Correlation between STEP BULATS Speaking and TOEIC® Scores

by Michihiro Hirai (Waseda University and Kanagawa University)

Abstract

The author analyzed the results of STEP BULATS Speaking Tests administered to a group of 493 Japanese (predominantly businesspeople) from April 2005 to January 2009 and found the correlation coefficient between their scores and the TOEIC® scores to be **.66.** Moreover, the STEP BULATS Speaking levels observed were appreciably lower than those expected of competent international businesspeople, most notably in the upper-level group. The author attributes the relatively low performance in the STEP BULATS Speaking Test primarily to the test-takers' lack of experience and/or training in real-life business speaking skills.

Keywords: test score correlation, TOEIC, BULATS, business speaking test, business speaking skills

要旨

STEP BULATS スピーキング・テストを 2005 年 4 月から 2009 年 1 月に受験した日本人（主に社会人）のうち TOEIC® も受験した 493 人について、両テスト点数の相関を調べ、相関係数 **.66** を得た。さらに、STEP BULATS スピーキング成績が、国際ビジネスパーソンに期待されるレベルよりかなり低いこと、またその傾向が特に上級者の間で顕著であることが判明した。筆者はそれが、ビジネス現場に即したスピーキングの経験または訓練の欠如ないし不足に起因すると考える。

キーワード: 試験点数の相関、TOEIC、BULATS、ビジネス・スピーキング・テスト、ビジネス・スピーキング能力

In recent years, the pressing need to improve the English skills of corporate employees amid the accelerating trend toward globalization of the economy has boosted public interest in English tests, especially those which are designed, or claim, to measure work-place or business English skills. Despite the growing interest, however, there still remains a significant gap in realization among the public between the skills really needed in business and the skills actually tested in most companies. Whereas an increasing number of people realize the importance of productive speaking and writing skills in real-life business situations (Hirai, 2002a, p. 34), the vast majority of business entities, especially those in Japan, still rely heavily on receptive (passive) skill tests, most notably the Test of English for International Communication (TOEIC).

This gap in perception seems to result partly from the dearth of reports highlighting the inappropriateness of depending on receptive skill tests as the sole measure of employees' language performance in actual business situations, where language production and familiarity with business vocabulary play more significant roles. In the first place, the amount of statistical data showing the degree of correlation between receptive skill test scores and productive skill test scores

has been rather limited. Taking, for example, the correlation between the TOEIC scores and the speaking skills, most of the reports published so far originate from Educational Testing Service (ETS), the organization which developed and administers the TOEIC. The main reason why such reports are limited in number appears to be, quite understandably, that it is difficult for any organization other than the one which administers a test to collect statistically meaningful score data. Among the few reports on the issue published independently is one regarding the correlation between the scores of a company-internal speaking test and the TOEIC (Hirai, 2002b). Let us now take a closer look at some of the TOEIC research by the ETS and other sources.

Review of Prior Research

ETS has published a series of reports on the correlation between direct speaking measures, especially LPI ratings, and TOEIC scores, as summarized in Table 1. Note that LPI stands for Language Proficiency Interview, a criterion-referenced direct measure of oral language proficiency developed by the Foreign Service Institute of the U.S. Department of State. The test-taker's speaking proficiency is reported on a scale of 0 to 5, with augmentations ("+") added to 0, 1, 2, 3, and 4 to indicate a level of performance close to the next level up. Further, although not shown in Table 1, the *TOEIC Technical Manual* (The Chauncey Group International, Ltd. 1998) also cites Wilson's data (Wilson, 1993) (#3 in Table 1).

Table 1
Correlation Coefficients between LPI Ratings and TOEIC Scores (Reported by ETS)

| #	Author	Year	N	Countries	Cor. Coef.	with:	Page(s)
1	P. E. Woodford	1982	100	J	.83	L	9-15
2	K. Wilson	1993	285	J	.76	T	6
3			393	J, F, M, SA	.74	T, L	8
4	K. Wilson	1999	56	F	.65	T	14
5			1045	J, F, M, SA, K, T	.70	T	18
6	K. Wilson	2001	149	K	.69	T	10
7			738	J, F, M, SA, K, T	.61	T	22

Country Code: J: Japan, F: France, M: Mexico, SA: Saudi Arabia, K: Korea, T: Thailand
Legend ("with"): L: Listening, T: Total

The correlation coefficient between the two measures varies fairly significantly, from a low of **.61** to a high of **.83**. Generally, in analyzing and interpreting the score data of any language test, one needs to exercise caution as to how closely the sample (the group of people who actually took the test) represents the population (the entire group of people who are or will be under consideration), since any mismatch between them has a potentially significant impact on the validity of the data in this context. Thus, it is critically important to check how the test-takers were selected, especially when there were relatively few and also during the early phase of introduction of the test. It would therefore be worthwhile to review ETS's data given in Table 1 from this perspective.

In general, evaluation studies of a language test, especially during the early phase of its introduction, may tend to have an inherent bias toward collecting higher scores. The author is tempted to call this phenomenon an "initial-survey bias." This bias is likely due to sampling error and is considered distinct from any expectancy effect. A major factor contributing to the initial-survey bias would be the process of selecting the subjects (i.e., sampling). When the test organization wishes to conduct an evaluation study of a newly-introduced test, it needs to find schools or other organizations which are eager or willing to extend cooperation.

In the case of an English test, they are typically language schools or departments of universities with a number of students majoring in English or seriously engaged in English lessons, or companies with a number of employees already engaged in international business. In either case, they are likely to be above average in terms of English knowledge and skills and therefore should not be considered to represent the general public, especially in a country such as Japan where English has a special status because the vast differences between English and the native language serve as a significant social barrier. As a result, the data collected in such studies tends to be skewed toward the higher end, which effectively increases the correlation coefficient. As the test becomes more popular over time, with increases in the diversity and number of test-takers, the statistics tend to become more representative of the entire population.

A case in point is Wilson's data in Table 1 (Wilson, 1993), which is cited in many other ETS reports. One hundred and sixty-three out of the 285 subjects in his study were Japanese at a graduate-level business school near Tokyo. They were spending all day studying international business and trade in a closed residential environment. Their average Total TOEIC scores ranged from 608 to 640, 10% to nearly 20% higher than the national average. Moreover, their average LPI levels ranged from 1.78 to 1.93. Wilson (1993, p.8) also shows the distribution of the subjects' LPI levels (see Figure 1). Figure 2, on the other hand, shows the LPI level distribution among Japanese test-takers for 2008 (The Institute for International Business Communication, 2009). This can be considered a typical distribution for an ordinary year (i.e., not in an early phase of introduction).

Mean LPI Level = **1.86**   N = 285
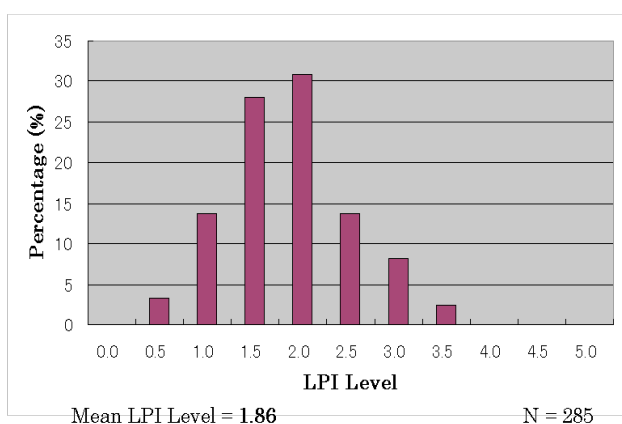
The test takers' TOEIC (Total) scores ranged from below 300 to over 900, with a mean of 618.

Figure 1. LPI Level Distribution among Japanese Test-takers (Wilson 1993)

Mean LPI Level = **1.21**   N = 485
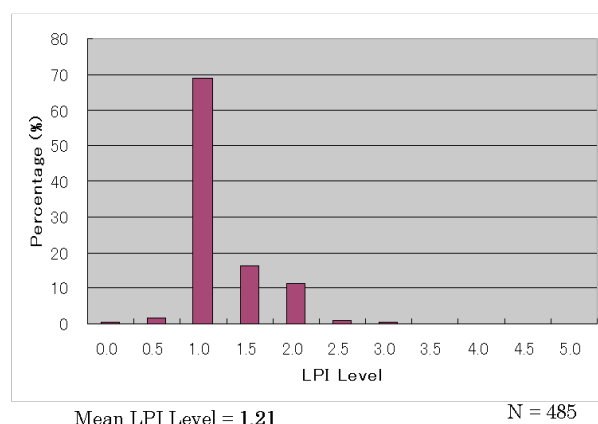
Source: LPI Score Report for Individual Test Takers FY2008, The Institute for International Business Communication (IIBC), Japan

Figure 2. LPI Level Distribution among Japanese Test-takers (Open Test FY2008)

A comparison of Figures 1 and 2 reveals a striking difference between the mean LPI scores of the two samples. The difference is so great as to prompt questions such as whether the LPIs used in the two sets of data were really the same and whether the subjects in Wilson's survey were representative of the general public. Assuming that the LPI's reliability is reasonably high, let us focus on differences, if any, between the two groups of test-takers.

The test-takers in the former group demonstrated markedly better performance, both in terms of the mean (1.86 vs. 1.21) and the distribution pattern, than those in the latter. These differences suggest that many of the test-takers in Wilson's 1993 report belonged to an extraordinary group of English speakers with significant amounts of training in speaking skills offered at the graduate-level business school. Generally, the presence of an exceptionally high-performing group of a disproportionate size in the sample tends to distort it by inflating the correlation coefficient (see the argument below about Woodford's study). Conversely, if we have an average group of test-takers selected randomly from the general public, we should expect a somewhat lower correlation coefficient than that (.76) claimed in Wilson's 1993 report.

Similarly, it would be advisable to exercise caution when interpreting the results of ETS's very first survey (Woodford, 1982), by considering how this kind of survey is usually initiated and how its participants are selected in Japan.

Furthermore, Woodford's 1982 study has a few peculiarities worth investigating. First, the data set used in his initial validity study of the correlation between LPI scores and TOEIC Listening scores was not randomly selected from the original sample of 2,710 subjects who took the TOEIC test in December 1979. Instead, he used a fixed convenience sample of 500 examinees, with 100 at each of the following five approximate score levels: 950, 765, 580, 315 and 45 (Woodford, 1982, p. 9). These groupings are neither contiguously nor equidistantly spaced. (Note that he then narrowed down the examinees to 100 using a procedure not explained.) Selecting the same number of people at each score bracket effectively flattens the score distribution, weighing the higher and lower ends disproportionately high. Such a manipulation generally serves to inflate the correlation coefficient. Furthermore, in calculating the correlation coefficient, he recorded a plus in the LPI level as 0.7, whereas most other ETS reports appear to record it as 0.5. This difference in the treatment of a plus in the LPI level also affects the resulting correlation coefficient, albeit to a lesser extent.

All in all, one should keep in mind that statistical data represents nothing other than the sample, instead of the entire population from which the sample was taken. Therefore, before making any generalization based on statistical data, one should carefully examine how closely (i.e., how free from biases) the sample represents the population, by analyzing differences in characteristics between the two. According to classical test theory, generalization is considered credible only when (a) the sample is large enough in statistical terms and (b) the sample replicates the characteristics of the population without bias. To meet the second requirement (b), random sampling should be employed. The sampling procedures employed in ETS's initial studies, however, appear to be far from random, as illustrated above. Therefore, one should take ETS's early claims with a grain of salt. Thus, by taking the apparently biased data out of consideration, it would be reasonable to

assume from Table 1 that the correlation coefficient between LPI ratings and TOEIC (Total) scores is likely to lie between **.60** and **.70** for most Japanese businesspersons.

Meanwhile, to provide more independent data on this topic, Hirai (2002b) conducted a statistical study on the correlation between productive skills as measured by an in-house test and TOEIC scores. Based on data collected at an internal corporate language institute, he reported a correlation coefficient of **.78** between the institute's interview test and TOEIC scores ($n = 475$). The interview test employed at the time consisted of 10 questions of varying degrees of difficulty on familiar or general social issues and used a set of evaluation criteria mainly focusing on linguistic features such as vocabulary, grammar, pronunciation, aural comprehension, and fluency. It appeared to be less rigorous than the LPI and to produce a much more linear distribution of scores on a scale of 0 to 100, in other words closer to a normal distribution, than the LPI, which tended to produce score distribution patterns that were significantly skewed toward the lower end for the research sample used, as shown in Figure 2. Whereas the correlation coefficient appeared to be relatively high, Hirai offers a caveat as to its heavy dependency on demographic sampling, i.e., the characteristics of specific group of test-takers (Hirai, 2002b, pp. 5-6). In the present study the author is going to investigate how well business-speaking skills as measured by the STEP BULATS Speaking Test run parallel with TOEIC Test performance. In addition, he will attempt to shed light on plausible reasons for the relatively weak and significantly varying performance in business speaking, especially among advanced learners of English in Japan.

Evaluation Methods Employed

The author analyzed the score data of 779 individuals (original sample) collected by The Society for Testing English Proficiency (STEP) who had taken both the TOEIC and the STEP BULATS Speaking Test between April 2005 and January 2009.

The STEP BULATS Speaking Test is a part of the BULATS Test Suite developed by the University of Cambridge Local Examinations Syndicate (UCLES), which is designed to test the foreign language proficiency of individuals in English, German, French, or Spanish. The emphasis is on measuring the test-taker's English skills in work situations and environments, a significant departure from general English tests. The Speaking Test consists of three parts: interview, presentation, and information exchange and discussion. A more detailed description of the BULATS Test Suite can be found in *BULATS Candidate Handbook* (Business Language Testing Service 2007). Appendix A also presents two sample Speaking Test tasks. It should be noted that the BULATS Test Suite is now administered in Japan by STEP as the STEP BULATS Test Suite.

In compiling meaningful score data, care was taken to eliminate items which can be considered practically duplicate, namely the scores of the same individuals (i.e., repeaters) recorded more than once within the timeframe in question, in which case the best STEP BULATS Speaking level (score) was taken and the rest were discarded. Also, out of consideration for reliability, the data of individuals who took the STEP BULATS Speaking Test and the TOEIC Test more than 24 months apart were eliminated.

As a result, the score data of 493 test-takers out of the entire sample were finally selected for this study. Of the total of 493, 118 were female; 92 were college students and 401 were adult non-students (mostly employed by various organizations including local as well as central governments); 224 were in their 20s, 199 were in their 30s, and 70 were over 39. While the test-takers were not asked to reveal their nationalities, the overwhelming (more than 90%) majority were considered Japanese.

The results of the STEP BULATS (as well as the BULATS) Speaking Test are reported in discrete levels ranging from 0 to 5 (more specifically, from levels 0, 0+, 1–, 1=, 1+ to levels 5–, 5=, and 5+) instead of scores. For the sake of statistical handling, these levels are converted to numbers by assigning –0.3 to "–" and +0.3 to"+": for example, Level 1+ is represented as 1.3 and Level 5– as 4.7.

Results

Figure 3 illustrates how the BULATS Speaking level and the TOEIC score are distributed in a two-dimensional space, with the two variables represented by the vertical and horizontal axes. The correlation coefficient was found to be **.66** (between .61 and .71 at a confidence level of 95%), which is considerably lower than the correlation coefficient of **.83** reported by Woodford (1982; see Table 1 #1) but which fits in well with the range of **.61** to **.70** reported by Wilson (1999 and 2001; see Table 1 #4 – #7).
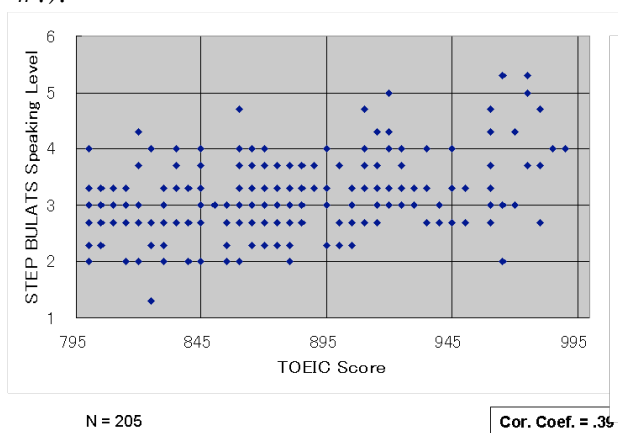

N = 205   Cor. Coef. = .39

Figure 5. Correlation between STEP BULATS Speaking Levels and TOEIC Scores (TOEIC ≥ 800; predominantly Japanese businesspersons)
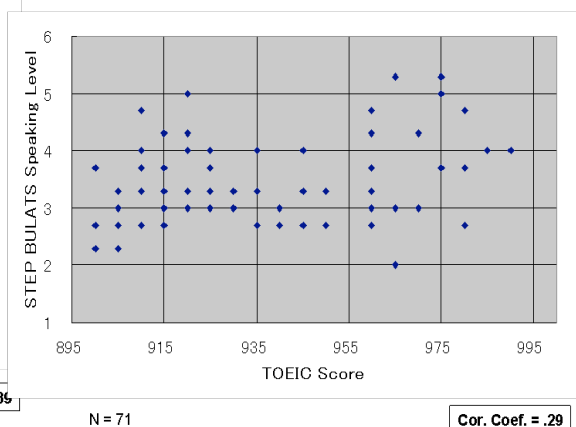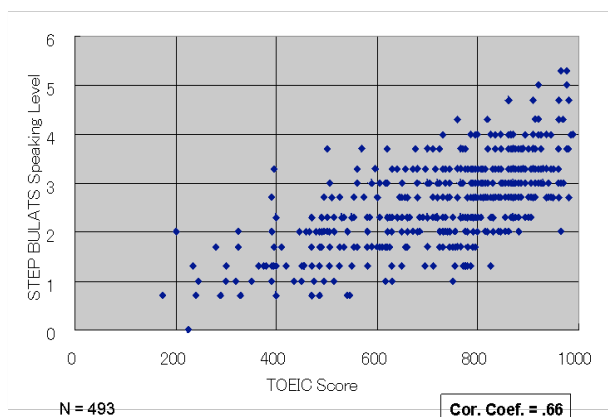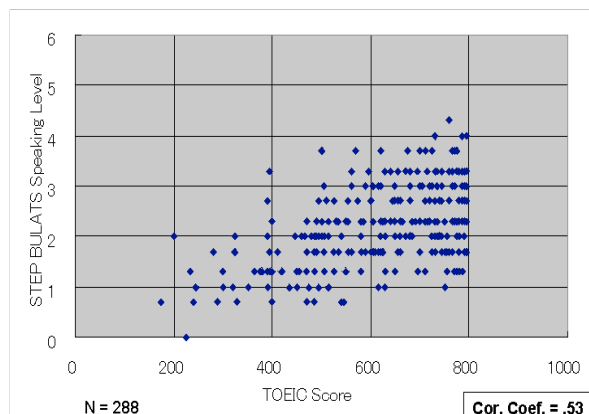
N = 71   Cor. Coef. = .29

Figure 6. Correlation between STEP BULATS Speaking Levels and TOEIC Scores (TOEIC ≥ 900; predominantly Japanese businesspersons)

When the entire sample was divided into two groups according to TOEIC scores, the correlation coefficient tended to decrease for each group. For example, when the entire sample was divided at a TOEIC cut-off point of 800, the correlation coefficient decreased somewhat to .53 (between .44 and .61 at a confidence level of 95%) for the group with TOEIC scores below 800 and it decreased significantly to **.39** (between .27 and .50 at a confidence level of 95%) for the group with TOEIC scores at or above 800, as shown in Figures 4 and 5. Furthermore, when the cut-off point was raised to a TOEIC score of 900, the correlation coefficient dropped to as low as **.29** for the top group (with TOEIC scores more than or equal to 900) as shown in Figure 6.

17

Figure 3. Correlation between **STEP BULATS Speaking Levels**
**and TOEIC Scores** (overall; predominantly Japanese businesspersons)

Figure 4. Correlation between **STEP BULATS Speaking Levels**
and TOEIC Scores (TOEIC < 800; predominantly Japanese businesspersons)

Figures 7 and 8 illustrate the distributions of STEP BULATS Speaking levels at TOEIC scores of 800 and 900, respectively. As can be seen, the STEP BULATS Speaking levels varied significantly even at these TOEIC scores, which are considered very high in terms of general English.
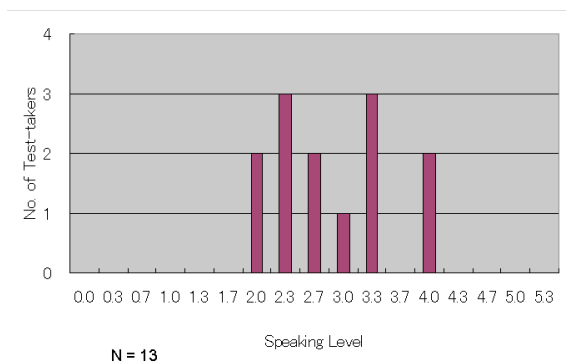
Figure 7. Distribution of STEP BULATS Speaking Levels (TOEIC = 800)
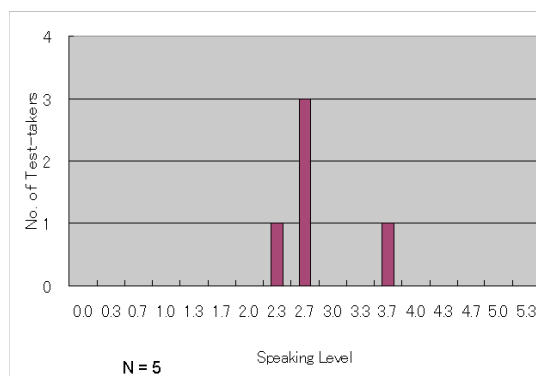
Figure 8. Distribution of STEP BULATS Speaking Levels (TOEIC = 900)

To check the effect of non-uniform sampling as mentioned earlier, the author conducted two experiments on how pruning the present sample would affect the correlation coefficient. In Experiment A, in an attempt to replicate Woodford's procedure of selecting his sample (Woodford, 1982) as closely as possible, the author randomly selected 20 subjects in the TOEIC score bracket of 900 to 990, 20 in the bracket of 715 – 810, 20 in the bracket of 530 – 625, and 10 in the bracket of 265 – 360, so that the brackets roughly corresponded to the "approximate score levels of 950, 765, 580, 315, 45" (Woodford, 1982, p. 9), which are neither contiguously nor equidistantly spaced. As a result, the correlation coefficient increased from **.66** (n = 493) to **.75** (n = 71).

In Experiment B, the author randomly selected 20 subjects each in the contiguously and equidistantly spaced TOEIC score brackets of 900 – 990, 800 – 895, 700 – 795, 600 – 695, 500 – 595, 400 – 495, and 300 – 395 and included all the eight subjects below 300 as they were, to make the distribution as flat as possible, and recorded a correlation coefficient of **.79** (n = 148). These results suggest that pruning the sample in such a way as to flatten the score distribution pattern tends to significantly distort (*inflate* in the present case) the correlation coefficient.

Analysis and Discussion

Analysis of Correlation Coefficients

Based on the data thus obtained, a few observations are in order. To begin with, the fat spindle shape of Figure 3 indicates that the two variables, the BULATS Speaking level and the TOEIC score, are not tightly related to each other. Indeed, the overall correlation coefficient of **.66** should not be considered high enough to justify the use of TOEIC Test scores as a meaningful indicator of business speaking skills as measured by the BULATS.

Among individuals with high TOEIC scores (above or equal to 800 or even more so with 900), the correlation tended to become progressively weaker. These drops in correlation coefficient for test-takers with high TOEIC scores, however, should not be conclusively interpreted by themselves as a sign of the correlation being particularly weaker among high-performing groups, since the correlation coefficient tends to decrease when the sample is split according to the value of one of the two variables in question (in this case the TOEIC score), as demonstrated by Hirai (Hirai, 2008, pp. 41–43).

Viewed from another angle, the STEP BULATS Speaking levels for high-performing groups spanned broad ranges, i.e., from 1+ to 5+ for the group with TOEIC scores of 800 and above, or from 2= to 5+ for the group with TOEIC scores of 900 and above. At TOEIC cut-off points of 800 and 900, the test-takers exhibited widely ranging STEP BULATS Speaking levels, as Figures 7 and 8 attest. These wide variances partially explain the low correlation coefficients and suggest, within narrow score ranges, that the TOEIC score does not particularly indicate the speaking skills that the STEP BULATS Speaking Test is designed to measure, especially at advanced levels.

As demonstrated by Experiments A and B, pruning the sample in such a way as to flatten the score distribution pattern tends to significantly distort (*inflate* in the present case) the correlation coefficient. Therefore, as touched on in Review of Prior Research, ETS's initial validation survey (Woodford, 1982) should not be taken at face value. If we eliminate Woodford's and Wilson's early data (Wilson, 1993), which appear to be somewhat affected by the initial-survey biases, the correlation coefficient of **.66** obtained in the present study is in fact in good agreement with most other reports from ETS (Wilson, 1999 & 2001) (#4 – #7 in Table 1), although the STEP BULATS Speaking test and LPI are substantially different in format, content, and evaluation criteria.

Apart from the nominal differences gleaned from a study of their respective manuals, namely *BULATS Candidate Handbook* (2007) and *Language Proficiency Interview: Manual for Testers* (2005), the author's own experience of taking both tests suggests that, compared with other similar tests, the former seems to place emphasis on task accomplishment, whereas the latter seems to be extremely meticulous about linguistic features such as pronunciation and fluency (in other words, how closely the test-taker sounds like an average native speaker in these respects). This may partially explain the low-end skew of the LPI level distribution in Figure 2. These features, though distinctly different from each other, seem to be contributing to their relatively weak correlations with the TOEIC scores, which reflect primarily the test-taker's receptive skills including vocabulary and

grammar knowledge.

STEP BULATS Speaking Test Performance Levels

Figure 9 is based on Figure 3, with a regression line and another line added that indicates the level desired for international businesspeople. The regression line represents the center line around which the population of the sample is balanced, and thus can be considered, for a given value of one variable, to represent the average expected value of the other variable. It should be noted that the regression line has a slope that is significantly gentler than the diagonal line connecting the lowest possible values to the highest possible values of the two variables (the STEP BULATS Speaking Test level and the TOEIC score), which can be regarded as a balanced performance line (Note that the word "balanced" is used here to mean the speaking skill is proportionate to the receptive skill as represented by the TOEIC score). It is clear from Figure 9 that the average speaking levels were significantly lower than those expected from the balanced performance line, most notably among test-takers with high TOEIC scores.
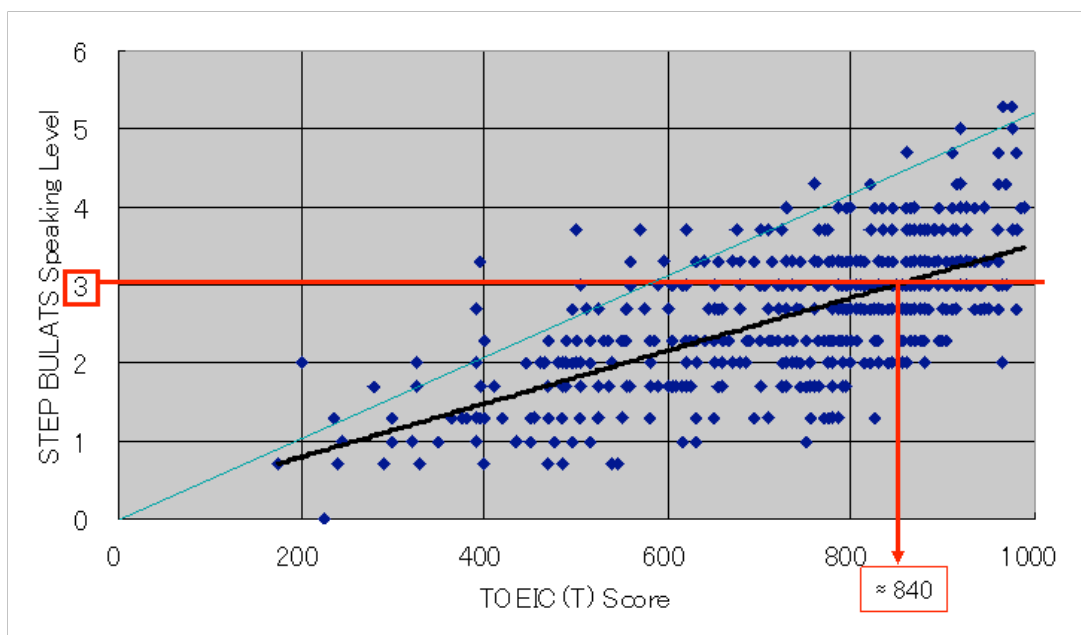
Figure 9. Regression Line between STEP BULATS Speaking Levels and TOEIC Scores

Koike et al. (2008) point out that the CEFR level (Council of Europe, 2001) most often cited in a recent poll of 7,354 Japanese people as the minimum level desired for competent international businesspeople is B2, which corresponds closely to Level 3 in the STEP BULATS Test (BULATS, 2007). This level typically enables the test-taker to "deal with clients and resolve most problems in their own field" (STEP, 2004). From Figure 9, the average TOEIC score corresponding to Level 3 was found to be about 840. Similarly, Figure 5 suggests that the majority (54.6%) of test-takers with TOEIC scores 800 and above were rated Level 3 or below (35.6% were below Level 3).

The regression line in Figure 9 was found to be flatter than that for the relationship between the STEP BULATS Standard Test scores and the TOEIC Test scores (Hirai, forthcoming), indicating that the downward deviation from the balanced performance line, especially among the holders of high TOEIC scores (advanced learners), was significantly greater for the STEP BULATS Speaking Test than for its receptive skills test. The author attributes this tendency to the fact that business speaking is a composite skill involving a variety of subskills and kinds of language knowledge, such as logical thinking, rhetoric, organization, and a large active vocabulary. In addition, as shown in Appendix A, the STEP BULATS Speaking Test places emphasis on task accomplishment, which is an important factor in real-life business and which calls for extra attention beyond mere language skills. All of these demand more experience and ad-hoc training than the sort of receptive skills measured by the TOEIC.

Conclusions

STEP BULATS Speaking levels were found to correlate moderately with TOEIC scores, with a correlation coefficient of **.66** for the entire sample (*N* = 493). This relatively low correlation coefficient compared with the author's previous report (Hirai, 2002b) on the correlation between general speaking test scores and TOEIC scores (a correlation coefficient of **.78**) can be attributed to the fact that the STEP BULATS Speaking Test emphasizes the skills of speaking in business contexts to accomplish various tasks, involving, in addition to pure linguistic skills, some experience and/or training in real-life business situations.

If we ignore ETS's early reports, which appear to be considerably affected by initial-survey biases, the correlation coefficient obtained in the present study agrees with ETS's other reports suggesting that correlation coefficients between direct speaking measures and TOEIC scores generally lie between **.60** and **.70**. This similarity can be considered coincidental, since the direct speaking measures used in ETS's studies were not as specific to the business domain as the STEP BULATS Speaking Test is.

The author considers the overall correlation coefficient of **.66** not high enough to justify the use of the TOEIC Test score alone as a meaningful indicator of business speaking skills. The inappropriateness of the use of the TOEIC Test score for such a purpose is

> *"the overall correlation coefficient of **.66** not high enough to justify the use of the TOEIC Test score alone as a meaningful indicator of business speaking skills."*

also evident from the wide spreads of the BULATS Speaking levels at certain (particularly high) cut-points of the TOEIC scores (e.g., 800 and 900).

It was also revealed that the STEP BULATS Speaking levels observed were appreciably lower than those expected of competent international businesspeople, most notably in the upper-level group. The author attributes the test-takers' relatively low performance in the STEP BULATS Speaking Test, especially toward the higher end of the spectrum, also to the content and format of the STEP BULATS Speaking Test emphasizing the skills of speaking in business contexts to accomplish various tasks as mentioned above, and further to their general lack of experience and/or training in those skills in real-life business situations.

The author suggests that in assessing business speaking skills, a test designed specifically for the business community, with practical business situations in mind, be employed, instead of general-purpose tests focusing on receptive skills such as the TOEIC.

References

Business Language Testing Service (BULATS), University of Cambridge ESOL Examinations. (2008). *The BULATS test.* Retrieved July 3, 2008 from http://www.bulats.org/tests/index.php

Business Language Testing Service (BULATS), University of Cambridge ESOL Examinations. (2007). *BULATS candidate handbook.* Retrieved July 3, 2008 from http://www.bulats.org/handbook/index.php

Business Language Testing Service (BULATS), University of Cambridge ESOL Examinations. *Speaking test sample question paper.* Retrieved June 7, 2005 from http://www.bulats.org/samples/index.php

Chauncey Group International. (1998). *TOEIC Technical Manual.* Princeton, NJ: Author.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press.

Hirai, M. (2002a). *Bijinesupaason no tame-no eigo chou-kouritsu benkyouhou.* [Maximally Efficient Approaches to Studying English for Businesspeople]. Tokyo: Nihon Jitsugyo Shuppan.

Hirai, M. (2002b). Correlations between active skill and passive skill test scores. *Shiken: JALT Testing & Evaluation SIG Newsletter, 6 (3)*, 2–8. Retrieved June 12, 2008 from http://jalt.org/test/hir_1.htm

Hirai, M. (2008). Correlation between STEP BULATS writing and TOEIC® scores. In T. Newfields, P. Wanner, & M. Kawate-Mierzejewska (Eds.), *Proceedings of 7th Annual JALT Pan-SIG Conference.* Kyoto, Japan: Doshisha University Shinmachi Campus. Retrieved July 8, 2009 from http://jalt.org/pansig/2008/HTML/Hirai1.htm

Institute for International Business Communication. (2009). *Tesuto ninteishou atacchi letaa 2008 nendo deeta.* [LPI score report for individual test takers FY2008]. Tokyo: Author.

Koike, I. et al. (2008). *Kigyou ga motomeru eigo-ryoku chousa houkokusho: Daini gengo shuutoku kenkyuu wo kiban to suru shou, chuu, kou, dai no renkei wo hakaru eigo kyouiku no sendou-teki kiso kenkyuu (Kenkyuu kadai bangou 16202010: Heisei 16 nendo – Heisei 19 nendo kagaku kenkyuu-hi hojokin (Kiso kenkyuu (A))).* [Report on English skills required by companies: Pioneering basic research on English education aimed at providing coordination between elementary schools, junior high schools, high schools, and universities on the basis of research on second language acquisition (Research project No. 16202010: Scientific research subsidy for fiscal years 2004–2007 (Basic research (A)))]. Faculty of Languages and Cultures and Graduate School of Applied Linguistics, Meikai University.

Peace Corps (2005). *Language Proficiency Interview: Manual for testers* (Information collection and exchange publication No. T0130), Retrieved May 16, 2009 from
http://multimedia.peacecorps.gov/multimedia/pdf/library/T0130_English_LPI.pdf

Society for Testing English Proficiency. (2004). *Sekai kijun no bijinesu eigo nouryoku tesuto STEP BULATS* [Global-standard business English test STEP BULATS]. Tokyo: Author.

Wilson, K. (1993). *Relating TOEIC scores to Oral Proficiency Interview ratings.* Princeton, NJ: Educational Testing Services. Retrieved June 6, 2009, from
http://www.ets.org/research/researcher/TOEIC-RS-01.html

Wilson, K. (1999). *Validity of global self-Ratings of ESL speaking proficiency based on an FSI/ILR-referenced scale.* Princeton, NJ: Educational Testing Service. Retrieved June 6, 2009, from
http://www.ets.org/research/researcher/RR-99-13.html

Wilson, K. (2001). *Overestimation of LPI ratings for native-Korean speakers in the TOEIC testing context: Search for explanation*, Princeton, NJ: Educational Testing Service, Retrieved June 6, 2009, from http://www.ets.org/research/researcher/RR-01-15.html

Woodford, P. E. (1982). *An introduction to TOEIC: The initial validity study.* Princeton, NJ: Educational Testing Service. Retrieved June 6, 2009, from
http://www.ets.org/research/researcher/TOEIC-RS-00.html

Appendix A. Sample of STEP BULATS Speaking Test (Version: EN40)

<u>Source</u>: *Speaking Test Sample Question Paper*, University of Cambridge ESOL Examinations

Part 2 Task Sheet

<u>INSTRUCTIONS</u>

Please read all three topics below carefully.

Choose one which you feel you will be able to talk about for one minute.

You have one minute to read and prepare your talk.

You may make notes.

Topic A

Describe an important business meeting you attended.

You should say:

- ■ where it was
- ■ what it was about
- ■ why it was important.

What were the most interesting moments?

Topic B

Describe someone you particularly enjoy working with.

You should say:

- ■ what this person does
- ■ what sort of work you do with this person
- ■ why you like working with this person.

Would you change anything about this person? Give reasons for your answer.

Topic C

Describe the best workplace you have ever had.

You should say:

- ■ where the workplace was
- ■ what you were doing there
- ■ why you liked to work there.

Would you change anything about it? Give reasons for your answer.

Part 3 Information Exchange and Discussion

Conference arrangements

You have one minute to read through this task.

Information Exchange

You are making the arrangements for a one-day conference at a local hotel.
The Examiner is the Conference Organiser for the hotel
and is visiting you to discuss the conference.

Find out this information:

(i) the size of the largest conference room

(ii) the cost for that room

(iii) equipment available.

Do you think the hotel is offering you a good service for the price it is charging?.

Discussion

Now discuss this topic with the Examiner:

■ What makes a successful conference?