

Point-to-Point

A Reaction to Brown and Yamashita “English Language Entrance Exams at Japanese Universities: What Do We Know About Them?”

Barry O’Sullivan

Faculty of Education, Okayama University

The article by Brown and Yamashita (*JALT Journal* 17(1), pp. 7-30) serves to highlight the lack of published accounts of empirical research in the area of university entrance test evaluation in Japan. In attempting to achieve their stated purposes

- a) to describe current testing practices at major institutions, and
- b) to establish a baseline of information so that change or lack of change in testing practices of such universities can be monitored in future years, (p. 11)

the authors express a desire to “help English teachers in Japan prepare students for taking such [entrance] tests and help their students in deciding which test to take” and, in addition, to “aid those responsible for creating entrance examinations to prepare high quality tests” (p. 7).

While the article is a welcome and long overdue look at university entrance examinations, there are quite serious problems with it that cause it to lose much of its value, and to fail to achieve its above quoted goals.

The design of the study severely reduces the possibility of using the data for the purpose of either classroom- or research-driven decision making. Though it is useful as a guide for those who may have no previous experience of the entrance examinations, the data presented here lacks the type of information necessary for any teacher to formulate hypotheses. An experienced teacher will agree that decisions such as which test to take, strategy planning and item prediction can only be made when a study is made of individual tests over a number of years. Such a study would consist of the type of descriptive analysis attempted here, and also of an analysis of the language and content of the individual test items.

Though a large amount of information is provided by the data presented by Brown and Yamashita, it is clear that its lack of detail and 'history' ensures that its usefulness as an adequate baseline for guiding pedagogy (i.e. test-taking pedagogy) and generating research hypotheses or testing decisions is called into question.

In practical terms, the Japanese teacher of English must ask

- a) Can I, as a teacher, confidently predict from the data the nature and organisation of any one of the described tests for the coming year?
- b) Can I suggest to my student any specific test-taking strategies based on the data for particular tests presented here?

The data provided here cannot allow for a positive response to either of these questions, making the findings of the paper of limited use to that teacher. Therefore, the decision to use a cross-sectional study, combined with the purely descriptive nature of the data appear therefore to ensure that the first goal, that of formulating an adequate description of the tests, is clearly not met.

The other audience for the paper appears to be those involved in setting the tests. While the data here may highlight some areas of possible worry, such as the high Flesch-Kincaid and Fog readability indices for some tests—Keio, Kyoto, Tokyo and Yokohama are singled out (p. 26)—it provides no empirical evidence of problems with validity and reliability in any of the tests reviewed—even with the tests quoted above we can not be sure if they deliberately and consistently use passages with high readability indices or if the examples from the year surveyed were in some way unusual.

Though the use of inferential statistics was explicitly avoided in the study (p. 13), the authors proceeded to make generalizations in the conclusions that one would expect to have been generated from such statistics. This seriously affects the usefulness of the conclusions, and must surely limit the effectiveness of the appeal to the "universities and the language professionals who write the tests" to attempt to avoid the "problems" highlighted (p.28).

If there is a question mark over the validity and reliability of entrance tests it is better that studies focus on the provision of concrete evidence relating to the existence of these problems. While it is accepted that the limited availability of test scores severely restricts the extent to which tests can be analysed, a study of the design and construction of any of the tests referred to here over a number of years would allow the researcher to reach conclusions that could be used as

'real' evidence of the existence of problems. This, I believe, would be seen as a more constructive form of criticism and would have a far greater chance of reaching those test makers referred to above.

While the thrust of this reply has taken a rather negative view of the Brown and Yamashita article, it is not meant merely to criticize what is a valuable and solid first step in the process of evaluating Japanese university entrance tests. In opening a debate on the reliability and validity of these examinations the article has confronted an issue of growing importance, and has raised a series of questions which researchers should now strive to answer with empirical evidence. These questions include:

- Is there evidence of a topic awareness bias in some tests?
- How harmful is the dependence on translation?
- Can we establish the content and construct validity of these tests?

The Authors Respond to O'Sullivan's Letter to *JALT Journal*: Out Of Criticism Comes Knowledge

James Dean Brown

University Of Hawaii At Manoa

Sayoko Okada Yamashita

International Christian University

We would like to begin by thanking Barry O'Sullivan for his criticisms of Brown and Yamashita (1995a), as well as for his words of praise.

Taking the criticisms first, as far as we can tell, his primary complaints are that there are "quite serious problems" with our study in that:

1. "the design of the study severely reduces the possibility of using the data,"
2. we do not provide enough "detail and 'history'," and