

**TOEIC® Washback Effects on Teachers:
A Pilot Study at One University Faculty**

英語教員におけるTOEIC®の波及効果について
(学部における一つの試験的研究)

by Tim Newfields

要旨

東洋大学経済学部では、TOEICはカリキュラムの一部であり、入学時の英語能力別クラス分けに導入されている。この論文は、能力別クラス分けとカリキュラムへのTOEICの導入が英語教員にどのような影響を与えているかについて調査・分析したものである。アンケートによる量的なデータ、更に、インタビューによる質的なデータの両方から、TOEICがカリキュラムとクラス分けの手段となったことについて、現場の教員の意見は異なる。類似するTOEICの波及効果に関する研究と比較した上で、TOEICの研究に関する今後の課題を提案したい。

キーワード：TOEIC研究、TOEICの波及効果、テストインパクト解析、能力別クラス

Abstract

This paper explores how the use of the TOEIC® as a streaming tool and curricular component at one faculty in a Japanese university has impacted EFL teachers there. Both quantitative and qualitative data reveal mixed reactions to the use of this test as a placement tool and part of the curriculum. After comparing the results of this study with related test washback research, suggestions for additional TOEIC research are offered.

Keywords: instructor washback, TOEIC research, washback studies, test impact

Since washback is the central theme of this paper, let us begin by clarifying that term. Washback has been defined by Alderson and Wall (1993) as "the way that tests are . . . perceived to influence classroom practices, and syllabus and curriculum planning" (p.117).

A key feature of this concept, they point out, is that it impels teachers as well as students ". . . to do things they would not necessarily otherwise do" (ibid.).

Bachman and Palmer (1996, pp. 29-35) regard washback as a feature of a wider phenomenon known as *test impact*. They suggest test impact should be viewed both in terms of its *micro* effects in a classroom as well as its *macro* effects on educational systems and societies at large. Just as micro and macro economics have synergistic patterns, a synergism often exists between micro and macro test impact. In many cases, tests both influence and are influenced by the social climates in which they are used. However, the TOEIC test itself has been relatively impervious to change for several decades. Until a revised form of this test was announced in July 2005, the test had been fossilized for decades.

A concept closely related to washback is *consequential validity*. Some authors regard these two as synonymous. Linn, Baker, and Dunbar (1991) describe consequential validity as the intentional as well as unintentional effects of an assessment tool on teaching and learning. In other words, to evaluate how effective a test is, they emphasize that we must also consider its consequences on students and course content. Since at many universities in Asia now offer classes explicitly to help students raise their TOEIC scores, the consequential validity of this test is worth considering. For many schools, the TOEIC is not just a possible measure of English proficiency: it is a core part of the English curriculum. Though the test was not designed for that purpose, Hilke and Wadden (1997) have pointed out that it is common for high-stakes tests to interface with the curriculum closely and become a facet of the curriculum.

Though the concept of washback is evolving as new research comes to light, in this paper we will work with Anderson and Wall's standing definition of washback. Those desiring a more complete discussion of this concept should refer to a recent volume by Cheng, Watanabe, and Curtis (2003).

Previous Research

Most information on TOEIC washback consists of anecdotal evidence rather than systematic data. Accounts concerning the impact of this test on teachers suggest mixed results. Ikeda (2005) affirms that most teachers at Yamaguchi University feel grateful to have a placement test such as the TOEIC to streamline classes. Using an external performance criteria rather than each teacher's subjective assessment may, in his words, "level the playing field" in

terms of English achievement. Noting the wide gap of English ability among entering freshmen, he echoes the TOEIC sales literature by stating that the TOEIC provides standardized proficiency criterion for streaming.

Iwabe (2005) states that many teachers at Yamaguchi University also believe that adopting a minimum TOEIC score as a graduation requirement sends an unambiguous message that students need to learn something during their years of English study. "Before adopting TOEIC score graduation requirements, students were pretty much expecting to pass a class just by attending the lessons" Iwabe added. "Now it has become clearer that they also need to learn something."

Iwabe (2005) also mentions a study in which students reported how much time they spent on homework before and after a TOEIC program was implemented at one Yamaguchi University faculty. The majority of the students claimed that their study time for English increased an average of 300% and that the time spent studying for other courses also rose. He conjectures that once students acquire positive study habits in one field, it washes back into other fields. However, with this sort of data it is wise to question the motives of both the respondents and researchers. Since a lot of money and prestige is at stake in hi-stakes tests such as the TOEIC, objective measures such as recorded behaviors or test scores should supplement self-reports. The practical difficulties involved in obtaining such data have made it rare to include clear behavioral data in most washback studies.

Far from all university teachers view the TOEIC as positively as Iwabe or Ikeda. In particular many native English speakers (who often prefer to focus on communicative content) question this examination's validity. Chapman (2003) voices some reservations by stating:

[The TOEIC] is still based on the structuralist, behaviorist model of language learning and testing that informed discrete-point testing. If ETS has accepted this model is no longer suitable as a basis for the TOEFL, why has TOEIC not been treated similarly? (p. 3)

Cunningham (2002) has also correlated TOEIC scores of fifty Japanese university freshman with an in-house direct test of listening, reading, and writing and found that TOEIC reading scores correlate negatively ($-0.3908, p=0.609$) with the direct test she employed.

On the other hand, TOEIC listening scores did yield a +0.8193 ($p=0.181$) correlation. Cunningham adds:

It would appear that students are much closer in ability when it comes to language competence than the TOEIC test scores would demonstrate. It also suggests that the TOEIC was not an accurate method for determining group levels for these learners. (p. 46)

These cautionary remarks are worth reflecting on. Over forty universities and junior colleges in Japan currently use the TOEIC as a placement tool (Tonegawa, 2005). Nall (2004) laments that many of the claims being made by ETS about the TOEIC are being taken at face value. There is not enough critical examination of this test. This paper examines how the TOEIC may be impacting teachers in one micro-environment: the faculty of one university in Tokyo.

Research Questions

This study explores ways that English teachers in the Faculty of Economics at Toyo University perceive the TOEIC. In this paper the term "teachers" refers to the 24 instructors responsible for first year English courses within that program. The term "students" refers specifically to the 689 first year students within that program in 2004-2005. Specifically, this paper investigates teacher attitudes and behaviors about these issues:

(1) *Placement/Streaming* –

How useful is the procedure of dividing first year English classes according to TOEIC scores?

(2) *Classroom Time Allocations* –

How much time do teachers generally spend per class on various TOEIC-related activities?

(3) *Content Appropriacy* –

Do teachers believe TOEIC study material is appropriate for their classes?

(4) *Student Interest* –

Do teachers feel most students are keen on raising their TOEIC scores?

(5) *Consequential Validity* –

Do teachers believe that focusing on the TOEIC enhances overall English ability?

(6) *Pedagogical Changes* –

Since the TOEIC was adopted by this faculty, how have teachers' classroom practices changed?

(7) *Future Changes* –

What changes, if any, do teachers recommend in the faculty's TOEIC policies?

Methods

Participants

All instructors ($N=24$) teaching first year English classes in the Faculty of Economics at Toyo University were subjects in Phase One of this study. In Phase Two of this study, a convenience sample of eight faculty were selected.

Instruments

The instrument for Phase One consisted of the self-response survey appearing in Appendix 1. Seven of the survey items were forced-choice questions; the remaining items were open-ended. All survey items were in a bilingual Japanese/English format.

The instrument for Phase Two consisted of six core interview questions as noted in Appendix 2. Using qualitative methodology, clarification concerning responses to the Phase One questions were sought. The average time frame for each interview was five minutes. Two of the interviews were conducted in entirely in English, two entirely in Japanese, while the remaining involved considerable language switching.

Procedure

The survey form in Appendix 1 was distributed to all participants in December 2004. Responses were either mailed their completed response forms or handed them directly to the faculty office within three weeks of administration. The response rate was 75% ($N=18$).

The interviews in Phase Two were conducted in April-July 2005 and responses were recorded on an interview sheet promptly following each interview.

Results and Discussion

Question 1

The first question concerned whether the respondents felt the TOEIC was useful as a classroom streaming tool. 78% of the respondents ($N=14$) responded positively to this question. 14% of the respondents ($N=3$) responded negatively and seven respondents offered no response. This data suggest there is general support for the TOEIC as a streaming tool.

The qualitative interviews, however, reflected a greater degree mixed feelings the ability of the TOEIC to discriminate between learners. Half of the respondents noted the relatively minor difference between the TOEIC scores of most students. For that reason, though the TOEIC did help in separating some of the most proficient EFL students from the least, there is some question as to whether the test is sufficiently "fine-tuned" for the school's student population. One interviewee noted that class placement is merely one of the reasons that the TOEIC is being conducted: some institutions may adopt the TOEIC for reasons unrelated to class screening.

From a quantitative standpoint, since the mean score difference between the top one-third and bottom one-third of all freshmen examinees in April 2005 was less than 100 points, it would appear that the ability range in the population is too narrow for the TOEIC to be of much predictive value. We should remember the TOEIC is a normative test devised to measure competence in "business communication" among a general population and the mean score for the most recent issue of the TOEIC was 542.3 - well above the Japanese university average (IIBC, 2005). The standard deviation for the June 2005 TOEIC was 166 points and measurement error about 30 points. Anyone who understands basic statistics should question whether the TOEIC is an appropriate tool for screening incoming Japanese university freshmen.

Questions 2-4

The next three questions concerned how much time the teachers reported spending per class on various parts of the TOEIC. The results are summarized in Figure 1.

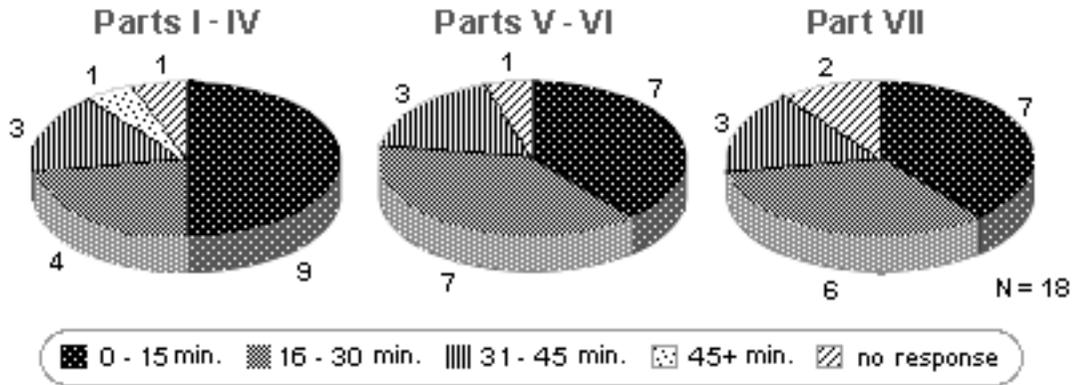


Figure 1. The amount of time per 90 minute class the university teachers indicated devoting to TOEIC study.

If the data in Figure 1 represents an accurate picture of what is actually happening in class, it would appear that only a minority of teachers devote most of class time to explicit TOEIC instruction. The majority focus on other materials, spending most of the lesson on activities other than explicit test preparation.

The qualitative interviews shed more light in this area. Five of the interviewees said that the amount of time they spent on the TOEIC depended on which class they were teaching. When teaching students had "relatively higher" TOEIC scores, the respondents indicated it was possible to devote a fair portion of the lesson to TOEIC-related practice. However, several teachers noted that when teaching students had "relatively lower" TOEIC scores, it seemed necessary to simplify the material and slow down the pace. As a result, students in such classes received less TOEIC exposure.

Qualitative data further suggests that the amount of time devoted to TOEIC teaching per class may not be consistent throughout the semester. Three teachers, for example, mentioned

how they "scaled back" TOEIC usage towards the end of the semester because students either felt it was too difficult or they were gradually losing interest. A research area worth exploring would be how the upper one-third and bottom one-third of the students respond to the TOEIC differently. Anecdotal evidence suggests that students with relatively high TOEIC scores tend to be pro-active in attempting to raise their scores further, yet those with low scores tend to perceive themselves as "bad English learners" and easily get stuck in a rut of ennui.

Question 5

Question 5 concerned whether or not respondents felt the TOEIC was too difficult for the students they teach. 46% of the respondents ($N=11$) either fully or somewhat agreed that the TOEIC was above the level of most students. However, 21% of the respondents ($N=5$) disagreed with that assessment. Three teachers gave variable responses, commenting that the TOEIC may be too difficult for some but not for others. They felt that the TOEIC was appropriate for students in the upper spectrum, but less so for those still lacking basic English skills.

Five of the qualitative interviews also pointed out students' responses were by no means uniform. One added, "Just because students are in a university does not mean that they are capable of university level English. Many of them are still having a hard time with things supposedly learned in junior high school."

Two interviewees also suggested that TOEIC difficulty might be a positive feature: students need to feel challenged. They criticized teaching practices which "dumb down" lesson content too much. "If the standards keep getting lower," one added, "performance will erode. This is what often happens when students pass a specified bench mark - they lose their incentive to study further. For such reasons is good to set high goals."

Question 6

Question 6 examined how interested respondents thought most students were in raising their TOEIC scores. 58% of the respondents ($N=14$) felt that students were either somewhat or very interested in boosting their scores. 17% of the respondents ($N=4$) disagreed with this assessment and another 25% ($N=6$) had no opinion.

On the issue of TOEIC score gains, in 1999 Robb and Ercanbrack investigated whether or not students at Kyoto Sangyou University benefited from formal TOEIC instruction over the course two semesters. What they found was the non-English majors taking classes twice a week showed no significant improvement in their listening scores, but a slight improvement in their reading scores. Interestingly, English majors who scored higher in the TOEIC at the onset and took an average of eight English classes per week did not exhibit any significant improvement in their listening or reading scores within the same time frame.

It should be remembered that TOEIC score gains are not unilinear. Saegusa (1985) did a study of how many hours of TOEIC instruction are needed for Japanese university students to exhibit score gains. He suggests 150 hours of instruction are generally required for most students to progress from a score of 300 to 400, but that 250 of instruction are necessary to move from a score of 500 to 600.

In the context of Toyo University, since only students receive less than 40 hours of English instruction between the April test and December retest, the average score gains are indeed modest. In 2002 the TOEIC European Service offered this advice on retesting:

The TOEIC programme generally recommends that learners whose native language is that of Western European origin do not take the TOEIC test until they have received at least 60 hours of English training and/or practice. Native speakers of languages from other origins should probably wait at least 100 hours. (p. 12)

In light of these comments, it is worth reflecting on why TOEIC re-tests occur in much shorter time frames. Templer (2004) cautions that market-driven drives to produce "quick results" may downgrade the effectiveness of some programs and place substantial burdens on both students and teachers.

Conversely, in the qualitative interviews two respondents noted how they felt some students do study harder because of the December retest: it creates a degree of stress, but that stress could be beneficial. "Without any retest there is less reason for students to focus much on the TOEIC," one teacher added. These comments make it clear that washback is far from a monolithic phenomena: it varies from student to student and is perceived by teachers in

various ways. Whereas lower level students may tend to give up, some who score well do feel motivated to raise their scores further. This concept is perhaps in tune with Bruner's (1960) hypothesis of optimal difficulty: material which is too easy tends to bore students, but that which is too difficult may lead students to give up. Optimal material would be slightly above the level of the majority of students. At a time when much discussion is devoted to raising educational standards, it is wise to remember Kohn's (1999) adage that "maximum difficulty isn't the same as optimal difficulty."

Question 7

Question 7 explored whether or not the respondents thought focusing on the TOEIC enhanced the overall English ability of the students. 83% ($N=15$) believed that focusing on this test does indeed improve general English ability. Only three expressed doubt as to whether studying for the TOEIC had such an effect. What is interesting about these responses is the way that they reveal a cluster of beliefs about the TOEIC. The majority of teachers seemed to voice general support for this norm-referenced, multiple-choice, standardized test. Only a minority expressed critical concerns about what the TOEIC was actually measuring and the methodology of the test itself. As a result, most of the respondents who answered negatively to Question 7 also answered the same way to Questions 5 and 6. Conversely, the majority who answered positively to Question 7 also answered the same way to Questions 5 and 6. Figure 2 hints at this correlation.

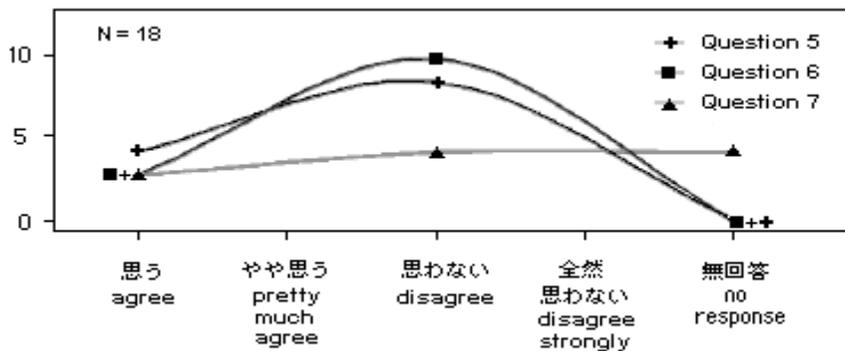


Figure 2. A comparison of responses to survey questions 5 – 7.

Examining this data, it is tempting to conjecture that there may be distinct groups of "believers" and "disbelievers" in the test. However, a more thorough factor analysis as well as a larger research sample would be needed to validate or disprove that conjecture.

Question 8

Question 8 addressed the issue of incentives for freshmen taking the TOEIC retest. At the time this survey was conducted, the Faculty of Economics required incoming freshmen to take the TOEIC test both in April and December. However, since about 22% of the freshmen did not take the winter retest, the issue of test incentives was raised. This was an open-ended suggestion and respondents were allowed to write as many comments as they wished. Four of the respondents gave either no response or a response unrelated to the question. Another four respondents suggested offering one class attendance credit for those who took the retest. Three respondents questioned the value of a retest in such a time frame. The remaining responses were idiosyncratic and ranged from using the retest for second year placement to offering no special incentive for the retest.

Five of the quantitative interviewees expressed no particular opinion about this issue. Two echoed the belief that class attendance credit should be offered for those who took the retest. Only one questioned the whole nature of the retest procedures. That person added:

The TOEIC was not designed to be a summative test measuring learned content. As Childs pointed out a decade ago, it is not an appropriate way to gauge what individuals learn over a period of time. To evaluate what students may have learned, a more appropriate method would be to adopt a criterion-referenced test.

Unfortunately, many teachers are not very clear about how normative and criterion-reference tests differ. Issues of practicality and face validity are likely to weigh more heavily in the minds of non-experts than concerns about construct validity when making test

planning decisions.

Question 9

Question 9 was another open-ended question which explored how the TOEIC has impacted the teaching style of respondents. One-third of the respondents ($N=8$) indicated that the TOEIC had had no impact on their teaching approach: their classroom practices had not changed since this test was introduced in 2003. 17% of the respondents ($N=4$) indicated that their approach to teaching the TOEIC listening section had in fact changed. Unfortunately, none specified precisely how it changed in the survey. 8% of the respondents ($N=2$) said they were now devoting more classroom time to the explicit TOEIC instruction. Two respondents also commented that teaching had also become easier since streaming was introduced: the bandwidth of abilities was narrowed.

During the qualitative interviews, six of the respondents said that the TOEIC had little, if any, direct effect on their teaching style. Other factors such as the overall level of the students appeared to have a more tangible effect. This is congruent with the comments by Cheng (2003), who suggested what teachers teach may be modified readily, but how they teach is much more impervious to change.

Question 10

The final question concerned what changes the respondents thought should occur within the faculty regarding TOEIC use. One-third of the respondents ($N=8$) offered no response. 17% of the respondents ($N=4$) indicated that they favored existing policies. 13% ($N=3$) felt that over-emphasizing the TOEIC may be counter-productive. Two teachers expressed a desire to have second year classes be streamed according to TOEIC retest scores given seven months after the first test. All other responses were idiosyncratic.

In the qualitative interviews, a range of ideas were aired. Three felt that the TOEIC, whatever its imperfections, was giving a sense of focus to the curricula and a useful taste of business English. Instead of relying on teachers' impressionistic judgments about student ability, it provides an objective score. Three other interviewees expressed a sense of fatalism about the TOEIC: for whatever reason, it was recognized as a feature of the educational

landscape and something both teachers and students must learn to live with. One interviewee commented, "There are strong macro-economic factors across the world pulling schools to accept the TOEIC. It is one concrete symbol of increasing globalization and the commercialization of knowledge. As schools become increasingly market-driven, the factors impelling large-scale testing are likely to increase."

Conclusion and Implications

Before offering any conclusions, some of the limitations of this pilot study should be acknowledged.

First of all, the data collection procedure for the quantitative section of this study was problematic: it was possible to identify 18 of the participants by their handwriting and/or envelope markings. This surely led to a kind of data contamination known as subject expectancy (Brown 1988, pp. 33-34) in which respondents may have felt inclined to write what they felt researchers wanted to hear.

It is also good to remember all of the data in this pilot study is based on self-reports rather than observed behaviors. What teachers report doing may not reflect what actually happens in class. Consciously or unconsciously, self-reports are prone to cosmetic alteration. As Weiner and Cohen (2003) point out, teacher behavior is often so ingrained that it is nearly unconscious: teachers might not even notice many of their set behaviors.

Third, some of the questions in the survey should have been more neutral. As Hajipourmezhad (2004) suggests, it is very difficult to filter out researcher subjectivity. That is why researchers should be candid about their own biases and attempt to minimize them. With a larger survey population size, researcher subjectivity could be reduced by using multiple versions of a survey form with positive and negative questions evenly balanced. For a pilot study of this size, however, a degree of researcher subjectivity is inevitable.

Finally, as with many bilingual studies, the Japanese and English survey questions were not entirely congruent. As Griffe (1998, pp. 15-17) makes it clear, participants often answer questions differently depending on the language in which they appear. Moreover, some English concepts do not translate well in Japanese. For example, hedge expressions such as "most" or "generally" appear in the English but not Japanese version of the Phase One survey. In

Japanese they sound inappropriate, but in English they add useful specificity. I chose not to adopt a monolingual survey form because of anticipated lower response rates. Whenever the task load in conducting a survey increases, the response rates tend to drop unless there is a strong incentive to complete the task.

With such limitations in mind, what this survey suggests is that a broad consensus exists in favor of the use of some sort of placement test by teachers, though there are doubts as to whether the TOEIC is the ideal tool. Hirai (2002, p. 3) maintains that though the TOEIC may have some value in discriminating between extremely proficient and extremely non-proficient EFL learners, it is of questionable use in discriminating among intermediate level students. The majority of Japanese university incoming students fall into that category.

Concerning the impact of the TOEIC on the department curriculum, the results are certainly mixed. Given the limitations of this study, it is perhaps best to describe those results impressionistically rather than in numerical data. Table 1 highlights some widespread respondent attitudes.

Table 1. *A summary of perceptions of TOEIC washback effects by English teachers at the Faculty of Economics of Toyo University (2005).*

+	-
Motivates some students to study English	Low level students may be discouraged
Practice in class may lead to improvement in TOEIC scores	Possible test fatigue and performance anxiety
Provides a standard-referenced content and clear curricular focus	Students may focus less on language skills unrelated to the TOEIC
Business focus is often apt for economics majors	Some test preparation strategies unrelated to English proficiency

Since there is currently no minimum TOEIC score graduation requirement and teacher evaluations are not based on average classroom TOEIC score gains, the pressure to teach TOEIC related skills at Toyo University is less intense than it is at some universities where such features exist. No respondents, for example, spent extensive time in class teaching guessing hints to bolster TOEIC scores. Such behavior tends to increase when there are strong

external incentives to have students pass a given test and/or if teacher evaluation becomes linked to actual score gains.

Since the TOEIC is likely to remain a facet of English education for some time to come, further research is in order. Specifically, these questions merit additional exploration:

1. Teacher Impact Elsewhere

What impact is the TOEIC having on teachers at other institutions? It would be especially interesting to expand this pilot study and compare schools with hi-stakes, hi-intensity TOEIC programs (such as Yamaguchi and Hiroshima universities) with those which have relatively low-stakes, low-intensity programs (such as Toyo and Tokai universities). In hi-stakes settings students are required to obtain a specific TOEIC score to graduate and/or teacher evaluation is measured at least in part on the basis of score gains.

2. Student Impact

What impact is the TOEIC having on students at other institutions? More information about student backwash effects would be worth investigating. A hypothesis to explore is that the TOEIC provides a positive incentive for students with higher scores, but may lead weaker students to develop negative attitudes.

3. Meta-Learning Strategies

How do students who perform well on the TOEIC differ from those who don't? What meta-learning strategies do more competent students use that differs from less successful ones?

4. Correlation and Validation Studies

How do TOEIC scores correlate with other English proficiency test scores? It may be useful to replicate some previous correlation studies in a Toyo University context to make sure that tests results reported for different populations also apply here. Since many TOEIC research studies were conducted with small populations and/or have design errors it would also be worth validating some previous.

Since the TOEIC holds a dominant position in English language testing in Asia, these questions are worth exploring. Over four million people have taken that exam since its inception, 72% of them Japanese (ETS, 2004). When we examine the volume of research that has been conducted on the TOEIC and compare it with other large-scale tests such as the TOEFL or IELTS, it becomes evident that much work remains to be done - particularly in light of the 2005 revised TOEIC. Since this version of the test appears to be more difficult than the previous version, its backwash patterns are likely to differ.

Washback thrives in educational environments which emphasize measurement driven learning (Shohamy, 1993, p. 4). In such learning contexts, tests have a significant role in shaping curricular content. Instead of devising a test to fit an existing curricula, essentially what happens is people change the curriculum to fit a test. Measurement driven learning is prevalent in conservative educational environments in which an increase in specific test scores is regarded as a sign that significant learning has taken place. One problem of this view is that many language skills are not precisely measurable. Filling out a multiple-choice exam is quite different from actually communicating in a foreign language. Many students who are relatively good at guessing the correct answers in multiple choice formats are nonetheless inept at genuine communication. For that reason the degree that the TOEIC measures actual communication skills should be questioned. The TOEIC is reputedly a measure of "communication skills . . . in an international environment" (ETS, 2005). What it likely measures are a few salient facets of language. This leads to what Bachman (2004, p. 28) refers to as an underspecification error, which occurs when factors influencing an outcome are simply ignored.

Used in conjunction with other measures, the TOEIC may give us valuable insights into the language proficiency of an examinee. However, as a sole yardstick of language proficiency, it is subject to marked distortions. Educators concerned about promoting international communication also be careful to avoid what Hirai (2005) describes as the "Clothes Makes the Man Syndrome": a case in which obtaining specific test qualifications is considered to equate with language proficiency.

Acknowledgements

*Many thanks to Kondoh Hiroko and Katou Osamu for help
in constructing the Phase One Survey.*

References

- Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14 (2), 115-129.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, U.K.: Oxford University Press.
- Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. Cambridge: Cambridge University Press.
- Brown, J. D. (2000, January). University entrance examinations: Strategies for creating positive washback on English language teaching in Japan. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 3 (2), 4-8. Retrieved February 27, 2005, from jalt.org/test/bro_5.htm.
- Bruner, J. S. (1960). *The process of education*. Cambridge : Harvard University Press.
- Chapman, M. (2003). TOEIC: Tried but undertested. *SHIKEN: JALT Testing & Evaluation SIG Newsletter*, 7 (2) 2-5. Retrieved February 21, 2005, from jalt.org/test/cha_1.htm.
- Cheng, L., Watanabe, Y., & Curtis, A. (2003). *Washback in language testing: Research contexts and methods*. Mahwah, NJ / London: Lawrence Erlbaum Assoc.
- Childs, M. (1995). Good and bad uses of TOEIC by Japanese companies. In J.D. Brown & S. O. Yamashita (Eds.). *Language Testing in Japan*. (pp. 66-75). Tokyo, Japan: JALT.
- Cunningham, C. R. (2002). The TOEIC test and communicative competence: Do test score gains correlate with increased competence? Unpublished M. A. Dissertation. University of Birmingham. Retrieved March 15, 2005, from www.cels.bham.ac.uk/resources/essays/Cunndiss.pdf

Educational Testing Service. (2005). *TOEIC Technical Manual*. Retrieved August 27, 2005, from www.toeic.cl/download/toeic_tech_man.pdf

Flyman-Mattsson, A., & Burenhult, N. (1999). Code-switching in second language teaching of French. *Lund University, Dept. of Linguistics Working Papers*, 47, 59-72. Retrieved August 24, 2005, from www.ling.lu.se/disseminations/pdf/47/Flyman_Burenhult.pdf

Griffee, D. (1998, October). Can we validly translate questionnaire items from English to Japanese? *Shiken: JALT Testing & Evaluation SIG Newsletter*, 2 (2), 15-17. Retrieved April 15, 2005 from jalt.org/test/gri_1.htm

Hajipournezhad, G. (2004). An approach to the validation of judgments in language testing in T. Newfields, S. Yamashita, A. Howard, & C. Rinnert (eds.) *2003 JALT Pan-SIG Conference Proceedings*. Tokyo: JALT Pan-SIG Committee. (p. 80 - 84). Retrieved June 5, 2005, from jalt.org/pansig/2003/HTML/HajiPourNezhad.htm.

Haladyna, T., Nolan, S., & Hass, N. (1991, June/July). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20 (5), 2-7.

Hilke, R., & Wadden, P. (1997). The TOEFL and its imitators: analyzing the TOEFL and evaluating TOEFL-prep texts. *RELC Journal*, 28 (1), 28-53.

Hirai, M. (2002, September). Correlations between active skill and passive skill test scores. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 6 (3), 2 - 8. Retrieved July 6, 2005, from jalt.org/test/hir_1.htm.

Hirai, M. (2005, September 2). Clothes make the man: A viewpoint of a lifelong English learner. Presentation at the 9th Annual JLTA Conference. Faculty of Information Studies, Shizuoka Sangyo University (Fujieda Campus).

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Ikeda, M. (2005, March). Untitled Presentation. TOEIC Training Seminar sponsored by the TOEIC Kenkyuu-kai. Osaka, Japan.

Institute for International Business Communications. (2004, January). *TOEIC Newsletter #84* (Digest Version). Retrieved March 1, 2005, from www.toeic.or.jp/toeic/data/pdf/NewsL87.pdf

Institute for International Business Communications. (2005, July). *TOEIC Deeta / Kakushuu Shiryou: Kore Made no TOEIC Koukai Heikin Sukoa*. [TOEIC Data / Misch. Resources: Prior Public TOEIC Scores]. Retrieved August 24, 2005, from <https://www.toeic.or.jp/toeic/data/data01.html>

Iwabe, K. (2005, March). Untitled Presentation. *TOEIC Training Seminar sponsored by IIBC*. Osaka, Japan.

Kohn, A. (1999, September 15). Confusing harder with better. *Education Week*. Retrieved June 20, 2005, from www.alfiekohn.org/teaching/edweek/chwb.htm

Nall, T. (2004). TOEIC: A discussion and Analysis. *ELT Two Cents Cafe*. Retrieved May 22, 2005, from www.geocities.com/twocentseltcafe/teach/TOEIC.html

Robb, T., & Ercanbrack, J. (1999). A study of the effect of direct test preparation on the TOEIC scores of Japanese University students. *TESL-EJ*, 3 (4) A-2. (p. 1-22). Retrieved March 12, 2005, from www-writing.berkeley.edu/TESL-EJ/ej12/a2.html

Saegusa, Y. (1985). Prediction of English proficiency progress. *Musashino English and American Literature*, (18) Tokyo: Musashino Women's University. Cited in Pro Lingua (2000). TOEIC Info. Retrieved August 22, 2005, from www.prolingua.co.jp/TOEIC.html.

Shohamy, E. (1993). A collaborative/diagnostic feedback model for testing foreign languages. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research*. (pp. 185-202). Alexandria, VA: TESOL Publications.

Templer, B. (2004, March). High-stakes testing at high fees: Notes and queries on the international English proficiency assessment market. *Journal for Critical Education Policy Studies*, 2 (1). Retrieved August 22, 2005, from www.jceps.com/index.php?pageID=article&articleID=21.

4. 90分の授業中に、TOEICに関連するリーディング (Part VII) のアクティビティに、どのぐらいの時間を要しますか。Generally, how much time per class do you usually spend on TOEIC-related Part 7 reading activities?

(イ) 0-15分 (ロ) 16-30分 (ハ) 31-45分 (ニ) 45分以上

5. TOEICの試験は東洋大学経済学部の学生にとって、難し過ぎると思いますか。
Do you think the TOEIC test is too difficult for most of our department students?

(イ) 思う (ロ) やや思う (ハ) 思わない (ニ) 全然思わない

6. 学生は自分のTOEICスコアを真剣にアップ化たいと思っていますか。
Do you feel most students are seriously interested in raising their TOEIC scores?

(イ) 思う (ロ) やや思う (ハ) 思わない (ニ) 全然思わない

7. TOEICを重視する現在の教育は、学生の英語力向上に結びついていると思いますか
Do you feel focusing on the TOEIC® enhances the English ability of students?

(イ) 思う (ロ) やや思う (ハ) 思わない (ニ) 全然思わない

PART 2: (Open Questions) Answer the questions in the blank space.

8. 1年生が12月のTOEIC®を受験することに対する、インセンティブ(報償、または優遇措置)があると思いますか。必要があるとお考えの場合は、何が必要だと考えますか。(例:出席回数の1つに加える。など) What incentives do you feel there should be for 1st yr. students to retake the TOEIC test in December?
(Example: Award one attendance credit, etc.)

9. 16年度以降、TOEICによるクラス分けを実施してから、ご自分のTOEIC教授法や教え方に何か変化がありましたか。 In what ways, if any, has your approach to teaching the TOEIC changed since 2003?
10. 経済学部のTOEICに関するアプローチで、改善すべき点があると思いますか。「ある」と思われる場合は具体的にご記入下さい。 What changes, if any, would you like to see in our faculty's TOEIC policies?(If you do have any suggestions, please mention below in detail.)

Appendix 2: Qualitative Core Interview Questions

事前の同意事項：私はこの学部の英語教員の皆さんに与えるTOEIC®の波及効果を研究しています。無記名でご回答いただき、内容については統計としてだけ使用します。以下の質問に対してお答え頂けるでしょうか？ はい ・ いいえ

Preliminary Consent Statement: *I am conducting research on TOEIC® washback effects among teachers in this faculty. All information will remain confidential and be used only for statistical purposes. Is it okay to proceed with this interview?* YES NO

Date of Interview: _____ Language of Interview: _____ Respondent: _____

1. 入学後のTOEIC・IPテストが1年生のクラス分けに役立っているとお考えですか？
How useful do you feel the TOEIC is as a placement tool for 1st yr. students?

2. あなたは 90 分の授業のうち TOEIC 対策にどのくらいの時間を配分していますか？
Altogether, about how much time per class do you devote to the TOEIC?
3. あなたは生徒の TOEIC の高得点獲得を目指す為に、特別なテクニックを 教えていますか？あればどのようなテクニックですか？Do you teach any explicit TOEIC test taking skills or guessing techniques in your classes? (If so, what techniques?)
4. あなたが半期の期間中に TOEIC を教えるにあたり、期間中に時間配分に変化がありますか？Does the amount of time you spend on the TOEIC change over the course of an academic semester? (If so, specifically how?)
5. レベル別のクラスで TOEIC を教える時間に変化（差）がありますか？ Does the amount of time you spend on the TOEIC vary with higher and lower classes? (If so, specifically how?)
6. この学部の現在の TOEIC 方針についてどう評価していますか？How would you evaluate our faculty's existing TOEIC policies?

www.newfields.info/Articles/washback.htm

Copyright (c) 2005 by Tim Newfields