

The Author Responds: More on Test-Types

Akihiro Ito

Hiroshima University

A few *JALT Journal* readers of my article "Japanese EFL learners' test-type related interlanguage variability" (Volume 19, No. 1, May 1997, pp. 89-105) have raised questions. One, Takao Imai, has made his opinions public.

Let me begin with the issue of the relative order of the participants. The results of rank order correlations between the tests showed moderate to relatively high correlation between each pair of the three tests. Spearman's rank order correlation coefficients (r_s) among the three tests¹ ($N = 41$) are: 1) CP-GJ, $r_s = 0.571$, $p < 0.001$; 2) CP-SC, $r_s = 0.702$, $p < 0.001$; and 3) GJ-SC, $r_s = 0.734$, $p < 0.001$. I cannot conclude that the correlation coefficients were high enough to ignore the variability of test scores manifested by the difference of test-types. Even though the relative orders were moderate to high, the orders were not totally in accordance. If the purpose of a placement test is to determine only the relative order of our students, we can accept the correlation coefficients as sufficiently high. In my study I did not discuss the purpose of the relative clause tests. I think that people can use the tests according to their needs such as placement, achievement, proficiency, or diagnosis.

Second, I think it's common to assert that grammaticality judgment tests are effective for measuring subjects' linguistic intuition and metalinguistic knowledge if the test requires correction of grammatical errors. Like Imai, I think it's possible for subjects to take in ungrammatical sentences. However, as I explained (Ito, 1997a, pp. 94-95), after reviewing recent articles on the acquisition of relative clauses, I selected "typical errors in relative clause formation." Therefore, I would like to think that some students might raise their grammatical awareness in comprehending and forming sentences with relative clauses. However, I have found one problem with the use of the Grammaticality Judgment test. In Ito (1997b), I describe how higher level learners performed better in Sentence Combining than Grammaticality Judgment, possibly due to *hypercorrection*.

Third, I have re-examined the data of the participants' overall English language proficiency (Ito, 1996). In this study, I measured subjects' overall

English proficiency level through the 50-item multiple-choice cloze test used in Shimizu (1991), administered to the participants a week before the research reported in *JALT Journal* (Ito, 1997a) was conducted. According to the results, as the proficiency level increases, the magnitude of variability is smaller because subjects at a higher level mark high accuracy rates in every test-type. In this regard, the Ohba's (1994) hypothesis, which I reviewed (Ito, 1997a), seems reasonable.²

Finally, while I would like to undertake a discussion of the validity of the test types, this is beyond both the scope of the questions raised and the space available to comment.

As a concluding remark, I would again like to emphasize the importance of research on test-type related interlanguage variability. We should pay attention not only to the variability of interlanguage performance but also the quality of tests utilized in order to: 1) evaluate our students more accurately in an educational evaluation sense; 2) construct more reliable and valid language tests in a language testing sense, and 3) to improve second language acquisition research methodology.

Notes

1. CP: Cloze Procedure, GJ: Grammaticality Judgment, and SC: Sentence Combining.

2. Off-prints of the article in *CeLeS Bulletin*, 26 (Ito, 1996) and copies of a yet unpublished replicative study with a larger number of participants under more controlled conditions (Tests as a second language research method: Their types, reliability, validity, and variable research results) are available on request. E-mail: akito@ipc.hiroshima-u.ac.jp

References

- Ito, A. (1996). A study on the variability of test performance of Japanese EFL learners: A combination of two theoretical frameworks. *CELES Bulletin*, 26, 227-234.
- Ito, A. (1997a). Japanese EFL learners' test-type related interlanguage variability. *JALT Journal*, 19(1), 89-105.
- Ito, A. (1997b). An analysis of test-type related variability of interlanguage performance among Japanese EFL learners. *JACET Bulletin*, 28, 29-45.
- Ohba, H. (1994). Task-related variability in interlanguage by Japanese EFL learners. *STEP Bulletin*, 6, 34-48.
- Shimizu, Y. (1989). A study on the correlations between subtests in the *STEP Test* and cloze tests. *STEP Bulletin*, 1, 103-116.