

## Teaching Materials

### Reliability of Scores From Teacher-Made Tests

#### 1. Short Answer Test Items

These items may be used to supplement the self-test in measuring student achievement of the module objectives or they may serve as points of discussion for the class or small groups of students. The items might serve as a pretest, also. Information about student entry levels can be obtained and the items could serve as advance organizers for students.

#### 2. Discussion Questions

This set of questions could be assigned as an out-of-class activity that would be supplemented by a classroom question-answer session. It might serve, also, as a means of reviewing the important points of the module prior to a unit test. Some of the questions might be used in class presentations by the instructor to stimulate students' thinking or to check on the extent of learning in the group.

#### 3. Additional Computational exercises

The questions and accompanying data could be used as illustrations in class presentations or as sample problems to be worked by students in class as a comprehension check. Often such in-class activities are successful when done with students in pairs. Of course, the exercises can be used for out-of-class assignments, either for all students or for those who wish to have additional practice. Finally, small data sets like these can be used to illustrate the effect of changing students' scores (modifying item difficulty and discrimination) on the reliability coefficient.

#### 4. Handouts (Draft Form)

One handout displays the computational formulas for three methods of internal analysis. These representations might be useful when presented on the overhead projector to demonstrate the similarities and differences between pairs of procedures. In addition, students who are not mathematically inclined sometimes profit from a verbal representation of a mathematical formula.

The second handout can be used to demonstrate the step-by-step procedures for computing  $K-R^2$ . The sample on one side can be a reassuring reference and the fill-in-the-blank approach on the other side tends to reduce the complexity of the arithmetic steps. Finally, students might be encouraged to keep a folder with multiple copies of the handout. In practice, having such a "form" to fill out may be sufficient leverage to develop the habit of computing reliability after each test has been scored.

## Short Answer Test Items

1. What kinds of measurement errors could affect multiple choice test scores but not essay scores?
2. What kinds of measurement errors could affect essay test score but not multiple choice scores?
3. What advantage does K-R21 have over K-R20 as a method of estimating reliability?
4. What misunderstanding about reliability is conveyed by the person who says, "This test is quite reliable"?
5. How does the idea of "content sampling" affect the reliability of a set of test scores?
6. How could an individual's physical handicap contribute systematic errors, but not random errors, to the person's test score?
7. Why might systematic errors have no effect on norm-referenced test score interpretations?
8. Why might domain-referenced score interpretations be affected by both systematic and random errors?
9. What does it mean if the variance of a test item is equal to 0.00? 0.25? 0.50?
10. How is the sum of the item variances related to the reliability coefficient of a set of scores?
11. If we computed K-R20, K-R21, and K-R21' for a typical teacher-made test, what relationship should we expect to find between the three estimates? Why?
12. What factors should a teacher consider in trying to decide if the scores from her test are reliable enough?
13. If the reliability of the scores from a classroom test is 0.51, for what reasons might we be willing to accept those scores as "accurate enough"?
14. How does the length of a test (# of items) relate to the reliability of the scores from it?
15. Why might an item of moderate difficulty contribute less to high reliability than an item on the same test that only 25 percent of the group answered correctly?
16. Why might the scores from an expertly-constructed test turn out to be fairly low in reliability?

## Discussion Questions

1. Under what circumstances might the scores from a certain classroom achievement test be highly reliable, yet the scores still are not useful to the teacher?
2. Measuring the length of each of several pieces of wood tends to yield more reliable "scores" than measuring the achievement of each of several students. What are the critical differences between the two situations that might explain the discrepancy in reliabilities?
3. How does the common usage of each of these terms relate to the meaning of reliability as a characteristic of a set of test scores: accurate, generalizable, dependable, consistent, stable, reproducible, and reliable?
4. How can deficiencies in item writing skill on the part of the teacher contribute to measurement errors when the test is administered?
5. What advice could a teacher give to students on the day before a test to help reduce the chance that certain measurement errors will affect the test scores?
6. Using the formula for computing Coefficient Alpha or K-R20, show why the reliability coefficient could never have a negative value. When would the reliability estimate equal zero?
7. What are some of the trade-offs faced by the teacher who, because of low score reliability, must decide to (a) eliminate some of the very difficult items from the test and recompute the scores, (b) reduce the weight of the scores in the course grade, or (c) disregard the scores completely and administer a new test.

Additional Computational Exercises

Student

Item No.	A	B	C	D	E	F	G	H	Total
1	1	1	1	1	1	1	1	1	8
2	1	1	1	1	1	1	1	0	7
3	1	1	1	0	1	1	1	1	7
4	1	1	1	0	1	1	1	1	6
5	1	1	1	1	1	1	0	1	7
6	0	0	1	1	1	1	1	1	5
7	1	0	1	1	0	0	0	0	3
8	1	0	1	1	0	0	0	0	2
Total	8	7	5	8	3	7	5	6	4

a. Compute both  $\Sigma S_i$  (Coefficient Alpha) and  $\Sigma pq$  (K-R20) for these data. Compare the two values and explain the result.

b. Which is a more accurate estimate for K-R20 for these data, K-R21 or K-R21?

2.

Student

Item No.	A	B	C	D	E	Total
1	1	1	0	1	1	4
2	1	1	0	1	1	4
3	1	1	1	1	0	4
4	1	0	0	1	1	3
5	1	0	0	1	0	2
6	1	1	1	1	1	5
7	1	1	0	1	1	4
8	1	0	1	0	1	3
9 (essay)	1	3	3	5	4	16
10 (essay)	8	8	7	9	8	40
Total	17	16	13	21	18	

a. Compute Coefficient Alpha for this 10-item test (8 multiple choice, one 5-point essay, one 10-point essay).

b. Compute K-R20 for the first eight items.

c. Compute K-R21 for the first eight items.

### Answer to Additional Computational Exercises

1.
  - a. Both values turn out to equal 1.172. The two methods are equivalent for estimating the sum of the item variances for dichotomously scored items. Discrepancies would be due only to rounding or computational error.
  - b. The sum of the item variances is 1.172, the mean is 5.625, and the variance of the scores is 2.484.  
  
The value of K-R20 is 0.602, K-R21 is 0.374, and K-R21' is 0.462.  
  
K-R21' provides a more accurate estimate of K-R20 than does K-R21.
  
2.
  - a. The sum of the item variances is 3.52 and the variance of the scores is 6.80. Coefficient Alpha is 0.535.
  - b. The sum of the item variances is 1.36 and the variance of the scores is 2.96. K-R20 is 0.616.
  - c. The mean is 5.8 and the variance of the scores is 2.96. K-R21' is 0.569. (K-R21 is 0.526).

## Verbal Computational Formulas Methods of Internal Analysis

$$\text{Alpha} = \frac{\text{No. of Items}}{\text{No. of Items} - 1} \left[ 1 - \frac{\text{Sum of Item Variances}}{\text{Test Variance}} \right]$$

$$\text{K-R20} = \frac{\text{No. of Items}}{\text{No. of Items} - 1} \left[ 1 - \frac{\text{Sum of Item Variances}}{\text{Test Variance}} \right]$$

$$\text{K-R21} = \frac{\text{No. of Items}}{\text{No. of Items} - 1} \left[ 1 - \frac{(\text{Mean})(\text{No. of Items} - \text{Mean})}{(\text{No. of Items})(\text{Test Variance})} \right]$$

$$\text{K-R21}' = 1 - \frac{(.8)(\text{Mean})(\text{No. of Items} - \text{Mean})}{(\text{No. of Items})(\text{Test Variance})}$$

## K-R21 Worksheet

$$K-R21 = \frac{k}{k-1} \left[ 1 - \frac{\bar{X}(k - \bar{X})}{(k S_t^2)} \right]$$

$k$  = \_\_\_\_\_, the number of items

$\bar{X}$  = \_\_\_\_\_, the test mean

$S_t^2$  = \_\_\_\_\_, the test variance

$$K-R21 = \frac{(\quad)}{(\quad - 1)} \left[ 1 - \frac{(\quad)(\quad - \quad)}{(\quad)(\quad)} \right]$$

$$K-R21 = (\quad) \left[ 1 - \frac{(\quad)}{(\quad)} \right]$$

$$K-R21 = (\quad) \left[ 1 - (\quad) \right]$$

$$K-R21 = (\quad) [ \quad ]$$

$$K-R21 =$$

## K-R21 Worksheet

$$k = \underline{20}, \text{ the number of items}$$

$$\bar{X} = \underline{11.74}, \text{ the test mean}$$

$$S_t^2 = \underline{19.80}, \text{ the test variance}$$

$$K-R21 = \frac{(20)}{(20 - 1)} \left[ 1 - \frac{(11.74)(20 - 11.74)}{(20)(19.80)} \right]$$

$$K-R21 = (1.053) \left[ 1 - \frac{(96.97)}{(396)} \right]$$

$$K-R21 = (1.053) \left[ 1 - (.2449) \right]$$

$$K-R21 = (1.053) [.7551]$$

$$K-R21 = 0.795$$