Brief Report

How Do Examinees Behave When Taking Multiple-Choice Tests?

Rand R. Wilcox
University of Southern California

Horst (1933) assumed that when examinees respond to a multiple-choice test item, they eliminate as many distractors as possible and guess at random from among those that remain. More recently, Wilcox (1981) proposed a latent structure model for achievement test items that was based on this assumption and that solves various measurement problems (see also Wilcox, 1982a, 1982b).

Suppose an item is administered according to an answer-until-correct (AUC) scoring procedure. That is, examinees choose a response and they are told whether it is correct. If incorrect, they choose another response, and this process continues until the correct response is selected. Now consider two specific distractors. If Horst's assumption is true, then among the examinees choosing these two distractors, the order in which they are chosen should be at random. Of course, for three distractors the same conclusion holds, only now there are six patterns of responses rather than two. An empirical investigation of this implication is described below.

As for previous tests, the final examination for students enrolled in an introductory psychology course was administered according to an AUC scoring procedure. For 26 items, examinees were asked to record the order in which they chose their responses. Bonus points were given to those examinees complying with this request. There were 236 examinees who took the first 13 items, and 237 examinees who took the remaining 13. All items had four alternatives.

For any two distractors, the null hypothesis of random order in responses can be tested with the usual sign test. Among the examinees choosing all three distractors, the chi-square test given by

$$\chi^2=\Sigma(x_i-N/6)^2/(N/6) \qquad [1]$$

was used where x_i is the number of examinees choosing the ith response pattern, and N is the number of examinees choosing all three distractors. Some exact critical values are given by Katti (1973) and Smith, Rae, Manderscheid, & Silbergeld, (1979), and they were used whenever possible. For larger values of N, the adjusted chi-square test was used (Smith et al., 1979).

For each item the responses to all pairs of distractors were tabulated. For $N < 5$, no test was made because it is impossible to reject the null hypothesis at the .1 level. For the first test form, 29 tests were

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 7, No. 2, Spring 1983, pp. 239–240

0146-6216/83/020239-02$1.10

made and the hypothesis of random choices was rejected three times at the .1 level. For the second test form, 25 tests were performed, and again H_0 was rejected three times. Next, an analysis was performed on those responses where all three distractors were chosen. Again, no test was made for $N < 5$. The largest value for N was 75. At the .1 level, H_0 was rejected for 5 of the 12 items on the first test form, and for the second test form the rejection rate was 3 of 11.

The question remains as to the relative extent to which responses are not random when H_0 is rejected. For the case where all three distractors were chosen, this quantity was measured with

$$w=(X^2-X^2_{min})/(X^2_{max} - X^2_{min}) \qquad [2]$$

where $\chi^2_{\max}$ and $\chi^2_{\min}$ are the maximum and minimum possible values of χ^2. From Smith et al. (1979), $\chi^2_{\max} = 5N$, and $\chi^2_{\min}$ is given by Dahiya (1971). The quantity w has a value between 0 and 1 inclusive. The closer w is to one, the more unequal are the cell probabilities in a multinomial distribution.

Marshall and Olkin (1979) suggest that when measuring inequality, a certain class of functions (called Schur functions) should be used. Writing Equation 1 as a function of Σx_i^2 (Dahiya, 1971) and noting that Σx_i^2 is just Simpson's measure of diversity, it follows from results in Marshall and Olkin (1979) that w is a Schur function.

Note that using the w statistic is similar to using Hays' ω^2 (Hays, 1973). That is, rejecting the null hypothesis does not indicate the extent to which the cell probabilities are unequal. It may be, for example, that the cell probabilities are not equal but that for practical purposes they are nearly the same in value.

For the first test form where H_0 was rejected, the w values were found to be .074, .183, .286, .167, and .137. For the second test form they were .098, .125, and .133. Thus, even when H_0 is rejected, Horst's assumption appears to be a tolerable approximation of reality in most cases. Of course, there will probably be items where this assumption is grossly inadequate. In this case, the measurement procedures proposed by Wilcox (1981) may be totally inappropriate.

References

Dahiya, R. D. On the Pearson chi-squared goodness-of-fit test statistic. *Biometrika*, 1971, *58*, 685–686.

Hays, W. *Statistics for the social sciences*. New York: Holt, Rinehart, & Winston, 1973.

Horst, P. The difficulty of a multiple-choice test item. *Journal of Educational Psychology*, 1933, *24*, 229–232.

Katti, S. K. Exact distribution for the chi-square test in the one-way table. *Communications in Statistics*, 1973, *2*, 435–447.

Marshall, A., & Olkin, I. *Inequalities: Theory of majorization and its applications*. New York: Academic Press, 1979.

Smith, P. J., Rae, D. S., Manderscheid, R. W., & Silbergeld, S. Exact and approximate distributions of the chi-square statistic for equiprobability. *Communications in Statistics—Simulation and Computation*, 1979, *B8*, 131–149.

Wilcox, R. R. Solving measurement problems with an answer-until-correct scoring procedure. *Applied Psychological Measurement*, 1981, *5*, 399–414.

Wilcox, R. R. Some empirical and theoretical results on an answer-until-correct scoring procedure. *British Journal of Mathematical and Statistical Psychology*, 1982, *35*, 57–70. (a)

Wilcox, R. R. Some new results on an answer-until-correct scoring procedure. *Journal of Educational Measurement*, 1982, *19*, 67–74. (b)

Author's Address

Send requests for reprints or further information to Rand R. Wilcox, Dept. of Psychology, University of Southern California, Los Angeles CA 90089, U.S.A.