Assessment Literacy Self-Study Quiz #3
by Tim Newfields

This column mentions a few basic questions about testing, statistics, and assessment in a quiz format to help readers become more familiar with some of the jargon and concepts in the field and also think critically about some testing practices. Suggested answers for the items below are available at http://jalt.org/test/SSA3.htm.

Part I: Open Questions

1. What statistics should be made public about a commercial test? If too little information is offered, a loss of face validity might occur. On the other hand, general readers might be baffled by too much information. Take a look at what was offered to the public about some recent TOEIC® scores from the Institute for International Business Communication website:

The number of people who took this test in July 2007: 57,131

	Listening	Reading	Total
Maximum Score	495	495	990
Minimum Score	5	5	10
Mean Score	316.0	264.1	580.1
Standard Deviation	86.9	92.4	170.4

Source: IIBC. (2007). *Koushiki Deeta: Heikin Skoa Shosai*. [Public Information: Mean Score Details]. Retrieved on August 10, 2007 from http://www.toeic.or.jp/toeic/data/ data_ave01.html?score_id=0000000018

If you could include one additional statistic about this test, which would probably be most helpful to general readers? What other statistics about this test should probably be mentioned to the public?

2. A teacher notices an error in her school entrance exam after the exam was administered: one multiple choice question had two possible correct answers. What, if anything, should be done?

3. Recently a rating was developed by Isao Kobayashi of Kawai Juku of the "Good Places to learn English" [*Eigo no Yoi Mon*] in Japan based on his personal analysis of an unspecified number of 2005 university entrance exam problems. He concluded that the Osaka University of Foreign Studies and Niigata University were the best places to learn English based on the quality of their entrance exam questions. By contrast, Sophia University was ranked at 50th place because that author considered their entrance examination questions poorly constructed.

Any problem with this ranking system?

Source: Kobayashi, Isao (2006). Jukensei ga Ochi-iriyasui Machigatta Eigo Gakushuu-hou wo Tadasu. In K. Shimizu (Ed.) *2007 Nen Daigaku Rankingu*. Tokyo: Asahi Shinbun. p. 60-63.

4. At a university in Japan a placement test was developed to stream science majors into different classes for an EFL reading program. Examinees read a 369 word passage about Alfred Nobel, then answered a series of true/false statements about the passage. One statement from that test was -

"Alfred Nobel lived his whole life in Sweden."

The response format was to circle either "true" or "false" in the answer sheet. Any problematic points concerning this question? Also, any issues with using T/F response formats for this sort of test?

Source: Weaver, C. (2007). A Rasch-based evaluation of the presence of item bias in a placement examination designed for an EFL reading program. In T. Newfields, I. Gledall, P. Wanner, & M. Kawate-Mierzejewska (Eds.). *Second Language Acquisition - Theory and Pedagogy: Proceedings of the 6th Annual JALT Pan-SIG Conference*. Retrieved September 21, 2007 from at http://jalt.org/pansig/2007/HTML/Weaver.htm

5. To compare the *mean* of a particular sub-group to the mean of a larger group that is within the same population, what test(s) should be performed?

Part II: Multiple Choice Questions

NOTE: *At least one of these questions has more than possibly "correct" answer, so be careful!*

1. Which of the following is not a common procedure to investigate test reliability?

    (a) test-retest method   (b) equivalent scores method
    (c) split-half method   (d) Kuder-Richardson method

2. When reporting test scores to students, which of the following is/are unethical?

    (a) Stating the mean score and SD for that test.
    (b) Mentioning who scored at the top and bottom.
    (c) Indicating how many students were in each score band.
    (d) Mentioning the SEM for that test and what it doesn't measure.

3. To determine a EFL student's progress toward mastery of a classroom content area, a should be used.

    (a) grammar-translation   (b) placement
    (c) formative   (d) norm-referenced

4. What does this statistical symbol denote: R^2?

    (a) Pearson's rank correlation coefficient
    (b) Pearson's product-moment correlation coefficient
    (c) coefficient of determination
    (d) multiple correlation coefficient

HTML: http://jalt.org/test/SSQ3.htm   PDF: http://jalt.org/test/PDF/SSQ3.pdf