Assessment Literacy Self-Study Quiz #6
by Tim Newfields

To encourage readers to become more familiar with key concepts and terminology in the field of language testing, this column explores questions about testing, statistics, and assessment in a quiz format Suggested answers to the problems below are online at http://jalt.org/test/SSA6.htm.

Part I: Open Questions

1. Consider the following graph from the Japanese Ministry of Education, Culture, Sports, Science and Technology website, reputedly describing the "reasons for selecting a particular university" among high school students in Japan who appear to have successfully entered a Japanese university. What key information is missing from this graph? What problems are likely to arise in interpreting this graph? How should this graph be revised to minimize such problems?

Figure 2.3. Reasons for Selecting a Particular University
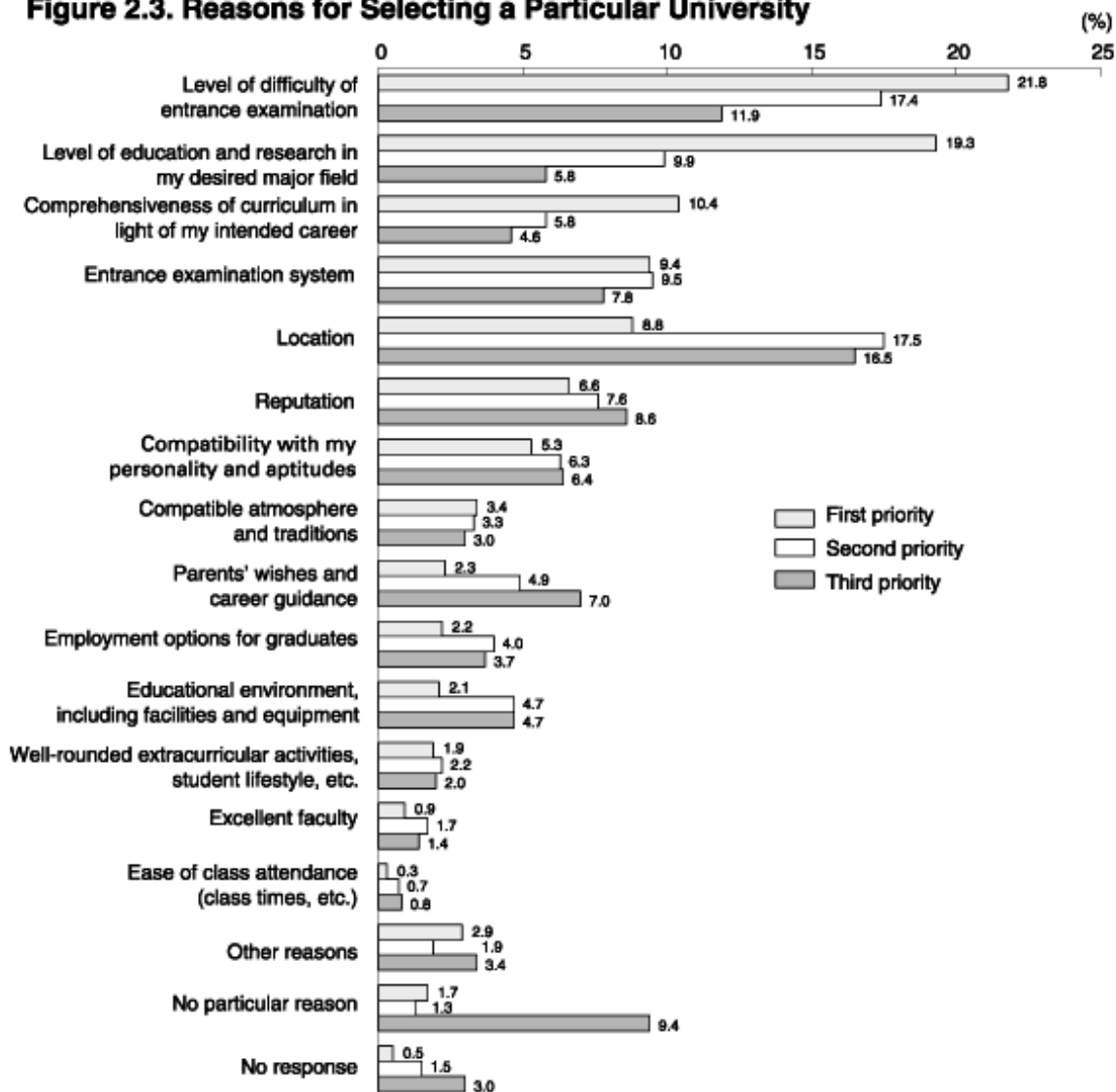
Note: Responses have been abridged.

Source: Ministry of Education, Culture, Sports, Science and Technology of Japan. (n.d.).
Remaking Universities: Continuing Reform of Higher Education: Chapter 2, Section 1, Sub-Section 3:
Approaches to the Improvement of University Entrance Examinations. Retrieved December 1, 2008 from
http://www.mext.go.jp/b_menu/hakusho/html/hpae199501/hpae199501_2_017.html

2. In the context of multiple-choice testing, what does "proportionality" refer to and why is this concept important to test designers?

3. A well-known publisher and test developer has attempted to explain how its test correlated to two other famous tests by offering the following instructions:

Follow these steps to convert Longman English Assessment percentage scores in *Listening, Writing,* and *Reading* into an approximate TOEFL® or TOEIC® score.

Note that you must take all 3 Longman English Assessment (LEA) skill tests (Listening, Writing and Reading) to obtain a TOEFL or TOEIC correlation, and that the 3 skill tests must all be at the *same* level (Beginning, Intermediate, or Advanced).

1. Gather your LEA score for each skill test. For example:

  Beginning Listening percentage score 70%
  Beginning Writing percentage score 80%
  Beginning Reading percentage score 90%

2. Add the three scores to obtain a sum of the percentage scores.

  Sum of the percentage scores 240

3. Divide the sum by 3 to obtain an average of the percentage scores.

  Average of the percentage scores 80%

4. Use the correct table (depending on whether you took the Beginning, Intermediate, or Advanced skill tests) to determine your approximate TOEFL or TOEIC score.

  Approximate TOEFL score 440
  Approximate TOEIC score 475

LEA / TOEFL / TOEIC Correlation Tables

Average of Percentage Scores on **Beginning Level** LEA Skill Tests (Listening, Writing, Reading)	Approximate Overall TOEFL Score	Approximate Overall TOEIC Score
100%	480+	575+
90%	460	525
80%	440	475
70%	420	425
60%	400	375

(abridged)

Source: Pearson Education, Inc. (2003). Correlations between Longman English Assessment and TOEFL and TOEIC. Retrieved December 1, 2008 from www.pearsonlongman.com/ae/multimedia/pdf/leacorrelation.pdf

What ethical issues have been compromised in reporting this information? What information needs to be included to enable teachers and school administrators to make an informed decision about how well the exam developed by Pearson Education actually correlates with the other two ETS exams?

4. Mention one concrete situation for which each of the following tests would be appropriate:
        (A) A mastery test
        (B) A power test
        (C) A speed test
        (D) A -ze test

Part II: Multiple Choice Questions

1 Which of these statements below is generally true of ANOVA tests? (Hint: Think a bit and don't expect a single, clear answer.)

        (A) They examine only how groups and items interact.
        (B) They point out pronounced items which differ markedly from the norm.
        (C) They examine proportional score differences across total score categories.
        (D) They are arbitrary as far as ability levels are concerned.
        (E) They can help show whether or not there is a differential pull in terms of specific distracters between groups.

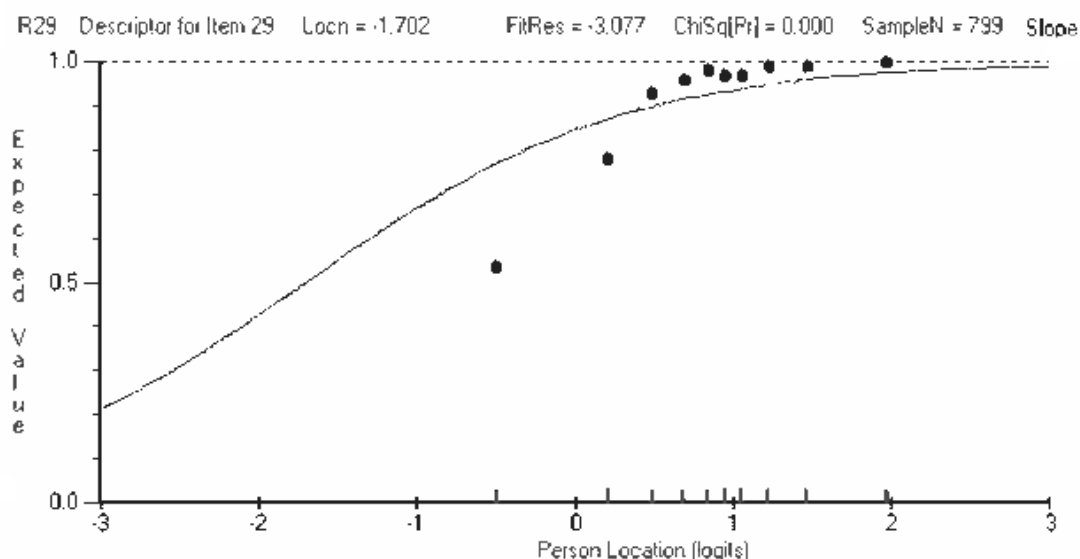2. Which statement about the test item below is true?

Figure 2. An examination of item R29 and its item characteristic curve (ICC) in the 2006 [Keio University] pilot placement test

        (A) It discriminates fairly well across all ranges of language ability.
        (B) It discriminates fairly well only for mid- and upper- ability students.
        (C) It discriminates fairly well only for mid- and lower- ability students.

(D) It does not discriminate well across various language ability ranges.

Source: Nakamura, Y. (2007). A Rasch-based analysis of an in-house English placement test. 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 December 16, 2008 from
http://jalt.org/pansig/2007/HTML/Nakamura.htm

3. Please look at the survey item below:

What are the **three** main reasons that you have for studying English?
Rank the most important one as 1 and the least important as 3.

    ☐ Because I want to be able to understand Western music
    ☐ Because I want to be a '*kokusai-jin*' [English: "international person"]
    ☐ Because I want to speak English when I go traveling
    ☐ Because I have to as part of my course
    ☐ Because I want to be able to teach my children English
    ☐ Because I enjoy taking English classes
    ☐ Because I am interested in other cultures by learning English
    ☐ Because I want to be able to understand western movies
    ☐ Because if I speak English other people will respect me more

What type of scale is it based on?
  (a) a Thurstone ratio scale   (d) an ordinal rank-order scale
  (c) a Guttman interval scale   (e) a categorical dummy-variable scale
  (c) There is no way to tell by merely looking

Also, how should this scale be revised to enhance its validity?

Source: Jenkins, A. C. (2005). EFL motivations and teacher preferences: A survey of Japanese university students in Niigata. Unpublished master's thesis, Department of Linguistic, Cultural & Translation Studies, School of Arts, University of Surrey, Guildford, Surrey, U.K. Retrieved December 16, 2008 from
http://www.surrey.ac.uk/ALRG/dissertations/pdf/Jenkins_A_2005.pdf

4. In most 3-parameter IRT tests, the third parameter refers to _____?

  (a) item difficulty   (c) rater difficulty
  (b) item discrimination   (d) guessing

5. At what point can a Pearson product-moment correlation be interpreted as "strong"?

  (a) If it is within a range of .1 (anywhere from $\pm.90$ to ± 1).
  (b) If it is within a range of .3 (anywhere from $\pm.70$ to ± 1).
  (c) If it is within a range of .5 (anywhere from $\pm.50$ to ± 1).
  (d) If it is within a range of .7 (anywhere from $\pm.30$ to ± 1).

HTML: http://jalt.org/test/SSQ6.htm / **PDF**: http://jalt.org/test/PDF/SSQ6.pdf