

# Testing ESL in Japan : Comparing TOEFL ITP and Cambridge B2 First

江 口 均

# Testing ESL in Japan : Comparing TOEFL ITP and Cambridge B2 First

Hitoshi EGUCHI

## Contents

Abstract  
Introduction  
Background  
Method  
CEFR  
Cambridge English  
Examination  
Results of the Questionnaire  
Results of TOEFL ITP  
Results of Cambridge English  
Examination B2 First  
Conclusions  
Acknowledgements  
References  
Appendices

## [Abstract]

This study analyzed student scores of two English language proficiency tests by measuring correlation, linear regression analysis, and other statistical analyses (such as cluster analysis). Upon comparing the scores of students on the Cambridge B2 First and the Test of English as a Foreign Language Institutional Testing Program (TOEFL ITP) tests, the paper aimed to examine the characteristics of the students and the differences between the test scores. In addition to these analyses, a questionnaire was designed to investigate students' perceived difficulties regarding these two tests. The scores of the tests showed very high correlations ( $r = 0.76$ ) among students whose score was higher than 153 on Cambridge B2 First ( $r = 0.75$ ) but low correlations for those who scored under 153 ( $r = 0.24$ ). These findings seem to suggest that introducing criteria that differ from the Common European Framework of Reference for Languages (CEFR) levels would be needed for higher education admissions so as to identify groups whose language proficiency is lower than B1. Moreover, students' responses in the questionnaire showed that the Cambridge B2 First test, which focuses on everyday language knowledge, measured what they were actually learning in their English courses.

## Introduction

This study will analyze the scores of two English proficiency tests, Cambridge First Certificate in English (FCE), which is now called "B2 First," and TOEFL ITP (TOEFL), by measuring correlation, linear regression analysis, and other statistical analyses (such as cluster analysis), to investigate the characteristics of the students through the two tests and the differences between the two tests' scores.

Testing is a hot topic in English education in Japan, especially with regard to the entrance examinations of universities, since there is a debate about the possibility of replacing the Center Test, the national standardized university exam, with other commercially available English proficiency test engines, the so called "private English tests" (Lowe, 2018) such as TOEFL, TOEIC, and STEP. Testing should be concerned with teaching objectives and

---

Key words : Admission, Cambridge B2 First, CEFR, Language Assessment, Proficiency Tests, TOEFL ITP

students' achievements: the discussion is therefore about whether or not the presently used tests are measuring what students should have learned at each proficiency level. For this reason, all the schools are reexamining teaching objectives and methods.

All the test makers presented a score conversion table based on the Common European Framework of Reference (CEFR), and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) is encouraging Japanese universities and colleges to include such scores as an admission requirement. "The Cambridge English Exam" is one such test, but Japanese educators are not so familiar with it, compared to other tests. This study will examine what the score of this test shows about the proficiency of the students and analyze the score conversion by comparing TOEFL scores and Cambridge English Examination scores. This will be done by looking at the results of a sample of university students from Hokusei Gakuen University (HGU).

The English Department of HGU offers two opportunities to take TOEFL ITP in the first year and one in the second year and third year to see students' progress in learning English. Therefore, in this study, TOEFL ITP scores will be the basis of analysis, and they will be compared to Cambridge FCE scores.

From previous research comparing TOEFL ITP and TOEIC L&R (Eguchi, 2011), it seems to be very difficult to make any score conversion or a score conversion formula without having a large sample of participants in the study. In the present study, 50 students volunteered to take Cambridge FCE in 2017 and 2018, and investigating their scores only is not enough to generalize anything about the tests. The significance of a study like this might lie in the fact that the researchers know the test takers better, since they are their students, and it is possible for them to be more informed about their learning environment and individual characteristics.

This study does not attempt to generalize its findings by doing a simple analysis of the scores of the two tests. Its objective would be to provide some information about the proficiency of the students at each English language teaching institute and provide these institutes with more insight into how they could analyze the test scores and characteristics of their learner's English proficiency.

## Background

English teachers in Japan, including college teachers, are experiencing a drastic change due to the debate going on about the English test in university entrance exams, nationwide. The plans are aimed at including tests on productive skills such as speaking and writing. The ultimate purpose of this change is to revolutionize the whole system of English education in Japan, whose reputation is low with regard to the development of communicative competence. "The Ministry of Education, Culture, Sports, Science and Technology (MEXT) has embarked on an ambitious program of reform to improve the country's practical English language skills" (Green, 2016).

MEXT has tried to overturn Japan's negative reputation by including different rules in their "Teaching Guide for the Japanese Course of Study," changing the university entrance test content, and issuing action plans such as "An Action Plan to Cultivate Japanese with English Abilities." This series of drastic reforms started in the late 1990s.

English has been taught as a subject in elementary schools since 2018. This is why teachers at all education levels from elementary schools to universities are discussing how English should be taught at school, and how to test it.

## Method

The participants of the study were 50 students of HGU. They took the B2 First test and the TOEFL ITP, paper-based TOEFL. The correlation coefficient between the scores in the two tests were then analyzed. In addition, the correlation coefficient of the two tests and between each section, as well as regression analysis, were conducted.

The data collection included an online questionnaire concerning the students' perceived difficulties with regard to each section of the B2 First Test. It asked them to make a comment about how well or poorly they performed and to explicitly compare their impressions to the ones they had while taking different tests like TOEFL and TOEIC. It was administered through an LMS, and 35 volunteers out of 50 students participating in the study responded to it.

## CEFR

CEFR aims to provide "a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe. The Framework also defines levels of proficiency which allow learners' progress to be measured at each stage of learning and on a life-long basis" (Council of Europe, 2001), and "providers of language proficiency tests refer to the six levels (A1-C2) of the CEFR proficiency scales because of their international recognition" (Papageorgiou et al., 2015).

This Framework has had a huge impact on the English language teaching policy in Japan as well. The levels of English proficiency to achieve for native Japanese speakers learning English have been set on the basis of this framework by MEXT (see [http://www.mext.go.jp/b\\_menu/houdou/30/03/\\_icsFiles/afieldfile/2019/01/15/1402610\\_1.pdf](http://www.mext.go.jp/b_menu/houdou/30/03/_icsFiles/afieldfile/2019/01/15/1402610_1.pdf)), and the selected English exam providers also presented the score conversion formula for their scores to be used for university admission qualifications (see Appendix 1).

Native Japanese speakers' English proficiency used to be discussed by referring to TOEFL or TOEIC scores; however, referring to the CEFR might give a better indication of the actual proficiency level. It should be welcome to have a different measurement for English proficiency level among Japanese learners whose content and difficulty level reflect what the Japanese learners are actually taught and what they can do. However, the CEFR was made in Europe, where the geographical, historical, social, and educational background is different from

that of Japan, and it is not realistic to apply it to the context of Japanese students learning English as a foreign language.

The expected proficiency level for high school students is CEFR A2 level, which, according to STEP, corresponds to Eiken top half of Semi Grade 2 and the bottom half of Grade 2 (which has a score range of 1700–1949). It seems that the CEFR ranges are too wide, especially for A1 and A2 levels, to evaluate the students clearly, while shorter ranges should be presented to use it for admission purposes. Different scales would probably be needed in order to distinguish the proficiency and achievements of Japanese students, reflecting what they are taught at school, instead of applying the CEFR levels.

## Cambridge English Examination

Cambridge English Examination is one of the proficiency tests that have been chosen to prove English qualification for Japan's university entrance examinations by MEXT. There are five levels of the test; (1) A2 Key, (2) B1 Preliminary, (3) B2 First, (4) Advanced, and (5) C2. The first two letters represent the CEFR levels. Cambridge English Examination is one of the most popular English tests globally (Roca-Varela & Palacios, 2013) and it has a hundred years of history of English proficiency assessment (Cambridge Assessment, retrieved from <http://www.cambridgeassessment.org.uk/insights/100-years-cambridge-english/>).

Exam Format	Content	Purpose
Reading and Use of English (1 hour 15 minutes)	7 parts/52 questions	Shows you can deal confidently with different types of text, such as fiction, newspapers, and magazines. Tests your use of English with tasks that show how well you can control your grammar and vocabulary.
Writing (1 hour 20 minutes)	2 parts	Requires you to be able to produce two different pieces of writing, such as letters, reports, reviews, and essays.
Listening (about 40 minutes)	4 parts/30 questions	Requires you to be able to follow and understand a range of spoken materials, such as news programs, presentations, and everyday conversations.
Speaking (14 minutes per pair of candidates)	4 parts	Tests your ability to communicate effectively in face to face situations. You will take the Speaking test with one or two other candidates.

(Table 1:Cambridge Assessment English, 2015)

## Results of the Questionnaire

Some of the comments provided by the participants in this study through the questionnaire about the characteristics of the tests were really interesting. First of all, many

students mentioned that the Cambridge B2 First test seems to be measuring more practical language knowledge and skills by using words and expressions they hear more often in their everyday life than those used in the TOEFL exam. On the contrary, TOEFL is designed to measure the students' preparedness for studying in American universities; therefore, the language is similar to what students studying in American universities often hear and use in academic settings (GlobalExam, 2016).

As for the writing section of the test, some mentioned that they performed well by using what they learned in the writing classes they took. Looking at those comments, it might be safe to conclude that Cambridge English Examination asks more general questions than TOEIC or TOEFL, which tend to be about more specific situations.

With regard to the listening section, there are two distinctive features in Cambridge English Exam that the participants mentioned. The first is that the recording is played twice, while it is only played once in TOEFL. Therefore, some students perceived Cambridge English Examination to be easier. Knowing the topic of the conversation and listening to the same recording again after getting a grasp of the content from the first listening might make comprehension become easier, as it makes it possible to review the learners' guessing.

The second feature deals with the variety of English. Out of 35 respondents, 22 said the listening part was difficult because British English was used in the test, while the English variety they usually hear while studying for TOEFL is American English.



(Graph 1: Perceived Difficulty Because of British English)

It is not true that all they heard was British English. Judging from the sample test provided by Cambridge Assessment (see Cambridge Assessment English, 2015) it seems that different varieties are used. However, students might have perceived that the different varieties which did not sound familiar to them were all British English. In fact, all the instructions in the test are announced in British English. This might have given them the impression that they had heard British English only. Moreover, different varieties being used one after another can be confusing for learners.

As for the speaking section, there is a feature that other tests do not have. Cambridge English Examination requires face-to-face communication, asking the candidates to interact with another test taker and an examiner. Such speaking tests require "synchronous

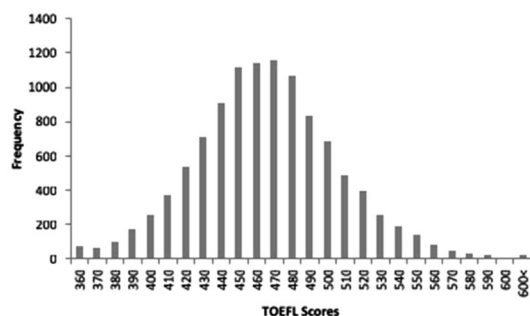
communication” (Ockey et al., 2015), which is closer to what communication is like outside the classroom, in real life situations. One of the test takers has to start talking, and the interlocutors (as mentioned, another test-taker and an examiner) have to give a response based on what the first test taker said.

On the contrary, when the expected response is based solely on an initial verbal and/or written stimulus, and further responses from the interlocutors are not expected, this kind of communication requires an “asynchronous response” (Ockey et al., 2015). TOEFL iBT, TOEIC Speaking, and GTEC use this type of communication exercise in the speaking section, where the test takers are speaking to a computer and further responses from the interlocutors are not expected (Ockey et al., 2015). Because of this, it seems that the students got the impression that they were able to perform in the speaking test according to what they had learned in the process of learning English.

Overall, almost all the respondents to the questionnaire (33 out of 35 respondents) said that it was a good opportunity to take the test and that it seemed to be measuring what they were actually learning. In all the sections, students did not have to rush in order to answer questions in time, so it most probably did not affect the performance much, as it does with TOEFL ITP (in which many students cannot manage to answer all the questions). To sum up, the core feature of the Cambridge English Examination test might be the fact that it measures general English rather than a specific area of English.

## Results of TOEFL ITP

The HGU English Department has been using TOEFL to see the students’ proficiency and their progress since 1996. For this study, all the data stored on Floppy Disks, CDs, and paper were collected. There are some years whose data are missing. Between 2000 and 2009, all the freshmen students took the TOEFL twice, in April and January, and once in the second, third, and fourth year. Since 2010, TOEIC L&R has been conducted for 100 volunteer students every year, and it has become optional for the fourth-year students to take TOEFL.



(Graph 2: Distribution of All the Scores of TOEFL ITP at HGU)

	<i>S1</i>	<i>S2</i>	<i>S3</i>	<i>Total</i>
Mean	46.7685049	46.1659949	45.8541189	462.635939
Standard Error	0.04509474	0.04787674	0.05134656	0.39174721
Median	47	46	46	460
Mode	47	46	46	463
Standard Deviation	4.71148354	5.00214537	5.36393406	40.9239923
Sample Variance	22.1980771	25.0214583	28.7717886	1674.77315
Kurtosis	1.04925857	0.72033516	0.5072305	0.60692235
Skewness	0.07543256	-0.0007059	-0.42807	0.21706471
Range	36	42	41	333
Minimum	29	26	26	310
Maximum	65	68	67	643
Sum	510525	503948	500406	5048746
Count	10916	10916	10913	10913

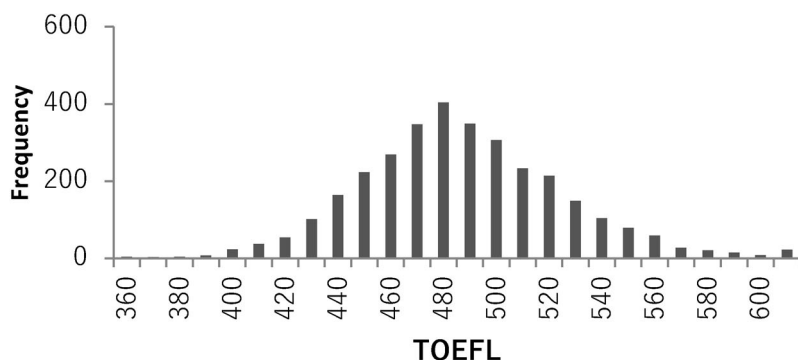
(Table 2: Descriptive Statistics of All the Scores of TOEFL at HGU between 1996 and 2018)

Graph 2 and Table 2 show the data of all the scores of TOEFL ITP, which include multiple scores of individual students. Therefore, it might represent the average English proficiency of the students cross-sectionally, as measured by TOEFL ITP. The data concerning the lowest scores and highest scores of individual students, instead, represent how much they learn through English education at HGU.

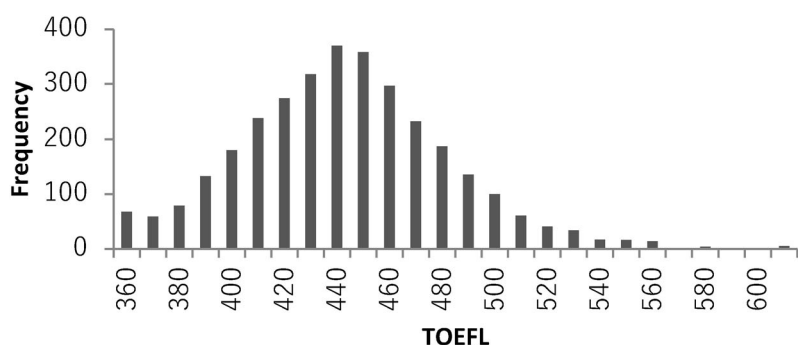
Lowest Scores of Students					Highest Scores of Students				
	S1	S2	S3	TOTAL		S1	S2	S3	TOTAL
Mean	44.8	43.9	43.2	439.6	Mean	48.6	48.3	48.2	483.7
Standard Error	0.1	0.1	0.1	0.7	Standard Error	0.1	0.1	0.1	0.7
Median	45.0	44.0	44.0	440.0	Median	48.0	48.0	48.0	480.0
Mode	44.0	44.0	44.0	440.0	Mode	47.0	47.0	48.0	473.0
Standard Deviation	4.7	5.1	5.7	39.7	Standard Deviation	4.5	4.8	4.7	38.6
Sample Variance	22.4	26.3	32.5	1575.6	Sample Variance	20.3	23.0	22.5	1491.1
Kurtosis	1.0	0.7	-0.1	0.8	Kurtosis	0.7	0.7	1.4	0.8
Skewness	-0.1	0.0	-0.3	0.3	Skewness	0.4	0.2	-0.4	0.4
Range	35.0	37.0	37.0	323.0	Range	33.0	42.0	41.0	300.0
Minimum	30.0	26.0	26.0	310.0	Minimum	32.0	26.0	26.0	343.0
Maximum	65.0	63.0	63.0	633.0	Maximum	65.0	68.0	67.0	643.0
Sum	144463.0	141628.0	139394.0	1418264.0	Sum	156907.0	155672.0	155509.0	1560301.0
Count	3226.0	3226.0	3226.0	3226.0	Count	3226.0	3226.0	3226.0	3226.0

(Table 3: Comparison of the Lowest Scores and Highest Scores of the Students)



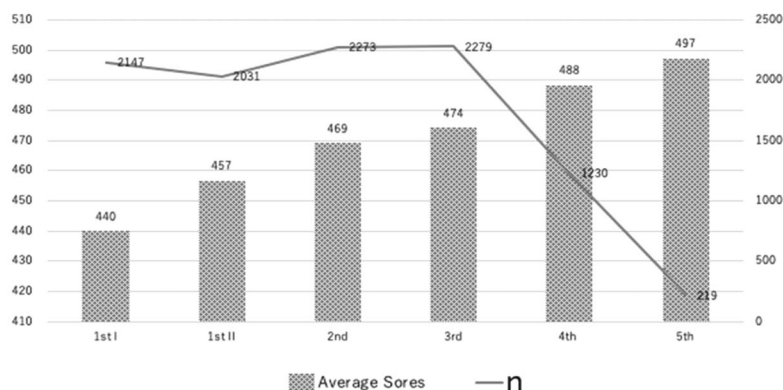


(Graph 3: Distribution of the Lowest Scores of TOEFL ITP at HGU)



(Graph 4: Distribution of the Highest Scores of TOEFL ITP at HGU)

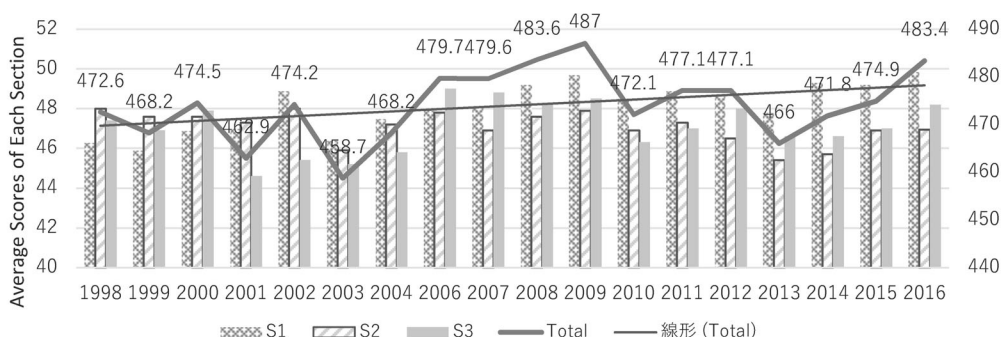
Graph 3 and Graph 4 show the histograms of the lowest scores and the highest scores of the students. The lowest scores in many cases are of the first TOEFL ITP they took at the very beginning of the first year after admission, though there are some exceptions. Because it is the first attempt for almost all the freshmen at taking the TOEFL, the biggest hikes (about 20 points) in scores were seen in the one taken at the end of the academic year in January (see Graph 4). It shows a steady progress after their admission.



(Graph 5: Average Scores of TOEFL ITP at Each Testing)

There is a 30 point improvement on average between the first TOEFL and the one in the third year. HGU is not offering TOEFL ITP to everyone during their fourth year or later, and only the students who volunteer actually take the test. Almost all the students who took it in their 5th year were those who experienced a school exchange program or students who spent a year in an English-speaking country. Such students' TOEFL scores are higher, though the number of such students is small. Judging from the score comparison among the third-year students only, there is a slight increase over 17 years.

The core feature of the English Department education program is the first two-year intensive English education taught by native speaking English teachers only. Two oral communication classes, one writing class, and one reading class are offered during the first two years as mandatory courses. From the third year, various content classes are offered in English, one discussion class is mandatory, and more advanced oral communication classes such as public speaking classes are offered as electives. There is always a chance to use English for practical reasons, and communication in English is very natural and common at school. Therefore, there is an improvement in speaking that cannot be measured by TOEFL ITP, since it is a paper-based proficiency test.



(Graph 6: Average Scores of Third Year Students)

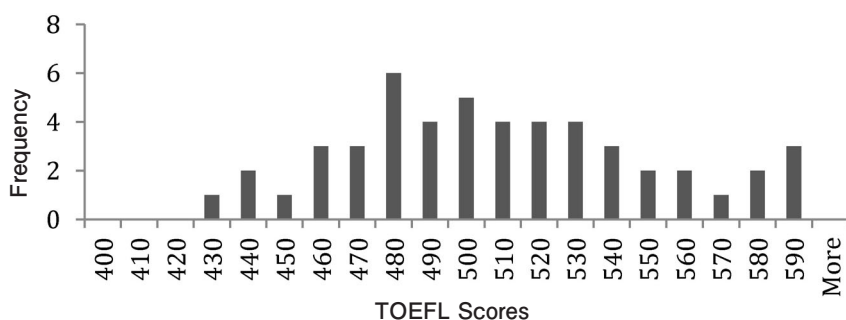
Graph 6 shows the scores of the third-year students from different admission years. It is often said that the academic level of the students is declining along with the decline of the youth population in Japan. As seen in this graph, though, there is not an evident decline in the TOEFL scores even though there are arbitrary ups and downs every year in total scores. Scores of Listening Comprehension (S1) show that students are the strongest at listening nowadays, in contrast with the ones admitted in the early 2000s. From 2007 up to 2016, listening scores in the third year are the highest.

There could be different interpretations. One is that students are more competent in listening than in the other sections (grammar and reading), and, consequently, there is no evident decline in the TOEFL score. The other is that the TOEFL ITP score does not really represent what they have learned in high school because the content of the test items is far more difficult than the contents learned in high school. Therefore, through college education, students can perform as well as the students in the past.

It is necessary to use a proficiency test which is able to measure the four skills to see what students can actually do and what changes are taking place in them. The present students of HGU seem to be far better than the students of 30 years ago, in terms of listening comprehension and speaking. It is important to measure English proficiency levels in a different way, and not to claim that their levels are declining by measuring it with traditional measurement tools.

## Results of Cambridge English Examination B2 First

Descriptive Statistics of Cambridge English (CE) Examination B2 First



(Graph 7: TOEFL Scores of CE Participants)

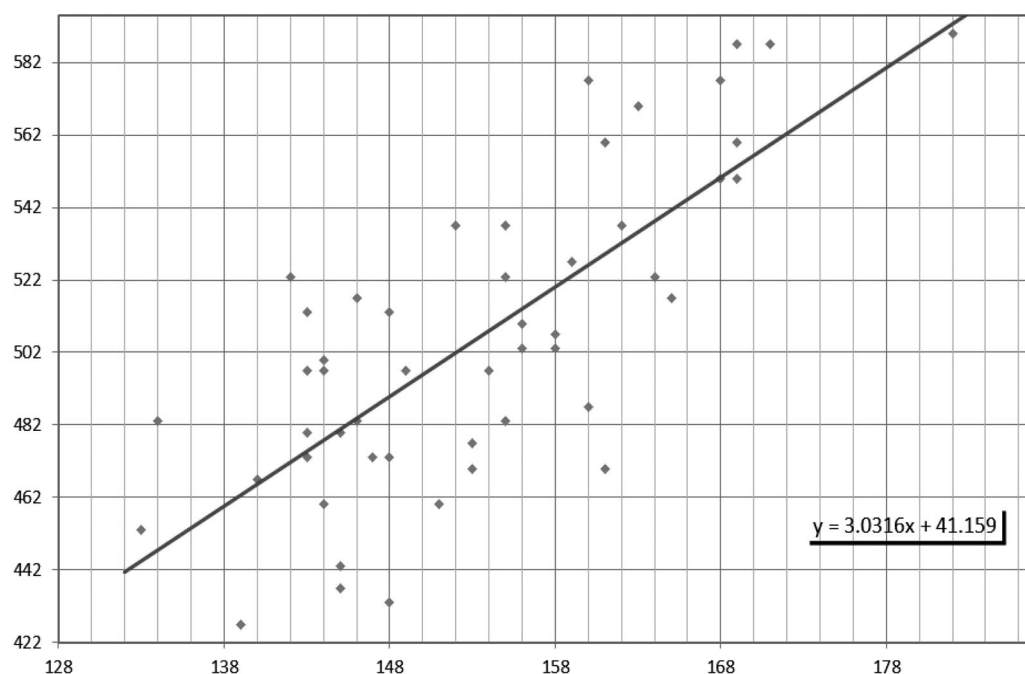
	<i>Overall Score</i>	<i>Reading</i>	<i>Use of English</i>	<i>Writing</i>	<i>Listening</i>	<i>Speaking</i>
Mean	153.3	148.6	152.6	157.0	145.5	163.2
Standard Error	1.5	1.7	1.9	1.6	2.0	1.6
Median	153.0	146.0	152.0	154.0	141.0	165.0
Mode	143.0	143.0	152.0	150.0	138.0	173.0
Standard Deviation	10.6	12.1	13.5	11.1	14.2	11.6
Sample Variance	111.8	146.6	183.1	124.2	201.0	134.7
Kurtosis	− 0.2	− 0.1	− 0.5	− 0.3	0.4	− 0.4
Skewness	0.4	0.4	0.2	0.5	0.7	− 0.3
Range	49.0	55.0	55.0	48.0	68.0	52.0
Minimum	133.0	123.0	127.0	136.0	122.0	138.0
Maximum	182.0	178.0	182.0	184.0	190.0	190.0
Sum	7666.0	7431.0	7632.0	7848.0	7276.0	8161.0
Count	50.0	50.0	50.0	50.0	50.0	50.0

(Table 4: Descriptive Statistics of B2 First)

As mentioned earlier, 50 students volunteered to take Cambridge English (CE) Examination for this study, and Graph 7 shows that among them there are a lot of students who scored higher than 500 on TOEFL ITP. It may indicate that students who are confident in their English skills might be more willing to be measured by different tests. So, the

participants in this study are very eager in learning English and their scores are also higher than that of the regular students.

As Table 4 indicates, the highest score was 182, which is in the range of C1 in the CEFR scale; 14 students were in the range of B2, 32 were in the range of B1, and 3 were not reported in CEFR scales, because the score was not good enough for B1 level.



(Graph 8: Correlations Between TOEFL ITP and CE B2 First)

	Overall Score	Reading	Use of English	Writing	Listening	Speaking	TOEFL S1 L	TOEFL S2 G	TOEFL S3 R	TOEFL
Overall Score	1.00									
Reading	0.77	1.00								
Use of English	0.86	0.63	1.00							
Writing	0.77	0.42	0.60	1.00						
Listening	0.87	0.63	0.70	0.63	1.00					
Speaking	0.80	0.47	0.62	0.61	0.65	1.00				
TOEFL S1 L	0.64	0.46	0.55	0.53	0.58	0.58	1.00			
TOEFL S2 G	0.62	0.48	0.65	0.52	0.48	0.40	0.48	1.00		
TOEFL S3 R	0.59	0.44	0.65	0.36	0.48	0.58	0.59	0.44	1.00	
TOEFL	0.76	0.56	0.76	0.57	0.62	0.62	0.81	0.81	0.83	1.00

(Table 5: Correlations Between Each Section of TOEFL ITP and CE B2 First)

Graph 8 indicates that the scores of TOEFL and CE B2 First scatter between the two, and the red line indicates the linear regression of the two tests. The correlation between the two tests for the 50 students participating in this study was very high ( $r=0.76$ ), and it

indicates that both tests are likely to measure the same things.

The correlation for the reading section is somewhat noticeable: it is measuring the same language modality, yet the correlation is low ( $r=0.44$ ). It is difficult to understand what causes the low correlation. Reading seems to be difficult for students, and this is the area students need to focus on in order to improve their English. Looking at the correlation of the scores of each section of both tests with the average scores of all sections, reading scores are lower in both tests.

CE	Reading	Use of English	Writing	Listening	Speaking
Higher	13	24	38	5	47
Lower	36	26	12	44	3
TOEFL	S1	S2	S3		
Higher	35	23	19		
Lower	15	25	29		

(Table 6: The Number of the Students Who Scored Higher or Lower Than the Average of All)

Listening seems to be the strongest of all sections (35:15) on TOEFL ITP. On the other hand, listening is the weakest (5:44) on CE. Considering that the productive skills (speaking and writing) are stronger than the receptive skills (reading and listening), it could be said that the students are better at productive skills.

In speaking 47 students got a higher score on CE, 22 got 10 points more than the average scores of all the sections, 13 got even 15 points more, and 3 got even 20 points more. Therefore, it is evident that speaking cannot be measured by measuring other skills, and it is important to test students directly with speaking tests, if speaking is considered important in learning English.

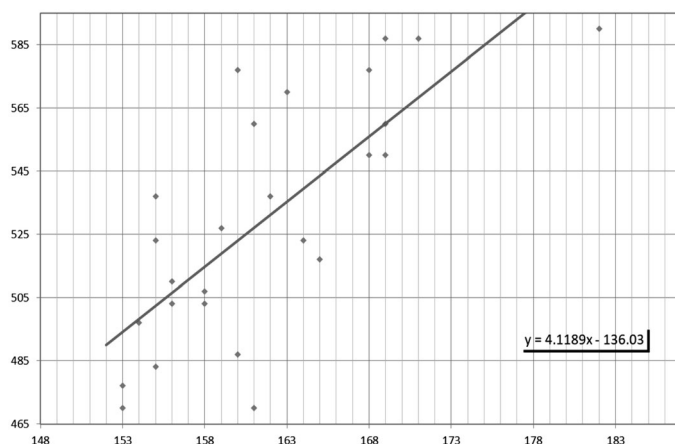
Furthermore, 38 students got higher scores than the average scores of the other sections in writing on CE, too. So, the productive skills, speaking and writing, are the ones that HGU English Department students are better at.

As seen in the scatter chart of the two scores, there are more scores spread in the score range below CE 150, and TOEFL ITP 485. Both tests are measuring English proficiency, even though CE is measuring different modalities (all four skills). When the group is divided into two, High Group ( $\geq 153$ ,  $n=26$ ) and Low Group ( $< 153$ ,  $n=24$ ) of CE, there is a clear difference in correlations.

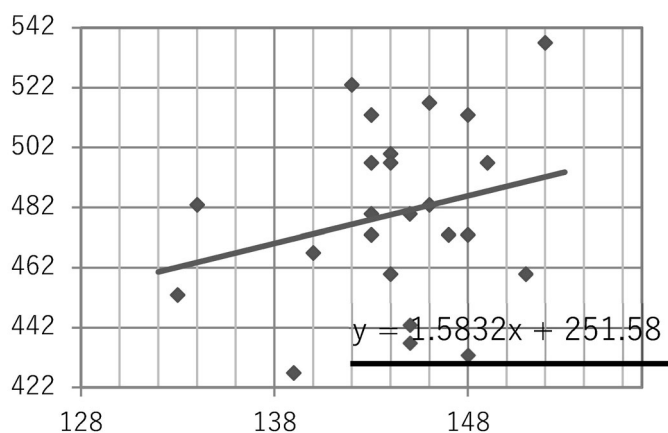
	n	r
ALL	50	0.76
153= $<$	26	0.75
153 $>$	24	0.24

(Table 7: Correlations Between CE and TOEFL)

It is possible that one of the two tests does not measure well the proficiency of the students with low scores, though it is difficult to decide which test. The total scores of both tests are the average scores of all the sections (the average scores are multiplied by 10 in TOEFL ITP). Therefore, it is possible to see how high and low points of each section are contributing to the total points. In TOEFL ITP, 32 out of 50 got higher scores in S1 (Listening Comprehension) while in CE, only 4 got higher scores in listening. It cannot be concluded which test is more difficult than the other, 38 and listening is not in CE.



(Graph 9: Correlations Between CE and TOEFL)



(Graph 10: Scatter Chart of <153 of CE and TOEFL ITP)

## Conclusions

In this study, the test scores of Cambridge English (CE) Examination (B2 First) were examined by comparing the score of TOEFL ITP, which the Hokusei Gakuen University English Department has been offering regularly to measure its students' proficiency and their achievements as a result of study while they are enrolled at HGU.

There is a strong correlation between the two tests ( $r=0.76$ ) in terms of total scores of both tests. However, the correlation with CE's Reading ( $r=0.44$ ) is not so high considering that they are supposed to measure the same skills. It is necessary to use a test to measure the students with low proficiency better.

The post-test questionnaire results indicate that students have a strong impression that CE measures what they are learning in class and it motivates them to study more and take the exam again. This is because CE measures all the four skills directly, and it gives the students an impression that their English is measured more accurately.

Unfortunately, TOEFL ITP is not as useful, especially for those whose score is not high. TOEFL measures the preparedness for studying in American universities and the items used in the test are not necessarily what Japanese students need to learn to survive in today's rapidly globalized society. TOEFL is now offering TOEFL iBT and TOEFL junior, which also measure speaking, but there is a problem in accessibility and affordability of the tests for a school that offers to a large number of students a way to measure their English. There is the same problem with Cambridge English Examination.

English proficiency among Japanese students is said to be very low, but it may not be so, if measured properly by employing different tests. Japanese students nowadays have more exposure to media in English, such as movies, TV programs, radio programs, and even classroom materials. English proficiency is changing along with the advancement of technologies and society's evolving needs. So, it is important to measure students' proficiency by measuring what they are actually learning instead of adopting traditional methods.

Further research in this area is still necessary to examine the testing methods offered by private firms in Japan, and this small research project appears to provide meaningful insights in this framework.

## Acknowledgements

This research was supported by the Special Research Funds of Hokusei Gakuen University for 2017–2018.

*I would like to thank Editage ([www.editage.jp](http://www.editage.jp)) for English language editing.*

## References

- Cambridge Assessment English. (2015). *B2 First exam format*. Retrieved February 2018, from: <https://www.cambridgeenglish.org/exams-and-tests/first/exam-format/>
- Council of Europe. (2001). *Common European framework of reference for languages (CEFR)*. Retrieved February 2018, from: <https://rm.coe.int/1680459f97>
- Eguchi, H. (2011). An examination of correlations between TOEFL and TOEIC scores among Hokusei Gakuen students: An attempt to find a score conversion formula. *Hokusei Review, the School of Humanities*, 2, 35–44.

- Global Exam (2016). *TOEIC and TOEFL in US universities*. Retrieved September 2016, from: <https://global-exam.com/blog/en/toEIC-and-toeFL-in-us-universities/>
- Green, A. (2016). Testing four skills in Japan. *British Council New Directions in Language Assessment: JASELE Journal (Special Edition)*, 135-144.
- Lowe, J. (n.d.). *Reforms for university admissions tests*. Retrieved February 2018, from: <https://www.garneteducation.com/changes-in-japanese-english-tests/>
- Ministry of Education, Culture, Sports, Science and Technology.(2003). *Action plan to cultivate “Japanese with English abilities”*. Retrieved March 2007, from: [http://www.mext.go.jp/b\\_menu/houdou/a15/03/03033101/001.pdf](http://www.mext.go.jp/b_menu/houdou/a15/03/03033101/001.pdf)
- Ockey, G. J., Koyama, D., Setoguchi, E., & Sun, A. (2015). The extent to which TOEFL iBT speaking scores are associated with performance on oral language tasks and oral ability components for Japanese university students. *Language Testing*, 32 (1), 39-62. doi:10.1177/0265532214538014
- Papageorgiou, S., Morgan, R., & Becker, V. (2015). Enhancing the interpretability of the overall results of an international test of English-language proficiency. *International Journal of Testing*, 15 (4), 310-336. doi:10.1080/15305058.2015.1078335
- Roca-Varela, M. L., & Palacios, I. M. (2013). How are spoken skills assessed in proficiency tests of general English as a foreign language? A preliminary survey. *International Journal of English Studies*, 13 (2), 53-68. doi:10.6018/ijes.13.2.185901



Appendix 1

Score conversion formula for university admission qualifications

各資格・検定試験とCEFRとの対照表

文部科学省 (平成30年3月)

CEFR	ケンブリッジ 英語検定	実用英語技能検定 1級-3級	GTEC Advanced Basic Core CBT	IELTS	TEAP	TEAP CBT	TOEFL IBT	TOEIC L&R/ TOEIC S&W
<b>C2</b>	230   200 (230) (210)	各級CEFR 算出範囲	各試験CEFR 算出範囲	9.0   8.5				
<b>C1</b>	199   180 (190)	3299   2600 (2630) (2599)	1400   1350 (1400) (1280)	8.0   7.0	400   375	800	120   95	1990   1845
<b>B2</b>	179   160 (170)	2599   2300 (2304) (2299)	1349   1190 (1304) (1280)	6.5   5.5	374   309	795   600	94   72	1840   1560
<b>B1</b>	159   140 (150)	2299   1950 (1949) (1980)	1189   960 (1080)	5.0   4.0	308   225	595   420	71   42	1555   1150
<b>A2</b>	139   120 (120)	1949   1700 (1728) (1728)	959   690 (840) (1278)		224   135	415   235		1145   625
<b>A1</b>	119   100 (100)	1699   1400 (1699) (1456) (1400)	689   270 (770)					620   320

○表中の数値は各資格・検定試験の定める試験結果のスコアを指す。スコアの記載がない欄は、各資格・検定試験において当該欄に対応する能力を有していると認定できないことを意味する。  
 ※ケンブリッジ英語検定・実用英語技能検定及びGTECは複数の試験から構成されており、それぞれの試験がCEFRとの対照関係として測定できる能力の範囲が定められている。当該範囲を下回った場合にはCEFRの判定は行われず、当該範囲を上回った場合には当該範囲の上限に位置付けられているCEFRの判定が行われる。  
 ※TOEIC L&R/ TOEIC S&Wについては、TOEIC S&Wのスコアを2.5倍して合算したスコアで判定する。  
 ※障害等のある受検生について、一部技能を免除する場合等があるが、そうした場合のCEFRとの対照関係については、各資格・検定試験実施主体において公表予定。

□は各級合格スコア

※定額内の数値は、各試験におけるCEFRの対照関係として測定できる能力の範囲の上限と下限

Retrievable at: [http://www.mext.go.jp/b\\_menu/houdou/30/03/\\_icsFiles/afeldfile/2019/01/15/1402610\\_1.pdf](http://www.mext.go.jp/b_menu/houdou/30/03/_icsFiles/afeldfile/2019/01/15/1402610_1.pdf)

## Appendix 2

Common Reference Levels: global scale (Council of Europe, 2001)

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics, which are familiar, or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.