

**THE UNIVERSITY OF BIRMINGHAM**

**Centre for English Language Studies**

**MA TEFL/TESL**

**Module 6 Assignment**

**September 2008**

**Andrew Joseph Lawson**

**Testing the TOEIC: Practicality, Reliability and Validity  
in the Test of English for International Communication**

**TS/06/05**

Describe an English language test with which you are familiar and discuss how valid and reliable the test appears to be. (If possible, include illustrative examples from the test itself.) Describe any procedures you would use to establish its validity and reliability. (You should not carry out these procedures unless they are quick and simple to complete.)

# Contents

1. Introduction	1
2. 'Testing a test'	1
2.1. Practicality	2
2.1.1. 'Dump the baby into the bath'	3
2.2. Reliability	3
2.2.1. Random factors	5
2.2.2. Test method facets	5
2.2.3. Personal attributes	5
2.2.4. Reliability coefficients	6
2.3. Validity	6
2.3.1. Construct validity	7
2.3.2. Content validity	7
2.3.3. Criterion-related validity	8
2.3.4. Face validity	8
3. The TOEIC Test	9
3.1. Origins of the TOEIC	9
3.2. The Objectives of the TOEIC Test	9
3.3. The Format of the TOEIC Test	10
3.4. Current popularity & usage	10
3.5. The 'New' TOEIC Test	12
4. The practicality of the TOEIC Test	13
5. The reliability of the TOEIC Test	13
5.1. Test method facets	13
5.2. Personal attributes	14
5.3. Procedures to establish further reliability	14
6. The validity of the TOEIC Test	15
6.1. Construct validity	15
6.2. Content validity	16
6.3. Criterion-related validity	16
6.3.1. Predictive validity	16
6.3.2. Concurrent validity	17
6.4. Face validity	17
6.5. Procedures to establish further validity	18
7. Misuse of the TOEIC Test	19
8. Backwash	19
9. Conclusion	20
9.1. TOEIC Speaking and Writing Test	22
9.2. Need for further research	22
Appendices	23
References	30

## 1. Introduction

When Japanese people over school-age discuss English language proficiency, the measurement of ability they are most likely to mention is ‘TOEIC score’. This refers to the Test of English for International Communication, which was first administered in Japan in 1979. The test has continued to grow in popularity and influence throughout the world, but nowhere more so than in Japan, where it is administered by the Institute for International Business Communication (from herein referred to as IIBC), a non-profit organization under the auspices of the Japanese Ministry of International Trade and Industry. This paper seeks to examine the quality of the TOEIC Test, via the three ‘classic criteria’ for ‘testing a test’: practicality, reliability and validity (Brown, H.D., 2001: 385). Prior to this, there will be an overview of these criteria.

## 2. ‘Testing a test’

In questioning the quality of a language test, Bachman and Palmer argue that a model of test usefulness is “the essential basis for quality control throughout the entire test development process” (1996: 17). This model of test usefulness may be expressed as follows:

$$\text{Usefulness} = \text{Reliability} + \text{Construct validity} + \\ \text{Authenticity} + \text{Interactiveness} + \text{Impact} + \text{Practicality}$$

*Figure 2.1: Usefulness (Bachman & Palmer, 1996: 18)*

Bachman and Palmer also emphasise the need for test developers to avoid considering any of the above test qualities independently of the others. Trying to maximise all of these qualities inevitably leads to conflict. Thus, the test developer must strive to attain an appropriate balance of the above qualities for each specific test. Obviously, there is no perfect language test, much less one single test which is perfectly suitable for each and every testing situation.

## 2.1. Practicality

$$\text{Practicality} = \frac{\text{Available resources}}{\text{Required resources}}$$

If practicality  $\geq 1$ , the test development and use is practical.

If practicality  $< 1$ , the test development and use is not practical.

*Figure 2.2: Practicality (Bachman & Palmer, 1996: 36)*

Before we even examine the content of the test, we must ask if it is feasible. A good test must be practical. Whether a test is practical or not is a matter of examining available resources. It may be possible to develop a test which is highly valid and reliable for a particular situation, but if that test requires more resources than what are available, it is doomed. It is possible that the test may initially be used if test development does not exceed available resources. But ‘cutting corners’ in administering or marking the test, in order to make savings in time or money, will immediately lead to an unacceptable deterioration in reliability. Test practicality involves the “nitty-gritty” (H.D. Brown, 2001:386) of man-power, materials and time. We can only make the best use of what is available to us. Examples of these resources are listed in Figure 2.3.

---

### 1 Human resources

(e.g. test writers, scorers or raters, test administrators, and clerical support)

### 2 Material resources

*Space* (e.g. rooms for test development and test administration)

*Equipment* (e.g. typewriters, word processors, tape and video recorders, computers)

*Materials* (e.g. paper, pictures, library resources)

### 3 Time

*Development time* (time from the beginning of the test development process to the *reporting of scores* from the first operational administration)

*Time for specific tasks* (e.g. designing, writing, administering, scoring, analyzing)

---

*Figure 2.3: Types of resources (Bachman & Palmer, 1996: 37)*

### **2.1.1. ‘Dump the baby into the bath’**

Indeed, in many respects, practicality is the whole *raison d’être* of testing. Owen (2001: 8-9) provides the brutal but succinct analogy: “for the complete bath water test, just dump the baby into the bath and see what happens”. For the complete test of an employee’s English communication skills in a business context, a firm should send the employee out to work overseas and do business in English. If he/she is able to do the work required, the company may deem the employee’s language abilities to be sufficient. If there are lots of costly errors due to communication problems, and the employee becomes chronically depressed as a consequence of being unable to do their job properly, make friends or even make a dental appointment where they are living, then the company may deem the employee’s language skills to be insufficient.

Clearly, there are reasons why companies do not do this (just as there are obvious reasons why we don’t test bathwater temperature by tossing in our infants, and waiting to see if they boil or not). The costs to a company of sending staff into foreign language-speaking business situations without examining their readiness beforehand would often be financially colossal, as well as morally questionable. “A test is a more economical (and hopefully more reliable) substitute for a more complete procedure” (Owen, 2001: 9).

## **2.2. Reliability**

“How much of an individual’s test performance is due to measurement error, or to factors other than the language ability we want to measure?”

*(Bachman, 1990: 160)*

A good test must be consistent. If the same group of subjects take the same test on two different occasions, results should be similar, both in individual scores, and in the rank order

within the group (H.D. Brown, 2001: 386; Bachman & Palmer, 1996: 20). If the same written answer in a test is scored by two different markers, the two different scores should be similar (Bachman, 1990: 24). If two forms of the same test are created which are intended to be used interchangeably, an individual should obtain very similar scores on both versions (Bachman & Palmer, 1996: 20).

The crucial point here is that results should be *similar*. Given the vast array of factors affecting reliability, particularly the human element involved, we must accept that results are highly unlikely to ever be 100% reliable (Hughes, 2003: 36; Bachman & Palmer, 1996: 20). Marker reliability is also often likely to play a significant role. Modern computerised marking has greatly reduced the human error element in many objectively scored test questions, but technological advances are unable to provide much assistance in the process of scoring more subjective test items which assess productive skills.

The following path diagram (Figure 2.4) from Bachman (1990: 165) illustrates the factors which affect language test performance:

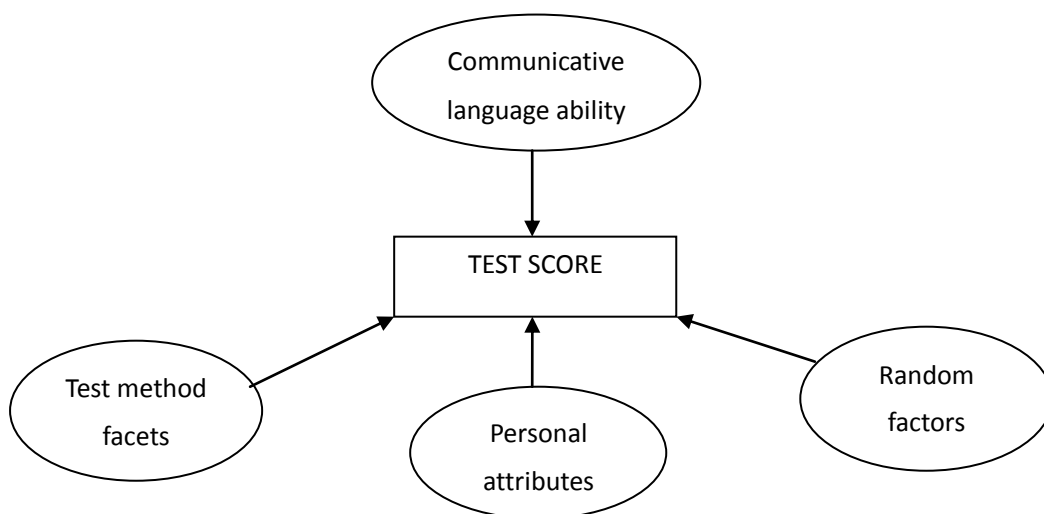


Figure 2.4: Factors that affect language test scores (Bachman, 1990: 165)

### **2.2.1. Random factors**

There is little we can do to control whether a test-taker has caught a cold, slept well the night before or had a fight with his/her partner at breakfast on the day of the test. These are just a few of the random factors, which Bachman (1990: 164) refers to as “unsystematic”, as they are unpredictable and beyond the test creator’s control. But we must strive to minimise any inconsistencies within our control.

### **2.2.2. Test method facets**

Bachman (1990: 119) provides an extremely detailed framework of test method facets, which he cautions is by no means comprehensive (1990: 117). This is later refined to a ‘framework of task characteristics’ (Bachman & Palmer, 1996: 49-50), and it is at this stage that the overlap between reliability and validity becomes most apparent. It is beyond the scope of this paper to fully discuss Bachman’s lengthy, detailed framework of test method facets (Bachman’s list of *Categories of test method facet* is reproduced in Appendix A), but the five major categories are as follows:

1. Characteristics of the testing environment
2. Characteristics of the test rubric
3. Characteristics of the input
4. Characteristics of the expected response
5. Relationship between input and response

### **2.2.3. Personal attributes**

The third group of factors affecting language test scores are those referred to in Figure 2.4 as ‘personal attributes’. A test-taker’s age, gender, cognitive style and background may all have an impact on test performance, and should be taken into consideration when analysing results. Both

personal attributes and test method facets are considered by Bachman (1990: 164) to be “systematic” sources of error, as they are likely to affect an individual regularly.

#### 2.2.4. Reliability coefficients

Reliability coefficients are a method of quantifying the reliability of tests, in order to allow us to compare them. A test’s reliability coefficient is between zero and 1, with the two hypothetical extremes meaning absolutely no reliability and complete reliability respectively (Hughes, 2003: 38-39). The level of reliability which we might reasonably expect depends on the type of language test. Lado (1961, in Hughes, 2003:39) suggests the following levels:

Type of Test	Reasonable Reliability Coefficient
Vocabulary, structure & reading	.90 - .99 range
Auditory comprehension	.80 - .89 range
Oral production	.70 - .79 range

*(adapted from Hughes, 2003: 39)*

We must also be aware that the required level of reliability for a test depends on how the results of that test are being utilized. The higher the stakes, the higher the required level of reliability (Bachman & Palmer, 1996: 135; Hughes, 2003: 39).

#### 2.3. Validity

The concerns of reliability and validity can ... be seen as leading to two complementary objectives in designing and developing tests: (1) to minimize the effects of measurement error, and (2) to maximize the effects of the language abilities we want to measure.

*(Bachman, 1990:161)*

While test reliability is concerned with minimising the effects of measurement error, test



validity involves ensuring that a test truly tests the abilities which we are looking to assess. Owen (2001: 20) describes validity as the “sine qua non of language testing” and laments the lack of consensus and the “dispiriting” number of aspects of validity mentioned in the standard literature. There is clearly a great degree of overlap among the various aspects (Owen, 2001: 22), and both Bachman (1990:238) and O’Sullivan (2006: 96) support Messick’s view of validity as a “unitary though multifaceted concept” (Bachman, 1990: 238). Here is a brief overview of the most commonly highlighted aspects:

### **2.3.1. Construct validity**

Several authors have commented on the difficulties involved in trying to untangle construct and content validity (Owen, 2001: 22; Moritoshi, 2001: 9). Construct validity is related to the question of whether the test meets certain theoretical requirements (Oller, 1979: 50; Jafarpur, 1987: 199). How much is the test’s content based on the theoretical construct which it is claimed to be? And thus, how much does a test score actually allow us to make an appropriate interpretation of an individual’s ability? (Bachman & Palmer, 1996: 21)

### **2.3.2. Content validity**

Content validity, on the other hand, is more concerned with the question of whether the test covers a fully representative sample of the theoretical construct (Jafarpur, 1987: 200; Moritoshi, 2001: 9). If the test actually requires the test-taker to perform all the skills that it claims to measure, it can be said to have content validity (H.D. Brown, 2001: 388; Hughes, 2003: 26; Oller, 1979: 50).

### **2.3.3. Criterion-related validity**

Another aspect of test validation involves ensuring that it gives the same results as another independent, reliable assessment of the testee's abilities (Hughes, 2003: 27; Owen, 2001: 21). There are two forms of criterion-related validity. "Concurrent validity is established when the test and the criterion are administered at about the same time" (Hughes, 2003: 27). Hughes uses the idea of a proposed shortened oral test being measured against the existing longer test to exemplify concurrent validity. However, Owen (2001: 21) describes concurrent validity as "irritating", and makes the salient point that given the vast majority of new tests are created precisely because the test-developer does *not* have confidence in those which are already available, it is only rarely that true concurrent validity can be established. Predictive validity is a measure of how well the test is able to accurately predict the testee's future performance. The difficulty in dealing with predictive validity is that so many other factors may be involved in determining future performance. As a consequence, "a validity coefficient of around 0.4 (only 20 per cent agreement) is about as high as one can expect" (Hughes, 2003: 30).

### **2.3.4. Face validity**

If a test *appears* to do what it claims, it is more likely to be accepted by all concerned. Jafarpur (1987: 199) considers face validity to be subordinate to the other types listed above. Indeed, it is a qualitative measure, not a scientific one. However, in light of harsh, economic realities, I would take issue with Jafarpur's claim. Face validity is of the utmost importance in the development and adoption of a new test. If those paying for the test (whether it be an individual, a company or an educational authority) are unimpressed by what they see, it is highly unlikely that the test will be widely-adopted. If teachers or students fail to perceive the value of the test, it is likely to have a detrimental impact on preparation and scores.

### **3. The TOEIC Test**

We now proceed to a more detailed overview of the TOEIC test.

#### **3.1. Origins of the TOEIC**

The Test of English for International Communication (TOEIC® test) was the brainchild of Kitaoka Yasuo, a former Vice President of the Institute for International Business Communication (IIBC), who saw the need for Japanese businesses to have personnel with better English communication skills, in order to thrive in the increasingly-competitive global economy. He approached the New Jersey-based Educational Testing Service (hereafter ‘ETS’), who created the TOEIC test, based upon its already popular Test of English as a Foreign Language (TOEFL).

The TOEFL itself was deemed unsuitable for businesspeople due to its emphasis on academic English, and this new test was intended to be a more accurate assessment of English ability in everyday business contexts. Kitaoka’s vision was supported by the Japanese Ministry of International Trade and Industry (MITI), who were dissatisfied with the effects of the non-profit Society of Testing English Proficiency (STEP) test, known as *Eiken*. The Eiken test had been established in 1963 with the support of the Japanese Ministry of Education (*Mombusho*).

#### **3.2. The objectives of the TOEIC Test**

The test creators claim that the TOEIC test is “an English language proficiency test for people whose native language is not English”. They also claim that “it measures the everyday English skills of people working in an international environment. TOEIC test scores indicate how well people can communicate in English with others in the global workplace” (ETS, 2007).

According to ETS, the TOEIC test “offers accurate assessment of English skills, from the beginner stage to those at advanced levels.” (IIBC, 2006: 2). The test involves “the concept of making a direct measurement of passive skills (listening and reading) to obtain an indirect measurement of productive skills (speaking and writing).” (ibid: 3)

### **3.3. The format of the TOEIC Test**

The TOEIC is a norm-referenced test, and in its current format is comprised of a Listening Section and a Reading Section. The duration of the test is two hours, and although the Listening Section is considerably shorter than the Reading Section (45 minutes compared to 75 minutes), both parts contain 100 questions and have the same range of possible scores (5 to 495 points), resulting in a testee’s total score being between 10 and 990 points. All 200 questions on the test are multiple-choice, and the test is entirely in English (ETS, 2006a).

There are two ways to take the TOEIC test; the Secure Program (SP), which is fully administered by IIBC at specified dates and venues, and the Institutional Program (IP), which is conducted at the time and location of the administering company, school or organisation’s choosing (IIBC, 2008).

### **3.4. Current popularity & usage**

Since it was introduced in 1979, the TOEIC test has developed into one of the world’s most dominant forces in English language assessment, with over 4.5 million people taking the test each year (Shimizu, 2006; IIBC, 2008). It is recognized by over 5,300 multinational companies. Approximately 1.5 million of these tests are taken in Japan, where it is now “widely held to be *the* test for assessing business English skills” (H. Brown, 2006: 1177). There are now over 2700

Japanese companies, organizations and institutions using TOEIC to assess English proficiency (IIBC, 2008). These range from corporate heavyweights including Toyota and Matsushita (Shimizu, 2006), to prestigious learning institutions such as Meiji University (Rebuck, 2003: 30). In the corporate world, it is used for making significant personnel decisions such as hiring or promotion, and English language schools and academic institutions are using it as a placement tool or to assess improvement and achievement. The test producer gives a list of what it considers to be appropriate uses of the TOEIC test scores:

- Hiring of applicants for an open position within a corporation or organization where workplace/everyday English is a required job skill
- Placement of applicants or test takers within a corporation or organization where workplace/everyday English is a required job skill
- Promotion of test takers within a corporation or organization where workplace/everyday English is a required job skill
- Measurement of workplace/everyday English proficiency levels of students in secondary schools and universities
- Measurement of individuals' progress in workplace/ everyday English proficiency levels over time"

*(English Testing Service, 2007: 12)*

Additionally, a separate ETS document, intended for language schools in Europe, states that the TOEIC test can be used for placement, demonstrating progress, evaluating course progress, motivating learners, and validating language proficiency (ETS, 2006b).

### 3.5. The ‘new’ TOEIC Test

In 2006, the TOEIC test was redesigned for the first time since its inception a generation earlier, with the producers claiming an underlying concept of ‘more authentic’ questions. The main changes involved introducing a wider selection of native speaker accents in the listening section (adding Australian, British, Canadian and New Zealand accents to U.S.), the elimination of the error recognition section, and an increase in the average length of some of the reading and listening stimuli. The changes in the question format are summarised in Figure 3.

Part	Old TOEIC	Redesigned TOEIC
<b>Listening Comprehension Section 1</b>		
1	Photographs: 20 questions	Photographs: 10 questions
2	Question-Response: 30 questions	Question-Response: 30 questions
3	Short Conversations: 30 questions (30 conversations, 1 question each)	Conversations: 30 questions (10 conversations, 3 questions each)
4	Short Talks: 20 questions (6-9 talks, 2-4 questions each)	Short Talks: 30 questions (10 talks, 3 questions each)
<b>Reading Comprehension Section 2</b>		
5	Incomplete Sentences: 40 questions	Incomplete Sentences: 40 questions
6	Error Recognition: 20 questions	Text Completion: 12 questions (4 reading sets, 3 questions each)
7	Reading Comprehension: 40 questions	Reading Comprehension: 48 questions (single passages: 28 questions; double passages: 20 questions)

Figure 3: Comparison between TOEIC and redesigned TOEIC (Adapted from Educational Testing Service, 2008)

The creators insist that the previous levels of reliability and objectivity have been maintained (IIBC, 2006: 9) and that “the new TOEIC has also been developed to make a comprehensive assessment of English communication proficiency through the testing of listening and reading skills.” (IIBC, 2006: 10)

#### 4. The practicality of the TOEIC test

Other things being equal, it is good that a test should be easy and cheap to construct, administer, score and interpret.

*(Hughes, 2003:56)*

There is no question that the TOEIC test's greatest strength lies in its practicality. The test is conducted entirely in multiple choice question format, allowing for large numbers of students to be tested simultaneously, and the papers to be marked electronically. All scoring is exact, and there is no need to worry about difficult judgement calls that arise in SEMAC scoring. Administration is quite straightforward and the test is relatively inexpensive.

#### 5. The reliability of the TOEIC Test

The creators and administrators of the TOEIC test frequently claim high reliability regarding their test (IIBC, 2006: 9; ETS, 2007: 5). And these claims do appear to have some merit. ETS report the reliability of the test to be over 0.9 (based on the Kuder Richardson formulae), and "well within the generally accepted limits for measurement of individual achievement" (Woodford, 1982: 8).

	<b>KR-20 Reliability coefficient</b>	<b>Standard Error (score units)</b>
<b>Listening comprehension</b>	0.916	25.95
<b>Reading comprehension</b>	0.930	23.38
<b>Total test</b>	0.956	34.93

##### 5.1. Test method facets

The IIBC make every effort to ensure that the facets of the testing environment are as reliable as possible. TOEIC test venues are in colleges and universities, and are tests always held

on the same day of the week, and at the same time. In an informal survey of Japanese acquaintances who have taken the test, all expressed satisfaction with the testing environment (venue, temperature, lighting, quality of audio component etc).

Douglas (2000: 232) highlights the “thoroughness that is a hallmark of the ETS approach to language testing” and lauds the TOEIC test for the completeness of its rubric “in terms of objective, time allotment, distinction between tasks, and response procedures (ibid: 233). Perkins (in Chapman, 2003) also supports ETS’ claims of reliability.

## **5.2. Personal attributes**

While ETS have obviously been unable to devise a test which takes into account the age, gender, cognitive style and background of each individual testee, such factors are considered. Questionnaires are given to all test takers, and data concerning performance in relation to attributes such as employment, educational background and English-language experience is published (IIBC, 2008). Unfortunately, there is little research available which analyses these statistics.

## **5.3. Procedures to establish further reliability**

It would be of interest to examine results in relation to testees’ personal data, to determine how much personal attributes affect test reliability, and further establish test fairness. It would also perhaps be worthwhile to further examine the notion of test format familiarity. Many people in Japan take the TOEIC test three or four times per year. Given the highly standardised test format, these students must develop a high level of familiarity with the test. How much can the score improvements achieved by these individuals be attributed to enhanced test familiarity?



## **6. The validity of the TOEIC Test**

A common criticism of standard language proficiency tests is that they fail to assess the communicative competency of the testee, and are thus invalid as measures of proficiency in the true sense of the word (H.D. Brown, 2001: 387). Several scholars (Douglas, 2000; Cunningham, 2002; O’Sullivan, 2006) have questioned the validity of the TOEIC test on these grounds. It fails to assess two of the four related aspects in Canale and Swain’s model of communicative competence ; sociolinguistic and strategic competencies (H.D. Brown, 2007: 219-220).

### **6.1. Construct validity**

There are clearly some problems in establishing construct validity for this test. The TOEIC test claims to assess overall English communication skills, yet it does so by only testing listening and speaking skills. This would imply that the TOEIC is constructed upon the theory that an individual’s productive language abilities are proportional to his/her receptive abilities (Miyata, 2004: 61). Hughes (2003: 31) suggests that it is unnecessary to define construct validity for direct tests of some commonsense constructs, such as ‘reading ability’ or ‘writing ability’, but:

Once we try to measure such an ability indirectly, however, we can no longer take for granted what we are doing. We need to look to a theory of writing ability for guidance as to the form an indirect test should take, its content and techniques.

*(Hughes, 2003: 31)*

Has ETS designed its test based on a flawed construct, or in the words of Moritoshi (2001: 9) has it merely “tended to skirt around the issue”? There is the distinct possibility that in spite of its advertising claims, the test designer has defined the construct of ability in a far narrower sense (i.e. the capacity to study for and pass a specific test), manipulating the construct for “scholastic,

economic, and social stratification purposes” (Ross, 2008: 8). Interestingly, the test administrator seems to believe that an element of concurrent validity is sufficient to claim construct validation (Chauncey Group International Ltd., 1998: III-1).

## **6.2. Content validity**

Given the uncertain nature of the construct validity of the TOEIC test, it is unsurprising to learn that its content validity has also been called into question. Oller asserts that content validity ensures that the examinee has to “perform tasks which are genuinely the same or fundamentally similar to tasks one normally performs in exhibiting the skill or ability the tests purports to measure” (Oller, 1979: 51). If we accept this, and ETS’ claim that the TOEIC measures English communication skills in a business context, there appears to be little content validity. Without content validity the test is unlikely to be accurate, and likely to have a harmful backwash effect (Hughes, 2003: 27). Douglas (2000:236) also states that as it is “unlikely that the reading tasks engage the test takers in any genuinely communicative behavior or in genuinely specific purpose use”, should TOEIC be even considered to be a genuine test of Language for Specific Purpose?

## **6.3. Criterion-related validity**

### **6.3.1. Predictive validity**

ETS assessment specialist Trina Duke claims that the revised TOEIC test is more than a mere knowledge test: “The new TOEIC test assesses how valid that knowledge is in the real world, that is, it tests for a more authentic processing capability” (IIBC, 2006: 4) But there is no further explanation, and there appears to be no evidence to back this bold assertion, or any other claims of predictive validity.

### **6.3.2. Concurrent validity**

It was mentioned above that the test administrator depends on concurrent validity to establish construct validity for the TOEIC test, and the correlations between the TOEIC and a host of other speaking, listening, reading and writing tests are provided (Chauncey Group International Ltd., 1998: III-2 – III-5). But Bachman dismisses such an idea on the grounds that it “simply extends the assumption of validity to these other criteria, leading to an endless spiral of concurrent relatedness” (Bachman, 1990: 249). Moritoshi (2001: 10) also examines the correlations provided by the test developer, and discovers ‘various shortcomings’, including unvalidated tests, subjective scoring, unsubstantial sample sizes and lack of negative evidence (as recommended by Bachman 1990: 259).

### **6.4. Face validity**

In the modern, consumer-driven world, image is everything. And judging by its pervasive nature within Japanese society, and its presence in the global language testing sphere, the TOEIC test appears to enjoy a remarkably high level of face validity. The product has been extremely well-marketed. Ihara and Tsuroka (in Rebeck, 2003: 24) argue that initially marketing solely to companies and firmly establishing itself as “the Company English Test” was a masterstroke in being later able to promote itself to individuals and universities.

In 1991, Ministry of Education reforms gave universities greater freedom to decide their own curriculum and graduation conditions. Universities were also allowed to accept Ministry-recognised qualifications as university credits. Initially, only National examinations (*kokka shiken*) were accredited, but in 1999, the Ministry declared “TOEIC and other tests which had received wide recognition by society” deserved to be accredited too. (Rebeck, 2003: 30).

Whether TOEIC's exalted status in Japan led to its accreditation, or whether it was a consequence of Ministry support in the first place is a matter for debate.

TOEIC face validity is remarkably high among companies, universities and individuals. It appears that the group least satisfied with the TOEIC test are English teachers.

### **6.5. Procedures to establish further validity**

Clearly, a great deal of research is required to comprehensively establish overall validity. I would suggest that it is impossible to establish genuine construct validity, in light of ETS' claims about the test, and the actual current format. A genuine assessment of communicative ability must surely include a speaking component.

However, it may be possible to establish greater predictive validity, if an in-depth study was undertaken to analyse TOEIC scores and later language performance. ETS could perhaps work with a major multinational corporation to examine employee TOEIC scores prior to being posted overseas, and then analyse their communicative competence within their overseas daily work routine. Of course, such assessments would have a highly subjective element.

Some relatively straight-forward steps could be taken to establish greater content validity. The simplest of these would be to introduce Asian Englishes to the listening component. Given the nationalities and locations of the majority of TOEIC test candidates, it seems rather absurd that Indian and Singaporean Englishes have not been included in the revised Listening Section. Japanese, Korean and Chinese testees (who account for the majority of candidates) are far more likely to need to be able to comprehend an Indian speaker than a Canadian or New Zealand speaker.

## **7. Misuse of the TOEIC test**

Due to the remarkably high face validity of the TOEIC test in so many quarters, and the aggressive marketing style of the test developer, there are many situations where the TOEIC test is clearly being misused.

Given that it is a norm-referenced test and students are unable to compare their score directly against their previous attempts, the TOEIC is unsuitable for use as an achievement test. Nevertheless, it continues to be used as such within many organisations and by many individuals (Childs, 1995). The IIBC accepts that the TOEIC is not a diagnostic test, but they do recommend its use as a progress test (Chapman, 2004).

One of the most troubling aspects of TOEIC misuse is the number of lower-level students who take the test. Miller (2003) notes the inappropriateness of the test for low-level students, as well as the emergence of low-level TOEIC preparation books which often do nothing to prepare students for the actual test.

## **8. Backwash**

Sometimes referred to as ‘washback’, backwash is “the effect of testing on teaching and learning”, whether good or bad (Hughes, 2003:1). As a consequence of its popularity in Japan, the TOEIC test must have a backwash effect on numerous classrooms, students and teachers. Unquestionably, some of that backwash is positive. Students being motivated to study more in order to attain a higher score in the test should be considered positive. However, negative backwash is also clearly in evidence too.

In their desperation to help students attain higher TOEIC scores prior to graduation and entering the employment market, more and more universities throughout Japan are replacing English proficiency classes with TOEIC preparation classes. Communicative competence is patently not the primary objective of these classes, and instruction time is focused on learning discrete grammar items and mastering test-taking strategies (Miller, 2003).

In a commercial environment, teachers are likely to have to deal with conflicting pressures, between giving the students what the teacher perceives they need, and giving the clients what they want (the clients being corporations, institutions or individuals). Brown’s study focusing on graded reading for TOEIC preparation (2006) gives a vivid example of the attitudes teachers face. Table 8 highlights students’ lack of connection between reading helping English and helping TOEIC score, and the improvement in perception after a mock TOEIC test was administered. “Even though learners felt the reading was improving their reading speed and fluency, they did not see a connection to TOEIC success.” (H. Brown, 2006: 1176)

Survey Statement	Average Likert Scale Response	
	Initial	6 month
Reading this book helped my TOEIC score.	2.1	4.1
Reading this book helped my English.	3.6	4.3
I read faster now than I did when I started this book.	3.3	4.1
I understand more of what I read now than I did when I started this book.	3.4	4.3

*Table 8: Student’s reactions to graded reading (H. Brown, 2006: 1180)*

## 9. Conclusion

In summary, the TOEIC test appears to be quite reliable, and has a very high level of face validity, particularly in Japan. It is also extremely practical. However, validity remains a genuine

concern. There is confusion as to the underlying construct of TOEIC (O’Sullivan, 2006), and this leaves it susceptible to damning criticism from Bachman:

To refer to a test or test score as valid, without reference to the specific ability or abilities the test is designed to measure *and* the uses for which the test is intended, is therefore more than a terminological inaccuracy. At the very least, it reflects a fundamental misunderstanding of validity; at worst, it may represent an unsubstantiated claim about the interpretation and use of test scores.

*(Bachman, 1990: 238)*

The intention here is not to dismiss the TOEIC test as worthless. It satisfies a huge market for a practical, fair and objective test of receptive language skills. The 2006 revised version of the test includes some commendable improvements. The sample questions available (see Appendix B) suggest that there is a greater frequency of business-related stimuli (emails, business letters etc). Greater variation in the length of reading passages partially answers criticisms of those in the older version (that they were too short and largely decontextualised), suggesting a better assessment of comprehensive reading ability. The introduction of a wider variety of accents in the listening section is also an improvement.

But some fundamental issues remain. The TOEIC test was devised in the 1970s, based on a structuralist model of language learning which pre-dates the communicative era (O’Sullivan, 2006; Douglas, 2000; Chapman 2003). Despite claims to the contrary from ETS, it appears that the current form of the test is still based on the original model. Multiple choice question items alone “can never engage the candidate in the kind of cognitive processing evident in listening or reading in a business domain” (O’Sullivan, 2006; 22).

## **9.1. TOEIC Speaking and Writing Test**

Hughes (2003: 32) asserts: “When in doubt, where it is possible, direct testing of abilities is recommended.” ETS are clearly aware of these criticisms, as they have produced the new TOEIC Speaking and Writing Test. But this remains entirely separate from the main Listening and Reading test, and appears to be a long way from achieving mainstream status. To date, this author has yet to hear of any first-hand accounts of this test. Recent interviews of Human Resources Managers at Japanese branches of five major international companies revealed that all firms placed a premium value on TOEIC scores, yet none had even heard of this new Speaking and Writing Test.

## **9.2. Need for further research**

Since its inception in 1979, there has been precious little independent research into the TOEIC test. And other than an opinion piece by Chapman and Newfields (2008), there has been nothing on the new version. At the very least, the test producers need to provide updated versions of the results given in Woodford (1982). Given the status and widespread use of this test in Japan, in-depth research is long overdue. The major stumbling block is that the TOEIC is a closed test, and that the producer appears highly unlikely to release detailed test material and results for analysis. Thus, in the current climate, we are unlikely to see any major research which would allow us to establish greater validity and reliability. The TOEIC test firmly remains the market leader for assessment of business English. But with the emergence of new tests targeting the same market (Arita, 2003), it would be wise not to be complacent.



## Appendix A

### Bachman's Categories of test method facet

<p><b>1 FACETS OF THE TESTING ENVIRONMENT</b></p> <p><b>Familiarity of the place and equipment</b>  <b>Personnel</b>  <b>Time of testing</b>  <b>Physical conditions</b></p>	<p><b>4 FACETS OF THE EXPECTED RESPONSE</b></p> <p><b>Format</b>  <i>Channel (aural, visual)</i>  <i>Mode (productive)</i>  <i>Type of response (selected, constructed)</i>  <i>Form of response (language, nonlanguage, both)</i>  <i>Language of response (native, target, both)</i></p> <p><b>Nature of language</b>  <i>Length</i>  <i>Propositional content</i>          Vocabulary (frequency, specialization)          Degree of contextualization (embedded/reduced)          Distribution of new information (compact/diffuse)          Type of information (concrete/abstract, positive/negative, factual/counter-factual)          Topic          Genre</p> <p><i>Organizational characteristics</i>          Grammar          Cohesion          Rhetorical organization</p> <p><i>Pragmatic characteristics</i>          Illocutionary force          Sociolinguistic characteristics</p> <p><b>Restrictions on response</b>  <i>Channel</i>  <i>Format</i>  <i>Organizational characteristics</i>  <i>Propositional and illocutionary characteristics</i>  <i>Time or length of response</i></p>
<p><b>2 FACETS OF THE TEST RUBRIC</b></p> <p><b>Test organization</b>  <i>Saliency of parts</i>  <i>Sequence of parts</i>  <i>Relative importance of parts</i></p> <p><b>Time allocation</b></p> <p><b>Instructions</b>  <i>Language (native, target)</i>  <i>Channel (aural, visual)</i>  <i>Specification of procedures and tasks</i>  <i>Explicitness of criteria for correctness</i></p>	<p><b>5 RELATIONSHIP BETWEEN INPUT AND RESPONSE</b></p> <p><b>Reciprocal</b>  <b>Nonreciprocal</b>  <b>Adaptive</b></p>
<p><b>3 FACETS OF THE INPUT</b></p> <p><b>Format</b>  <i>Channel of presentation (aural, visual)</i>  <i>Mode of presentation (receptive)</i>  <i>Form of presentation (language, nonlanguage, both)</i>  <i>Vehicle of presentation ('live', 'canned', both)</i>  <i>Language of presentation (native, target, both)</i>  <i>Identification of problem (specific, general)</i>  <i>Degree of speededness</i></p> <p><b>Nature of language</b>  <i>Length</i>  <i>Propositional content</i>          Vocabulary (frequency, specialization)          Degree of contextualization (embedded/reduced)          Distribution of new information (compact/diffuse)          Type of information (concrete/abstract, positive/negative, factual/counter-factual)          Topic          Genre</p> <p><i>Organizational characteristics</i>          Grammar          Cohesion          Rhetorical organization</p> <p><i>Pragmatic characteristics</i>          Illocutionary force          Sociolinguistic characteristics</p>	

Figure 5.1 Categories of test method facet

(Reproduced from Bachman, 1990: 119)

## Appendix B

# Sample Questions

With 200 questions, the TOEIC test measures a wide range of English proficiency. The following sample questions do not indicate the full range of difficulty you will find in an actual TOEIC test.

### General Directions

*The following general directions are taken directly from the test book. After the general directions are specific directions for each part of the test, along with sample questions for each of the parts.*

This test is designed to measure your English-language ability. The test is divided into two sections: Listening and Reading.

You must mark all of your answers on the separate answer sheet. For each question, you should select the best answer from the answer choices given. Then, on your answer sheet, you should find the number of the question and fill in the space that corresponds to the letter of the answer that you have selected. If you decide to change an answer, completely erase your old answer and then mark your new answer.

### Section I: Listening

In the Listening test, you will be asked to demonstrate how well you understand spoken English. The entire Listening test will last approximately 45 minutes. There are four parts, and directions are given for each part. You must mark your answers on the separate answer sheet. Do not write your answers in your test book.

### Part I: Photographs

**Directions:** For each question in this part, you will hear four statements about a picture in your test book. When you hear the statements, you must select the one statement that best describes what you see in the picture. Then find the number of the question on your answer sheet and mark your answer. The statements will not be printed in your test book and will be spoken only one time.

Look at the example item below.



#### Example

*You will hear:* Now listen to the four statements.

- (A) They're leaving the room.
- (B) They're turning on the machine.
- (C) They're standing near the table.
- (D) They're reading the newspaper.

Statement (C), "They're standing near the table," is the best description of the picture, so you should select answer (C) and mark it on your answer sheet.

## Sample Questions (continued)

## SAMPLE QUESTIONS

## Question 1

You will see:

1.



You will hear: 1. Look at the picture marked number 1 in your test book.

- (A) He's checking his watch.
- (B) He's wearing a jacket.
- (C) He's adjusting his tie.
- (D) He's folding his clothes.

## Question 2

You will see:

2.



You will hear: 2. Look at the picture marked number 2 in your test book.

- (A) She's speaking into a microphone.
- (B) She's putting on her glasses.
- (C) She's studying from a book.
- (D) She's using a microscope.

## Part 2: Question-Response

**Directions:** You will hear a question or statement and three responses spoken in English. They will not be printed in your test book and will be spoken only one time. Select the best response to the question or statement and mark the letter (A), (B), or (C) on your answer sheet.

## Example

You will hear: Where is the meeting room?

- You will also hear:
- (A) To meet the new director.
  - (B) It's the first room on the right.
  - (C) Yes, at two o'clock.

The best response to the question "Where is the meeting room?" is choice (B), "It's the first room on the right," so (B) is the correct answer. You should mark answer (B) on your answer sheet.

## SAMPLE QUESTIONS

## Question 1

You will hear: 11. Did you make a dinner reservation?

- (A) I prefer fish.
- (B) Flight 261 to Osaka.
- (C) Yes, it's at 7 o'clock.

## Question 2

You will hear: 12. Who takes the packages to the post office?

- (A) Turn left at the corner.
- (B) Martin usually does it.
- (C) No, I didn't.

## Question 3

You will hear: 13. This software is difficult to use, isn't it?

- (A) Yes, it's very complicated.
- (B) Yes, I often wear it.
- (C) No, but she used to.

## Question 4

You will hear: 14. Why don't you have a seat while you wait?

- (A) Two in the same row, please.
- (B) It didn't weigh very much.
- (C) Thanks, I think I will.

## Question 5

You will hear: 15. They've reduced the price of these cameras.

- (A) Then let's go ahead and buy one.
- (B) I don't think we've been introduced.
- (C) I never win anything.

## Appendix B

### Sample Questions *(continued)*

---

#### Part 3: Conversations

**Directions:** You will hear some conversations between two people. You will be asked to answer three questions about what the speakers say in each conversation. Select the best response to each question and mark the letter (A), (B), (C), or (D) on your answer sheet. The conversations will not be printed in your test book and will be spoken only one time.

#### SAMPLE QUESTIONS

*You will hear:* Questions 41 through 43 refer to the following conversation:

*(Woman)* I think I'll have to take the train to the regional sales meeting up in the city next week.

*(Man)* Don't you usually drive when you go to those meetings? I thought you didn't like to take the train.

*(Woman)* I don't, but the highway's being repaired, and I'm afraid I might be late if I have to make a detour through an area I don't know very well.

*(Man)* You're right. And it'll be expensive to park up there, too.

*You will then hear:* 41. Why is the woman going to the city?

*You will read:* 41. Why is the woman going to the city?

- (A) To attend a sale
- (B) To go to a meeting
- (C) To get her car repaired
- (D) To go on a tour

*You will hear:* 42. How will she get there?

*You will read:* 42. How will she get there?

- (A) By car
- (B) By bus
- (C) By train
- (D) By airplane

*You will hear:* 43. What is the problem?

*You will read:* 43. What is the problem?

- (A) The trains are often late.
- (B) The meeting may be canceled.
- (C) The tour is expensive.
- (D) The roads are being fixed.

#### Part 4: Talks

**Directions:** You will hear some talks given by a single speaker. You will be asked to answer three questions about what the speaker says in each talk. Select the best response to each question and mark the letter (A), (B), (C), or (D) on your answer sheet. The talks will not be printed in your test book and will be spoken only one time.

#### SAMPLE QUESTIONS

*You will hear:* Questions 71 through 73 refer to the following announcement.

*(Woman)* Thank you for calling Central Electric Services, your local power company. We are still experiencing problems with power failures caused by the recent storm. We expect service to be back up in most areas early this evening, although some homes in the northwest will be without electricity until tomorrow morning, and it may be tomorrow evening before service is fully restored in all areas. We apologize for the inconvenience this has caused our customers, and we thank you for your continued patience as we work to resolve these problems. This service announcement will be updated throughout the day. It was last updated at 6:00 A.M. on Sunday, April third.

*You will then hear:* 71. Where would this announcement be heard?

*You will read:* 71. Where would this announcement be heard?

- (A) On the television
- (B) On the radio
- (C) Over the telephone
- (D) In a company meeting

*You will hear:* 72. What is the purpose of the announcement?

*You will read:* 72. What is the purpose of the announcement?

- (A) To tell people about an approaching storm
- (B) To provide information about electric services
- (C) To apologize for staffing shortages
- (D) To describe the reorganization of a company

## Sample Questions (continued)

*You will hear:* 73. When is the problem expected to be completely resolved?

*You will read:* 73. When is the problem expected to be completely resolved?

(A) This morning  
(B) This evening  
(C) Tomorrow morning  
(D) Tomorrow evening

## Section II: Reading

In the Reading test, you will read a variety of texts and answer several different types of reading comprehension questions. The entire Reading test will last 75 minutes. There are three parts, and directions are given for each part. You are encouraged to answer as many questions as possible within the time allowed.

You must mark your answers on the separate answer sheet. Do not write your answers in your test book.

### Part 5: Incomplete Sentences

**Directions:** A word or phrase is missing in each of the following sentences. Four answer choices are given below each sentence. Select the best answer to complete the sentence. Then mark the letter (A), (B), (C), or (D) on your answer sheet.

### SAMPLE QUESTIONS

101. A late fee will be applied to your account \_\_\_\_\_ payment is not received by March 31.  
(A) and  
(B) whether  
(C) but  
(D) if
102. The award is given to an individual who has made \_\_\_\_\_ contributions to the community through volunteer work.  
(A) detailed  
(B) significant  
(C) secure  
(D) updated
103. Last year, Andrea Choi \_\_\_\_\_ the Choi Economic Research Center at Upton University.  
(A) to establish  
(B) established  
(C) was established  
(D) establishing
104. Ms. Ikeda and Mr. Arroyo are the final candidates under \_\_\_\_\_ for the position of director of development.  
(A) consideration  
(B) elimination  
(C) recognition  
(D) confirmation
105. Of the two animated films released today, \_\_\_\_\_ is certain to be popular with children, while the other will appeal more to adults.  
(A) neither  
(B) it  
(C) one  
(D) another

## Appendix B

### Sample Questions (continued)

#### Part 6: Text Completion

**Directions:** Read the texts that follow. A word or phrase is missing in some of the sentences. Four answer choices are given below each of the sentences. Select the best answer to complete the text. Then mark the letter (A), (B), (C), or (D) on your answer sheet.

#### SAMPLE QUESTIONS

Questions 141–143 refer to the following e-mail.

To: All Employees  
From: Camille Raynes  
Date: December 14  
Re: Performance bonus

Dear Employees,

As you know, the past year was a great success for us. To reward you for your excellent performance, the Board of Directors has approved a bonus for all employees. This bonus will be \_\_\_\_\_ in your next paycheck.

141. (A) involved  
(B) joined  
(C) composed  
(D) included

\_\_\_\_\_, we are now calculating wage increases for the upcoming year. Each employee's performance

142. (A) Instead  
(B) In addition  
(C) Beforehand  
(D) Otherwise

will be examined carefully as we determine the appropriate increase. All full-time employees are eligible for this increase. Your supervisor \_\_\_\_\_ you of the amount of your increase during the first week of January.

143. (A) informed  
(B) to inform  
(C) will inform  
(D) was informing

Thank you again for making last year such a success!

Sincerely,

Camille Raynes  
Human Resources

#### Part 7: Reading Comprehension

**Directions:** In this part you will read a selection of texts, such as magazine and newspaper articles, letters, and advertisements. Each text is followed by several questions. Select the best answer for each question and mark the letter (A), (B), (C), or (D) on your answer sheet.

#### SAMPLE QUESTIONS

Questions 153–154 refer to the following notice.

#### ATTENTION ART ENTHUSIASTS

Becksvooort Gallery is holding a

#### PUBLIC SALE

in the Grand Ballroom at the Hotel Luxe  
299 Norman Street, Melbourne

Come and see our range of exquisite pieces.  
Original oils \* Watercolors \* Tapestries \* Crafts

**Friday, July 15**

**8:30 A.M. – 4:00 P.M.**

For descriptions and pictures of select items  
visit our Web site at <http://www.becksvooort.com/market>.

153. What is the purpose of the notice?  
(A) To announce a sale of artwork  
(B) To advertise the opening of a hotel  
(C) To offer a discount on painting lessons  
(D) To publicize a photography exhibition
154. According to the notice, what can people do online?  
(A) Purchase selected items  
(B) Order tickets to an event  
(C) Register for art classes  
(D) View some pieces of art

## Appendix B

### Sample Questions *(continued)*

Questions 181–185 refer to the following letter and document.

January 15

Jeanne Sokol  
14 Jubilee Street  
Brighton

Dear Ms. Sokol:

I have received your letter of January 7 concerning your recent visit to our restaurant in London. I am very sorry that you did not have an enjoyable experience. I agree that you should not have had to wait over thirty minutes for a table when you had made a reservation in advance. Please know that this was an isolated incident and that our customers usually remark on the promptness of our staff.

In an attempt to remedy this unfortunate situation, I am sending you a complimentary voucher for use at any of our restaurants. I encourage you to try us again and experience our award-winning cuisine. If there is any way in which I can be of further assistance, please do not hesitate to contact me by telephone or mail.

Sincerely,

Simon Jenkins  
Director  
Guest Relations

### Winchester Falls Café

The finest dining in Britain!

**The bearer of this certificate is entitled to £25 off the price of a meal at Winchester Falls Café.**

Café Locations: London, Birmingham, Manchester  
*Expires: December 31*

For customer service or inquiries, contact Winchester Falls' Guest Relations Department by telephone at 0616 555-4239 or by mail at 77 Twickenham Road, Manchester WF1 JK2.

181. Why did Simon Jenkins write to Jeanne Sokol?
  - (A) To thank her for visiting a restaurant
  - (B) To apologize for poor service
  - (C) To inquire about her dining experience
  - (D) To invite her to an awards dinner
182. In the letter, the word "reservation" in paragraph 1, line 5, is closest in meaning to
  - (A) hesitation
  - (B) supply
  - (C) doubt
  - (D) appointment
183. What is suggested about Jeanne Sokol?
  - (A) She has recently traveled to Manchester.
  - (B) She spoke to Simon Jenkins on the telephone.
  - (C) She wrote a letter of complaint.
  - (D) She frequently dines in London.
184. What is Simon Jenkins offering?
  - (A) A refund on a previous purchase
  - (B) A discount on a future purchase
  - (C) Replacement of a damaged product
  - (D) Special assistance in ordering a product
185. Where does Simon Jenkins most likely work?
  - (A) In Manchester
  - (B) In Brighton
  - (C) In Birmingham
  - (D) In London

### Answer Key

PART 1	PART 2	PART 3	PART 4	PART 5	PART 6	PART 7
1. B	11. C	41. B	71. C	101. D	141. D	153. A
2. D	12. B	42. C	72. B	102. B	142. B	154. D
	13. A	43. D	73. D	103. B	143. C	
	14. C			104. A		181. B
	15. A			105. C		182. D
						183. C
						184. B
						185. A

## References

- Arita, E.** (2003) 'New tests challenging TOEIC stronghold.' *Japan Times Online*. Accessed online 20<sup>th</sup> September 2008, from <http://search.japantimes.co.jp/cgi-bin/nn20031025b3.html>.
- Bachman, L.** (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman L. & Palmer, A.** (1996) *Language Testing in Practice*. Oxford: Oxford University Press.
- Bradford-Watts, K., Ikeguchi, C. & Swanson, M. (Eds.)** (2006) *JALT 2005 Conference Proceedings*. Tokyo: JALT.
- Brown, J.D. & Yamashita, S.O. (Eds.)** (1995) *Language Testing in Japan*. Tokyo: JALT
- Brown, H.** (2006) 'Learner perceptions of TOEIC test results and language skill improvements: "I don't want to study English, I want to study TOEIC."' In Bradford-Watts et al. (2006), pp. 1176-1181.
- Brown, H.D.** (2001) *Teaching by Principles* (2<sup>nd</sup> edition). New York: Longman.
- Brown, H.D.** (2007) *Principles of Language Learning and Teaching* (5<sup>th</sup> edition). New York: Pearson.
- Chapman, M.** (2003) 'TOEIC<sup>®</sup>: Tried but undertested.' *Shiken: JALT Testing & Evaluation SIG Newsletter*, Vol. 7. No. 3, Autumn 2003, pp. 2-5.
- Chapman, M.** (2004) 'Voices in the field: An interview with Kazuhiko Saito.' *Shiken: JALT Testing & Evaluation SIG Newsletter*, Vol. 8. No. 2, August 2004, pp. 10-12.
- Chapman, M. & Newfields, T.** (2008) 'Opinion Piece: The New TOEIC<sup>®</sup>.' *Shiken: JALT Testing & Evaluation SIG Newsletter*, Vol. 12. No. 2, April 2008, pp. 30-35.
- Chauncey Group International Ltd.** (1998) *TOEIC<sup>®</sup> Technical Manual*. Accessed online 4<sup>th</sup> September 2008, from [http://www.toeic.cl/images/toeic\\_tech\\_man.pdf](http://www.toeic.cl/images/toeic_tech_man.pdf).
- Childs, M.** (1995) 'Good and bad uses of TOEIC by Japanese companies. In J.D. Brown & S.O. Yamashita (Eds.) (1995), pp. 66-75
- Cunningham, C.** (2002) 'The TOEIC test and Communicative Competence: Do Test Score Gains Correlate With Increased Competence?' *Unpublished Masters dissertation*. Accessed online 20<sup>th</sup> September 2008, from [www.cels.bham.ac.uk/resources/essays/Cunndiss.pdf](http://www.cels.bham.ac.uk/resources/essays/Cunndiss.pdf) .



- Douglas, D.** (2000) *Assessing Languages for Specific Purposes*. Cambridge, England: Cambridge University Press.
- Educational Testing Service** (2006a) *FAQ TOEIC® Europe*. Accessed online 13<sup>th</sup> September 2008, from [http://www.eu.toeic.eu/index.php?id=3058&no\\_cache=1&print=1&](http://www.eu.toeic.eu/index.php?id=3058&no_cache=1&print=1&)
- Educational Testing Service** (2006b) *TOEIC and Language Schools*. Accessed online 13<sup>th</sup> September 2008, from <http://www.eu.toeic.eu/toeic-sites/toeic-europe/language-schools/>
- Educational Testing Service** (2007) *TOEIC User Guide – Listening & Reading*. Educational Testing Service.
- Educational Testing Service** (2008) *The Redesigned TOEIC® Listening and Reading Test*. Accessed online 14<sup>th</sup> September 2008, from <http://www.ets.org/portal/site/ets/menuitem.22f30af61d34e9c39a77b13bc3921509/?vgnextoid=1b5ed87dcc9ae010VgnVCM10000022f95190RCRD>.
- Hughes, A.** (2003) *Testing for Language Teachers* (2<sup>nd</sup> edition). Cambridge: Cambridge University Press.
- Institute for International Business Communication** (2006). *TOEIC Newsletter* No.92, April 2006. Institute for International Business Communication.
- Institute for International Business Communication** (2007). *TOEIC Score Descriptor Table*. Accessed online 18<sup>th</sup> September 2008, from [http://www.toeic.or.jp/toeic\\_en/toeic/score\\_descriptors.html](http://www.toeic.or.jp/toeic_en/toeic/score_descriptors.html)
- Institute for International Business Communication** (2008). *TOEIC Test Data and Analysis 2007*. Accessed online 18<sup>th</sup> September 2008, from [http://www.toeic.or.jp/toeic\\_en/media/pdf/TOEIC\\_DAA2007.pdf](http://www.toeic.or.jp/toeic_en/media/pdf/TOEIC_DAA2007.pdf).
- Jafarpur, A.** (1987) ‘The short-context technique: an alternative for testing reading comprehension.’ *Language Testing* 4 (2), pp. 195-220.
- Miller, K.** (2003) ‘The Pitfalls of Implementing TOEIC Preparations Courses.’ *Yoshinaga Review*. Accessed online 22<sup>nd</sup> September 2008, from <http://www2.shikoku-u.ac.jp/english-dept/pitfalls.html>.
- Miyata, R.** (2004) ‘A Review of TOEIC: A Critical Analysis of the Structure, Methods, and Purpose of TOEIC, Including its Effect on Teaching and Learning English as a Second or Foreign Language’. *Bulletin of Beppu University Junior College* 23, pp. 59-69.

- Moritoshi, P.** (2001) 'The Test of English for International Communication (TOEIC): necessity, proficiency levels, test score utilisation and accuracy.' *Unpublished essay*. Accessed online 24<sup>th</sup> September 2008, from [www.cels.bham.ac.uk/resources/essays/Moritoshi5.pdf](http://www.cels.bham.ac.uk/resources/essays/Moritoshi5.pdf)
- O'Sullivan, B.** (2006) *Issues in Testing Business English (Studies in Language Testing 17)*. Cambridge: Cambridge University Press.
- Oller, J.W.** (1979) *Language Tests at School*. London: Longman.
- Owen, C. et al** (2001) *Testing*. Centre for English Language Studies, Birmingham University.
- Rebuck, M.** (2003) 'National Standards. The Use of TOEIC by Companies in Japan.' *NUCB Journal of Language, Culture and Communication*, Vol. 5, No.1, 23-32. Accessed online 8<sup>th</sup> September 2008, from <http://www.nucba.ac.jp/cic/pdf/njlcc051/02REBUCK.PDF>.
- Ross, S.J.** (2008) 'Language testing in Asia: Evolution, innovation, and policy challenges.' *Language Testing* 25 (1), pp. 5-13.
- Shimizu, K.** (2006) 'TOEIC revisions mean big change in English study.' *The Japan Times Online*. May 1<sup>st</sup>, 2006. Accessed online 12<sup>th</sup> September 2008, from <http://search.japantimes.co.jp/cgi-bin/nn20060501f1.html>.
- Woodford, P. E.** (1982) *TOEIC Research Summaries - An Introduction to TOEIC: The Initial Validity Study*. Educational Testing Service.