

Assessment Literacy Self-Study Quiz #9

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/SSA9.htm>.

Part I: Open Questions

1. What does the term *effect size* actually mean? How is it commonly measured? Also, when can an effect size be justifiably regarded as “large”?
2. What is the difference between the *basket* and *Angoff* rating methods? What are the pros and cons of each procedure? When should they be employed?
3. What is the university entrance exam item below probably attempting to measure? How could this item be improved?

I 次の1～10のうち、誤った英語表現を含んだ部分がある場合には a～d から誤りを1つ選び、誤りがない場合には e を選んでマーク解答用紙にマークせよ。

3. It was customary in that country, the couple decided to get married only after having received the permission of both of their entire families. NO ERROR
- a b c d e

Source: Waseda University Faculty of Social Science (2010, February 22). 2010 Waseda Daigaku Shakai Gakubu Eigo Nyuushi: Dai Ikka. [2010 Waseda University Faculty of Social Science English Entrance Exam, Section I]. Retrieved on March 1, 2010 from <http://nyushi.yomiuri.co.jp/10/sokuho/>

4. What is the *John Henry effect*? How does it differ from the *Hawthorne effect*? How can researchers minimize both these effects?
5. What steps could be taken to improve the *differential validity* of a school entrance exam? How often are such steps seldom taken at institutions that you are familiar with?

Part II: Multiple Choice Questions

1. Q: Which of the following procedures are best suited for comparing data from two 5-point Likert scales from the same sample in a pre-test/post-test research design?
(A) A t-test
(B) A Mann-Whitney U test
(C) A Spearman test
(D) Somer's D coefficient
(E) A Kruskal-Wallis H-test
(F) Multiple linear correlations
(G) Pairwise multiple ANOVAs
(H) Other: _____

2. **Q:** In the field of statistics, which of the following terms correspond most closely with an "observed variable"? (Hint: More than one of choice below fits.)
- (A) a dependent variable (D) the predictor variable
(B) an independent variable (E) an extraneous variable
(C) a criterion variable (F) outcome variable
3. **Q:** Arranging the blocks of a test on the basis of an estimate of what should allow examinees to gain the maximum number of points in the least amount of time is an example of _____.
- (A) efficiency ordering (C) difficulty ordering
(B) facility ordering (D) reactive ordering
4. **Q:** What does "truncation" generally refer to in test equating?
- (A) Using just the sections of the respective tests being compared.
(B) Assigning scaled scores in a way that ignores the very highest or lowest raw scores.
(C) The shifting of data from an equivalent-groups design into a single-groups design.
(D) A form of statistical censoring that occurs when a given value is outside the range of the measuring instrument.

(Suggested answers online at <http://jalt.org/test/SSA9.htm>.)

HTML: <http://jalt.org/test/SSQ9.htm> / **PDF:** <http://jalt.org/test/PDF/SSQ9.pdf>