《研究ノート》

Occasion to Implement an English Proficiency Test and Possibility to Pick Up Industrious Students: A Case Study in Japanese National University

Yuzuru Utsunomiya

Abstract

The purpose of this note is to investigate when a placement test should be implemented and possibility to pick industrious students up at their entry. Using TOEIC score data collected in a Japanese national university, we estimate occasion when the students' score raise up and test difference of the score.

We found two results in the following. First, scores of the industrious students raise up between in their freshman and sophomore dramatically. Ones between in their sophomore and junior raise up less than that of their youth. Second, at the students' entry, we cannot separate the industrious students.

We conclude that solutions to improve students' English proficiency in their sophomore and upper should be considered.

Keywords: English proficiency: TOEIC: Industrious student

Introduction

Purpose and background

The purpose of this study is to investigate students' English proficiency in terms of TOEIC score. Particularly, we focus on the following topics.

1. How does a trend of students' score transform by proficiency section; listening, reading, and total.
2. Whether is it possible to select " industrious " students at their entry?

To develop students' English proficiency, Japanese universities have introduced a lot of solutions such as streaming class, examination fees support, additional training course throughout 2010s. Streaming class is one separated in accordance with a result of placement test. To operate an English class easily and effectively, or to follow Japanese government's intention, the streaming class method have been introduced in many Japanese universities with variation (Tomioka, 2010; Saegusa, 2014). If a student obtain a remarkable score, the students is supposed to assign a class with high-level content. If not, the students are supposed to learn a fundamental content.

Viewing from a supervisor in charge of English proficiency development, regarding the solutions, it is likely that there are two issues exist; finding industrious student and occasion to start a so-called streaming class.

Finding an industrious student is crucial matter. The industrial student is a student taking tests and attending classes without any absence. When a university / faculty implement the solution, university should accumulate and analyze the students' score data for the measurement so that it measures long-term performance of the solutions above. If targeting students would be

absent or not take a test, the data collection will not be able to remain any more. As a result, the university cannot analyze the solutions and compensate social responsibility of the university.

In reality, many students have dropped out even though a university provide privileges such as testing fee support. As a result, a university cannot often collect data and analyze long-term trend and characteristics of the score. A university have to make effort to increase the number of the industrial students.

Second, occasion implementing a placement test matters. Streaming classes have been implemented to improve students' English proficiency more effectively. Such a class is established on the basis that the students are separated by their proficiency. So to open such a class, a university should implement a placement test before opening the class. Somehow the test have been implemented in April, at the beginning of fiscal year. In the month, other activities are implemented at a time and the university implement the test while spending insufficient period. As a result, some students may lose opportunities to take a class with a sufficient level. Moreover, some students are not interested in learning English. To open such a class for the students is waste of time and human resource. For a university and professors relating the test, the test is a heavy burden. They should implement the test, mark its result, analyze it within a couple of days, separate the students in accordance with their proficiency, and declare the result to the students within a few of weeks. In the period, there are other works and sometimes the professors cannot distribute their working hours to the placement procedure sufficiently. If the test can be done in other period, a university can consider to switch the test to other months. We, however, do not know when the most appropriate period is.

Prior research

For the placement test, many types of tests such as TOEIC, TOEFL, and IELTS have been applied. This time, we would like to focus on TOEIC by ETS (Educational Testing Service). According to prior researches, TOEIC is dependable to use as an English proficiency test (In'nami, 2011; Zhang, 2006). In detail of TOEIC, please refer to the following website (http://www.ets.org/toeic).

Researchers in Japanese universities have reported their experiences on TOEIC in relation to their English education. Ogasawara reports effects of streaming class (Ogasawara, 2011; Ogasawara, 2012). Using G-TELP, an English proficiency test equivalent to TOEIC, he tried to evaluate the effects of streaming class toward development of English proficiency in some faculties. According to the results of his trial, the class have succeeded in terms of separating the students by proficiency. He translated the G-TELP score to TOEIC in the process of evaluation. This means that a university as well as students and other principals pay attention to TOEIC score. Students' English proficiency should be evaluated by tests no other than TOEIC score. The survey, however, contains some faults on statistical analyses. He uses so-called t-test to evaluate the effects of streaming class without considering effect of faculty. Maruyama reports situation regarding English proficiency in a faculty of Japanese university. According to the survey, the faculty have not implemented inspection of students' English proficiency so far until 2010s' (Maruyama, 2011). Students' proficiency level is around TOEIC C level.

We understand that many Japanese universities make use of English proficiency tests such as TOEIC, TOEFL, and G-TELP as a placement test. In addition, the tests are implemented in a specific period while spending a

limited period. The tests are regarded as a part of general education, not specific education. We should note that there are some limitations. First, we, however, cannot find controversy discussing on effective implementation of the test and application of the test's results. When should we implement the placement test? Who is a target of the test? The prior researches focus on freshmen or sophomore students. Probably learning English belongs to general education and many trials have been done in the period. Nowadays, however, the proficiency at Japanese higher education becomes more crucial matter. In current situation, in every faculty or department, a university should consider to grow students' English proficiency up. A university do not have to fix occasion to learn English just in a general education.

Second, the prior researches have treated the students as a mass. We, however, should treat to separate the students based on their characteristics and behavior carefully. Suppose comparison industrious students' scores with ones of lazy students who take the test intermittently. If we mix the two or other groups with different characteristics, we cannot measure entire growth of English proficiency precisely.

Object and method

Our target is TOEIC score that students in a Japanese university have obtained. The score have been collected from 2012 to 2015. In the university, since FY 2011, the test have been implemented once a year for students. We particularly focus on students entering in FY 2012 and FY 2013. Regarding the testing schedule in each of the generation, please refer to table Table 1.

Using the data, we generate two type of dataset; a panel data of industrious students' scores belonging to the two generations. To confirm their long-

Table 1 Testing calendar by year (Y: Period when the tests have been inplemented)

FY	Grade	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.
2012	Freshmen										Y		
	Sophomore										Y		
	Junior										Y		
	Senior												
2013	Freshmen	Y					Y				Y		
	Sophomore	Y					Y				Y		
	Junior	Y					Y				Y		
	Senior						Y				Y		
2014	Freshmen	Y			Y								
	Sophomore	Y											
	Junior	Y											
	Senior												
2015	Freshmen				Y								
	Sophomore				Y								
	Junior				Y								
	Senior												

term growth, we employ Friedman's rank sum test and pairwise comparison method while using the panel data. We set that significant levels of the statistical procedures are 5% each.

Another is a cohort data of students' score at their entry. We sort original score data by the entry year and frequency of taking the tests. As for the frequency, logically, the following eight groups can be generated.

- Group 1: students taking all the three tests during the period. We call the students belonging to this group as" industrious students ".
- Group 2: students taking two tests at freshmen and junior.
- Group 3: students taking to tests at freshmen and sophomore.
- Group 4: students taking two tests at sophomore and junior.
- Group 5: students taking an test at junior.

- Group 6: students taking an test at freshmen.
- Group 7: students taking an test at sophomore.
- Group 8: Students not taking any tests. We cannot find such students because no records regarding their score exist.

To find a difference of their scores by the group, we employ Kruskal-Wallis rank sum test and the multiple comparison. Using the methods, we estimate a possibility whether it is possible to predict students' future behavior in terms of initial score.

For the analyses demonstrated above, we use R (Ver.3.2.0), a free and reliable statistical analysis environment.

Results

The number and score trend of industrious students

First, we would like to confirm the number of students taking all the targeting tests by the group above. We can understand that most of the students have not taken the tests (Table 2). Students do not always take the tests every time even though the university gives promotion of free taking. Unless we would not separate by the groups, we would not be able to compare to analyze their score precisely. For example, Students entering in FY 2012, only 8 students have taken the exams fully. Rest of the 394 students have suspended one or two tests. In FY 2013, 19 students have taken the tests. Rest of the 388 students have suspended one or two tests. Considering total number of the students by year, the N of industrial students is extremely low. At least, although the university has given opportunities to take the tests without any payment, their ambition has not been enhanced. Also,

regarding the generation and the groups, we can observe a statistically significant difference (Fisher's Exact Test, $p<2.2e-16$). By generation, pattern of distribution of the group is different.

Table **2** the nuimber of students taking exam by generation and pattern

	Groups						
Generation	1	2	3	4	5	6	7
FY2012	8	41	30	0	1	321	1
FY2013	19	256	5	0	10	117	0

The industrious students' score increase year by year. Particularly, between their freshman and sophomore, it grow up more than one between sophomore and junior. Total median (mean) score of FY 2012 grow up from 447.5 (408.1) to 502.5 (498.8) between freshman and sophomore. One of FY 2013 grow up from 370 (379.5) to 420 (450.8). Regarding the total score between the periods, there are statistically significant differences at 5 % level. (Table 3). Considering results of multiple comparison (Table 4), a difference exists. In FY 2012, there is not statistically significant difference between grades. In FY 2013, however, not only between freshman and sophomore but also freshman and junior, there are statistically significant differences. These results supports that the students' total score grow up between their freshman and sophomore. Moreover, we cannot regard that the score grow up between their sophomore and junior.

Regarding growth in students' listening proficiency score, we can confirm the same pattern as total score. Namely, the score increase between freshman and sophomore as well as between freshman and junior (Table 3, Table 4).

Scores representing students' reading proficiency demonstrate a different

Table 3 Friedman's rank sum tests results

Treat	Entry year	Chi-squraed statistics	df	p
Total	FY2012	4.8667	2	0.08774
Reading		6.6452	2	0.03606
Listening		2.25	2	0.3247
Total	FY2013	13.68	2	0.000107
Reading		13.162	2	0.000139
Listening		14.48	2	0.000717

Table 4 Multiple comparison by generation and testing year, and proficiency results

FY2012								
Total			Reading			Listening		
	F	S		F	S		F	S
Sophomore	0.32		Sophomore	0.45		Sophomore	0.32	
Junior	0.32	0.95	Junior	0.45	0.67	Junior	0.50	1.00
FY2013								
Total			Reading			Listening		
	F	S		F	S		F	S
Sophomore	0.03		Sophomore	0.09		Sophomore	0.01	
Junior	0.00	0.06	Junior	0.00	0.09	Junior	0.00	0.50

movement. Certainly, according to the Friedman's rank sum tests results, there are statistically significant differences at 5% level in both FY 2012 and FY 2013 (Table 3). This means that the scores are difference between the years. Considering results of multiple comparison, there is a significant difference merely between freshmen and junior of FY2013. This fact means that it needs longer period / effort to grow the reading proficiency up in comparison with the rest two proficiency.

Initial score

We cannot hardly separate the students into industrious students and others at their entry.

Between the groups, regarding scores of the three proficiency, there are

statistically significant differences at 5% level (Table 5). However, considering the scores' distribution while referring to results of multiple comparison (Table 6), we cannot but hesitate to separate them using the test results. In total and listening scores of FY 2012, we can observe the significant difference merely between group 2 and group 6, and group 3 and group 6. As for the reading score, we cannot observe the difference. In total, we can observe the difference merely between group 2 and group 6. Similarly, as for the reading score, we cannot observe the difference. Between rests of the group, there is no significant differences.

Table **5** Kruskal-Wallis rank sum tests results

Treat	Entry year	Chi-squraed statistics	Df	p
Total	FY2012	17.354	3	0.000598
Reading		12.364	3	0.006234
Listening		14.346	3	0.00247
Total	FY2013	9.3651	3	0.02461
Reading		6.7113	3	0.08169
Listening		8.2139	3	0.04179

Table **6** Multiple comparison by generation and group, and proficiency results

Total				Reading				Listening			
FY2012	G1	G2	G3		G1	G2	G3		G1	G2	G3
Group 2	1.00	-	-	2	1.00	-	-	2	1.00	-	-
Group 3	1.00	1.00	-	3	1.00	1.00	-	3	1.00	1.00	-
Group 6	0.32	0.03	0.02	6	0.67	0.09	0.07	6	0.80	0.04	0.04
FY2013	1	2	3		1	2	3		1	2	3
Group 2	1.00	-	-	2	1.00	-	-	2	1.00	-	-
Group 3	1.00	1.00	-	3	1.00	1.00	-	3	1.00	1.00	-
Group 6	1.00	0.02	1.00	6	1.00	0.07	1.00	6	1.00	0.03	1.00

Discussion and Conclusion

The purpose of this study is to investigate when students' English proficiency raise up and possibility to separate the students into industrious one and others. For analyses, we use the students' TOEFL scores obtained from 2012 to 2015 in a Japanese national university. To predict when the score raise up and fall down, we employed Friedman's rank sum test and multiple comparison for industrious students, ones who have taken the exams without absence. To discuss on the possibility on separation, we employed Kruskal-Wallis rank sum test and multiple comparison.

We observed the following results. First, as for the scores they raise up between their freshman and sophomore most. The score, however, raise slightly up between sophomore and junior. Second, regarding their scores of the three proficiency, among groups generated by the students' frequency of taking tests, there is no critical difference. The fact means that we cannot separate the industrious students from the beginning. In other words, to pick such students, we have to wait until their proficiency will grow up.

We would like to conclude that treatment for sophomore and upper students is also vital issue to increase the students' English proficiency continuously. The students' score, even by the industrious students, increase as they grow up. This means that between their sophomores and junior, there must be some reasons to inhibit their growth. At least, the existing solutions work as expected.

A university cannot stop supporting the students' English learning. According to our survey, we cannot find industrious students, contributing for university's analyses, cannot separate after finishing entire tests. Unfortunately, we cannot distinguish such students and others in terms of the

tests' scores. We should seek such students while spending a few years. Otherwise, we are not able to find them and consider the next data-oriented solutions to improve the students' English proficiency.

It should be notes that the conclusion merely mention English proficiency which is able to measure by means of the test. Of course, some other proficiency which is NOT able to measure using such tests. In other words, the measurable proficiency is a necessary condition to build the students' total English proficiency. A university should not just focus on the tests' score to raise up its reputation, not students' proficiency.

Finally, we would like to note further researches. First, we would like to investigate why the scores fall down between the grades. If we can make it clear, we are able to propose more effective solution to raise up the score and grow up their English proficiency.

Acknowledgement

This study is partly supported by JSPS KAKEN Grant (No. 26370730)

Appendices A Descriptive statistics of TOEIC score by entry year and group

Total

Entry Year	Group	n	Min.	Median	Mean	Max.	SD	SE
FY2012	1st.	8	240	447.5	408.1	460	76.0	26.9
	2nd.	41	225	390	404.9	600	88.3	13.8
	3rd.	30	265	420	417.3	580	97.4	17.8
	6th.	321	165	360	361.3	8,555	91.3	5.1
FY2013	1st.	19	265	370	379.5	535	79.1	18.1
	2nd.	256	195	380	383.2	610	82.7	5.2
	3rd.	5	240	455	396	480	101.9	45.6
	6th.	117	175	340	354.8	620	94.8	8.8

Reading

Entry Year	Group	n	Min.	Median	Mean	Max.	SD	SE
FY2012	1st.	8	100	185	181.2	203	48.8	17.2
	2nd.	41	80	180	179.5	330	55.6	8.7
	3rd.	30	90	190	182.3	285	49.5	9.0
	6th.	321	50	155	157.3	385	54.5	3.0
FY2013	1st.	19	100	160	166.3	280	44.6	10.2
	2nd.	256	65	160	170.1	340	51.6	3.2
	3rd.	5	110	165	173	230	46.4	20.8
	6th.	117	60	150	154.8	280	51.0	4.7

Listening

Entry Year	Group	n	Min.	Median	Mean	Max.	SD	SE
FY2012	1st.	8	140	225	226.9	305	46.6	16.5
	2nd.	41	140	225	225.4	330	42.2	6.6
	3rd.	30	115	220	235	330	58.0	10.6
	6th.	321	75	210	204.1	470	50.7	2.8
FY2013	1st.	19	130	200	213.2	305	47.6	10.9
	2nd.	256	90	210	213.2	325	45.0	2.8
	3rd.	5	130	225	223	315	69.3	31.0
	6th.	117	80	195	200	355	57.5	5.3

Appendices B Descriptive statistics of TOEIC score by entry year and grade

Total

Entry Year	Grade	n	Min.	Median	Mean	Max.	SD	SE
FY2012	Freshman	8	240	447.5	408.1	460	76.0	26.9
	Sophomore	8	265	502.5	498.8	675	123.1	43.5
	Junior	8	180	527.5	480	720	167.7	59.3
FY2013	Freshman	19	265	370	379.5	535	79.1	18.1
	Sophomore	19	325	420	450.8	765	111.1	25.5
	Junior	19	380	455	485.5	765	97.4	22.3

Reading

Entry Year	Grade	n	Min.	Median	Mean	Max.	SD	SE
FY2012	Freshman	8	100	185	181.2	230	48.8	17.2
	Sophomore	8	100	232.5	226.9	285	59.5	21.0
	Junior	8	75	222.5	208.8	320	87.3	30.9
FY2013	Freshman	19	100	160	166.3	280	44.6	10.2
	Sophomore	19	130	190	191.3	360	54.7	12.5
	Junior	19	150	190	215	365	59.8	13.7

Listening

Entry Year	Grade	n	Min.	Median	Mean	Max.	SD	SE
FY2012	Freshman	8	140	225	226.9	305	46.6	16.5
	Sophomore	8	165	280	271.9	390	71.5	25.3
	Junior	8	105	287.5	271.2	400	87.2	30.8
FY2013	Freshman	19	130	200	213.2	305	47.6	10.9
	Sophomore	19	165	235	259.5	405	65.1	14.9
	Junior	19	210	250	270.5	400	50.4	11.6

References

In'nami, Y., & Koizumi, R. (2012). Factor structure of the revised TOEIC®test: A multiple-sample analysis. *Language Testing*, *29*(1), 131-152.

Zhang, R., & Ishino, F. (2009). *Key to effective english remedial education: Intimation derived from multiple regression*. Paper presented at the AIP Conference Proceedings.

Zhang, R., Zhang, J., Steele, D., & Feng, L. (2015). *Findings about students with great discrepancy at a web-based English remedial program*. Paper presented at the Proceedings of 2014 International Conference on Interactive Collaborative Learning, ICL 2014.

Zhang, S. (2006). Investigating the relative effects of persons, items, sections, and languages on TOEIC score dependability. *Language Testing*, *23*(3), 351-369.

小笠原真司. (2011). 英語習熟度別クラスの効果と G-TELP による成績分析－工学部総合英語Ⅲのデータを中心に－.『長崎大学大学教育機能開発センター紀要』, 2, 9-19.

小笠原真司. (2012). 英語習熟度別クラスの効果的運用について－工学部総合英語Ⅲの G-TELP データによる分析.『長崎大学大学教育機能開発センター紀要』, 3, 9-20.

三枝和彦. (2014). 北見工業大学における習熟度別授業: 英語講読ＩＡ・ＩＢへの導入の成果と課題.『人間科学研究北見工業大学編』, 10, 35-55.

富岡龍明, 髙橋玄一郎, 金岡正夫, 村山陽平. (2011).『鹿児島大学英語教育改革報告書Ⅱ: 平成21年度－平成22年度』.

丸山真純. (2011). 長崎大学経済学部生の英語習熟度 (1) 二つの英語試験と TOEIC 得点の観点から.『経営と経済』, 91(1-2), 93-113.