OPINION PIECE:
The 'New' TOEIC®
by Mark Chapman and Tim Newfields

May 2006 saw the first significant changes to the TOEIC since its launch back in 1979. Over the last 27 years this test has become the de-facto standard measure of English proficiency in many parts of Asia, at least in business contexts. According to a 2008 Japan Institute of Lifelong Learning report, 64% of the 162 universities colleges in Japan described in their study use the TOEIC for streaming incoming students – a use for which this test was never designed. Moreover, in line with MEXT's (2003) Action Plan to develop "Japanese people who can use English", since 2005 the Prefectural Boards of Education of at least half a dozen prefectures in Japan have required English teachers to obtain TOEIC scores of at least 730 (or equivalent TOEFL® or STEP-Eiken scores) to obtain certification (MEXT, 2005). Worldwide the test is now taken by over 4.5 million candidates annually. Japan and South Korea account for 87% of total administrations of the secure-format of this test (ETS, 2004, p. 3). According to The Institute for International Business Communication, the organization responsible for administering the TOEIC in Japan, "the new TOEIC was developed after a close examination of the latest theories relating to language proficiency. The tasks in the assessment were refined to make them more authentic," (IIBC, 2006. p. 4) This opinion piece will outline what changes have been made and call for some further suggested improvements.

The 2006 Revision

The principle changes in the 2006 new TOEIC are an adoption of a variety of English accents (US, British, Canadian, Australian and New Zealand) in the listening section, which was formerly recorded using only North American accents. Asian English varieties are not represented on the TOEIC, yet a large percentage of the TOEIC test population is Asian and they are more likely to inteact with other non-native English speakers than they are with speakers from the 5-6 national dialects currently on this test. According to Finster (2004, pp. 9-10), 80% of the in real-life interactions around the world in English are now conducted among non-native speakers of English.

The average length of some of the listening and reading stimuli has also increased, making the new test more challenging for EFL learners. The other changes are summarized in Table 1.

Table 1 *A Summary of the Main Differences between the Former and Current TOEIC Test* (Adapted from ETS, 2007)

	Old TOEIC	New TOEIC
Part 1	20 4-option MC photo statements	10 4-option MC photo statements
Part 2	30 short 3-choice MC question-responses	(no change in format)
Part 3	30 short conversations, one 4-option MC Q each	10 longer conversations, three 4-option MC Qs each
Part 4	6-9 short talks, 2-4 MC Qs per talk (20 Qs total)	10 short talks, 3 MC Qs per talk (30 Qs total)
Part 5	40 4-option MC blank word sentences	(no change in format)
Part 6	20 sentence-level MC error recognition exercises	12 4-option MC blank word sentences embedded in text
Part 7	40 single-passage MC reading Qs	28 single-passage & 20 double-passage MC reading Qs

What's Still the Same

Although we laud changes made in the 2006 reversion of the TOEIC, in our opinion the changes have not been comprehensive enough. Indeed, what's remarkable about the new version of this test is how much is unaltered. The test in its entirety remains in a multiple-choice format. In 50

> "Although we laud changes made in the 2006 reversion of the TOEIC, in our opinion the changes have not been comprehensive enough"

of the 100 Listening Section questions, applicants can read the questions as well as the possible responses, making one wonder about the extent it is actually measuring listening skills. Over half the questions in this test still focus on sentence-level comprehension rather than discourse-level input. It is precisely for such reasons that the construct validity (Buck, 2001; Hirai, 2002, pp. 6-8), content validity (Douglas, 1992) and consequential validity (Chapman, 2005) of the original TOEIC have been criticized. Each of these factors should be reconsidered in light of the re-launch of the new TOEIC.

Buck (2001, p. 214) questions the TOEIC for its failure to assess essential aspects of listening comprehension required in real-life communication. These include ". . . indirect speech acts, pragmatic implications or other aspects of interactive language use" (p. 214). He also disparages the way the Listening Section of this test lacks the natural hesitations, fast speech, phonological shifts,

33

and negotiations of meaning between interlocutors (p. 216). The new TOEIC material which has come out so far does not suggest that the concerns raised by Buck have been addressed.

Douglas has also been critical of the narrow construct measured by the original TOEIC. He claimed that the original TOEIC failed to measure textual, illocutionary, or sociolinguistic knowledge. The changes to the final part of the test, with longer reading passages and some double passages seem to partially answer his criticisms. There is now more credibility to the claim that the TOEIC is a valid measure of reading comprehension and not just of grammar and vocabulary. However, as Lee, Yoshizawa & Shimabayashi (2006, p.154) suggest, one ongoing problem with content validity of the TOEIC is that the test does not measure a specific business English domain because a significant amount of the newly revised test material still focuses on general content that is not directly related to business or commerce.

Moreover, Alderson (2000) notes that the TOEIC still does not employ authentic, "real-life" methods of testing reading comprehension. The new format TOEIC employs only cloze and multiple choice items as measures of reading comprehension; test methods which are held up as bearing "little or no relation to the text whose comprehension is being tested nor to the ways in which people read texts in normal life," (Alderson, 2000, p. 248).

The final issue is the fundamental one of construct validity. The TOEIC still claims to be a measure of communication skills (IIBC, 2006). The argument continues to be made (IIBC, 2006, p. 10) that the TOEIC makes "a comprehensive assessment of English communication proficiency through the testing of listening and reading skills." This claim is made alongside the belief that the new TOEIC is now aligned with current language proficiency theories. It would be of great interest to see a theoretical case made for a current language proficiency theory in support of the claim that a complex, multifaceted construct such as communication proficiency can be comprehensively assessed through the testing of only receptive language skills. To the authors' knowledge that argument has never been made in the public domain. It would be a major step forward for the credibility and validity of TOEIC if ETS could provide a public account of a theoretical and / or data-driven construct validation of this test as a measure of communicative English proficiency.

What Needs to Change

If ETS is earnest about developing a more communicatively oriented TOEIC test, we suggest the following specific measures take place:

1. *Include more varieties of Asian English* – With over 90 million English speakers in India and 45 million in the Philippines (Wikipedia, 2008), not to mention millions more in places such as Pakistan, Malaysia, and Singapore, shouldn't more varieties of English be offered in the next revision of the TOEIC? Australian English, with just twenty million speakers, is included in the 1996 TOEIC revision. In light of shifting population demographics, is it wise for ETS to foster the "native speaker myth" by restricting the English used on the TOEIC to a narrow sample of the varieties that are spoken worldwide?

2. *Move from a solely descriptive focus to a broader narrative/descriptive focus in Part 1* – Instead of using solitary "snap shot" photos and statements that only ask respondents to guess what is statically happening, a richer use of language can be obtained through multi-frame picture sequences depicting stories or showing contrasts. This type of format could be a springboard for a wider range of tasks and certainly a richer amount of language. The STEP-Eiken Level 2 test already uses such sequences (Obunsha, 2006) as does ALC's Standard Speaking Test (ALC, 2006).

3. *Avoid printing the questions/answers in Parts 3 and 4* – If the TOEIC is really designed to measure listening skills, then the amount of reading material should be kept to a minimum. As it stands now, skilled test takers can pre-read questions and quickly skim through answers even before hearing them, making Parts 3 and 4 of the TOEIC in fact a composite reading-listening task rather than a listening task.

4. *Adopt alternative response formats* – Rather than have all sections of the TOEIC test in standard multiple-choice format, we feel the test would have more authenticity if a wider variety of response formats were used. Viable alternative formats could include constructed-response, multiple matching, or for Part 7, even a scrambled paragraph format (Mid-continent Research for Education and Learning, 2008).

5. *Move more from sentence-level to paragraph level exercises* – Since students are more apt to remember material when it is in larger chunks, there is a strong rationale for shifting away from isolated sentences to thematically related sentence clusters. Without denying that sentence level test items have value, why not also have exercises that require paragraph level interpretative skills? A concrete example of what that might look like for a possible TOEIC Part 5 prototype is online at http://jalt.org/cha-newEx.htm. Before we can say that this prototype is a viable alternative, naturally extensive trialing and revision is needed. Nonetheless, we feel some of the skills these exercises represent the sort of creative exploration that is needed in the TOEIC.

6. *Allow limited note taking* – The current and past version of the TOEIC is unrealistic by allowing absolutely no note-taking. As a result, test performance is overly dependent on memory, which is not necessarily a language skill. Since note-taking is a common practice in real life, why shouldn't it be permitted on this test? Security concerns can be met by limiting the paper permitted for notes and requiring examinees to return all of their notes after the test administration.

7. *Provide a compulsory section that tests a productive language skill* – We are familiar with the TOEIC Speaking and Writing Test, which is currently an option to the standard TOEIC. In our view, if the TOEIC is being marketed as a measure of communicative language proficiency, then the Speaking and Writing Test should be an integral part of the entire test package rather than an option. Currently there seems to be a gap between what the sales literature for the TOEIC claims is being measured and the content of the listening/reading sections of the TOEIC. TOEFLiBT® now incorporates both speaking and writing sections. If these sections are necessary for the validity of the TOEFL, another major English proficiency test operated by ETS, why wouldn't they be necessary for the validity of the TOEIC?

> *"Though it seems safe to say that the 2006 revision of the TOEIC represents some small steps in the right direction, in our opinion that this test still remains far short of being a valid test of English proficiency as required in real-life communication."*

We also acknowledge that further research into each of these brief proposals is needed. Though it seems safe to say that the 2006 revision of the TOEIC represents some small steps in the right direction, in our opinion that this test still remains far short of being a valid test of English proficiency as required in real-life communication.

Acknowledgement

References

ALC (2006). Intabyuu Houhou > Stage 2. Retrieved March 3, 2003 from http://www.alc.co.jp/edusys/sst/interview.html

Alderson, C. (2000). *Assessing Reading*. Cambridge, UK: Cambridge University Press.

Buck, G. (2001). *Assessing Listening*. Cambridge, UK: Cambridge University Press.

Chapman, M. (2005). A case study of the need for change in the language testing policies of a Japanese corporation. *JLTA Journal* (8). 51-67.

Douglas, D. (1992). Test of English for International Communication. In Kramer, J., & Conoley, J. C. (Eds.). *The Eleventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Educational Testing Service. (2004). TOEIC: Report on test takers worldwide. Retrieved April 10, 2007 from http://ets.org/Media/Research/pdf/TOEICTT03.pdf

Educational Testing Service. (2007). What's new about the new TOEIC test? Retrieved April 8, 2007 from http://www.ets.org/portal/site/ets/menuitem.c988ba0e5dd572bada20bc47c3921509/?vgnextoid=40e02cfcb983b010VgnVCM10000022f95190RCRD&vgne

Finster, G. (2004). What English do we teach our students? In A. Pulverness (Ed.), IATEFL 2003 Brighton Conference Selections (pp. 9-10). Canterbury: IATEFL.

Hirai, M. (2002). Correlations between Active Skill and Passive Skill Test Scores. *Shiken: JALT Testing & Evaluation SIG Newsletter*. *6*(3), 2-8. Retrieved April 9, 2007 from http://jalt.org/test/hir_1.htm

Institute for International Business Communication. (2005, November). Shin TOEIC Tesuto. [The new TOEIC test]. *TOEIC Newsletter 92*. Tokyo: Author. Retrieved April 9, 2007 from http://www.toeic.or.jp/sys/letter/News92_0139.pdf

Japan Institute of Lifelong Learning. (2008 March 14). Daigaku nihonjin eigo kyoiku katsudou ni kansuru genjouchosa. [Report on English education at colleges]. Retrieved March 27, 2008 from http://www.shogai-soken.or.jp/ htmltop/toppage.files/kyoin_18.pdf

Lee, S., Yoshizawa, K. & Shimabayashi, S. (2006). The content analysis of the TOEIC and its relevancy to language curricula in EFL contexts in Japan. *JLTA Journal (9)*. 154-173.

List of countries by English-speaking population. (2008). From *Wikipedia, The Free Encyclopedia*. Retrieved March 23, 2008 from http://en.wikipedia.org/wiki/ List_of_countries_by_English_speaking_population

MEXT. (2006). Heisei 18-Nen Kyouin Saiyou Youto no Kaizen ni Okiru Torikumi Jirei > II Senkou Shakudo no Ougenka. [Sample Measures to Improve Policies for Hiring Teachers in the 2006 Fiscal Year > II Selective Pluralization Measures]. Retrieved March 27, 2008 from http://www.mext.go.jp/a_menu/shotou/senkou/06083114/003.htm

Mid-continent Research for Education and Learning. (2008). Scrambled Paragraphs. Retrieved February 28, 2008 from http://www.mcrel.org/compendium/ activityDetail.asp?activityID=169

Obunsha (Ed). (2006). *Eiken nikyuu zen monadaishu*. [The Complete STEP Eiken Level 2]. Tokyo: publisher.

HTML: http://jalt.org/test/cha_new.htm / PDF: http://jalt.org/test/PDF/newTOEIC.pdf