 spring students (See Table 6)

Table 6. Paired-sample t-tests of spring and fall students' two times' TOEFL scores

	Paired Differences							Sig. (2-tailed)
	M	SD	SE	95% Confidence Interval of the Difference		t	df	
				Lower	Upper			
Pair 1 Score1_Mid_2018Spring Score2_End_2018Spring	2.89	34.73	6.81	-11.14	16.91	.42	25	.68
Pair 2 ScoreA_Mid_2018Fall ScoreB_End_2018Fall	-9.80	25.86	5.17	-20.47	.87	-1.90	24	.07

7. Discussion

The results indicated that the TOEFL class had a mixed washback effect on students' attitudes and TOEFL ITP score improvements. Overall, results from the first research question indicated that students in both spring and fall 2018 semesters had positive attitudes toward the class. Students generally believed that the TOEFL class was helpful, informative, and necessary in the IEP. Students also believed that they have improved their receptive language skills. This shows that TOEFL ITP yielded some positive washback on learning TOEFL-related language skills such as grammar, vocabulary, and listening. However, student open-ended survey responses and interview data revealed that their positive attitudes toward the class were mainly because the TOEFL ITP was a test they must pass in order to go to the university and they thought that the TOEFL strategies and exercises were necessary to prepare them for the exam. Students' positive attitudes did not stem from the enjoyment of the TOEFL class or the improvement of English language skills, but rather the importance of obtaining a high TOEFL score.

Compared with the TOEFL class, students believed that the non-TOEFL classes were more helpful in improving their language abilities. Golubovich et al. (2018) indicated that TOEFL ITP did not directly evaluate productive skills and this may be one of the limitations of the test. As illustrated in Golubovich et al. (2018), because TOEFL ITP did not assess speaking or writing skills, teachers did not spend time practicing these skills, which were necessary for students to succeed in their academic studies at the university. This finding is concerning because students without speaking and writing skills may not be able to meet the demands of university courses and may end up struggling in and failing their classes. As a result, there was an unintended negative washback of TOEFL ITP on students' learning. For instance, the class became test-oriented and there was little opportunity for students to learn authentic English and develop productive language skills.

TOEFL classes can be improved, not only through implementing effective teaching strategies to improve student scores, but also through incorporating authentic English into the class. In this way, the class can be less drill-oriented and more authentic. Teachers and language program directors should reconsider the objectives of the TOEFL class and design it to not only improve students' test scores but develop their language learning skills as well. The TOEFL class should not be seen as "something else" or "different" from the other language courses, but should be embraced as an integral part of IEP so that even for students who did not need to take TOEFL can benefit from it in the long-term.

7.1 Students "Want More" From the TOEFL Class

The qualitative data provided more information that was not available in the quantitative results. First, students believed that the TOEFL class time was "too short." Students constantly reported in the interviews and survey responses, "we need more!" and "it is not enough!" They wanted more exercises, homework, strategies, mock-tests, and explanations of answer choices. They also requested more time spent reviewing the strategies, and more time for the TOEFL class in general. It is interesting that on one hand, students disliked that the class was filled with exercises, yet on the other hand, they still sought to have more practice because they saw the importance of the TOEFL ITP score. . Additionally, because class grades were only based on students' mid and end-of-semester TOEFL scores, there was no accountability for students to do homework and for teachers to grade them.

The IEP can improve by finding ways to maximize TOEFL class hours and adjust its grading system by adding a homework component in the midterm and final exam grades. In this way, teachers can assign and check homework, take more class time to explain each answer and review instead of "rushing" through skills and exercises. ELPs that already incorporate TOEFL classes can also reflect on their practices and find similar ways to meet student needs. This "want more" phenomenon also showed that students were motivated to improve their scores and that the TOEFL class is important. ELPs should take students' suggestions and needs into consideration when designing TOEFL classes.

7.2 The Choice of TOEFL IBT or TOEFL ITP Preparation Courses

Students' comments from the surveys and the interviews indicated that the TOEFL class is "very necessary" and "helpful" to prepare for TOEFL ITP. Thus, language programs that only offer ESL language courses should consider incorporating TOEFL courses into the curriculum. One interesting finding was that students suggested changing the paper-based TOEFL ITP class to the Internet-based TOEFL IBT class. There are two main advantages to a TOEFL IBT class. First, TOEFL IBT assesses speaking and writing skills; therefore, students will be able to practice the two skills in class and are likely to learn more practical English skills rather than solely doing drills/exercises in a TOEFL ITP class. Second, the majority of American universities accept IBT but not ITP scores as an admission requirement. Having a TOEFL IBT course can help students prepare for the IBT and give them more opportunities to enroll in other universities. It provides more opportunities for student academic success. However, there are some practical concerns for making this change. TOEFL ITP is easy for institutions to administer and is less expensive than IBT (Golubovich

et al., 2018). Some IEP students also indicated that they did not want to take IBT because it is a much harder exam. ITP provides them more chances to pass and enter the university. Finally, TOEFL ITP courses are already difficult for students at the IEP. It is questionable whether level 3 and level 4 students can handle the materials of TOEFL IBT. Therefore, teaching TOEFL IBT to the beginning- or intermediate-level IEP students may not be practical.

7.3 Positive Washback on Score Improvement

Findings from the second research question demonstrated that students in general improved their scores over time. More specifically, students' TOEFL ITP scores statistically improved from the first to the third time taking the TOEFL test for the spring group. This significant result aligned with Choi and Papageorgiou's (2014) findings which show that students in both Thailand and Turkey performed significantly better three to four months after receiving English lessons in the TOEFL ITP. The improvement was almost as significant for the fall group in the current study. This result may be due to the small sample size to detect statistical significance. Additionally, because there was only a two-month gap between the four times students took the test, significant score differences were not found within one semester period. This result indicated that it is difficult to see a significant increase in scores in two months even with TOEFL preparation training. Teachers and English language programs should anticipate this result and consider administering the TOEFL ITP test only once per semester or extending the test-taking times to be every four months instead of two months.

8. Limitation and Future Directions

Limitations of the study must be acknowledged. First, the student attitude survey used in the study was primarily perception items that focused on how students think and feel which are subjective (Lavrakas, 2008). Future studies can include behavior items that are objective, such as "I spend much a lot of time outside of the class preparing for TOEFL" and use this item as an indication for the importance of TOEFL. Second, because students who take non-TOEFL classes at the IEP are also required to take TOEFL classes, there is no comparison group that only takes non-TOEFL classes and do not take TOEFL classes. Therefore, it is not certain whether there will be a difference in students' TOEFL score improvement between the ones who take the TOEFL class and the ones who do not. In addition, students take three regular ESL classes while also taking the TOEFL class. Thus, the three regular ESL classes may also had an effect on students' TOEFL score improvement, however, the results were based on data collected from real-life situations where students take the TOEFL classes simultaneously with the other ESL courses offered at the program. Furthermore, the qualitative data helped address this limitation and provided evidence on students' perceptions toward the impact of the TOEFL class on their score improvement and English language learning. Although the generalizability of the study result is limited, the study shed light on program development and pedagogies regarding TOEFL courses. Results and suggestions made based on students' attitudes toward the TOEFL class and their score improvement over time can facilitate other English language programs' curriculum design and development. ELPs that consider incorporating TOEFL classes should consider

lengthening the TOEFL class duration, integrating authentic English materials and instruction, as well as offering both the paper-based and internet-based TOEFL classes.

References

- Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, 6(1), 7. <https://doi.org/10.1186/s40468-016-0030-z>
- Barnes, M. (2016). The Washback of the TOEFL iBT in Vietnam. *Australian Journal of Teacher Education*, 41(7), 158-174. <https://doi.org/10.14221/v2016n41.10>
- Cheng, L. (1998). Impact of a Public English Examination Change on Students' Perceptions and Attitudes toward Their English Learning. *Studies in Educational Evaluation*, 24(3), 279-301. [https://doi.org/10.1016/S0191-491X\(98\)00018-2](https://doi.org/10.1016/S0191-491X(98)00018-2)
- Choi, I., & Papageorgiou, S. (2014). Monitoring students' progress in English language skills using the TOEFL ITP assessment series. *Research Memorandum No. RM-14-11*. Princeton, NJ: Educational Testing Service.
- Chou, M. H. (2017). Impacts of the Test of English Listening Comprehension (TELC) on teachers and teaching in Taiwan. *Asian-Pacific Journal of Second and Foreign Language Education*, 2(1), 5.
- Creswell, J. W., & Garrett, A. L. (2008). The "movement" of mixed methods research and the role of educators. *South African Journal of Education*, 28(3), 321-333. <https://doi.org/10.15700/saje.v28n3a176>
- ETS. (2018). *About the TOEFL ITP assessment series*. Retrieved from www.ets.org/toeflitp
- Gan, Z. (2009). IELTS preparation course and student IELTS performance: A case study in Hong Kong. *RELC Journal*, 40(1), 23-41. <https://doi.org/10.1177/0033688208101449>
- Golubovich, J., Tolentino, F., & Papageorgiou, S. (2018). Examining the Applications and Opinions of the TOEFL ITP® Assessment Series Test Scores in Three Countries. *ETS Research Report Series*, 2018(1), 1-30. <https://doi.org/10.1002/ets2.12231>
- Green, A. (2006). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing*, 11(2), 113-134. <https://doi.org/10.1016/j.asw.2006.07.002>
- Green, A. (2007). IELTS Washback in Context: Preparation for academic writing in higher education. *Studies in Language Testing* 25. Cambridge: Cambridge University Press.
- Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39-51. <https://doi.org/10.6018/ijes.13.2.185891>
- Hamp-Lyons, L. (1998). Ethical Test Preparation Practice: The Case of the TOEFL. *TESOL Quarterly*, 32(2), 329. <https://doi.org/10.2307/3587587>

- Hawkey, R. A. H. (2006). Impact theory and practice: Studies of the IELTS test and ProgettoLingue 2000. *Studies in Language Testing* 24. Cambridge: Cambridge University Press and Cambridge ESOL.
- Huang, B. H. (2018). PD in Testing Preparation for University ESL. *The TESOL Encyclopedia of English Language Teaching*, 1-6. <https://doi.org/10.1002/9781118784235.eelt0886>
- Huang, B., Zhi, M., & Wang, Y. (2018, April). Investigating the Validity of a University-Level English Language Placement Assessment: A Mixed-Methods Study. *Paper presented at American Educational Research Association (AERA)*, New York City, New York.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Khodabakhshzadeh, H., Zardkanloo, R., & Alipoor, I. (2017). The Effect of Mock Tests on Iranian EFL learners' Test Scores. *International Journal of Education and Literacy Studies*, 5(3), 47-51. <https://doi.org/10.7575/aiac.ijels.v.5n.3p.47>
- Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage Publications, Inc. <https://doi.org/10.4135/9781412963947>
- Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & quantity*, 43(2), 265-275. <https://doi.org/10.1007/s11135-007-9105-3>
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256. <https://doi.org/10.1177/026553229601300302>
- Mickan, P., & Motteram, J. (2008). An ethnographic study of classroom instruction in an IELTS preparation program. *IELTS Research Reports* (p. 8). Canberra: IELTS Australia.
- Mickan, P., & Motteram, J. (2009). The preparation practices of IELTS candidates: Case studies. *International English Language Testing System (IELTS) Research Reports*, 10(1).
- Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis*. Beverly Hills.
- Muniz, H. (2017). *TOEFL requirements for college: what scores do you need?*. Retrieved from <https://www.prepscholar.com/toefl/blog/toefl-score-requirements/>
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33-49. <https://doi.org/10.1177/0265532209347148>
- Neal, J. W., Neal, Z. P., VanDyke, E., & Kornbluh, M. (2015). Expediting the analysis of qualitative data in evaluation: A procedure for the rapid identification of themes from audio recordings (RITA). *American Journal of Evaluation*, 36(1), 118-132. <https://doi.org/10.1177/1098214014536601>
- Rashidi, N., & Javanmardi, F. (2011). The IELTS Preparation Washback on Learning and Teaching Outcomes. *Cross-Cultural Communication*, 7(3), 132.
- Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage.

Saville, N., & Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. *Washback in language testing: Research contexts and methods*, 73-96.

Shih, C. -M. (2007). *Perceptions of the General English Proficiency Test and its washback: a case study at two Taiwan technological institutes*. Library and Archives Canada = Bibliothèque et Archives Canada, Ottawa.

Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155. <https://doi.org/10.1093/eltj/cci030>

Wall, D., & Horák, T. (2008). The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 2, Coping with change. *ETS Research Report Series*, 2008(2), i-105. <https://doi.org/10.1002/j.2333-8504.2008.tb02123.x>

Wall, D., & Horák, T. (2011). The impact of changes in the TOEFL exam on teaching in a sample of countries in Europe: Phase 3, The role of the coursebook phase 4, describing change. *ETS Research Report Series*, 2011(2), i-181. <https://doi.org/10.1002/j.2333-8504.2011.tb02277.x>

Wang, Y. (2019). The Impact of TOEFL on instructors' course content and teaching methods. *TESL-EJ*, 23(3).

Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, 30(1), 49-70. <https://doi.org/10.1177/0265532212442634>

Yu, G., He, L., Rea-Dickins, P., Kiely, R., Lu, Y., Zhang, J., ... Fang, L. (2017). Preparing for the Speaking Tasks of the TOEFL iBT® Test: An Investigation of the Journeys of Chinese Test Takers. *ETS Research Report Series*, 2017(1), 1-59. <https://doi.org/10.1002/ets2.12145>

Zhi, M., & Wang, Y. (2019). Washback of College Entrance English Exam on Student Perceptions of Learning in a Chinese Rural City. In *Chinese-Speaking Learners of English: Research, Theory, and Practice* (1st ed., pp. 26-37). <https://doi.org/10.4324/9780429290848-3>

Appendix

Appendix 1. Student Interview

1. Can you tell me a little bit about yourself? (your home country, how long have you lived in the U.S? your English learning experience)
2. Why are you taking the TOEFL course? And how long have you been taking the course at the program?
3. Think back and reflect, what did you learn from the TOEFL class?
4. What do you like and dislike about the class?
5. What do you think about the influence of the class on your TOEFL score?

6. What do you think about the influence of the class on your English learning?
7. How does TOEFL influence your English skills, speaking, reading, writing, and listening?
8. What do you think about the TOEFL test strategies taught in your class?
9. What do you think about the materials and activities used in the class?
10. Do you spend extra time outside of class to prepare for the TOEFL, if so, how?
11. What is your attitude towards the TOEFL test?
12. What is your attitude towards the TOEFL course?
13. Do you think TOEFL course in IEP is necessary? Helpful?
14. Do you have any suggestions for the TOEFL course in the ESL program?
15. Is there anything else you want to add or talk about the TOEFL course?

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>)