

Correlations between active skill and passive skill test scores

Michihiro Hirai (Director, Hitachi Institute of Foreign Languages)

Abstract

The correlations between speaking and receptive skills and between writing and receptive skills were studied, using a company-internal interview test, the BULATS Writing Test, and the TOEIC® test, administered to Hitachi employees who took two English courses of different levels. While the overall correlation coefficient between Hitachi's interview test and TOEIC scores was 0.78, endorsing ETS's research findings, the correlation coefficient for the intermediate level was as low as 0.49, supporting a widely-held perception that the TOEIC score is not representative of productive skills. The correlation coefficient between the BULATS Writing Test and TOEIC scores was 0.66, significantly lower than ETS's data showing the correlation between writing skill and the TOEIC reading test score. This discrepancy is attributed primarily to the difference in nature of the writing tasks between the BULATS Writing Test and the writing test used in ETS's study. Words of caution are offered in interpreting the correlations between productive skills and TOEIC scores.

Keywords: test score correlations, TOEIC, BULATS

It has long been of great interest to language teachers how strongly speaking and writing skills (productive skills) correlate with listening and reading skills (receptive skills). While many teachers seem to share the view that the scores of receptive skill tests do not accurately represent the test takers' productive skills (e.g., Gilfert, 1996; Brock, 1998, p. 35), the amount of statistical data showing the degree of correlation between these two types of skills has been rather limited.

Among the often-cited data in this field is a series of reports from Educational Testing Service (ETS), the organization that developed the Test of English for International Communication (TOEIC) and The Chauncey Group International, which administers the TOEIC. Woodford's study (1982, pp.14-15) cites a correlation coefficient of 0.83 between the direct Language Proficiency Interview (LPI) score and the TOEIC listening part score and a correlation coefficient of 0.83 between a "direct writing measure" and the TOEIC reading score, suggesting a "high degree of correlation" between productive skill scores and receptive skill scores. The *TOEIC Technical Manual* (1998, pp. 1-2) reports a correlation coefficient of 0.74 between a direct speaking measure and TOEIC score (both for the total test and listening part).

With a view to providing more independent data on the correlation between productive skills and receptive skills, the author conducted two studies: the first compares the scores of a company-internal interview test and the TOEIC, while the second compares the scores of the Business Language Testing Service (BULATS) Writing Test, which is part of a four-test suite developed by the University of Cambridge Local Examinations Syndicate (UCLES), and the TOEIC test.

In this report the author intends not only to present how these sets of skills correlate with each other, but also to shed light on how best to interpret the correlation data in a business context.

Materials and Method

The author employed the following three tests:

* A company-internal interview test, which, based on a series of questions, evaluates the test taker's aural comprehension, grammatical accuracy, vocabulary, pronunciation, and fluency. This test lasts 10 to 20 minutes, with 20 to 40 items covering about a dozen topics.

* The BULATS Writing Test, which evaluates the test taker's ability to produce short, well-organized business-

related memos/letters/e-mail with linguistic accuracy and appropriateness. The test lasts 45 minutes and consists of two tasks.

* The TOEIC, which is designed to evaluate the test taker's ability to read and aurally comprehend "real-life, business-type" English. The listening and reading parts have 100 items each, and the total test time is 120 minutes.

In order to study the correlation between speaking skills and receptive skills, the author collected the test scores of 475 students enrolled in an intensive, total immersion business English program (divided into Intermediate and Advanced levels, which students self-select) at the Hitachi Institute of Foreign Languages (HIFL) in Yokohama between October 1999 and September 2001. Each student took the interview test at the beginning and the end of the course. Each student's most recent TOEIC score before the start of the program was recorded, and all Intermediate level students took the TOEIC test at the end of the program.

To investigate the correlation between writing skill and receptive skills, the British Council and the author jointly administered the BULATS Writing Test to a total of 102 Hitachi employees, 90 of whom were students of either the Intermediate or Advanced course between September and November 2001. Twelve people, who were not students at HIFL at the time, took the test in April 2001. The correlation was calculated between the BULATS Writing Test scores and either the actual TOEIC scores on exiting the program (Intermediate students) or the most recent TOEIC scores (all others).

Results

When the scores of the Intermediate course students and the Advanced course students were combined, the interview scores were found to correlate fairly well with the TOEIC Total and Listening scores, as shown in Figures 1 and 2. The correlation coefficients of 0.78 for the Total and 0.73 for the Listening score were in line with the correlation coefficient of 0.74 between the direct speaking measure and the TOEIC Total (and also Listening) score as reported by The Chauncey Group International (1998, pp. 1-2).

When the two groups were examined individually, the correlation coefficient between the interview score and the TOEIC Total score dropped significantly (0.49 for the Intermediate course students and 0.65 for the Advanced course students), as shown in Figures 3 and 4. Figure 3 also reveals a distribution pattern for the Intermediate course students that is markedly different from that of the entire sample.

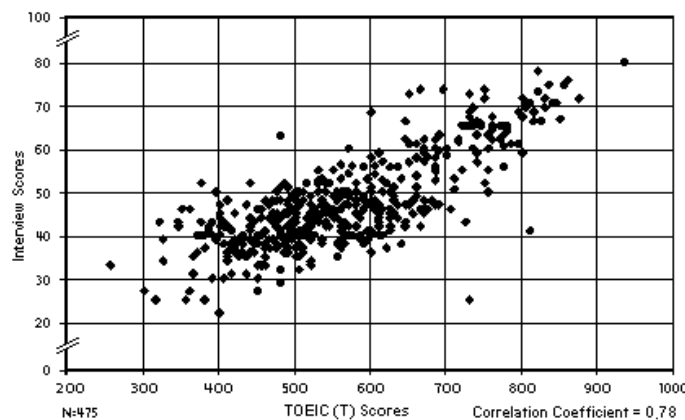


Figure 1. Correlation between Hitachi Interview scores and TOEIC composite scores

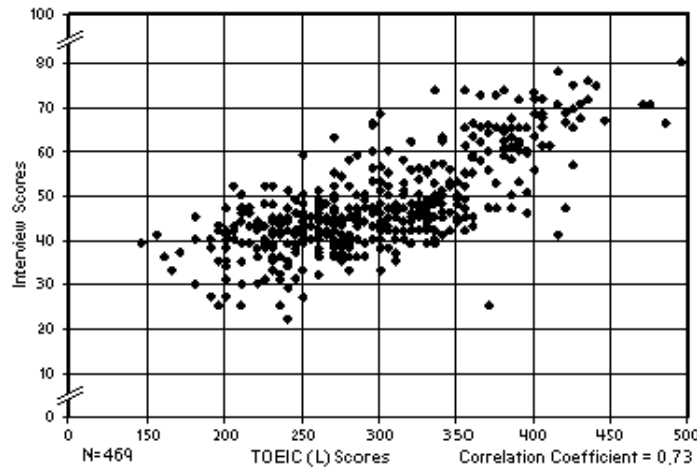


Figure 2. Correlation between Hitachi Interview scores and TOEIC listening scores

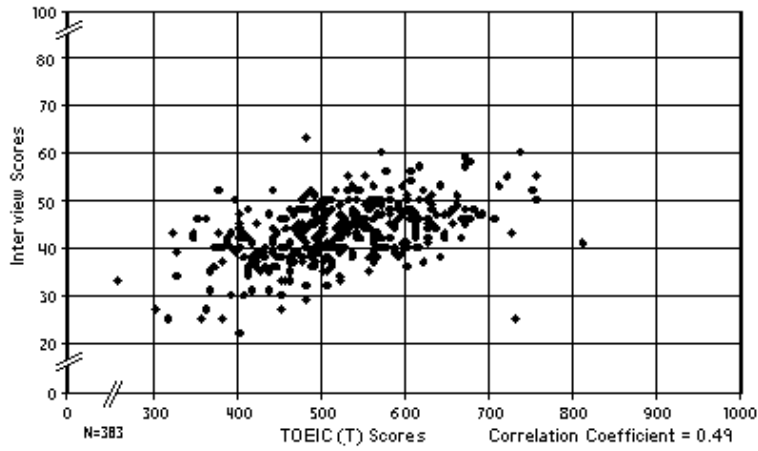


Figure 3. Correlation between Hitachi Interview scores and TOEIC composite scores among intermediate students before starting an intensive English program

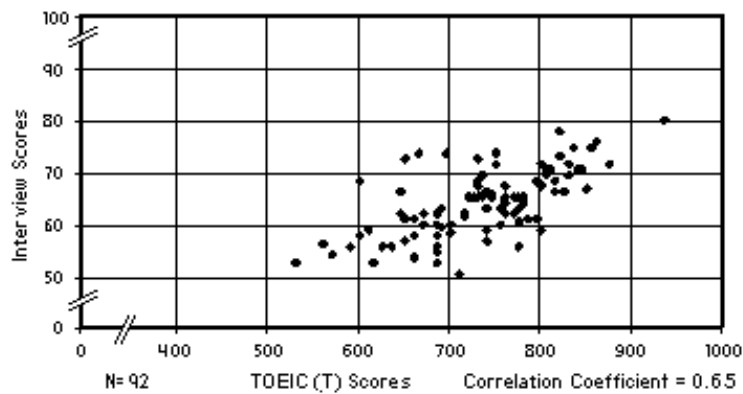


Figure 4. Correlation between Hitachi Interview scores and TOEIC composite scores among advanced students before starting an intensive English program

BULATS Writing scores were found to correlate a little more loosely than overall interview scores with the TOEIC Total and Reading scores, as shown in Figures 5 and 6. The correlation coefficient was found to be 0.66 for the Total and 0.59 for the Reading score, both considerably lower than the correlation coefficient of 0.83 as reported by Woodford (1982, p.15) and The Chauncey Group International (1998, pp. 1-2).

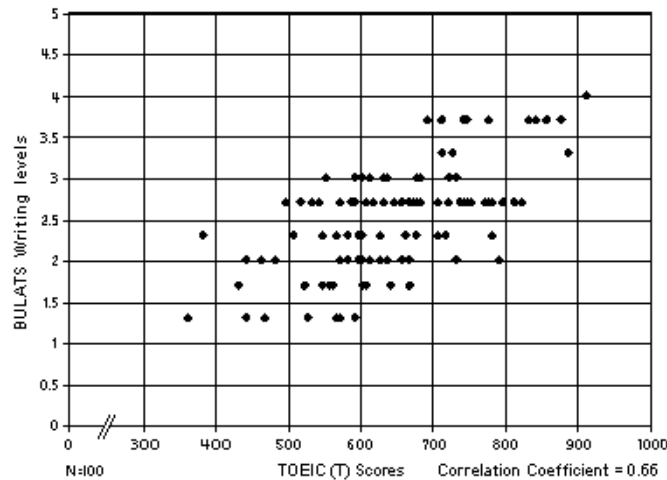


Figure 5. Correlation between BULATS Writing levels and TOEIC composite scores

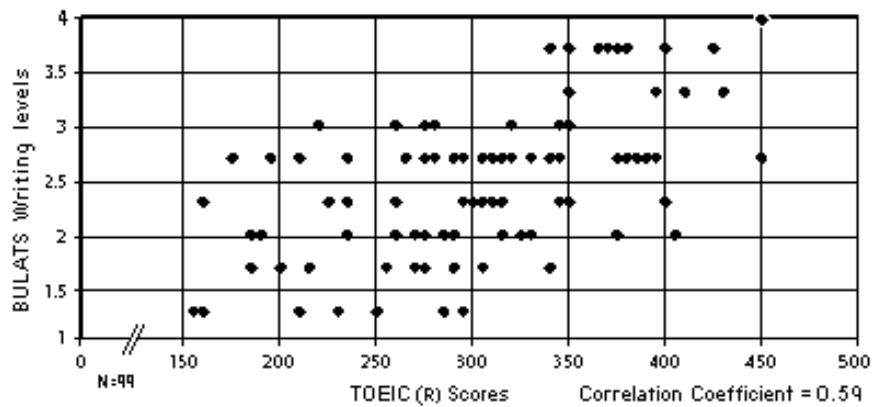


Figure 6. Correlation between BULATS Writing scores and TOEIC Reading scores

The results of the BULATS Writing Test are given in 17 levels instead of scores (from levels 0, 0+, 1-, 1=, 1+ to levels 5-, 5=, and 5+). For the sake of statistical handling, these levels have been converted to numbers: for example 1.3 represents a 1+ score and 4.7 represents a 5- score by assigning -0.3 to "-" and +0.3 to "+".

Discussion

(i) Hitachi Interview Test and TOEIC scores

With the scores of the Intermediate and Advanced course students combined, the distribution of the two sets of scores (interview and TOEIC) yielded relatively high correlation coefficients of 0.78 (for TOEIC total) and 0.73 (for TOEIC Listening part). Since the subjects in this study had a wide range of general English ability (with TOEIC scores of 255 to 935), they can be considered representative of the range of English learners in Japanese business environments. Superficially, therefore, the results point to a fairly high degree of correlation between speaking skills and TOEIC scores, at least among this group of Japanese businesspeople. In terms of general applicability, however, a few words of caution are in order. In interpreting statistical data, it is essential to check the degree of meaningfulness or reliability of the data in statistical terms, such as the sample size and the range or scope of the sample. In general, the larger the sample, the more reliable the data. Also, if the range of the sample does not match that of the population, then the statistical data does not accurately represent the characteristics of the population. In the present study, splitting the entire sample into two subgroups, where one group is made of the people with TOEIC (Total) scores of 730 or above and the other group is those with TOEIC (Total) scores of less than 730, reveals statistical features that are significantly different from those of the original sample. Both subgroups exhibit much lower correlation coefficients (0.45 and 0.63, respectively). In general, restricting the range results in lower correlation coefficients.

Likewise, it is of critical importance to check the nature of the data. Studying the data of an inherently biased sample often leads to an interpretation that is different from, or even contradictory to, that for the total population. In the present study, the Intermediate course students' interview scores flattened in the range of TOEIC 700 and above, as shown in Figure 3. This flattening effect should be attributed to the fact that the Intermediate course attracts employees with no or limited speaking experience, regardless of their TOEIC scores. The students came to the Intermediate course with an inherent bias toward poor speaking proficiency. As a result, the correlation coefficient was as low as 0.49. Note that the restricting of the sample's range to the Intermediate level was another contributing factor here. In contrast, the students of the Advanced course, which assume experience in an intermediate-level course and/or prior exposure to an English-speaking environment, exhibited a slightly higher correlation coefficient of 0.65.

This observation helps explain the apparent discrepancy between The Chauncey Group's report (and hence to some extent the composite correlation data in the present study) and the notion widely held by English educators that the TOEIC score is not a reliable measure of productive skills. Generally, company English courses, particularly low to intermediate-level courses, attract employees with no or limited speaking experience, regardless of their TOEIC scores. In terms of population, it is this group of low to intermediate-level course students that the majority of English teachers are assigned to and that is most often talked about. The above observation, therefore, seems valid as far as low to intermediate-level learners are concerned, but should not be extrapolated to speak of the entire range, which has different characteristics. By the same token, the relatively high overall correlation coefficient of 0.78 should not be considered applicable to any subset of the entire range such as low to intermediate levels. In general, statistical indices are valid only within the scope being studied. (Note that the term "validity" used in this context is different from the validity of a language test itself.)

(ii) *BULATS Writing Test and TOEIC Scores*

The relatively significant difference between the correlation coefficients in the present study and ETS/The Chauncey Group International's data can be attributed to the difference in nature between the two writing tests. The BULATS Writing Test consists of two tasks, each of which asks the test taker to compose from scratch a short e-mail message, letter, or memo. In contrast, the "direct measure" used by Woodford (1982, pp.10-11) has three tasks: dehydrated sentences, sentence translation, and a short (25-40

words) business letter, with weight factors of 0.3, 0.2, and 0.5, respectively. Obviously the first task tests receptive skills. The second task, which is sentence translation (as opposed to passage translation), may test the basic ability to put together words in grammatically correct order but does not test the ability to compose a passage that is acceptable in a business environment. Only the third task requires creative skills, which can be acquired or improved mainly through focused training.

Generally, it is creative writing skill that shows the greatest variance. In fact, Woodford's table summarizing the results of the "direct measure" tests (Woodford, 1982, p.11) shows a relatively large standard deviation of 3.211 against a mean of 5.859 (on a scale of 0 to 14) for the business letter part. On the other hand, the standard deviations for the other two parts were relatively small (7.243 against a mean of 37.824 on a scale of 0 to 50 for the dehydrated sentence part and 9.406 against a mean of 64.033 on a scale of 0 to 75 for the translation part). The standard deviation is a measure of variance in value of one quantity. Therefore, if the standard deviation of one quantity is large, then the correlation coefficient between this quantity and any other quantity is in general relatively small. In Woodford's study, however, the weighting of the three components effectively smoothed out the significant differences in variance among the three test components, producing the relatively high composite correlation coefficient of 0.83. Taking the letter writing part alone, it should be pointed out that there is a significant difference in elaborateness between the BULATS Writing Test and the direct measure test employed by Woodford. While the BULATS Writing Test gives the test taker two tasks, one between 50 and 60 words in length and the other between 180 and 200 words during 45 minutes, the letter writing part (creative part) of the direct measure test gives one task, to be completed in only 25 to 40 words in 20 minutes. It would be impractical to accurately measure the real writing skill with such a small task, and one should not draw too much significance in business context from the ostensibly high correlation coefficient of 0.83 reported by Woodford.

While the correlation coefficient is a general indicator of how closely two quantities relate to each other, one should be cautious about the potential pitfall of predicting the value of one quantity (e.g., writing skill level) from that of the other (e.g., TOEIC score) on the basis of the correlation coefficient, unless it is extremely close to ± 1 . Even for a narrow range of TOEIC scores, the writing level may vary significantly, if the correlation coefficient is not very close to ± 1 . For instance, in our sample with an overall correlation coefficient of 0.66, the BULATS Writing Levels of students with inclusive TOEIC scores from 550 to 595 – one of the most populous score brackets – spread randomly from 1.3 to 3.0 on a scale of 0 to 5.3.

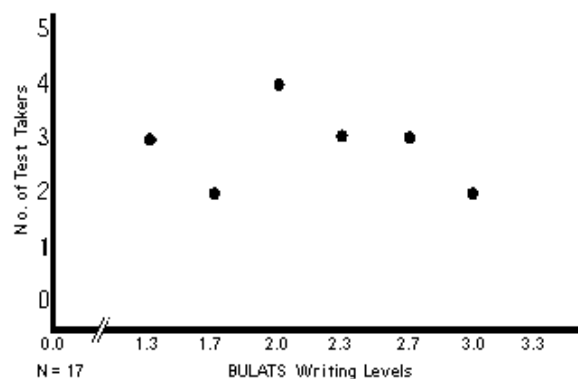


Figure 7. Distribution of BULATS Writing tests scores for those with TOEIC (T) scores between 550 - 595

While one can calculate the BULATS Writing Level's standard deviation for this slice of TOEIC score continuum to be 0.56 (against a mean of 2.14), the distribution is far from normal, as shown in Figure 7. This fact rendered the TOEIC score practically meaningless as a measure of writing skill for this sample.

All in all, it is worth pointing out that in interpreting writing test scores, proper attention should be paid to the nature and elaborateness of the test and that - apart from the correlation coefficient - a careful look at the distribution of scores of one test for any given score bracket of the other test would be essential in grasping how well the two sets of scores relate to each other. From the above analysis, the author maintains that TOEIC scores cannot be employed as a reliable measure of writing skills in business contexts.

Conclusion

Over a very wide range of TOEIC scores, interview and TOEIC scores were found to correlate relatively tightly, with a correlation coefficient of 0.78 (TOEIC Total) and 0.73 (TOEIC Listening) - well in line with ETS's research results. When two groups with different proficiency levels were taken separately, however, the correlation coefficient dropped significantly. For example, the Intermediate course students' interview scores showed a correlation coefficient of 0.49, primarily because they included employees with relatively high TOEIC scores who lacked experience speaking English. The BULATS Writing Level and the TOEIC scores were found to correlate more loosely than overall interview and TOEIC scores, with a correlation coefficient of 0.66, somewhat lower than ETS's findings. The author attributes this discrepancy to the difference in nature between the two writing tests. The author further suggests that TOEIC scores be interpreted cautiously by businesses. To assess business writing skills, an exam designed with a typical business environment in mind such as the BULATS Writing Test is recommended.

"... the author maintains that the TOEIC score cannot be employed as a reliable measure of writing skills in business contexts."

References

- Brock, R. L. (1998, August). The 64th TOEIC Seminar in Tokyo. *The Language Teacher*, 22 (8) Retrieved August 26, 2002 from <http://www.jalt-publications.org/tlt/files/98/aug/toEIC.html>.
- Chauncey Group International, Ltd. (1998). *TOEIC technical manual*. Princeton, NJ: Author.
- Gilfert, S. (1996, July). A Review of TOEIC. *The Internet TESL Journal*, 2 (8). Retrieved August 26, 2002 from <http://iteslj.org/Articles/Gilfert-TOEIC.html>.
- Woodford, P. E. (1982). *An introduction to TOEIC: The initial validity study*. Princeton, NJ: Educational Testing Service.

HTML: http://jalt.org/test/hir_1.htm / **PDF:** <http://jalt.org/test/PDF/Hirai1.pdf>