

## Assessment Literacy Self-Study Quiz #8

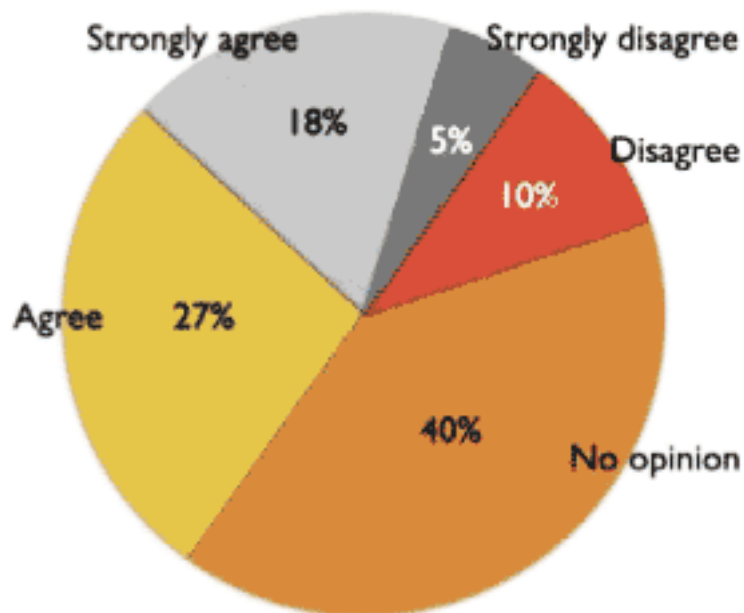
by Tim Newfields

This ongoing column features questions about testing, statistics, and assessment in a quiz format to promote greater assessment literacy. Suggested answers to the problems below are online at <http://jalt.org/test/SSA8.htm>.

### Part I: Open Questions

1. An article by Higginbotham (2009, pp. 15-18) in a recent issue of *The Language Teacher* attempted to highlight the value of extended task based instruction. Support for this instructional approach was based on a single questionnaire about a carol singing activity using that instructional framework. One month after the activity, 69 of the 80 students who participated in that activity completed a Likert questionnaire and vocabulary test with merely five words from the Christmas carols they had supposedly memorized. Based on those results, the author suggests activities such as this lead to “deep processing of the language” (p. 15).

What confounding issues are present in this study? How could the validity and reliability of studies such as this be enhanced? Also, how should the graph below probably have been revised for academic publication?



**Figure 1. I enjoyed the carol singing event**

Source: Higginbotham, G. (2009). Event-based learning: The benefits of positive pressure. *The Language Teacher*. 33 (1) 15-18.

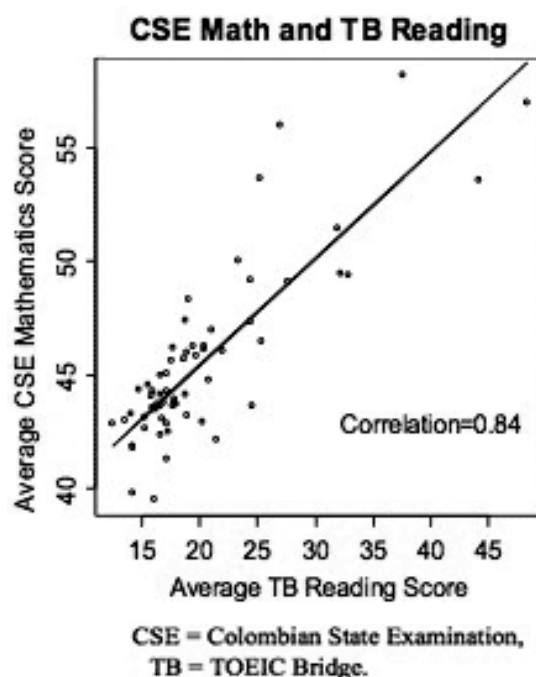
2. At most (if not all) universities in Asia, multiple entrance pathways exist. One widely used pathway is the *entrance by recommendation* system (推薦入学). In this system, a limited number of students from a select number high schools are allowed to recommend candidates for admission to a specific university. Those select schools are often, but not always, affiliated with each given university in some way. In many cases, admission decisions are based on a submitted essay and a very short screening interview. Applicants often have a month or so to compose their essays and they usually solicit the help of parents and/or teachers when composing and revising their essays. Moreover, most of the oral interview questions are known in advance and many students are able to churn out memorized answers to the most common questions.

What specific test validity issues are compromised through this process? What viable alternatives to the current entrance by recommendation (*suisen nyuushi*) system exist?

3. What are the pros and cons of balancing the correct answer choice sequences in the answer sheets of fixed-response exams? What alternative ways of organizing multiple-choice key answers currently exist?

4. In an ETS-funded research report, the TOEIC Bridge (Reading Section) was found to have a .84 correlation with a government run test of mathematics ability among 198 high school students in Columbia as the graph below suggests. What would be an appropriate interpretation of these results?

Source: Sinharay, S. et al. (2008). Establishing the Validity of TOEIC Bridge Test Scores for Students in Colombia, Chile, and Ecuador. *ETS Research Report 08-58*. Retrieved on November 12, 2009 from <http://www.ets.org/Media/Research/pdf/RR-08-58.pdf>



5. Briefly explain the difference between *item cloning* and *item anchoring*. In what situations are each of these practices employed?

## Part II: Multiple Choice Questions

1. One source of confusion in the field of language testing is that the symbol  $p$  is used to denote a variety of concepts. Which concepts below are commonly signified by  $p$  by different authors?

- (A) item difficulty (also symbolized as ID)
- (B) the probably of a chance occurrence
- (C) the population size
- (D) the proportion of students passing a given test

2. Which of these statement(s) about the KR-21 is/are considered true?

(Note: More than one correct response is possible.)

- (a) It is a good estimate of the KR20 if the range of item difficulty is relatively narrow.
- (b) It requires only one test administration.
- (c) Its reliability estimate is generally lower than the KR-20's.
- (d) It is robust if the unifactor trait is violated, provided the test is longer than 18 items.

3. Which of the following statements are not true about Wald tests? (Note: More than one correct response is possible.)

- (a) They can be used for dichotomous and continuous variables.
- (b) They provide a maximum likelihood estimate of a parameter(s) of interest .
- (c) They are generally preferred to likelihood ratio tests.
- (d) It does not require a large sample size in cases MANOVA models.

4. The following grading scheme suggested by Davies (2002) is an example of \_\_\_\_\_.

- |   |   |
|---|---|
| (a) confidence marking                            | For each question, candidates are awarded - |
| (b) elimination testing                           | +5 for 'very confident & right answer'      |
| (c) a liberal multiple-choice test grading scheme | +3 for 'fairly confident & right answer'    |
| (d) an order-of-preference grading scheme         | +1 for 'not confident & right answer'       |
|   | -2 for 'very confident & wrong answer'      |
|   | -1 for 'fairly confident & wrong answer'    |
|   | 0 for 'not confident & wrong answer'        |

Source: Davies, P. (2002). There's no confidence in multiple-choice testing. In M. Danson. & C. Eabry (Eds.), *Proceedings of the 6th CAA Conference*. Loughborough: Loughborough University, pp.119-130. [Cited in Ng, W. Y. & Chan, H.S. (2009). Different Methods of Multiple-Choice Test: Implications and Design for Further Research. S. I. Ao, et al (Eds.) *Proceedings of the International MultiConference of Engineers and Computer Scientists 20, Vol. II* March 18 - 20, 2009, Hong Kong. Retrieved November 12, 2009 from [http://www.iaeng.org/publication/IMECS2009/IMECS2009\\_pp1958-1963.pdf](http://www.iaeng.org/publication/IMECS2009/IMECS2009_pp1958-1963.pdf)]

5. Which of the following statements are **true** about parameter drift?

(Note: More than one correct response is possible.)

- (a) It may be caused by repeated exposure to a given test item.
- (b) It may be caused by a change in pre-test/post-test motivational levels.
- (c) It can be estimated by margin likelihood model.
- (d) It can be estimated by a Bayes modal procedure .

**HTML:** <http://jalt.org/test/SSQ8.htm> / **PDF:** <http://jalt.org/test/PDF/SSQ8.pdf>