

The TOEIC IP as a Placement Tool for the Faculty of Regional Studies

Robson, Graham*

Abstract

The TOEIC Test in its various forms has become widely accepted as a tool to measure general English proficiency. In Japan many organizations, including universities, use this test to screen their workers and students. This year the Faculty of Regional Studies in Toyo University has implemented the TOEIC IP test as a placement test for its first years. In previous years the Faculty had used the TOEIC Bridge; an easier and shorter test than the TOEIC IP. This paper looked at the effectiveness of the TOEIC IP Test as compared to the TOEIC Bridge Test as a placement tool. Using statistical analysis, the results showed the TOEIC IP to be more suitable than the TOEIC Bridge as a placement instrument. Therefore, the decision to change to the TOEIC IP can be justified.

Introduction

Over the last two decades in Japan the TOEIC Test, produced and maintained by the ETS organization, has become widely accepted as a tool to measure general English proficiency. As of 2010 six million people in 120 countries worldwide sat the TOEIC Test (ETS, 2010). In Japan, along with 2900 organizations, companies and school that use the TOEIC test, there are around 1.5 million individual test takers (Wikipedia, 2009). There are also many Japanese companies that require prospective applicants to have a score of at least 600 to go with their job application. Universities, too, including Toyo University has embraced the TOEIC IP (Institutional Program) to help make decisions regarding English.

This paper follows on from two previous papers (Robson, 2007, 2011) related to the instalment of a TOEIC IP within the Faculty of Regional Studies. Until 2011 that program had included a TOEIC Bridge Test score in the Tourism Department to stream students into ability-based classes at the beginning of the first year, followed by the TOEIC (IP) Test to stream students into ability-based classes for the second year. Lastly, students would take a final TOEIC (IP) test administered at the end of the second year when the Faculty's current compulsory English education finishes. The first program used the TOEIC Bridge at the beginning of the first year, as it was thought it would be easier than the TOEIC IP and fit the placement decisions it would be used for.

According to ETS (2012), both the TOEIC Bridge and IP are paper-based and have a reading and

* 東洋大学国際地域学部 : Faculty of Regional Development Studies, Toyo University

listening section with 100 questions in total. There are differences in the content, length and difficulty, however. The content for the Bridge is to measure an individual's ability to read and listen to everyday English, whereas the TOEIC IP is to assess the ability of takers to operate in English in business situations. This content for the Bridge appears in five sections, starting with listening, in which there is recognizing photographs, question and response, short conversations, talks and incomplete sentences. In the reading part students construct incomplete sentences and do reading comprehension items. The TOEIC IP is essentially the same, but the skill of text completion is added to the other sections. The lengths are also different; with the Bridge lasting one hour, compared to the IP, which lasts two hours. Lastly, the difficulty levels of the two tests are different. The length of the test, vocabulary, grammar, and text lengths make the IP much more difficult than the Bridge.

Since that first program, the Faculty, through the Foreign Language Committee and, indeed, the university in general, has wanted to make more use of the TOEIC IP. This has resulted in a plan to change the existing placement TOEIC Bridge at the beginning of the first year to the TOEIC IP, meaning that students will take the TOEIC IP at the beginning and end of the first year and at the end of the second year. The test at the beginning of the first year would be a placement test to stream students into classes of similar ability for at least the Tourism Department, making instruction easier. The decision to change all tests to the TOEIC IP reflects the influence of TOEIC in Japan. The Faculty's main reasons for this three TOEIC IP Test system are the following:

- 1) At present only the Faculty of Regional Studies would have such a system, and it would be seen to be setting the standard for other Toyo faculties to follow.
- 2) By the using the TOEIC IP test three times, students can clearly see improvements from both the beginning to the end of the first year, and from the end of the first year to the end of the second year. This may lead to increased motivation, and clearer goal-setting for students.
- 3) This helps the university and Faculty clearly measure the effectiveness of its first two years of compulsory English education. This data can then be compared to other faculties in Toyo, and nationwide averages.
- 4) It could be easier to make a system that requires students to reach a certain TOEIC score goal at the end of their compulsory English education. Without such a goal, they may not receive credits. Conversely, if they do reach that level earlier, they could be rewarded with credits towards their graduation.
- 5) The university or Faculty has a good selling point to appeal to incoming high school students. Many of the top private universities are already using TOEIC IP, and Toyo would fall into this category.

TOEIC IP as a placement tool

The TOEIC Test is described as a norm-referenced test (NRT). According to Brown (2005), these tests are used to classify or group students, measuring overall proficiency by covering a range of questions and skill levels that try to represent the full range of skills needed to master English (in the case of TOEIC IP that means in this Faculty reading and listening only). This ability to group students

is necessary for the placement of students at the beginning of the first year in the Tourism Department. Students will not know what appears on these NRT tests. Indeed TOEIC treats the data for its tests as a well kept secret. Further, students receive their scores as a percentile score, rather than a raw score. This percentile score ranks all students who took that particular form of the test on a continuum of ability that appears as a bell-curve with normal distribution. Most people will be centred around the mean, and as the scores move away from the mean, higher or lower, the number of people getting these scores becomes less. These NRT tests differ from criterion-referenced test (CRT) that measure specific learning objectives within an organization. The CRTs test are much more specific than NRTs because they are designed to foster learning. Students will know what appears on these tests, and would hope to get 100%, with, in the best case, scores being less of a continuum, and more bunched up around the top end (if instructional material has been employed affectively.)

The prevalence of TOEIC, and the clear delineation of NRT and CRT tests would make the decision to use TOEIC, an NRT, as a placement test at the beginning of the first year straightforward. However, as Brown (1995) states teachers and administrators should not expect too much from NRTs. He claims that NRT tests like the TOEFL (another test designed by the same company as the TOEIC for study abroad decisions) serve as good proficiency indicators, their use for placement decisions seems troublesome because of the broad nature of the test (p. 18), normed on a large population, compared to using it within the confines of an organization with considerably less test takers. Admittedly, the TOEFL is much more difficult than the TOEIC, but problems of not being able to separate the abilities of students may occur in difficult tests or test students have no experience of.

Another problem of the overall usefulness of the TOEIC was highlighted by Chapman (2003) who claims that there is a serious lack of credible articles on this test, compared to the TOEFL made by the same organization, other than those commissioned by ETS. He states that this lack of research can lead to mistrust and scepticism of the nature of the test. It means that companies are unable to challenge the validity of the test itself. Therefore, if Regional Studies Faculty students use the test for placement decisions, and score comparisons, the true nature of the scores may be difficult to evaluate.

Lastly, Koelbleitner, Gustavsen and Alderling (2002) compared scores on the TOEIC to an in-house placement test for Japanese university students and found that the TOEIC scores did not match by a long way the scores assigned to students from the in-house test. They also found that students with TOEIC experience have an unfair advantage over students that have not taken the test before because they have exposure to strategies and can budget their time better. They further state that students are rarely able to answer all the questions on the test. This may lead to a hypothetical situation where two students have the same score, but one may have finished, the other may have not. They would be in the same level class, but it would be difficult to state unequivocally they deserve to be studying at the same instructional level.

Despite the above points, TOEIC has enjoyed unparalleled success in Japan over the past 25 years since its inception. A study conducted by Tokunaga (2007) of 126 students from four different universities on the first day of their studies addressed perceptions of the TOEIC Test. Of the total 86%

agreed the TOEIC was important, and 94% said it was necessary for a job in the future. The students also had high expectations that teachers that taught them should know about the test, so they can teach the students appropriate strategies. This result was also mirrored in results from Faculty study (Robson, 2007). In that study, of 606 students from years one, two and three 89% of students agreed that having a good TOEIC score would be important. This result also matched with another result stating that 89% of students would make efforts to improve their TOEIC Test scores in the future. Although blindly following the ideas and perceptions of the students is inadvisable, Tokunaga does say that ignoring them can be a waste of valuable information, p. 258.

This short paper will analyze the ability of the TOEIC IP test to act as a placement test for students in the Department of International Tourism and Department of Regional Studies (at present only the International Tourism Department uses the test for placement decisions). This year is the first year that the Faculty will employ the TOEIC Test from the start of the first year for placement purposes, so close examination of results will shed light on the effectiveness of that decision.

Method

On the second of April 2012 first years from both departments sat the TOEIC IP test. In the Regional Studies Department 206 students sat the test and in the International Tourism Department 234 students sat the test. The results were put through the statistical package SPSS and were analyzed with reference to the functions in that program. The data was then compared to the total data set (two departments combined for a data set of 449) that constituted the TOEIC Bridge score for first years in 2011.

Results

The main descriptive results available through SPSS can be seen in Table one. There are a number of statistics to take note of to gauge the effectiveness of the TOEIC Test as a placement tool. The first is the standard deviation. The next is the Interquartile range. The last two are the z-scores for skewness and kurtosis.

Table one – Descriptive statistics for the total TOEIC Test scores in two departments

	Regional Dept.	Tourism Dept.
Number	206	234
Mean	388.67	385.43
St. Dev.	108.64	110.11
Min. Score	125	160
Max. Score	665	780
InterQRRange	150	136.25
Z-Skewness	-0.13	2.41
Z-Kurtosis	-0.99	1.15

The standard deviation is a measure of how much the scores are distributed either side of the mean (388.67 for the Tourism Department and 385.43 for the Regional Studies Department). The wider and more spread out the scores in the range from minimum (Min score) and maximum (Max. score), the larger the standard deviation will be. There are around three full standard deviations separating the mean from the end of a normal distribution in both directions. Therefore, if you add three standard deviations on to the mean in both directions you will reach somewhere near the minimum and maximum scores. Actually, if the data comes in a normal distribution then 99.74% of the data is explained by six standard deviations. The results confirmed that the data is well spread for both departments. In order to give meaning to the size of the mean and standard deviation, if they are compared to scores from the Bridge Test last year we find that for this year the standard deviation is around 28% of the mean for both departments. Last year this figure was around 12% meaning that the Bridge Test produced more scores located closer around the mean.

The next statistic to consider in placement decisions is the Inter quartile range (InterQRRange) in Table one. This statistic measures the distance between the 75th and 25th percentiles, and like the standard deviation is a measure of dispersion. It differs from the range because it does not take into account outliers in the data that can be described as exceptionally low or high scores that can effectively pull the data in one direction and give a distorted mean. When this year's data is checked against last year's figure we find that in the overall data last year's interquartile range was 16% of the mean, but this year it was between 35-39% for the two departments. This again shows this year's TOEIC IP Test dispersed the scores into more of a normal distribution than last year's Bridge Test.

The last two measurements of skewness and kurtosis explain the shape and dimensions of the normal distribution created by the scores. Skewness is the measure of symmetry in the normal distribution. The more symmetry a bell curve displays above and below the mean the more skewness is to a minimum. From this year's data the significance of the skewness can be obtained by the Z-skewness score, which is the skewness figure divided by the standard error of skewness. For Tourism students this comes out to 2.41 and it is -0.13 for Regional students. The figure should be as close to zero as possible, and figures that are +/- 1.96 standard deviations beyond this are significant. Therefore, the Tourism scores are significantly positively skewed, meaning that more people had scores below the mean. This could be interpreted that the test was quite difficult for the Tourism students. Incidentally, the skewness from the previous year's TOEIC Bridge Test was much higher than this figure with negative symmetry, meaning more people scored above the mean, making the test too easy for the sample. The second dimension measure is kurtosis, which is a measure of how flat the peak of the normal distribution is. The Z-kurtosis figure is again computed by dividing the kurtosis figure by the standard error of kurtosis. Scores for both departments fall inside +/- 1.96 range, meaning that the data is evenly distributed around the mean. The previous year's figure had significant negative kurtosis, with students bunched around the mean.

Along with the statistics above it is worth looking pictorially at the normal distributions created by the data in Graphs one, two and three. Graph one shows the Tourism department. There is a high

grouping of scores around the mean, and at the lower end, giving the high skewness figure, but it has spread students out fairly well. In Figure two, the data shows the Regional Department normal distribution. There is a semblance of normal distribution, with the scores dispersed well higher and lower than the mean. Figure three shows the data from the previous year's TOEIC Bridge Test. The normal distribution has clearly moved to the right, and is high in accumulate scores, represented by the significant skewness and kurtosis figures.

Figure one - Graph for Total IP TOEIC scores for the International Tourism Department 2012

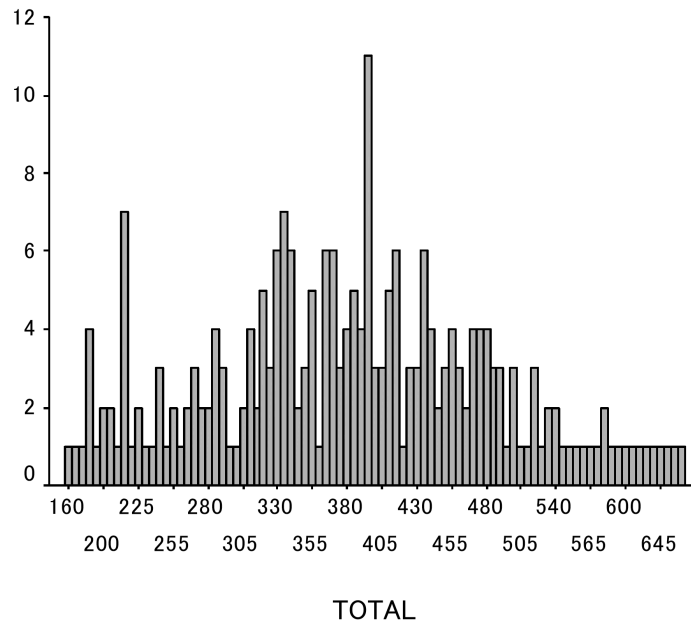


Figure two - Graph for Total IP TOEIC scores for the Regional Studies Department 2012

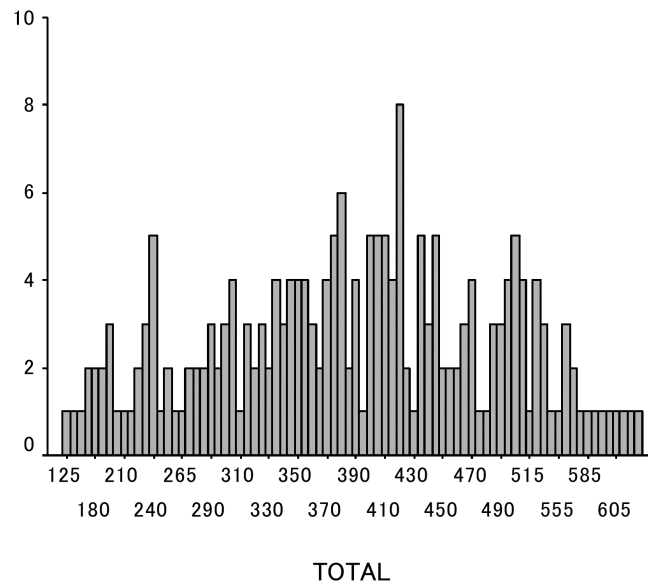
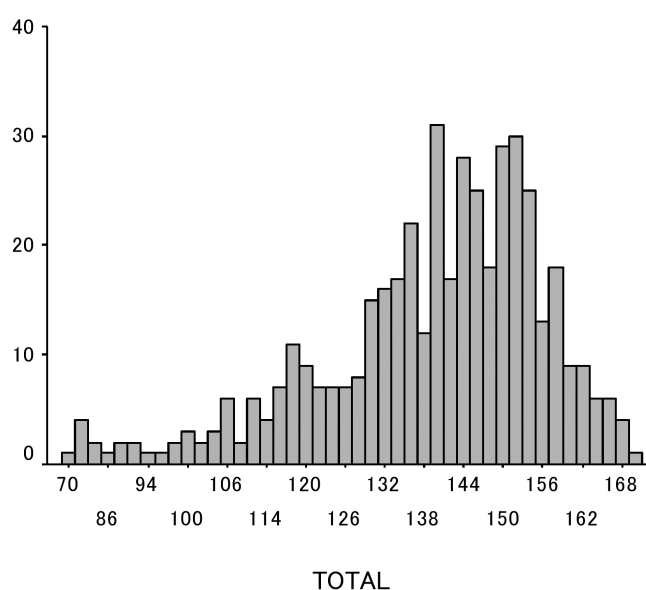


Figure three - Graph for Total Bridge score all students (Regional and Tourism combined) 2011



Conclusion

A decision to use the TOEIC IP test three times from the beginning of the first to the end of the second year was implemented in 2012. The reasons for this include an ability to measure student and university English performance. Part of this decision means that the TOEIC Bridge that had been previously used for class placement decisions would be replaced by the TOEIC IP to make those same decisions. At present only the International Tourism Department uses the scores for placement decisions, but data was collected from both departments on the first try for this new system and compared with data from the previous year's TOEIC Bridge class placement test. By comparing averages like the mean, standard deviation and interquartile range, and by assessing the ability of the data to form a normal distribution, and indeed, how that normal distribution is dispersed through measures of skewness and kurtosis, it was found that the TOEIC IP was difficult but could act as a useful placement decision tool. A normal distribution in the 2012 data was much better at discriminating student abilities for placement than the previous year's TOEIC Bridge Test. For the future, it is worth looking more at how the Faculty's student TOEIC scores change under different circumstances, including instruction, self access and study abroad.

Bibliography

Brown, J. D. (2005). *Testing in Language Programs: A Comprehensive Guide to English Language Assessment*. New York: McGraw Hill.

- Brown, J. D. (1995). Differences between norm-referenced and criterion-referenced tests. In *Language Testing in Japan*. (Eds.) James D. Brown & Sayoko Okada(pp. 12-20). Japan Association for Language Teaching, Tokyo.
- Chapman, M. (2003). Toeic: tried but undertested. Shiken: JALT Testing & Evaluation SIG Newsletter. JALT, Tokyo. Accessed on 22nd March from http://jalt.org/test/cha_1.htm.
- ETS (2012). ETS Website. Accessed on Mar. 21st from <http://www.ets.org/>
- Koelbleitner, C., Gustavsen, E. & Alderling, M. (2002). An examination of the proposed use of TOEIC at Asia University. Asia University. Accessed on 22nd March 2012 from http://www.asia-u.ac.jp/cele/cele_assets/An_Examination_C.Koelbleitner.pdf
- Robson, G. (2007). English language survey in Toyo Tourism Department. *Journal of Tourism Studies*. Vol. 6, pp.103-130.
- Robson, G. (2011). Analysis of TOEIC scores of Tourism Dept. *Journal of Tourism Studies*. Vol. 10. pp. 139-150.
- Official Toeic Website (2012). Accessed on 22nd March 2012 from http://www.toEIC.or.jp/toEIC_en/pdf/data/TOEIC_DAA2010.pdf
- Wikipedia (2009). TOEIC. Accessed on 22nd March 2012 from <http://en.wikipedia.org/wiki/TOEIC>

国際地域学部クラス分けテストとしての TOEIC IP テストについて

Robson, Graham

要約

種々の形態を持つ TOEIC テストは一般的英語能力を測る手段として広く受け入れられている。大学を含める日本の多くの機関において、自身の、また学生の英語能力審査にこのテストが使用されている。今年、東洋大学国際地域学部では TOEIC IP テストを 1 年生のクラス分けテストとして使用した。去年までは TOEIC IP テストより易しい TOEIC Bridge テストを使用していた。本稿では TOEIC IP テストの有効性を TOEIC Bridge テストと比較しながら調査した。調査結果を統計処理し、TOEIC IP テストは TOEIC Bridge テストよりクラス分けテストとしてより有効である、という結果が得られた。したがって、TOEIC IP テストへの変更は適正であった。

