

Proficiency Scores and Classroom Performance

Richard BLIGHT, Meredith STEPHENS

Introduction

In recent years, students' performance on various high stakes external exams (such as the TOEIC, TOEFL, Eiken) has been strongly emphasized at Japanese universities. There appear to be various reasons for this trend. Firstly, TOEIC (et al) test scores provide a clear differentiation between students' language levels, which fulfills an essential purpose for employers needing to quickly identify the most able recruits. They also assist to provide a form of institutional accountability (as well as individual merit) in terms of simple numerical scores. The adoption of high-stakes exams at colleges also tends to conform to the cultural direction by overtly valuing the formal accreditations that underlie such test administrations.

Yet this direction causes many concerns for teachers, who are frequently aware of what can be significant discrepancies between students' TOEIC scores and their observed performance in the classroom. In order to investigate these concerns, we conducted a classroom-based research study that compared a student's TOEIC score against results on three standard classroom tests. We wanted to investigate to what degree the results on the tests provided a similar (or different) measure of a student's English language ability.

Review of Current Issues

Using external exams such as the TOEIC to evaluate students' performance on university courses is controversial from a number of points of view. Perhaps central to this debate is that the TOEIC is a 'proficiency test' (i.e., it measures a student's general language abilities), while performance is traditionally measured in terms of 'achievement tests', which focus (more narrowly) on the content studied during classes.

Advantages of a TOEIC Focus

Edwards (2006) argues that many general benefits are

gained by incorporating forms of standardized testing in university curricula. Standardized tests (which include the TOEIC test) "provide a clear focus and specific goals for students to strive towards" (p.13). Such tests provide a common direction for teachers who may previously have been pursuing individual preferences for teaching content and testing systems. The clear focus that standard tests provide also tends to foster enhanced teamwork among teaching staff, which is another factor that has a demonstrated correlation with achievement gains. In addition, the transparency provided by numerical achievement scores allow for the adoption of relevant targets. Hence Edwards (2006) argues that the adoption of forms of standardized testing in university education importantly provides towards the critical goal of increased accountability, both on the part of the teachers and students.

Another major advantage of using the TOEIC over traditional university in-house exam systems is that the TOEIC gives equal weighting to listening and reading skills (each section comprises 50% of the final grade). Gaining a good TOEIC score hence requires students to improve their listening skills, a critical area of language acquisition that has traditionally been either undervalued or neglected. A significant washback effect from university testing back to the high school system has been previously reported, so listening ability would most likely also tend to become more seriously fostered in the preceding years of education.

Also important is that the incorporation of TOEIC testing provides for the continuation of a focus on test score performance. Such a direction has been heavily stressed in the preceding high school years, but tends to become less relevant and meaningful in the non-standardized testing environments currently found at many universities. Indeed, the students have been educated in a system that strongly emphasizes individual point scores to such an extent that they can easily underestimate the value in pursuing alternative goals (such as a focus on language fluency), particularly if

their performance cannot be simply represented in terms of numerical point scores.

Finally, TOEIC tests provide all students with an opportunity to demonstrate their individual skill and study application. The TOEIC puts all students on a 'level playing field', regardless of the name, ranking, or status of the tertiary institution they attend. Students can identify and develop various areas of their skills in a range of language contexts and structures. In this way, the TOEIC provides an important chance for students to achieve and demonstrate individual performance, rather than claiming upon the status of their university.

Disadvantages of a TOEIC Focus

Foreign language teachers working in Japan have cautioned against problems associated with placing too much emphasis on students' test scores (Guest, 2005). They argue that this direction ignores other vital language goals, and in particular fails to address the students' need to develop basic communication skills (Lyster, 2005). They see the preoccupation with TOEIC scores as perpetuating fundamental mistakes made in secondary curriculums, where the memorization of grammatical patterns and difficult vocabulary is typically prioritized over oral and written communication tasks. The teachers describe students who achieved good TOEIC scores, but who demonstrated difficulties holding the most basic of conversations. Brown discusses this paradoxical situation further:

If we are serious about teaching students to actually communicate in a second language, we must absolutely stop limiting ourselves to the relatively "easy" testing of their second language knowledge (as with grammar and vocabulary tests), or their receptive language skills (as in listening and reading tests). We will have to face the many challenges of testing their performance in written and oral communication. (Brown in Sunga 2003, p. 14)

One interesting study was conducted by Robb (1999), who investigated the effectiveness of TOEIC preparatory courses at a Japanese university. He followed two student groups (English majors vs. non-English majors) who took three different classes: TOEIC Preparation, Business English, and Four Skills English. At the end of the year, he checked for improvement in the students' TOEIC scores (in both the TOEIC reading

and listening sections). The hypothesis that there would be equal gains by students in all three groups on account of the additional English study (regardless of the course content) was confirmed for eleven of the twelve study groups, with the single exception being the non-English majors in the TOEIC Preparation class, who showed more substantial improvement on the TOEIC Reading section. Hence the benefits of the Test Preparation course were limited to assisting non-English majors to improve their reading score, and provided no additional benefits for English major students or on the listening section of the test, beyond the general benefits provided by other forms of English study.

'Putting the Horse before the Cart'

Zemach (2005) expresses frustration at test results (on the TOEFL) being considered more important than mastery of the language. Tests provide insight into a student's language abilities, but are a form of temporary measure of the student's learning progress, and never comprehensively represent the state of the student's learning. Indeed, we would describe this situation of prioritizing an inaccurate, temporary measure of learning over the process of learning as failing to "put the horse before the cart." Zemach consequently recommends offering TOEFL classes to "students who have already taken classes or demonstrated proficiency in other key skills" (2005, p. 17). She argues that a solid foundation of communicative skills is necessary to achieve good test scores, and certainly before students attempt to memorize less common points of grammatical distinction.

Similarly, Garant (2003) and Kobayashi (2001) indicate that in Japanese high schools the students' test scores are typically considered more important than their communicative skills. Garant cites the case of communicative English classes being cancelled at the end of Junior High School because of an approaching High School entrance examination. Furthermore,

Japanese teachers expressed that the goal of English education was to see that the students passed their English tests. The ability to communicate was seen as very important, but only if it could be accomplished without interfering with the examination process. (Garant, 2003, p. 123)

A further concern then becomes that the TOEIC

exam (taken at the end of university) may end up being assigned the same artificial significance as the university entrance exam (taken at the end of high school; see Matsuyama, 1978). So the problem of mistaking the fundamental goal of language acquisition with instead, the goal of achieving a good score on a test, looks set to be extended into the university years. Further, a common flaw of the test preparation courses is that students are repeatedly asked to tackle questions that are way beyond their current language level, and as a result they become discouraged when they consistently get the wrong answers. Similarly, the range of vocabulary that they are presented with in the TOEIC far exceeds their vocabulary knowledge. The 'test preparation' approach of persistently exposing students to materials at too high a difficulty level has a strong negative impact on motivation levels, and yet (as a reward for effort) provides only limited benefits. This approach also intuitively encroaches on common principles of learning, which generally assume that learning occurs best within protected stages of development. Frequently, 'over-exposure' to too much information tends to overload students (leading to processes of mental shutdown) rather than leading to the desired massive absorption of new information. Rather, effective learning requires measured increases in difficulty level, combined with repeated recycling and practice with the material.

Test preparation courses typically also involve students in practicing masses of previous test questions. But TOEIC questions are always 'narrow' in focus and lacking a general context; the common mistake then becomes for teachers to practise volumes of test questions 'in isolation,' rather than to master broad reading skills, which would ultimately lead to higher TOEIC scores. Class time is hence devoted to intensively studying 'form' rather than 'meaning.' By contrast, Krashen argues that 'meaning' is more fundamental to learning a language: "Language acquisition only happens when we understand messages" (2004b, p.1). Krashen also suggests that teaching grammatical rules should be delayed until students have reached an advanced level: "I would first give acquisition a chance, and then use conscious knowledge to fill in some of the gaps" (2004b, p.1). Unfortunately in Japanese education, the pressure of pending exams means it is very difficult for teachers to follow Krashen's advice to "give acquisition a chance".

Combining TOEIC Preparation with Modern EFL Pedagogy

There are alternative ways to boost TOEIC scores, rather than the methods used in test preparation courses. Krashen, for example, argues that developing reading skills promotes broad language improvement:

When children read for pleasure, when they get "hooked on books," they acquire, involuntarily and without conscious effort, nearly all the so-called language skills many people are so concerned about: They will become adequate readers, acquire a large vocabulary, develop the ability to understand and use complex grammatical constructions, develop a good writing style, and become good (but not necessarily perfect) spellers. (2004a, p.149)

Many of the skills that Krashen relates to 'good reading' are also required to perform well on the TOEIC test, including vocabulary knowledge and grammatical competence. Hence one way to facilitate an improvement in TOEIC scores is to get students actively involved in an extensive reading program. Fluent reading without recourse to a dictionary is necessary to enjoy reading, and the enjoyment of reading is necessary in order to persevere with the demanding task of learning. Harold Palmer, who was invited to Japan by the government in 1922 to promote TEFL, considered that "reading should be as fluent or natural as speaking or hearing, not the word for word puzzling out of meaning" (Masukawa, 1978, p.246). Unfortunately, this warning (originally made in pre-war Japan) is still equally as relevant today. A perusal of the questions in any TOEIC exam reveals the kind of written language that is likely to be found in newspapers. This language is considerably more sophisticated than the language used in graded readers. In order to succeed in the TOEIC exam, students must be familiar with the language that is typically found in newspapers and television broadcasts. However, in order to attain this level, students must first progress through the levels of difficulty such as are found in graded readers series. Novice readers are likely to waste their time launching straight into difficult newspaper English, without first developing a solid foundation for reading skill development.

Study Methodology

This study compares the results of 21 students (with two classes per week), who were being taught according to three distinct themes: A TOEIC preparation course, the popular *Let's Talk!* textbook (Leo Jones, CUP, 2002), and readers in the *Oxford Graded Reader* series (OUP, Level 1). The alternative tests used were published resources associated with the textbook materials. These tests covered vocabulary and listening dictations from the textbook materials.

This study examines the students' relative performance on the TOEIC course, which includes a wide range of possible test content, and on the OUP / CUP courses which feature more narrowly restricted test content because the associated tests aim to measure learning just on the textbook materials. In particular, we were interested to investigate how the students' performance on the various tests compared. Would students with high TOEIC scores also get high scores on the other tests? There seemed to be at least a good chance that this might NOT be the case, given the different teaching methodologies underlying the three sets of materials. We hoped that by comparing the results for the different tests, a pattern might be observable which could be used to predict (to some extent) the students' performance on future tests.

Given the nature of this project as a small-scale classroom investigation, it is appropriate to identify limitations associated with the research method. First, the results can only be verified to a limited degree, since they refer to single test administrations for a small group of students ($n=21$) on the tests. The capacity to generalize from the results is also limited to the degree that this sample of one university class is representative of other university classes. Further, due to the significant nature of the limitations just described, it was determined not to provide a statistical analysis of correlation coefficients between the test scores. Rather, the benefits of the present study would be realised by representing the test scores on a common graphical scale, which would then allow for any potential patterns to be observed. Teachers could then consider the results and how they might impact upon their teaching methods, and use the results to inform their teaching.

Summary of Results

The tests for the conversation textbook ('Let's Talk!') and two OUP readers ('Reader #1', 'Reader #2') were each marked out of 100 points, and so could be easily plotted on the same graphical scale. However, the TOEIC test is recorded on a different scale (10 pts–90 pts), so it was necessary to consider how to adjust the TOEIC test scores into a similar 100 point scaling, thereby enabling the TOEIC results to be plotted onto the same graph. We examined the students' TOEIC scores, and found that the test scores ranged from a low of 205 pts to a high of 535 pts. This was compared to the students' scores on the other three tests, which ranged from a low of 42 pts to a high of 96 pts. In order to adjust the TOEIC scores to the same 100 pt scale, the highest scores on the two scales were denoted as equivalent. Hence TOEIC 535 pts was set to 96 pts on the 100 pts scale, and the other TOEIC scores were adjusted according to the same mathematical computation. This provided the desired TOEIC scaling from a low of 37 pts to a high of 96 pts, which could then be plotted on the same graph as the other three test scores (see: Figure 1).

When we study these results, which are presented in the order of students' TOEIC scores (from low to high), we observe that first; the TOEIC scores have a wider range (37–96) than the other tests. The 'Let's Talk' and 'Reader #2' tests seem to be the most closely correlated, and grouped in a band around the 70–90 pts range. The 'Reader #1' results follow a similar general pattern, but are much more widely dispersed. By contrast, the TOEIC scores appear to be the least related to any of the other three tests.

There are also patterns observable for individual students. First, Student_21 achieved the highest TOEIC score (96 pts) and scored the second highest on the 'Reader #2' test (95 pts), but was average in the other tests (84, 80 pts). Similarly, Student_15 with the highest score on 'Reader #1' (96 pts) was more average on the other three tests (88, 85, 71 pts). Student_8 did very well on all three non-TOEIC tests (92, 86, 85 pts) but much worse on the TOEIC test (60 pts). Finally, Student_20 achieved the second highest TOEIC score (82 pts), but scored much lower on the three other tests (64, 63, 42 pts). At the low end of the TOEIC scale, Student_1 with 37 pts was low on the 'Reader #1' test (63 pts), but average on the other two tests (75, 80 pts).

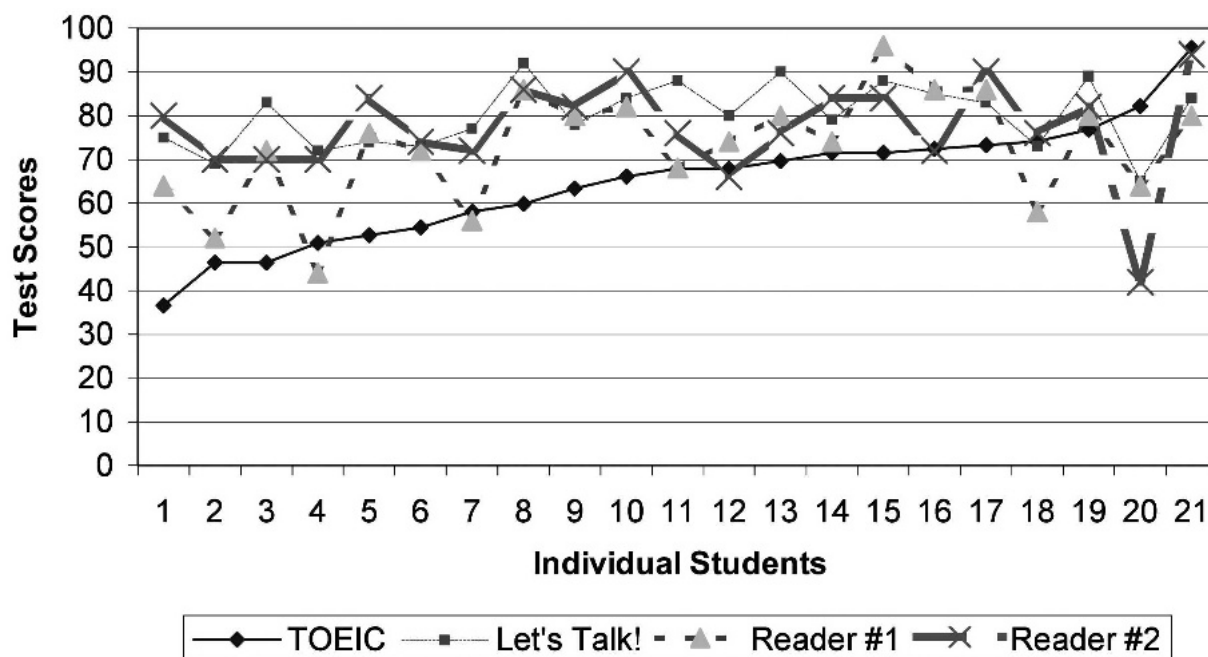


Figure 1 : TOEIC vs. Classroom Tests

Discussion of Results

While this study has not determined the causes of our findings, we would like to provide our interpretations where we believe these are relevant. The benefits of typical test preparation courses appear to us to be quite limited for a number of reasons. First, the test preparation courses practice test-taking “skills and strategies”, but the TOEIC (and similarly for the other major tests) is a “General Proficiency Test” that aims to measure a student’s “General Language Proficiency” (GLP). In Applied Linguistics, GLP is a complex scientific construct, which aims to provide a measure of broad language performance so as to provide insight into the nature of a student’s language abilities. Hence, in order to improve TOEIC scores, language exercises should be aimed at increasing GLP, a goal that contrasts markedly from the objectives of test preparation courses. Further, the inherent value to a student’s language acquisition in memorizing the answers to commonly occurring test questions, or in practicing test strategies (while ignoring the test’s primary focus on GLP) is questionable.

We are aware that it is difficult to achieve any form of substantial gain in GLP within the confines of a 15-week semester. However, it also seems somewhat incongruent to use such a program to simply aim for improvements on the margins, rather than for any form of serious learning. Teaching programs and methods which aim

at proficiency gains would appear more likely to have the added effect of increasing students’ test scores, which is a good example of the situation previously identified (re: “Putting the horse before the cart”). Instead of repeatedly studying test questions that might be used, students could be actively engaged in an extensive reading and/or listening program (to give some examples). Extensive reading can provide many substantial learning gains, including boosting knowledge of common vocabulary terms, which would also improve TOEIC scores, as well as increasing reading performance (another critical factor in test scores).

There is also a problem with the methods of study incorporated in test preparation courses. TOEIC workbooks consist of large numbers of test questions that are presented out of context. The students consequently are not provided with the opportunity to absorb important contextual cues, which in real-life activities tend to facilitate the learning process. Instead, natural learning processes are replaced by repetitious and tiring memorization exercises. Hence these workbooks tend to present masses of unrelated information, frequently in the guise of grammatical puzzles to solve, while at the same time blocking natural language acquisition processes.

Finally, the materials being presented in test preparation courses are generally not pitched specifically enough at students’ individual levels, as can, for example, be achieved with a graded readers program.

The other three tests used in this study were, by contrast, 'achievement tests' which examined knowledge of the materials studied during the preceding classes. Hence all students had an equal chance of learning the materials and achieving good results on the tests, and it is likely that this situation has contributed significantly to the students' performance on the three non-TOEIC tests being more closely related.

Conclusions

It appears that the current trend of increasing educational accountability means that TOEIC scores and other high-stakes English language tests will continue to grow in importance for university students. Clearly, teachers are obliged to do their best to help students attain the best possible scores, but need to simultaneously maintain a commitment to alternative pedagogical goals, such as communicative competence. Teachers should not, for example, revert to traditional grammar translation techniques in order to boost test courses, since the students' class time is seriously limited. Communicative methodologies should instead be adapted to foster general language competence gains simultaneously with the promotion of test scores.

We have discussed in this paper the types of discrepancies that teachers frequently observe in classrooms between TOEIC scores and classroom performance. It is intrinsic to the process of education that learning achievement needs to be measured, but it is simplistic to interpret a student's language ability as their most recent TOEIC score. Further, the resources that should be apportioned to pursuing the goal of test scores need to be carefully balanced against more fundamental language learning processes, which are inherently more difficult to measure. High scores do not necessarily correlate with English communication skills. Furthermore, students with good communication abilities may get low scores because they have not mastered basic grammar points. Proficiency scores and communicative competence are not necessarily compatible classroom goals, and program administrators need to perceive the inherent value in pursuing both directions in university courses.

References

- Edwards, N. (2006) "The historical and social foundations of standardized testing: In search of a balance between learning and evaluation." *Shiken*, 10(1), 8-16.
- Guest, M. (2005) "The problem of the talking sandwich." *The Daily Yomiuri*, October 21, p. 14.
- Chapman, M. (2005) "TOEIC: Claim and counterclaim." *The Language Teacher*, 29(12), 11-15.
- Garant, M. (2000) "EFL testing and university admission in Finland and Japan." *Asian Journal of English Language Teaching*, 10, 115-135.
- Krashen, S. (2004a) *The Power of Reading*. Portsmouth: Heinemann.
- Krashen, S. (2004b) "Applying the comprehension hypothesis: Some suggestions." Research presentation at *The 13th International Symposium on Language Teaching* (English Teachers Association of the Republic of China), Taipei, Taiwan, November 13.
- Kobayashi, Y. (2001) "The learning of English at academic high schools in Japan: students caught between exams and internationalisation." *Language Learning Journal*, 23, 67-72.
- Ichiyama, M., Critchley, M., Fiyouzat, R., Iijima, Y., & Suzuki, T. (2001) *Encounters*. Tokyo: Nanundo.
- Lyster, R. (2005) "Effective pedagogy for continued language growth in immersion." Research presentation at Ritsumeikan University, Nov 5.
- Masukawa, K. (1978) "The oral method: A critical survey" in Koike K., Matsuyama, M., Igarashi, Y. & Suzuki, K. (Eds.), *The Teaching of English in Japan*, pp 242-255. Tokyo: Eichosha Publishing.
- Matsuyama, M. (1978) "Entrance examinations: College entrance examinations and English education in Japan" in Koike, K., Matsuyama, M., Igarashi, Y., & Suzuki, K. (Eds.), *The Teaching of English in Japan*, pp 35-46. Tokyo: Eichosha Publishing.
- Robb, T. (1999) "A study of the effect of direct test preparation on the TOEIC scores of Japanese university students." *Teaching English as a Second Language*, 3(4), 67-76.
- Sunga, N. (2003) "An interview with James Dean Brown on language testing and assessment." *Guidelines*, 25(1), 13-16.
- Zemach, D. (2005) "From A to Z: TOEFL Drive." *Essential Teacher*, 2(3), 16-17.