

# **A Comparison of the Scores of Two TOEIC SW Tests of Hokusei English Majors: Facts and Analysis**

**Hitoshi EGUCHI**

# A Comparison of the Scores of Two TOEIC SW Tests of Hokusei English Majors: Facts and Analysis

Hitoshi Eguchi

## CONTENTS

1. INTRODUCTION
2. TOEIC SW TEST
  - Scoring
2. METHOD
3. TEST RESULTS
  - Comparison of the two TOEIC SW Tests
4. ANALYSIS
  - Other Variables
  - Comparison with TOEFL Scores
7. CONCLUSION

## INTRODUCTION

The unprecedented social and political demand towards English proficiency has meant that MEXT and schools in Japan are more eager than ever to impose better English education on curriculum. Oral English classes were introduced in 1994 in high schools, and the stereotypes of Japanese English learners being able to read and write well but not being able to speak well is beginning to change in this new context. Nonetheless, it is very difficult to grasp the true images of current English learners in Japan. Many still lament the poor English of Japanese people, and English proficiency level is measured on the basis of international standardized test engines, such as TOEFL PBT or TOEIC LR.

We also have to ask for what purposes Japanese teachers are teaching English; is it fair to judge people's achievement by the same scales and same methods when they are trying to achieve something different? For instance, MEXT is looking for communicative competence, and it is not being done for the purpose of getting a high TOEFL score. TOEIC seems closer to measuring communicative competence, but the English used in the test is the mainly for a business context.

From my classroom observations, although English levels among Japanese English learners have dramatically improved, it seems that Japanese people like to adhere to the stereotypical images of poor English learners. It might be because of the Japanese value of

---

Key words : Language Testing, Assessment, Test Engines, TOEIC SW, Productive Skills

modesty, but there could be other dynamics at play in such a low self-evaluation.

Even though people in general still rely on the scores of TOEFL PBT and TOEIC LR, they have been criticized (e.g., Irvin and Naggy, 2010) for not representing the communicative competencies that were developed in the second language acquisition studies in 1980's. Therefore, it is very important to measure current students' English proficiency in different tests to see what it is really like.

Even within the same institution, there are always discussions and debates about what tool is best in measuring general student achievement, and Hokusei Gakuen University (HGU) English Department is no exception. Malone (2010) suggests that the test score should be interpreted with regard to specific program needs (p. 633).

Therefore, in this paper, I strongly suggest that English proficiency should be measured by a tool, which directly tests students' productive skills, especially speaking. Furthermore, the achievement by the students at HGU should be measured by such tests as well.

In this study, two TOEIC SW tests were conducted to measure the development of English productive skills in a year. There were several purposes for this study: (1) how English majors improve their score by studying English (2) how TOEIC SW represents English proficiency of the students at HGU, and (3) what can be found by conducting TOEIC SW tests instead of TOEIC LR or TOEFL PBT.

From the data, several analysis were conducted: (1) how much improvement was seen (2) what variables correlate with the scores, and (3) how correlated the scores were with TOEFL PBT.

## TOEIC SW TEST

In brief, TOEIC SW Test is the English proficiency test made by ETS for productive skills to "directly assess the ability to speak and write in English in a workplace setting (ETS, 2009). Test takers speak into a headset and write on a keyboard attached to a computer. The

Task	Number of items	Response time	General outline of task
Read a text aloud	2	45 secs. per item (Prep time: 45 secs. per item)	Test-taker reads aloud a short text, such as an announcement or advertisement.
Describe a picture	1	45 secs. (Prep time: 30 secs.)	Test-taker gives a verbal description of a photograph.
Respond to questions	3	15 or 30 secs. (Prep time: none)	Test-taker responds to questions on a commonplace topic, as if responding in an interview.
Respond to questions using information provided	3	15 or 30 secs. (Prep time: none)	Test-taker responds to questions based on written information (such as a schedule of events) that appears on the screen.
Propose a solution	1	60 secs. (Prep time: 30 secs.)	Test-taker listens to a voice mail message describing a problem and gives a response. In the response, the test-taker indicates recognition of the problem and proposes a solution.
Express an opinion	1	60 secs. (Prep time: 15 secs.)	Test-taker expresses an opinion about a specific topic and the reasons for that opinion.
Write a sentence based on a picture	5	8 min. for 5 items	Test-taker writes one sentence based on a picture, using the two words or phrases provided.
Respond to a written request	2	10 min. for each item	Test-taker reads an e-mail message which is about 25 to 50 words long, and writes an e-mail in response.
Write an opinion essay	1	30 min.	Test-taker writes an essay expressing an opinion on the topic providing, giving reasons or examples to support the opinion.

(retrieved in 2012 from, [http://www.toeic.or.jp/toeic\\_en/sw/about.html#b1](http://www.toeic.or.jp/toeic_en/sw/about.html#b1))

data is sent online to a test center for assessment.

The speaking section consists of 6 different tasks each of which has 1-3 items and in total there are 11 items “to measure different aspects of speaking ability (ETS 2012).”

Explaining what TOEIC SW is like is not a purpose of this paper, so for more detail please check TOEIC homepage;

([http://www.toEIC.or.jp/toEIC\\_en/sw/index.html](http://www.toEIC.or.jp/toEIC_en/sw/index.html))

### **Scoring**

TOEIC SW is scored based on the criteria to scale students performance on each task, and the emphasis is placed on “task completion” of each item, instead of superficial fluency and accuracy (ETS Workshop, 2012) by just stating something general. Therefore, test takers are supposed to follow the instructions carefully and try to get the tasks done by speaking, and if the test takers gives only general information which can be an answer for any kind of questions, it won't be scored at all. Here is what is given by ETS on its homepage.

A Speaking test score and Writing test score will be given separately, each is reported on a scale of 0 to 200 in increments of 10. There are 8 proficiency levels for TOEIC Speaking test and 9 proficiency levels for the Writing test. Furthermore, Speaking test assessments relating to pronunciation as well as intonation and stress are also displayed in 3 levels respectively.

The rating process for the TOEIC Speaking and Writing tests takes place over ETS's Online Scoring Network (OSN), which is a highly secure internet-based system for evaluating test-takers' responses. Responses are sent to OSN, where they are scored by certified ETS test raters.

(retrieved from; [http://www.toEIC.or.jp/toEIC\\_en/sw/guide03.html#a](http://www.toEIC.or.jp/toEIC_en/sw/guide03.html#a))

## **METHOD**

Two TOEIC SW tests were conducted within an academic year, May 21, 2011 and February 2nd in 2012 with volunteer participants who were Hokusei Gakuen University English majors.

All of the participants (75 seats, maximum number possible within the budget) of 1st year, 2nd year and some of 3rd year students voluntarily participated in this study, were collected through online LMS (WebTube) in response to an announcement that they will be able to take TOEIC SW twice, on the condition that all the data will be used for research. In this way, the participants were all happy to take the test, which avoids an irrelevant variable due to their unwillingness to take the test. In this kind of research, it is important to control as many variables as possible to raise the reliability, and I attempted to lower the effect of unwillingness to take a test. The fourth year students are all those who experienced at least one-year study abroad, such as Hokusei Gakuen Exchange Program or Working Holiday in

Australia.

In reality, on the day of the first test, 5 students out of 75 who signed up, failed to show up. There were some excuses from oversleeping to being infected with the seasonal flu. For this reason, I sent a reminder to the participants when the second test was approaching.

10 out of 70 of the students claimed that they could not participate in the 2nd test for various reasons. Therefore, I collected an additional 15 students to fill the empty seats.

However, on the day of the second test, only 69 showed up. As a result, 55 students took the test twice. 55 is rather a small number to see the development of language proficiency by analyzing the difference of the two scores and the correlation between their score improvement and other relevant variables, such as attitude and motivation, or the correlation with another test engines, such as TOEFL and TOEIC RL. Therefore, it is very difficult to make any definitive conclusions. However, I believe it is worth of mentioning what we can see from the data and performance of the students in terms of testing productive skills, especially, speaking, which is rarely tested by any standardized tests.

## TEST RESULTS

From this test result, there is not much to be said, but comparing the speaking and writing crosssectionally in terms of years they are in, Hokusei students performed better in the writing test. The most frequent score of writing 150 in both Test 1 and Test 2, while those of both of speaking tests are 100 in test 1, and 130 in test 2 respectively. Although writing seems easier, it is more difficult to expect improvement in writing. The difference between 1<sup>st</sup> year students and 4<sup>th</sup> year students in speaking is much wider than the one of the writing,

**Date: May 21<sup>st</sup>, 2011**

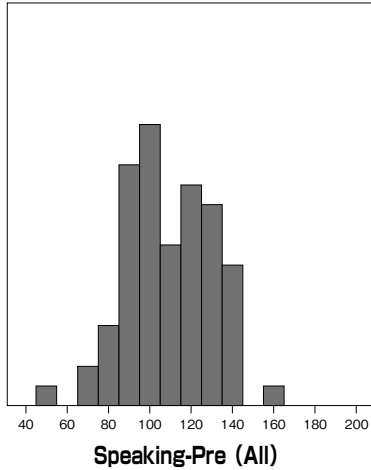
Test 1		Speaking		Writing	
		Mean	SD	Mean	SD
Total	70	109.0	21.0	132.6	22.4
1st	17	104.7	16.6	119.4	24.4
2nd	27	108.2	22.02	134.4	18.3
3rd	20	106.5	21.8	134.0	22.6
4th	6	133.3	8.2	156.7	8.2

[Table 1: Test Result of All Participants in Test 1]

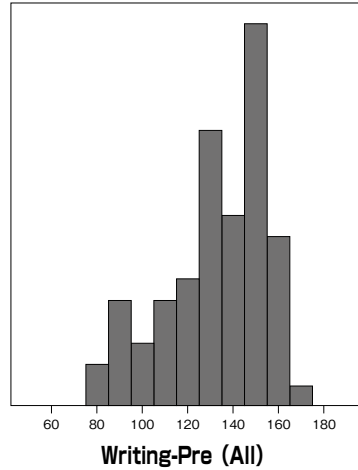
**Date: February 2<sup>nd</sup>, 2012**

Test 2		Speaking		Writing	
		Mean	SD	Mean	SD
Total	69	121.9	22.1	134.6	19.4
1st	18	112.8	22.4	127.2	16.7
2nd	28	126.8	19.1	138.6	17.2
3rd	20	120.5	25.0	136.0	23.5
4th	3	140.0	0.0	133.3	20.8

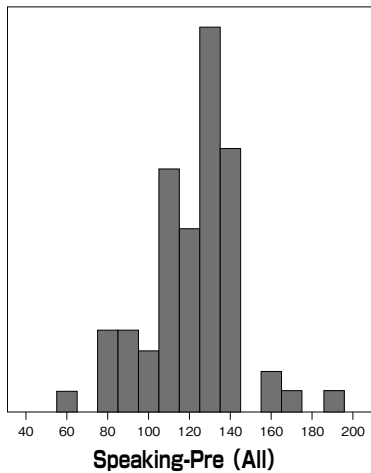
[Table 2: Test Result of All Participants in Test 2]



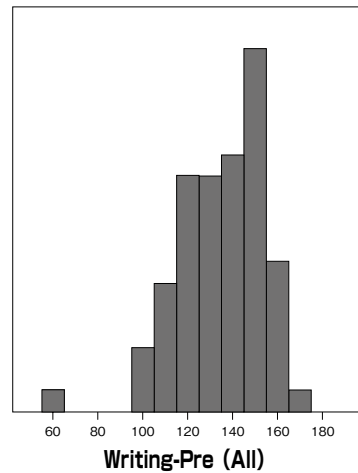
[Graph 1: Score Distribution of Speaking 1 of All]



[Graph 2: Score Distribution of Writing 1 of All]



[Graph 3: Score Distribution of Speaking 2 of All]



[Graph 4: Score Distribution of Writing 2 of All]

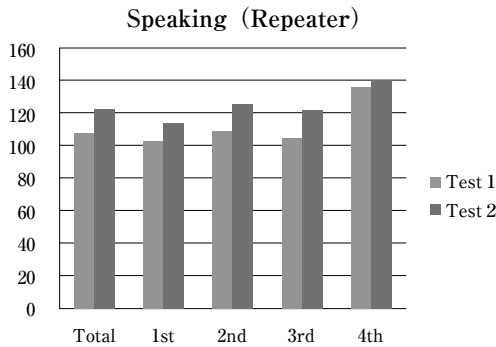
which may lead us to conclude that high school education might privilege the development of writing over other skills or that it takes more time to develop speaking skills and writing skill can be a threshold for developing speaking skill later. The results could also lead us to conclude that the education at Hokusei Gakuen University helps students to develop speaking skills

### Comparison of the two TOEIC SW Tests

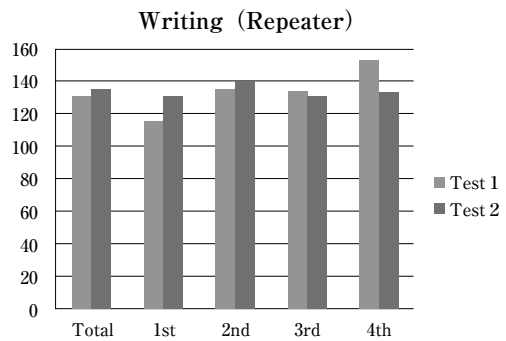
These graphs (5-6) indicate the cross-sectional scores of speaking test 1 and 2 and writing test 1 and 2 respectively. Interestingly, the average score of 2<sup>nd</sup> year students' was higher than that of 3<sup>rd</sup> year. There is a clear shift of the most frequently occurring score from around 100 to 130. Average score improvement is about 14 points, but taking the fact that some of their scores became lower, the score improvement is significant. 17 students experienced 30 points or higher score improvement, and 9 experienced even 40 point score

Repeater		Speaking			Writing		
		Test 1	Test 2	Difference	Test 1	Test 2	Difference
Total	55	108.0	122.4	+14.4	130.9	135.1	+4.2
1st	14	102.9	113.6	+10.7	115.7	130.7	+15.0
2nd	23	109.6	125.7	+16.1	135.2	140.4	+5.2
3rd	15	104.7	122.0	+17.3	134.0	131.3	-2.7
4th	3	136.7	140.0	+3.3	153.3	133.3	-20.0

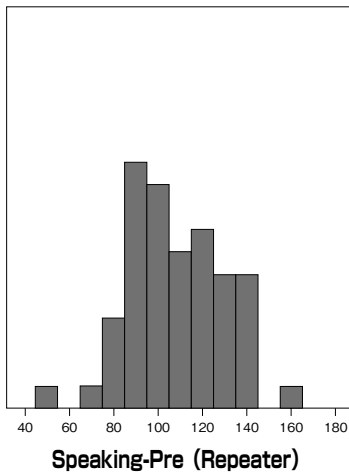
[Table 3: Test Scores of Two TOEIC Tests of the Participants in Both Tests]



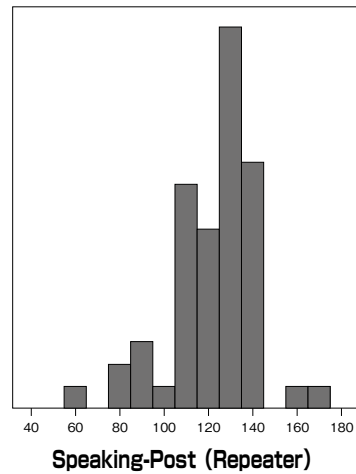
[Graph 5: Average Scores of Speaking Test 1 and 2 in Each Year]



[Graph 6: Average Scores of Writing Test 1 and 2 in Each Year]



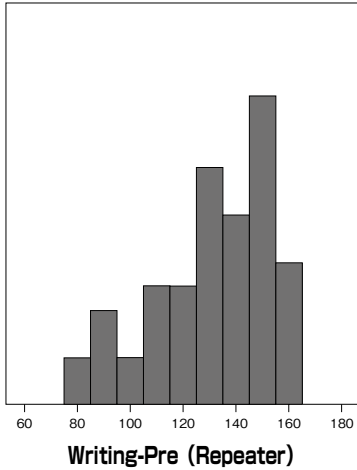
[Graph 7: Score Distribution of Speaking 1]



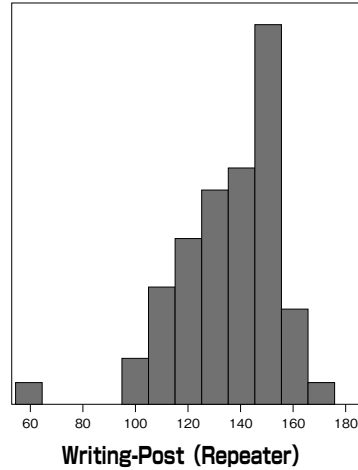
[Graph 8: Score Distribution of Speaking 2]

improvement.

Graph 5-8 show how the score distributed along the score line of both speaking and writing of both times. There were 8 months between the two tests. There are obvious shifts of the most frequent score from the left to the right (higher score).



[Graph 9: Score Distribution of Writing 1]



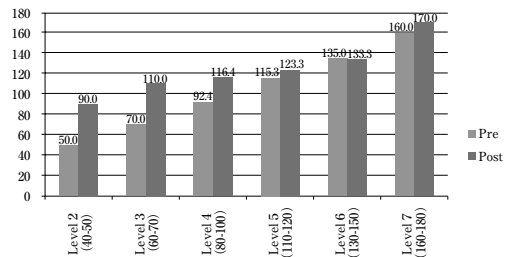
[Graph 10: Score Distribution of Writing 2]

## ANALYSIS

The score improvement in speaking is 14.4 points on average, while the biggest score improvement is 50 points. The score improvement in writing is 4.2 on average and the biggest score improvement is 50 points. In speaking, the score improvement was significant, taking into consideration the fact that there were 10 of negative improvement. It seems that it was easy to get a higher score when the score was low in the first test.

Here is the score comparison in each level (given by ETS) in the test 1 with the scores in the test 2. The levels are the score ranges in which “the types of general skills and abilities in speaking English that are common for the most people (ETS, 2012).”

Up to level 5 in Test 1, many (26 out of 42) experienced 20 points improvement in Test 2, but it seems very difficult to improve the score from level 6, and the lower the score is, the bigger the score improvement is. In



[Graph 11: Score Comparison of Average Score of Each Level in Test 1]

## Speaking

Levels in Test 1	N	Mean Score of Test 1	Mean Score of Test 2	Difference
Level 2 (40-50)	1	50.0	90.0	+40.0
Level 3 (60-70)	1	70.0	110.0	+40.0
Level 4 (80-100)	25	92.4	116.4	+24.0
Level 5 (110-120)	15	115.3	123.3	+8.0
Level 6 (130-150)	12	135.0	133.3	-1.7
Level 7 (160-180)	1	160.0	170.0	+10.0
Total	55	108.0	122.4	+14.4

[Table 4: Score Comparison of Average Score of Each Level in Test 1]



addition, the average scores of 4<sup>th</sup> year students, who all experienced one year study in English speaking country, are 133 and 140 in Test 1 and Test 2 respectively. There seems to be a big wall somewhere in level 6 (130-150). From my observation, students who have the ability to speak English naturally score higher than 140. Therefore, it might be very difficult for students who study English only in Japan to get 140 or higher, (though off-course there are many cases of exceptions) or expect further score improvement in 8 months.

Compared with speaking, there is only a 4.4 score improvement in writing. The average score of writing in Test 1 is 130.4, and it is already high. Therefore, we may conclude that English education in Japan is still in favor of developing writing skills, and they have already reached a plateau stage of language development in writing. However, the score improvement from Test 1 to Test 2 of the 1st year students is 15, and there is still significant score improvement in writing after they entered university. It is impossible to reach a definite conclusion without a multiple year study, but it is reasonable to say that students are learning quite a lot at college. The test scores also shows that there could be more development in writing too. Therefore, this kind of test can be a good tool to measure English proficiency of the students, but also to know what we still can do for students to develop more language skills.

### Other Variables

There could be many factors, which determine the scores and score improvement solely or working with other factors, such as experience studying abroad, motivation, washback effect, test wiseness, amount of time spent studying English or language learning aptitudes. I attempted to find some correlations between test scores and such variables by conducting a questionnaire survey to accompany the test.

Because there were only 55 participants in this study who took the test twice, it was very difficult to see a significant correlation between the score and such variables. The length of experience staying in English speaking countries seems to have a significant effect on the score, but the number of participants who have such experience is too small-8. So, it might be possible for me to say that the length of time spent in English speaking countries affects the scores based on my observation, but it is not clear enough to be statistically proven.

In addition, such factors cannot have sole effects on the scores of the participants in this study, because all the participants are English majors and have enough exposure to spoken and written English constantly throughout the year. Such factors as motivation would have more effect on the learners' behaviors and scores, if they are not studying English regularly at school.

Even though there is no strong correlation with the scores, more than 90% (50 out of 55) of the students claimed that they liked the TOEIC SW test, and 72.7% (40 out of 55) also claimed that knowing that they were going to take the TOEIC SW test had some effect on the way of studying English. People in general do not like the test, but when they know they are being tested on what they are good at, they might like the test. I think because they feel they

are tested for what they have been studying for, they like the test. I believe that TOEIC SW Test gives positive washback effect on how they get prepared for the test, and will motivate the students to study harder to get a better score.

After the first test was administered in May, many students came to me and claimed that they could do a better performance and promised to perform better for the next time. I believe that the TOEIC SW Test can be a motivational factor for development of productive skills, because it directly assesses the students' productive skills. Now students are more conscious about their fluency and practical skills rather than just a TOEIC or TOEFL scores. They are happy to be assessed on what they are eager to learn.

### Comparison with TOEFL Scores

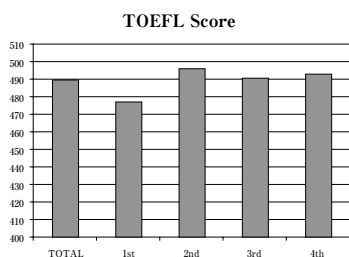
It was not a primary purpose of this study, but I attempted to see what can be seen by comparing the TOEIC SW scores and TOEFL ITP scores of the participants.

Hokusei Gakuen University is conducting TOEFL ITP every year with all students. In the academic year 2011, first year students took the test twice, in April and in January (2012), the second year in September, the third year in August. TOEFL was made by ETC and has been used as an admissions criterion for non-native speakers of English to higher educational institutions in North America (Taylor & Angelis, 2008). The first TOEFL was administered in 1964, and was developed into the latest version of TOEFL (iBT), which includes a speaking section (Anderson, 2009).

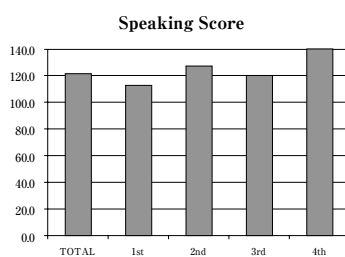
The version of TOEFL administered at Hokusei Gakuen University is TOEFL ITP, which uses new and previously administered TOEFL questions, and which has three sections; Listening Comprehension, Structure and Written Expressions, and Reading and Vocabulary with the same format with TOEFL PBT. There is no Speaking section, unlike TOEFL iTB,

Year	N	Mean	Mini	Max	SD
TOTAL	69	489.3	403	567	34.909
1st	18	477.2	410	520	33.128
2nd	28	495.8	403	567	39.018
3rd	20	490.7	410	540	31.468
4th	3	493.3	480	507	13.503

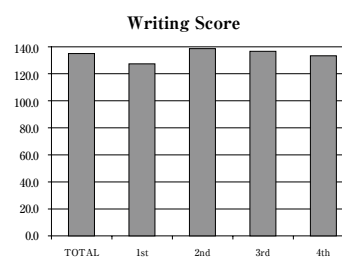
[Table 5: TOEFL Scores of the Participants]



[Graph 13: Average TOEFL Score]



[Graph 14: Average TOEIC Speaking Test 2]



[Graph 15: Average TOEIC Writing Test 2]

the most recent version of TOEFL.

As can be seen from the graph 13-15, there is a clear similarity among the average scores of each year. If we consider the average scores of the students of each year as a single score, besides the scores of 4<sup>th</sup> year students' (there is too small number of samples), there is a strong correlation between TOEFL Scores and the ones of Speaking (0.74) and between TOEFL Scores and Writing (0.92), but not so strong between Speaking and Writing (0.43, see Table 6). This must indicate that students at Hokusei Gakuen University are learning four skills in a well balanced manner.

However, if the correlations are sought from individual scores, they become not so strong (see Table 7)

These low correlations mean that the high TOEFL scores do not necessary mean high SW and vice versa.

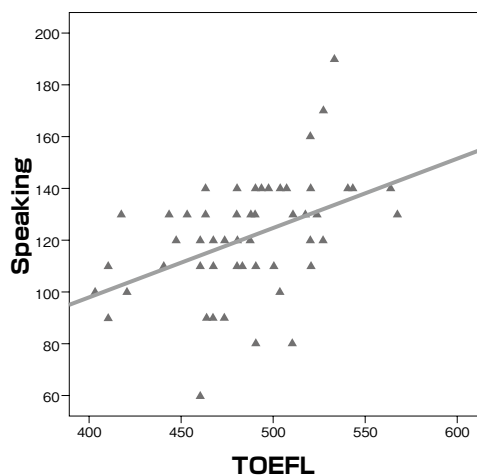
The graphs above show how the scores spread in relation between TOEFL and Speaking, and TOEFL and Writing. The lines in the graph shows the linear regression of the two scores. There is wider spread in speaking than writing, which shows clearly how those scores are correlated shown in Table 7. As I mentioned previously, there must be many factors which contribute to the scores of those tests, and the difficulty level of TOEFL might lead to the low correlations between speaking and TOEFL. However, based on my experience teaching English at Hokusei Gakuen University, there are many students who are fluent speakers with good listening comprehension skills, but without accuracy in forms. Because the scoring

	TOEFL	Speaking	Writing
TOEFL	1		
Speaking	0.74	1	
Writing	0.92	0.43	1

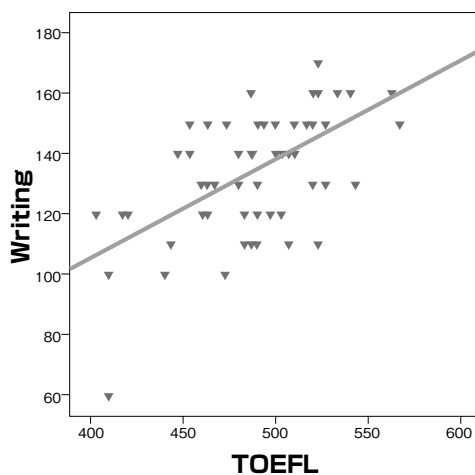
[Table 6: The Correlations with TOEFL Average Scores of Each Year]

	TOEFL	Speaking	Writing
TOEFL	1		
Speaking	0.42	1	
Writing	0.58	0.36	1

[Table 7: The Correlations of Individual Scores of TOEFL and TOEIC SW]



[Graph 16: Score Distribution between TOEFL and TOEIC Speaking]



[Graph 17: Score Distribution between TOEFL and TOEIC Writing]

criteria of speaking and writing test put more emphasis on task completion rather than the grammatical accuracy of the produced utterances, such students can score higher than the TOEFL score might indicate. These types of students can be the products of the change in teaching English in recent years. If the purpose of studying English is for acquiring practical skills by the use of English, the achievement should be measured by a test, which can directly assess what they have learned.

## CONCLUSION

Now the demand for productive skills in English is unprecedentedly high and English education has been changed because of that. Japanese peoples' characteristics as English learners have been changed along with the changes of education. It is especially important to assess the achievement fairly by a test which is able to measure the proficiency for which education is conducted.

In this study, two TOEIC SW tests were administered at the beginning and the end of an academic year, 2011, to see the development of productive skills as a result of education provided at Hokusei Gakuen University English Department by comparing the two scores. There was a hike in score improvement among those who scored low in the first test and there was a steady improvement in the higher score range too. It might reflect the general impression of language development that the higher the levels of proficiency become, the more difficult it is to experience the improvement.

If the score improvement can be attributed solely to the education provided at HGU English Department, it is easy, but language development is very complex and many factors must be contributing. In the history of second language acquisition research, the attempt to control many variables in research has been made, but it is very difficult to control them in a strict sense in this kind of research. In fact, it might be more preferable not to control them and to say as much as teachers can say based on their everyday observations in order to see the reality of complex language development.

Several attempts were made in this study to see correlations with several other factors, such as length of studying abroad or motivation by conducting a questionnaire along with the test. There should be bigger number of participants to see the relationship with score improvement. Such factors as motivation are not always correlated with language development, so it is difficult to see a certain tendency in this scale of research. More consistent and longer-term research should be done to see such a relationship with language development.

This study clearly shows that the students at Hokusei Gakuen University develop not only receptive skills, such as reading and listening, but also significant improvement in productive skills, such as speaking and writing. This seems to be justifying the approach by the HGU English Department English to teach for developing the four skills equally.

Through this study, it is also found that it is very important to assess the productive skills

by a test, which directly measures such skills, not by a test that measures them indirectly by testing receptive skills only. It is especially so when productive skills are emphasized in a language curriculum. Using an appropriate test also motivates students to study harder to be more prepared for a better score, regardless whether it will lead to score improvement.

TOEIC SW is a reliable test in terms of the content of the test, which is able to test a wide range of proficiency levels. The days are over when Japanese are studying English not knowing whether they will use it in the future. Japanese people need to learn English productive skills by using them. TOEIC SW will be a good candidate to support such practical English education in Japan.

## Acknowledgement

Data analysis and interpretation in this study were supported by the Institute for Internal Business Communication (IIBC)

## REFERENCES

- Alderson, J. (2009). Test review: Test of English as a foreign language™: Internet-based test (TOEFL Ibt®). *Language Testing*, 26 (4), 621-631. doi:10.1177/0265532209346371
- ETS. (2012). TOEIC Speaking and Writing Tests. Retrieved October 16, 2012 from [http://www.toeic.or.jp/toeic\\_en/sw/about.html#b1](http://www.toeic.or.jp/toeic_en/sw/about.html#b1)
- ETS. (2012). TOEIC SW Tests; Scores. Retrieved October 16, 2012 from [http://www.toeic.or.jp/toeic\\_en/sw/guide03.html#a](http://www.toeic.or.jp/toeic_en/sw/guide03.html#a)
- Irwin, B. & Nagy, P. (2011). The 2010 Revision of the TOEIC® Speaking Test. *JALT Testing and Evaluation SIG Newsletter*, 15 (2), 10-19
- Malone, M. E. (2010). Test Review: Canadian Academic English Language (CAEL) Assessment. *Language Testing*, 27 (4), 631-636.
- Powers, D. E., Kim, H., Yu, F., Weng, V. Z., VanWinkle, W., & Educational Testing, S. (2009). The TOEIC[R] Speaking and Writing Tests: Relations to Test-Taker Perceptions of Proficiency in English. Research Report. ETS RR-09-18. *Educational Testing Service*,
- Taylor, C. A., & Angelis, P. (2008). The Evolution of the TOEFL. In C. Chapelle, M. Enright, J. Jamieson (Eds.), *Building a Validity Argument for the Test of English as a Foreign Language* (pp. 27-54). New York, NY: Routledge.

[Abstract]

## A Comparison of the Scores of Two TOEIC SW Tests of Hokusei English Majors: Facts and Analyses

Hitoshi EGUCHI

In this study, two TOEIC SW tests were conducted to measure the development of English productive skills in one year to validate the education provided at Hokusei Gakuen University English Department. From the data, several analyses were conducted: (1) how much improvement was seen; (2) what variables are correlating with the scores; and (3) how correlated the scores were with TOEFL iTP. There was a clear score improvement among participants in this study, and it proves that students are developing not only receptive skills but also productive skills, while it was difficult to see strong correlations with other variables such as motivation. The author also emphasizes the importance of measuring the productive skills by a test designed for measuring directly those skills like TOEIC SW.

---

Key words : Language Testing, Assessment, Test Engines, TOEIC SW, Productive Skills

