Comparing TOEFL and teacher's assessment scores as a snapshot of student's English skills

Pritz Hutabarat
Universitas Kristen Maranatha
pritz.hutabarat@lang.maranatha.edu

Abstract

Standardized English tests have long been used to determine English learner's level of proficiency. Tests such as TOEFL (Test of English as a foreign language) and IELTS (International English language testing system) are arguably the most popular English tests in the world. On the other hand, English learners are consistently being assessed by their teachers through a performance-based assessment which provides a comprehensive description of student's learning achievement in learning English. The results of these two forms of assessments were hardly compared and correlated. Therefore, this paper investigates the correlation between student's TOEFL score and teacher score. Moreover, this study explores how teacher uses washback effects of TOEFL and class tests. Investigating 42 grade 12 students in an English immersion school in Indonesia, the result of this study indicated a positive but weak correlation between TOEFL and teacher's assessment scores ($r = 0.246$, $p = 0.116$). This weak correlation indicates that the two types of assessments measure student's achievement and performance differently. The washback effect of TOEFL is mainly in the form of teacher's encouragement to the students to be more serious when taking TOEFL test so that they will be about to get higher scores.

Keywords: IELTS; immersion; performance-based assessment; TOEFL

Introduction

Standardized tests have gained more popularity as recommended tests to use by teachers and schools within the field of English language teaching because it is believed that they have been carefully designed by reputable institutions and experts in the area (Liao, 2022a). One of the usages of standardized tests at university is that it is used as a mandatory test for students to graduate (Amelia & Harmaini, 2020). In their study, Amelia and Harmaini observed that TOEFL became a high-stake test and that nearly 400 students out of 1,220 had to take the test numerously before they could obtain the required TOEFL score. Ma and Chong (2022) on the other spectrum reports

the use of IELTS as a pre-entrance test for Chinese students going to study at a Sino-UK programme in a Chinese campus. To be eligible, the candidate to the programme needs to have IELTS score of 5.5 to start the course. However, in year 2 they are required to have at least band 6.0 in IELTS. If they fail to achieve it, they are not allowed to continue to the programme.

Standardized test measures student's proficiency level such as grammatical, skills and other linguistic knowledge within certain domain of language usage (Fulcher & Davidson, 2009). Weaver (2016) warns against the limitations of TOEIC as a standard test as its score does not provide a comprehensive reflection of the test taker's English level as it lacks in testing productive skills such as speaking and writing. Weaver (2016) recommends administering a separate speaking and writing tests to compliment TOEIC. Kokhan (2013) also points out the danger of using standardized tests as the sole predicator to place international students into their English as Second Language (ESL) programme because of the danger of misplacing the students into the wrong categories which may deny their right to get support in terms of developing their academic writing skills.

The TOEFL iBT assesses a student's capability to utilize and comprehend English in a university setting by evaluating their combined abilities in listening, reading, speaking, and writing for academic purposes. The scores on the Internet-based TOEFL (iBT) indicate the student's proficiency in English, ranging from low to intermediate to advanced (University Language Service, 2013). The components of the TOEFL iBT test are summarized in Table 1 on the ets.org website (ets.org, 2013).

Table 1 Sections in TOEFL

Section	Time Limit	Questions	Tasks
Reading*	60–80 minutes	36–56 questions	Read 3 or 4 passages from academic texts and answer questions.

Listening	60–90 minutes	34–51 questions	Listen to lectures, classroom discussions and conversations, then answer questions.
Break	10 minutes	—	—
Speaking	20 minutes	6 tasks	Express an opinion on a familiar topic; speak based on reading and listening tasks.
Writing	50 minutes	2 tasks	Write essay responses based on reading and listening tasks; support an opinion in writing.

Some studies investigate the correlation between TOEFL score and academic performance such as in Cho Y. and Bridgeman B (2012), Ginther A. and Yan X. (2018), and Rahmat et al. (2015). Cho and Bridgeman (2012) investigated how English proficiency level as indicated in TOEFL, GRE, GMAT, and SAT score of 2594 undergraduate and graduate students from 10 universities in the United States reflect their academic achievement. Their findings show that there is a positive but weak relationship between student's English proficiency test scores with their academic achievement as shown in their GPA

On the other hand, teachers assess their student's performance through various media such as day to day exercise, homework, weekly and monthly tests. Heaton (1990) argues that teachers utilizes tests as a device to support the learning process and motivate students or as a means of assessing student's performance in the language. Feron, Schil, and Weel (2016) investigating the potential of using teacher's score to predict student's performance in cognitive test for primary school students argue that teacher's subjective assessment is more reliable to place students into the correct class or programme compares to standard cognitive tests. Although the study was conducted in primary school setting, the result of the study validates teacher's subjectivity as a reliable source of measuring student's ability.

Akhter and Malik (2019) highlights the roles of teacher's beliefs in the assessment. Working with 344 secondary teachers in Pakistan, Akhter and

Malik conclude that teachers value communicative aspects of assessment and that assessment is a learning activity. They emphasize on the core value of an assessment is that it should inform the students, parents, and teacher where the students are in their learning progress. Another implication of the communicative aspect of an assessment is that it should inform the teacher what needs to be done to better help students in their learning process. In addition, teachers in Indonesia are required to teach based on the national curriculum or guidance given by the Ministry of National Education (MoNE). Therefore, their primarily responsibility is to ensure the success of their students in the national examination.

In this study, teacher scores are based on the student's performance in the classroom through formative and summative assessments. Formative assessments take place in nearly every class session and evaluate the student's performance and participation based on each lesson. Summative assessments, on the other hand, are given twice per semester, including a mid-term test and a final exam. These evaluations assess the student's overall English language skills, including speaking, listening, reading, and writing, through activities such as reading books and writing academic papers. The teacher score is represented as a single numerical score that reflects the student's performance in comparison to the school's standards in four areas: participation, homework, daily performance, and tests. A student is considered to have passed the subject if their total score is greater than 70.

Another aspect of testing is washback effects. Washback can be simply defined as the effects of a test to the teaching and learning process (Alderson & Wall, 1993). Teachers use the washback as an input to evaluate the effectiveness of teaching and learning process. A study of 12 English teachers from 12 secondary schools in Bangladesh shows the negative washback of a high stake test which questioned the relevant of such test to support the learning process (Rahman et al., 2021).Teachers may reflect upon the test results to help them plan for their future lessons or to decide whether or not to

use the same test again. Washback can be positive or negative. Positive washback means that the tests produce the expected test effects. For example, a test may motivate students to study more or may promote a connection between standards and instruction. Negative washback refers to the unexpected, harmful consequences of a test. For example, instruction may focus too heavily on test preparation at the expense of other activities (CAL, 2016).

In his work, Brown (2010) explores the concept of beneficial washback and highlights its many positive aspects. First, it is evident that beneficial washback has a positive impact on both the instructional content and methods employed by instructors. This influence extends to the learners themselves, influencing not only what they learn but also how they acquire it. Its capacity to provide students with the opportunity for comprehensive preparation enables them to approach their studies with a sense of preparedness. In addition, this process integrates feedback mechanisms that play a crucial role in fostering the language skill development of students. Furthermore, beneficial washback is characterized by its predominantly formative nature, which prioritizes ongoing assessment and development over final judgments. This is consistent with the notion that the process is intended to foster continuous growth rather than solely provide summative assessments. In addition, Brown emphasizes how this approach creates an environment conducive to learners' highest accomplishments and optimal performance (Brown & Abeywickrama, 2010). Through these multiple dimensions, the concept of beneficial washback demonstrates its capacity to positively shape the teaching and learning environments.

Moreover, Gokturk Saglam, A (2018) investigates 14 English language teachers at a university in Turkey who were participating in an English preparation programme for new students. The results showed that the language proficiency test had both positive and negative impacts on teaching. The positive washback was seen in the choice of teaching

materials, but the study also found negative washback in terms of a more limited curriculum. It was discovered that the degree and type of washback varied depending on the teacher, but the content and teaching methods were both affected.

It is this duality in the English testing which sparked questions by which this study was conducted. This study investigates how teachers of English in Indonesia make use of TOEFL test and their own test in informing their teaching practices. Therefore, this research is guided by two research questions below:

1. What is the correlation between teacher's assessments and TOEFL score?
2. How does the teacher use the information from the students' TOEFL and class scores?

Research Methodology

This study was conducted as mixed method that it combined both qualitative and quantitative research methods in a single study. According to Creswell (2018), mixed methods research is a research approach that involves the collection and analysis of both qualitative and quantitative data within a single study. Creswell defines mixed methods research as "a research design that involves collecting, analyzing, and mixing both quantitative and qualitative data within a single study or a multiphase study" (Creswell, 2014, p. 47). Creswell (2018) notes that the key feature of mixed methods research is the integration of both qualitative and quantitative data in a way that takes advantage of the strengths of each approach. This allows researcher to triangulate the data and gain a more comprehensive understanding of the research problem. Therefore, the data for this study was comprised by quantitative data in the form of student's TOEFL score and teacher's scores and qualitative data from semi structured interviews. The data was first analysed statistically using SPSS to see the correlation between student's TOEFL score and their teacher's score. The qualitative data gathered through

semi-structured interviews was analysed using thematic analysis. The results from the two analysis then were compared to see the consistency between the correlation test and teacher's perceptions of the use TOEFL and classroom score in assessing student's English level.

The participants of this study are 42 students of grade 12 senior high school and one English teacher teaching the class respectively. The school implements English medium instruction (EMI) programme. The teacher participated in this study is a native English speaker who at the time of data collection had been teaching at the school for three years. All of the students had learned at the school for more than two years which means that they are familiar with the English environment and can follow the instructions in the class.

Findings and Discussion

Findings

The findings of the study will be presented in two sections according to the research questions guiding this study.

1. Correlation between Teacher scores and TOEFL score

Data from class reports and TOEFL scores

TOEFL score is made up by four sections, namely reading, listening, speaking, and writing. Each of the section contributes 25 percent to the overall TOEFL score with the maximum score of 30 points for each.

Table 3 TOEFL score levelling

Section	Range of Scores	What It Means
Reading	0-30	0-14 (low)
		15-21 (intermediate)
		22-30 (high)
Listening	0-30	0-14 (low)
		15-21 (intermediate)
		22-30 (high)

Speaking	0-4 points, converted into a 0-30 scale	0-9 (weak)
		10-17 (limited)
		18-25 (fair)
		26-30 (good)
Writing	0-5 points, converted into a 0-30 scale	1-16 (limited)
		17-23 (fair)
		24-30 (good)

Student's TOEFL scores are given on Table 4 below.

Table 4 TOEFL Score

Student number	Reading	Listening	Speaking	Writing	Total
1	17	25	24	16	82
2	17	21	20	15	73
3	24	25	20	18	87
4	21	20	20	15	76
5	19	28	18	15	80
6	22	30	17	18	87
7	25	25	17	17	84
8	21	28	18	15	82
9	16	22	22	12	72
10	22	25	16	16	79
11	26	26	22	17	91
12	24	28	25	19	96
13	21	25	20	16	82
14	28	28	24	19	99
15	20	24	18	18	80
16	17	25	16	13	71
17	19	21	18	14	72
18	15	28	23	16	82
19	8	16	20	12	56
20	20	23	19	12	74
21	8	18	13	8	47
22	22	21	18	14	75
23	14	21	22	12	69
24	16	18	22	13	69
25	19	16	24	15	74

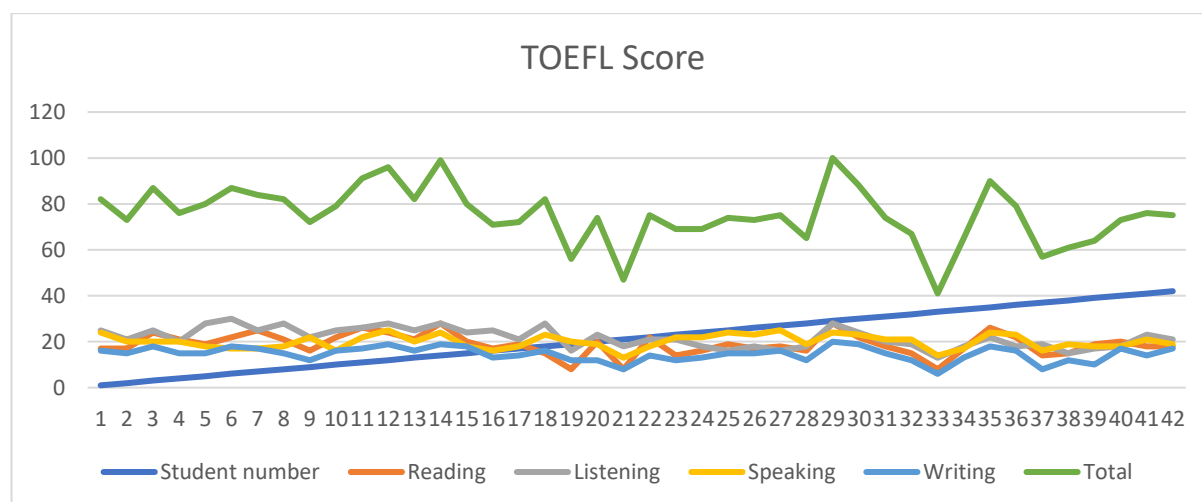| 26	17	18	23	15	73
27	18	16	25	16	75
28	16	18	19	12	65
29	28	28	24	20	100
30	22	24	23	19	88
31	18	20	21	15	74
32	15	19	21	12	67
33	8	13	14	6	41
34	17	18	17	13	65
35	26	22	24	18	90
36	22	18	23	16	79
37	14	19	16	8	57
38	15	15	19	12	61
39	19	17	18	10	64
40	20	18	18	17	73
41	18	23	21	14	76
42	18	21	19	17	75
Average	18,85714	21,7619	20,02381	14,64286	

Figure 1 TOEFL Score

The result shows that the students are intermediate users of English. As can be seen the average per skill set, the participants are high-intermediate for three skills namely reading, listening and speaking but low-intermediate in writing skill. This indicates that the students can communicate well in English. The detail scores of each section illustrate the balanced ability in the four

areas of English skills with listening and speaking scores are higher than reading and writing. This could be a result of their use of English for daily conversation. Writing and reading skills are known to be more academical, therefore, more difficult to be developed.

Teacher score

Teacher score are made up of four main categories as shown on Table 4 below. Student's participation is assessed through their contribution in class discussion, following instructions, and active participation in group or pair work activities. As can be seen from the table, student's participation scores are high ranging from 80 to 95. Secondly, the teacher also used homework to assess student's progress in learning. In this category, students also score high ranging from 80 to 100. Daily performance includes student's score from doing handouts or exercises from the textbook. Finally, test scores which include midterm test and final test.

Table 5 Teacher scores

Student number	Participation	Homework	Daily exercise	Test	Average
1	95	100	90	89	93,5
2	90	90	85	87	88
3	90	95	85	87	89,25
4	90	95	90	91	91,5
5	90	100	85	86	90,25
6	90	100	85	82	89,25
7	95	100	80	82	89,25
8	95	90	80	77	85,5
9	90	90	85	80	86,25
10	85	95	90	75	86,25
11	90	95	95	68	87
12	90	100	90	78	89,5
13	90	95	95	92	93
14	85	90	90	75	85
15	85	80	85	63	78,25
16	90	90	85	72	84,25

| 17	80	85	85	72	80,5
18	95	90	90	82	89,25
19	90	100	95	89	93,5
20	90	90	95	87	90,5
21	80	95	90	75	85
22	85	100	90	77	88
23	95	95	90	82	90,5
24	90	95	90	89	91
25	80	90	95	72	84,25
26	80	90	95	67	83
27	90	95	95	77	89,25
28	90	100	95	89	93,5
29	95	100	95	77	91,75
30	90	100	90	82	90,5
31	85	90	85	82	85,5
32	80	90	95	82	86,75
33	90	85	95	68	84,5
34	80	95	90	78	85,75
35	85	85	90	79	84,75
36	90	100	90	87	91,75
37	85	85	85	77	83
38	80	85	80	82	81,75
39	80	90	80	65	78,75
40	75	90	90	67	80,5
41	90	90	90	77	86,75
42	90	90	85	82	86,75
Average	87,61904762	92,97619048	89,04761905	79,214286	
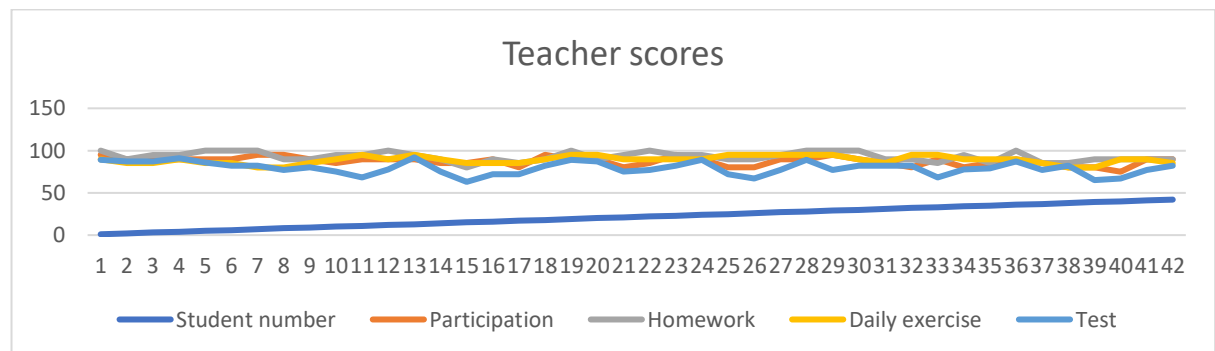
Figure 2 Teacher scores

From figure 2 it is clear that students perform worst in the test. Although it is worth mentioning that performing well on a test does not necessarily indicate good learning but should be treated as a partial story about the actual learning and teaching (Cheng & Curtis, 2004). The students seem to value their homework, and therefore score significantly higher than in the test. This is one indication of the teacher assessment as a formative rather than summative assessment.

Correlational test

SPSS version 2.2 was used to compute correlation ratio between student's TOEFL' scores and teacher's scores as can be seen in Figure 3.

Correlations

		Average	Total
Average	Pearson Correlation	1	.246
	Sig. (2-tailed)		.116
	N	42	42
Total	Pearson Correlation	.246	1
	Sig. (2-tailed)	.116	
	N	42	42

Figure 3 Pearson product moment correlation

It can be seen that the Pearson product-moment correlation coefficient was .246 which indicates positive relationship between students'

TOEFL score and their classroom scores. However, the correlation between the two variables are weak (r = .116). This result signals the need to further investigate how students and teacher perceive and use the test results to the teaching and learning process. Therefore, interviews were conducted to further investigate the roles of each tests in assessing student's progress in learning English.

2. Washback effect

The teacher was asked about how he used the information from class report and TOEFL scores in his teaching. He stated that he did not really care about the TOEFL scores, and that he only used the information from the class reports and his daily interactions with his students. He usually provides tutoring for those who are struggling in his class. In his assessment he always gives feedback to his students using casual language and pictures to help his students understand the message. In giving feedback he always tries to do it personally so that it will not embarrass the student. He also uses the class reports to adjust his year-long plan to ensure that his students really understand the materials. This adjustment is very important since the year-long is designed at the beginning of the school year so that there are assumptions made with regard to the student's performance throughout the upcoming course. Assessing the student's work will provide a snapshot of their performance which can be used to make necessary adjustments to the yearlong plan. The teacher also emphasized the importance of formative assessments to help monitoring student's progress. The formative assessments consist of four elements, namely class participation, daily performance, homework, and tests. He claimed that classroom assessment should not only be used merely as a progress-checking tool in which it sets the standard for the students to meet in order to pass the course, but more importantly, classroom assessments should seek to best support the students in their study.

Discussion

This study aims at investigating the correlation between teacher scores and TOEFL score and to explore how teachers utilize the standardized tests in his or her teaching practice. The first section of this chapter will discuss the relationship between TOEFL score and teacher's assessment. The second section will deal with the washback effect of the two kinds of assessments.

Referring to the Finding chapter, it is clear that there is very little or weak correlation between teacher score and student's TOEFL score. Such weak correlation means that there is no significant relationship between how the students perform in TOEFL test and in the class. Nonetheless, the comparison of TOEFL score to teacher's score is an interesting proposition and worth exploring. Although both types of scoring systems they are both different, yet they are related in many ways as it has been indicated in some studies (Cho & Bridgeman, 2012; Ginther & Yan, 2018; Rahmat et al., 2015)

The TOEFL or Test of English as a Foreign Language score is used to measure a person's mastery of the English language (Sawaki & Sinharay, 2018). It is an exam taken mainly by people whose native language is not English, and is generally required for entrance into universities in English-speaking countries, as well as for many internships and employment opportunities (Dang & Dang, 2021; Kokhan, 2013; Liao, 2022b, 2022c; Ma & Chong, 2022). The TOEFL score is based solely on the performance on timed tests

Teachers' scores, on the other hand, typically reflect a variety of methods. Many teachers use a combination of tests and quizzes, assignments, and class participation to evaluate a student's understanding of the English language. Some teachers may also use performance-based assessments, such as presentations or essays, to further gauge their student's abilities.

The TOEFL, or Test of English as a Foreign Language, is an international language examination that measures an individual's understanding, writing, speaking, and reading skills in English. The scores are widely used in university

settings to evaluate international students' language abilities and to assess their aptitude for enrolment (Kokhan, 2013).

The TOEFL score and teacher's score can be used to evaluate a student's level of language proficiency. The TOEFL provides a standardised language proficiency assessment that can be used to compare the English language ability of individuals from various countries who use English as a second language. One of the main TOEFL's primary purposes is to gauge if a student is ready for college-level work in English (Tweedie & Chu, 2019). It takes into account several language skills, such as reading, writing, speaking, and listening.

In comparison, a teacher's score assesses a student's language abilities within a classroom setting. Since it is based on a teacher's individual teaching style and methods, it may not necessarily reflect the student's overall language ability in other settings. For example, a student who produces a good essay may appear to understand the language, but may not necessarily be able to understand it in the context of a conversation or a listening activity. Thus, while a teacher's score can provide a good estimate of the student's language proficiency in the classroom, it should be viewed as limited in its accuracy.

Both the TOEFL score and a teacher's score are useful in assessing language proficiency in English. The TOEFL score is an internationally accepted evaluation for universities, and it assesses a student's language skills across multiple contexts. The teacher's score, on the other hand, is more tailored to a student's academic performance within a specific classroom setting. The two scores should be viewed as complementary, providing a view of language proficiency in different contexts.

The TOEFL score is a standardized measure of a student's language ability in a particular context, usually international. It is obtained by completing a computer-based multiple-choice or written essay test,

containing items designed to measure proficiency in reading, listening, speaking, and writing English. The TOEFL score is important in determining a student's English proficiency level and is often used as a benchmark in establishing eligibility for school admission or English-based jobs.

The teacher's score, on the other hand, is typically obtained by asking students to complete a series of academic tasks in class - such as writing essays, participating in class discussions or responding to prompts, taking part in debates, or completing practice exams. This score is based on a teacher's assessment of the student's language ability and proficiency level within particular academic contexts. It is important in giving feedback to the student on where their weaknesses and strengths lie, and how they can receive targeted feedback to improve their language proficiency.

The two scores are ideally compared and used in tandem to best assess a student's language proficiency in different contexts. The TOEFL score can provide a general metric of a student's ability to use English for academic and professional endeavors, while the teacher's score can provide insight on how a student expresses themselves in particular contexts related to the course in which they are enrolled. To obtain the most comprehensive view of a student's language capabilities, it is important to utilize and analyze both the TOEFL score and the teacher's score.

From the interview it is clear that the teacher uses TOEFL score to check his students' English level of proficiency. He said that he used TOEFL score to quietly assess his students' ability to follow the lessons. Having said this the teacher clarified that student's performance in TOEFL test does not always accurately represent their English ability in terms of their ability to follow the lessons. He said that probably because in general all of his students are competent English users because they have been at an EMI school for more than 6 years.

The use of standardized tests an initial English test is also mentioned in Ma and Chong (2022) and Liao(2022). In Ma and Chong (2022) IELTS was used to determine whether students are eligible to continue studying at the EMI programme or not. In Liao(2022), SAT was employed to place students into EFL class which is not the case of the current study. Here, TOEFL is used mainly to inform the teacher and students regarding student's level of proficiency. By knowing their scores, students are expected to be aware if they need to improve their English to certain level or they need to maintain their English level. This is important especially those who wish to study abroad because having a TOEFL or IELTS certificate is a mandatory in many universities abroad (Dang & Dang, 2021; Pearson, 2021).

The Washback effect of TOEFL test on the teaching and learning process is insignificant and indirect. The teacher did refer to the student's TOEFL score but only right after the test result was out to motivate the students to be more serious about improving their English through active participation in class. The teacher said *"Now that you know your TOEFL score I hope you will be more serious in your study. See if you want to study abroad you must have TOEFL score minimum 550."* Having said this the teaching did not change his yearlong plan that he made at the beginning of the academic year. It could be said that student's TOEFL score does not have any effect on the curriculum or topics that the teacher wants to cover. To this the teacher said that it was not his job to improve the student's TOEFL score. "It is completely their responsibility". Some students said that they were not too worry about their TOEFL score because they still have one whole year to improve it. One student said *"later I will take English course at XXX before I take TOEFL test after I finish school"*. The students did not have the expectation that their school teacher would help them with their TOEFL test. They seem to compartmentalized between school English and TOEFL.

Conclusion

In conclusion, the findings of this study shed light on several important aspects of the relationship between teacher scores and student TOEFL scores. The analysis reveals a relatively weak correlation between these two scores, implying that they gauge student performance from distinct vantage points. This divergence likely arises from the disparate emphases of the respective assessments: the teacher's score hinges on a performance-oriented evaluation, while the TOEFL score primarily scrutinizes the accuracy of the student's English language proficiency.

Furthermore, the role of the teacher in utilizing these scores diverges significantly. The teacher draws upon classroom performance information to offer feedback to students and to tailor lesson plans, harnessing the insights for instructional improvement. On the other hand, the student's TOEFL score serves as a tool to raise awareness of potential achievements or hurdles in pursuit of higher education abroad. It functions as an indicator of the student's readiness for further studies, if they opt to venture beyond their home country's borders.

A notable practice observed within the teacher's role is the issuance of personalized assessment reports to students. This approach facilitates a thorough monitoring of individual progress and challenges in the learning journey. By providing tailored feedback, the teacher creates an environment that promotes continuous improvement and targeted support.

Interestingly, it emerges that students and parents do not anticipate the teacher's involvement in preparing for the TOEFL test. Rather, the TOEFL examination is perceived as a distinct realm of learning, separated from the regular classroom instruction. This demarcation reflects a prevailing understanding that the TOEFL assessment serves as a standalone measure of English language proficiency for international educational pursuits.

In summation, the study underscores the nuanced nature of the relationship between teacher and student TOEFL scores, revealing how these scores function within distinct realms and fulfil diverse purposes. The teacher's emphasis on individualized feedback and the TOEFL's role in signalling educational potential contribute to a comprehensive view of the education landscape, where different assessments play unique and valuable roles.

Limitations

This study was conducted in an English immersion school where the medium of instruction is English and the student's level of English is most likely higher than those who study in non-English immersion school. Moreover, the teacher is a native English speaker which probably has different perspective of what a good English teaching is compare to the majority of Indonesian teachers. Therefore, the results of this study should not be generalized to other contexts.

Recommendations

This study correlates the TOEFL scores and classroom assessment. The weak correlation between the two could be the result of different categories of assessment in the scores compared. Future study should compare similar assessment criteria such as listening, reading, speaking and writing scores. Moreover, this study uses TOEFL scores from a single testing period which may not reflect an accurate measure of student's English level. A number of TOEFL testing might provide more accurate estimation of the student's English level thus increase the quality of the data.

References

Akhter, N., & Malik, S. I. N. (2019). Teachers' perceptions regarding multiple assessment paradigms in English teaching and its' influence on learning

achievement of secondary class students. *Journal of Research in Social Sciences*, *7*(2), 43–57.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, *14*(2), 115–129. https://doi.org/10.1093/applin/14.2.115

Amelia, R., & Harmaini, F. (2020). Development of web-based TOEFL learning media (E-TOEFL) using Moodle at Universitas Bung Hatta. *IOP Conference Series: Materials Science and Engineering*, *990*(1), 012002. https://doi.org/10.1088/1757-899X/990/1/012002

Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing*, *29*(3), 421–442. https://doi.org/10.1177/0265532211430368

Creswell, J. W., & Creswell, D. J. (2018). *Research design: qualitative, quantitative, and mixed methods approaches* (5th Ed.). Thousand Oaks.

Dang, C. N., & Dang, T. N. Y. (2021). The predictive validity of the IELTS test and contribution of IELTS preparation courses to international students' subsequent academic study: insights from Vietnamese international students in the UK. *RELC Journal*, 003368822098553. https://doi.org/10.1177/0033688220985533

Feron, E., Schils, T., & Weel, B. ter. (2016). Does the teacher beat the test? The value of the teacher's assessment in predicting student ability. *De Economist*, *164*(4), 391–418. https://doi.org/10.1007/s10645-016-9278-z

Fulcher, G., & Davidson, F. (2009). Test architecture, test retrofit. *Language Testing*, *26*(1), 123–144. https://doi.org/10.1177/0265532208097339

Ginther, A., & Yan, X. (2018). Interpreting the relationships between TOEFL iBT scores and GPA: Language proficiency, policy, and profiles. *Language Testing*, *35*(2), 271–295. https://doi.org/10.1177/0265532217704010

Gokturk Saglam, A. L. (2018). Can exams change how and what teachers teach? Investigating the washback effect of a university English language proficiency test in the Turkish context. *Eurasian Journal of Applied Linguistics*, *4*(2), 155–176. https://doi.org/10.32601/ejal.464094

Heaton, J. B. (1990). *Writing English language test*. Longman Inc.

Kokhan, K. (2013). An argument against using standardized test scores for placement of international undergraduate students in English as a Second Language (ESL) courses. *Language Testing*, *30*(4), 467–489. https://doi.org/10.1177/0265532213475782

Liao, Y. F. (2022). Using the English GSAT for placement into EFL classes: accuracy and validity concerns. *Language Testing in Asia*, *12*(1), 31. https://doi.org/10.1186/s40468-022-00181-6

Ma, H., & Chong, S. W. (2022). Predictability of IELTS in a high-stakes context: a mixed methods study of Chinese students' perspectives on test preparation. *Language Testing in Asia*, *12*(1), 2. https://doi.org/10.1186/s40468-021-00152-3

Pearson, W. S. (2021). The predictive validity of the Academic IELTS test. *ITL - International Journal of Applied Linguistics*, *172*(1), 85–120. https://doi.org/10.1075/itl.19021.pea

Rahman, Kh. A., Seraj, P. M. I., Hasan, Md. K., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, *11*(1), 12. https://doi.org/10.1186/s40468-021-00129-2

Rahmat, N., Min, L. S., Sungif, N. A., & Yusup, F. N. M. (2015). English language proficiency tests and academic achievement: a study on the Malaysian university English test as a predictor of technical programme undergraduates academic achievement. *Advances in Language and Literary Studies*, *6*(1), 114–119. https://doi.org/10.7575/aiac.alls.v.6n.1p.114

Sawaki, Y., & Sinharay, S. (2018). Do the TOEFL iBT ® section scores provide value-added information to stakeholders? *Language Testing*, *35*(4), 529–556. https://doi.org/10.1177/0265532217716731

Tweedie, M. G., & Chu, M.-W. (2019). Challenging equivalency in measures of English language proficiency for university admission: data from an undergraduate engineering programme. *Studies in Higher Education*, *44*(4), 683–695. https://doi.org/10.1080/03075079.2017.1395008

Weaver, C. (2016). The TOEIC IP test as a placement test: Its potential formative value. *JALT Journal*, *38*(1), 5–26.