Who and What Should Place Students in High-Stakes English Classes?

―Students Opinions and Results of IELTS and TOEIC International Testing―

いかにして高スキル英語クラスを編成すべきか?
―決定主体･基準と達成結果の相関性―

リンチ　ギャビン
Gavin LYNCH

〈Abstract〉

This research paper investigates what criteria should be used to place students appropriately in first year university English classes and also who should be involved in placing them, while investigating the ramifications of including students' opinions in their class placement. The research looks at data from two consecutive academic years, with the first group (Class A) being placed into a high-expectation class based on their performance on a pre-entry TOEIC test, while the second group (Class B) in the following year was chosen for such a class using their pre-entry test scores combined with the students' willingness to join the class (instead of an alternative, regular class). The groups were comparable in terms of pre-entry test-performance data. The students' performance in TOEIC and IELTS over one academic year was recorded, and it was found that both groups increased their TOEIC and IELTS scores over the year. Results showed that, over the course of one year, the second group performed better than the first group in TOEIC and that the second group showed a greater initial surge in performance in the IELTS. That gain was generally maintained over one year, although the gap between the groups narrowed slightly.

The study influenced the creation of a number of hypotheses, calling for further research into consequences of self-selection, test comparability, how knowledge of society and general knowledge influence English test results, the value of a score increase when taking the previous value (base value) into account, and performance over the longer term.

〈Keywords〉

language class placement, student voice, international testing

1　Introduction: Who and What Should Place Students Appropriately in English Classes?

In many universities in Japan, placement of students into English classes deemed appropriate for them are done using test scores such as those gained from placement tests. However, according to Westrick (2005), it is often the case that we do not look beyond those scores nor question their reliability, and instead "accept raw and converted scores as perfect reflections of students' abilities." This results in decisions that are "made without any second thoughts." Furthermore, Riches (2006) tells us that tests, even those such as globally accepted TOEFL tests or a test common in Japan, the TOEIC, "tell us nothing about how to appropriately place students in a course" (emphasis added) if other background data is not taken into account. The makers of those tests, ETS, admit that those tests do not actually measure "the fluency or communicative ability of the test taker at the time" they take the test (Chapman referenced in Riches, 2006), yet "ETS allows this misunderstanding to persist."

Colorado (2007) suggests that the problem of how to place students appropriately can be better managed by looking at a wider variety of data. He suggests including information such as students' language and literary

proficiency levels in their first language, knowledge and teaching strategies of each prospective class teacher, and expectations that the school or university has for different class levels. He goes on to add that information from the home should be included, such as by asking parents for information. All this, however, omits one very important source of information – the students themselves. Kellett (2009) tells us that young people have a valuable insider view of their world to which adults do not have clear access, no matter what tests we use. Lynch (2014) argues that giving students a "voice space", meaning the chance to have a say in their own education, can be motivational and confidence-improving as it is an empowering process for participants (Jacquez et al., 2013), allowing "students to control the resources that affect their lives" (Langhout and Thomas, 2010). However, according to Kellett (2009), this is seldom done in practice.

The above shows that, while many universities focus on the 'what' (such as what test, what score and what interpretation) when deciding on how to place students into appropriate classes, focus should also be put on the 'who' (who to include in decision making, who teaches, and who holds responsibility). That is to say, both 'what' and 'who' have a role to play in class placement decision making. Another factor is, of course, time and financial restrictions faced by every university. This research attempts to pull all these factors together.

2 Research Methods

2-1 Teacher Duty and Student Background

The research methods followed were not all planned from the beginning; rather they were developed by class observation (usually while teaching), brainstorming methods to improve the situation, implementing such methods, and observing results. The people involved were also teachers at a university so, as "it is a duty of a teacher to ensure ideal atmosphere" (Sengottuvel and Aktharsha, 2015), they were compelled to dynamically change things about the class to reflect new experience or current student needs. Because of this, there was no "control group", rather the research population consisted of a group of students who were taught in the 2013 Japanese academic year (from April 2013 to February 2014), and another group of students taught in the 2014 Japanese academic year (from April 2014 to February 2015). These, for the purposes of identification, are named Class A and Class B, respectively. An explanation of each class is given later.

Both groups of students were selected from a pool of about 450 students who entered a Japanese university, a number which represents the entire population of first year students in a department, and who are placed in an appropriate English class. They were economics or business majors, and had to take first year English classes as part of the university's set of obligatory classes. Class A was comprised of the students who scored the highest in a pre-entry placement test (TOEIC). Class B, on the other hand, was made up of students who scored a minimum acceptable level in their pre-entry placement TOEIC test AND were self-selected by answering a questionnaire. For the purposes of this study, only students who took tests at the beginning and end of the research period are included in the number counts in this paper (thereby excluding drop-outs, students who didn't turn up for tests, etc.).

2-2 Class A Composition

Class A was divided into two classes (A1 and A2), and totaled 52 students (again, just counting those students who took placement tests AND exit tests). The students were divided evenly in a pseudo-random way into two classes to facilitate lower teacher-student ratios that have been shown to allow "for greater flexibility for innovation in the classroom" (OECD, 2011).

Both classes were taught a general skills class twice per week by the same teachers, Teacher X and Teacher Y. Teacher X was a native speaker of English, and had completed all his education (to masters level in education) in Europe, mostly in Ireland. Teacher Y was Japanese, and had completed all his education (to PhD level in linguistics) in Japan. Teacher X used one weekly class (per group of 26 students) to teach mainly speaking and listening skills, while Teacher Y used the other one-weekly class to teach mainly reading and writing skills. The word mainly is used as the teachers agreed that to clearly and cleanly separate the skills would be unnatural in language pedagogy, therefore it would be sensible to allow either teacher some leeway to venture into the skills area taught by the other. The students also had the option of taking other classes outside the obligatory English class, and were expected to take a presentation/discussion class and an academic writing class.

The students were between 18 and 19 years old, and were all from the Hokuriku area of Japan. As mentioned

earlier, all were economics or business majors. About 450 students took a pre-entry TOEIC test and, from these, the 52 Class A students were chosen as they had scored the highest. The placement was decided by the university, i.e., students were not given a choice in this placement. Their test score data is shown below, in Table 1. N refers to the number of students, and scores shown are TOEIC points (max=990).

n	52
Overall Score (mean)	393
Median Score	390
Minimum Score	250
Maximum Score	545
Standard Deviation	73
Coefficient of Variation	18.54%

Table 1: Class A Pre-Entry TOEIC Score Data

2-3 Class B Composition

Class B consisted of one class of 19 students (again, just counting those students who took placement tests AND all exit tests while staying in the class). The teaching hours for Teacher X and Teacher Y were reduced to only one class for these students, so there was no subdivision of Class B (nor was it necessary). The teaching roles and methods teachers (Teacher X and Teacher Y) did not change greatly. Students in Class B were selected in a different way than Class A, in that the opinions of potential Class B students were taken into account. However, before explaining that, it is important to consider the scenario that would have occurred if the number of students in Class A had been the same as the number in class B. Table 2 shows the scores of the top 19 students of Class A (Class A Top-19).

n	19
Overall Score (mean)	453
Median Score	450
Minimum Score	370
Maximum Score	545
Standard Deviation	57
Coefficient of Variation	12.57%

Table 2: Class A Pre-Entry TOEIC Score Data, Top Nineteen Students (Class A Top-19)

The data in Table 2 shows higher scores, with a lower standard deviation and coefficient of variation, as should be expected when selecting a smaller group under the condition of students scoring in the top nineteen in their pre-entry (placement test) TOEIC.

Class B were selected based on their pre-entry TOEIC score, as well as the existence of a student-completed entry form in which they had answered questions and given logical reasons for wanting to be part of the class. There was also an interview where the students were given a chance to explain why they wanted to join the class and, after one semester, the students' motivation was checked at another interview. There were no credit or monetary incentives to joining the class, and students were made aware of the higher than average workload (compared to other classes) involved in the class. Initially, a minimum pre-entry TOEIC score of about 350 points was discussed, making it within 6% of the minimum pre-entry TOEIC score of the previous year's class. Eventually, however, a minimum score of 320 points was accepted. The eventual requirements for entry, including contents of some questions, are listed in Table 3.

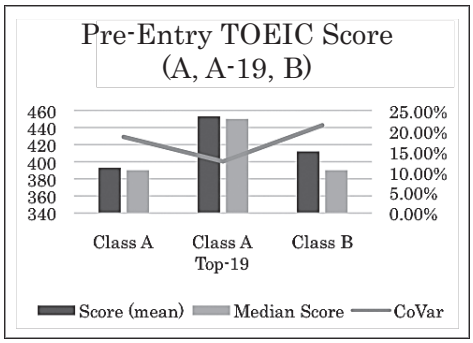
TOEIC Pre-Entry Min Score	320
Application Form	○
Desire to Study Abroad	○
Logical Reason for Joining	○
Willingness to Work	○
Interview	○

Table 3: Requirements to be accepted to Class B

In reality, almost all students who applied were accepted, with the main reasons for students not being accepted being a low pre-entry TOEIC score or not being able to write a logical reason (in Japanese) why they wanted to join. The logic of their reasons were not the only basis - a student only had to make some effort in writing sentences. Having students write a reason for joining (instead of just ticking a box) may have led to some students not bothering to apply, in effect screening out students who might claim that they would study but, in reality, were not highly motivated students (as having to write a form and think about reasons requires effort). The nineteen students who were accepted to Class B had the following characteristics (Table 4).

n	19
Overall Score (mean)	412
Median Score	390
Minimum Score	320
Maximum Score	615
Standard Deviation	88
Coefficient of Variation	21.38%

Table 4: Class B Pre-Entry TOEIC Score Data



Graph 1: Class A, A Top-19, B Pre-Entry TOEIC Score Data Mean, Median, CoVar

3　Class A and B Composition and Analysis

3-1　Selecting Suitable Groups for Comparison and Analysis

It can be seen that Class B had a very high coefficient of variation (CoVar) in the students' pre-entry TOEIC scores. In other words, a wide variety (ability-wise) of students joined the class. The data to be compared are not the minimum or maximum scores, as one student's result could change these figures. Instead, we should compare the mean (average), the median and the variation (CoVar) of the data. Comparing the three data sets of 1. Class A (n=52), 2. Class A's top nineteen students (n=19), and 3. Class B (n=19), gives us the following data, in Table 5. Standard deviation is left out, as it is included as part of the CoVar calculations.

	Class A	Class A Top-19	Class B
n	52	19	19
Score (mean)	393	453	412
Median Score	390	450	390
CoVar	18.54%	12.57%	21.38%

Table 5: Class A, Class A Top-19, Class B Pre-Entry TOEIC Score Data

The mean, median and CoVar data in Table 5 data are graphed below in Graph 1 (with the mean and median referencing the left Y-axis, and the CoVar referencing the right Y-axis) to allow easier comparison.

It can be seen that the composition of Class B is more similar to Class A than to Class A Top-19, showing that a composition of Class A and Class B is the way to proceed. On the other hand, comparing Class A-19 with Class B would not be a valid comparison, as the test data (and, therefore, the students) are not similar.

The above avoids a potential logic pitfall, which is that of comparing Class B with the same number of students (the top nineteen students) in class A due to the misconception that those students would be the best match for comparison. Instead, it would be more scientifically accurate to compare the nineteen students of Class B with the fifty two students of Class A. The median scores of both groups are same, and the mean and CoVar are similar.

3-2　Background of Factors for Comparison—Class A and Class B

Class A and Class B have the same median score, with the mean differing somewhat due to different minimum and maximum scores. The teachers (Teacher X and Teacher Y) are also the same. The question of why it was decided to make such an effort to change the selection process could be asked, especially in the light of the class pre-entry TOEIC scores being approximately the same. Such a decision connects to the discussion in the introduction of giving students a say in their own education – allowing them to become stakeholders and giving them responsibility for their successes or failures (Lynch and McKeurtan, 2011). It was noticed that some students in Class A complained about the high homework load, and some would not complete their assignments to a suitable (expected) quality. Furthermore, a lack of motivation also seemed to pervade some parts of the Class A atmosphere, although it should be made clear that, overall, Class A worked very hard and were, in general, serious about their studies. In the opinion of the teachers, they worked harder than any other of the English classes at the university. However, it was thought that making an opt-in system (and, as a consequence, a default opt-out system) to the classes taught by Teacher X and Teacher Y would result in students who signed up be less likely to complain

and more likely to think about what they wanted to achieve, leading to greater motivation, in turn connecting to improved ability and test results. The teachers decided to test this hypothesis as an educational experiment.

4 Results

4-1 Performance of Class A and Class B

The performance of the students was shown in many ways, and included an enhanced awareness of the English language and related culture and a greater willingness to try to improve their output skills such as writing and speaking, and not only focus on input skills such as reading and listening. However, such qualitative data was not collected in a scientific way. Quantitative data, however, was collected and includes increases in TOEIC scores, data from two IELTS tests, and in-class tests including end-of-semester examinations. As the latter were not standardized and, therefore, may be subjective, this paper discusses the results from the two standardized tests, TOEIC and IELTS.

Both Class A and Class B took a TOEIC test before embarking on one year of regular English classes at university. The classes were twice a week, as explained earlier. They were supplemented by other classes such as a weekly writing class and a presentation class, also taught by Teacher X and Teacher Y. The educational circumstances were basically the same for Class A and Class B.

4.1 TOEIC Results

Class A took a TOEIC test just before entering university in March 2013, and took one again at approximately the end of one academic year, in January 2014. The TOEIC they took is the standard TOEIC, which includes listening and reading sections only. The results (average of the whole group, n=52) are shown in Table 6 (note: some rounding errors may occur).

	Mar-13	Jan-14	Increase
Listening	229	257	28
Reading	165	192	27
Overall	393	449	56

Table 6: Class A TOEIC Test Scores, One Academic Year

It can be seen that there was an equal increase in both listening and reading skills. Similarly, Class B took a TOEIC test in March 2014, and again in January 2015. Their results are shown in Table 7.

	Mar-14	Jan-15	Increase
Listening	225	281	56
Reading	187	213	26
Overall	412	493	81

Table 7: Class B TOEIC Test Scores, One Academic Year

It can be seen that both Class A and Class B increased their TOEIC scores. The increase in reading ability (as judged by TOEIC) was similar for both groups, while the increase in listening scores was greater for Class B than for Class A. Overall, Class B increased their TOEIC scores more than Class A.

4.2 IELTS Results

The IELTS is a four-skills examination and "is the world's most popular high stakes English-language test for study, work and migration" (IELTS.org, 2015). It offers the advantages of allowing the students (and university) to be evaluated across four skills, and gives the students a globally-recognised certification. The university uses the IELTS as an evaluation tool for deciding if scholarships can be given to students who wish to study abroad long-term. In that sense, it is a high stakes examination for students who wish to study abroad while in university. To facilitate greater uptake of the examination, and to reduce the burden on students taking it, approximately 80% of the examination cost was paid by the university (of the examination cost of about 25,000 yen, the university paid about 20,000 and the students paid 5,000 yen).

The students took their first IELTS examination in June, two months after entering university. Then, it was taken again in the following January. A small number of students sat an additional examination at some point between those dates, but that data is not included in this study. A number of students were not able to take both examinations (for various reasons, none of which are related to the running of the classes). Due to this, the number of students in Class A taking both IELTS examinations was 48, while the number in Class B was 18. The IELTS results of Class A are shown in Table 8. Note that the individual skills added up and averaged do not exactly produce the overall result due to individual averaging of rounding results.

	Jun-13	Jan-14	Increase
Listening	4.24	4.2	−0.04
Reading	4.25	4.64	0.39
Writing	4.68	4.25	−0.43
Speaking	3.9	4.07	0.18
Overall	3.97	4.38	0.41

Table 8: Class A IELTS Test Scores (n=48), Approx. One Academic Year

It can be seen that there was an increase in reading and speaking scores, while listening and writing decreased. However, the overall score increased meaning the rise in reading and speaking was greater than the fall in listening and reading scores. Similarly, Class B took their IELTS tests in June 2014, and again in January 2015. Their results are shown in Table 9.

	Jun-14	Jan-15	Increase
Listening	4.33	4.14	−0.19
Reading	4.58	5.06	0.47
Writing	3.78	4.42	0.64
Speaking	3.81	4.5	0.69
Overall	4.17	4.53	0.36

Table 9: Class B IELTS Test Scores (n=18), Approx. One Academic Year

Class B increased its overall average reading, writing and speaking scores, as well as its overall IELTS scores. The listening score for Class B decreased.

Looking at the overall IELTS score results for Class A and Class B, it could be seen that the latter scored higher, and managed to maintain a higher score for the year. However, the progress (or lack thereof) in each skill needs to be analysed. This will be done as continued research in the future.

5 Discussion and Conclusions

Two classes, Class A and Class B were compared in this research. Class A were placed in a suitable class based on their pre-entry TOEIC scores only, while Class B were placed based on a combination of pre-entry TOEIC scores and their personal volition. The pre-entry TOEIC scores for both classes were similar and comparable. Over the course of one year, Class B outperformed Class A in terms of overall TOEIC score. Regarding IELTS scores, the first IELTS test was taken more than two months into the academic year, and Class B performed better than Class A overall. Subsequently, Class B managed to maintain its IELTS lead over one academic year.

The data analysis results indicated the following four conclusions:

Conclusion 1

University students can be expected to increase their TOEIC scores in listening and reading scores significantly over one academic year of a limited English programme. (Class A: +56 points, Class B: +81 points).

Conclusion 2

University students can increase their IELTS overall score over less than one academic year. (Class A: +0.41 points, Class B: +0.36 points).

Conclusion 3

When student choice is taken into account in class placement, students perform better overall in TOEIC over one academic year compared to students who were placed according to test scores only. The difference between the groups was attributed to an increase in listening scores.

Conclusion 4

When student choice is taken into account in class placement, students show a greater initial surge in performance in the IELTS, and perform better overall in IELTS over one academic year compared to students who were placed according to test scores only. The major difference between the groups was in the score of their first IELTS test, just over two months after entering university, suggesting a large initial educational results surge in the Class B group, with the initial gap being maintained but narrowing slightly after one year (see hypothesis 1, below).

The data suggested and influenced the following five hypotheses:

Hypothesis 1

Students who self-select themselves for classes advance their skills the most during the first two months (or one quarter) of an English programme.

This hypothesis comes from comparing the initial TOEIC scores of the classes (A and B), and the IELTS scores after just over two months. While the groups had similar TOEIC scores, the IELTS scores for Class B were significantly higher than those for Class A.

Hypothesis 2

The listening skills tested in TOEIC and IELTS are not comparable.

This hypothesis comes from comparing the TOEIC listening scores and IELTS listening scores of both classes. Over one academic year, Class B performed better than Class A in the TOEIC listening test (from March to January), but performed more poorly in the IELTS test (from June to January).

Hypothesis 3

Increasing IELTS skills requires knowledge of society and general knowledge. Therefore, test scores can fluctuate depending on what is asked and how it connects to the students' background knowledge.

Class A students showed decreases in IELTS listening and writing skills, while Class B showed a decrease in IELTS listening skills. Teachers X and Y were confident that students' skills had improved over the year, and this opinion was backed up by the improvement in the TOEIC scores. It could be that the students do not have enough knowledge to answer questions accurately, especially those questions that require background knowledge or inference. It may be possible that students may have made similar mistakes in the test even if it had been conducted in their mother tongue.

Hypothesis 4

Students who score highly in IELTS find it more difficult to increase their score by the same proportion as those who don't score as highly.

Class A increased their overall IELTS score by 0.41 to 4.38 from a base of 3.97 points. On the other hand, Class B increased their overall IELTS score by 0.36 to 4.53 from a base of 4.17 points. The hypothesis is that it is harder to increase a score from an existing higher level, especially for students who started out at the same TOEIC score starting point.

Hypothesis 5

In the longer term (more than one academic year), students who are given the chance to participate in their class placement process will outperform students who were not, in terms of IELTS scores (including listening scores).

This hypothesis is based on the higher motivation shown by Class B students towards English education. It is expected that, over time, this motivation will manifest itself in a larger gap in English ability (as shown in IELTS scores) than students who were not given the additional stimulus of having responsibility for their choices.

The conclusions above show that English education in university is worthwhile, as students continue to improve their abilities after entry. Furthermore, they show that students who are given a voice in their own education perform better than if they had not been asked their opinions. The number of hypotheses show that further analysis needs to be done to more clearly understand the situation of English education in university in terms of TOEIC and IELTS scores, student ability, and suitable placement into classes.

References

Chapman, M. (2003) 'TOEIC: Tried but Undertested', *Shiken: JALT Testing & Evaluation SIG Newsletter,* vol. 7, no. 3, pp. 2-5.

Colorín, C. (2007) *Placing English Language Learners in a Program of Instruction* [Online]. Available at http://www.colorincolorado.org/educators/assessment/placement/ (Accessed 17th March 2015)

IELTS.org (2015) About the IELTS Test [Online]. Available at www.ielts.org/test_takers_information/ (Accessed 17th March 2015)

Jacquez, F., Vaughn, L. M., Wagner, E. (2013) 'Youth as Partners, Participants or Passive Recipients: A Review of Children and Adolescents in Community-Based Participatory Research (CBPR)', *American Journal of Community Psychology,* vol. 51, no. 1-2, pp. 176-189.

Langhout, R. D., & Thomas, E. (2010) 'Imagining participatory action research in collaboration with children: An introduction', *American Journal of Community Psychology,* vol. 46, no. 1-2, pp. 60–66.

Lynch, G. (2014) 'Giving Young People a Voice Through Project Work: Hints for Japanese Schooling from the Singapore Model', *Kanazawa Seiryo University Human Sciences,* vol. 8, no. 1, pp. 83-86.

Lynch, G., McKeurtan, M. (2011) 'Motivation in the Curriculum: Effects of Making Students Stakeholders in Tertiary Level English Classes', *20th MELTA International Conference Proceedings,* pp. 386-397.

OECD (2011) Education at a Glance 2011: OECD Indicators [Online]. Available at www.oecd.org/education/skills-beyond-school/48631582.pdf (Accessed 17th March 2015)

Riches, D. (2006) Innovating English Language Education by Looking beyond the Syllabus of the Typical Japanese University English Program [Online]. Available at http://www.seijo.ac.jp/pdf/fasiv/1-2/Riches.pdf (Accessed 17th March 2015)

Sengottuvel, A., and Aktharsha, U. S. (2015) 'Teacher Effectiveness and Professional Competency in School Education', *Journal Impact Factor,* vol. 6, no. 1, pp. 181-190.

Westrick, P. (2005) 'Score Reliability and Placement Testing', *JALT Journal,* vol. 27, no. 1, pp. 71-94.