

# **ZEUS MANUAL**

**CLASSROOM  
TEST SCORING  
SYSTEM**

**TESTING CENTER  
SAN FRANCISCO STATE UNIVERSITY**

PART I  
THE ZEUS SCORING SYSTEM

A. GENERAL INFORMATION

The Zeus Test Scoring System utilizes the NCS 7005 to optically read marked test sheets. Two COBOL programs edit and sort the raw data; a third program scores the test data and produces the reports. There is no limit as to the number of different tests that may be put together in one run.

The following limitations apply to each test:

- Maximum number of QUESTIONS 200
- Maximum number of SUBTESTS 4

There can be only one correct answer (a/1, b/2, c/3, d/4, e/5) per question, and questions that are left blank on the instructor's key sheet will not be scored. Tests may be divided into subtests, which will be scored separately, using different scoring formulas, if so specified. Each form of a given test will be scored and statistically analyzed separately.

The following test information and statistics will be reported: (See sample in Appendix B of this manual)

- Student scores
- Number of items (questions) graded
- Number of omitted on key
- Number of students
- M e a n
- Standard deviation

- -Distribution of scores
- -Response-by-item report

Number and percent responses to each alternative

- -Item analysis

Percent correct

Biserial coefficient

T statistics

Kuder Richardson coefficient

## B. DETAILED INSTRUCTION FOR CLASSROOM TESTS

### 1. MATERIALS –

- a. KEY SHEETS (Instructor Form) --From Department via Corporate Express
  - one is needed for each test.
  - if a test has more than one form, one KEY SHEET is needed for each form.
- b. STUDENT ANSWER SHEET--Purchase in Bookstore
  - need as many as the number of students taking the test.
- c. USER'S MANUAL – Testing Center will send upon request (x82271)

### 2. INSTRUCTIONS FOR MARKING THE INSTRUCTOR'S KEY SHEET

- a. USE ONLY A NO.2 PENCIL OR SOFTER
- b. MARK ONLY ONE ANSWER PER QUESTION. IF YOU CHANGE A MARK, ERASE THOROUGHLY

Instructors for marking the  
Instructor's key sheet (cont'd.)

c. FILL IN THE FOLLOWING INFORMATION:

-NAME

Print your last name, first initial, and middle initial in the top row of  
Shaded area, fill in appropriate circle below each letter.

-COURSE SORT NO.

Enter the course sort number for your class as it appears in the class schedule or  
on your official class roster. It is critically important that the same number be  
used on the instructor's key sheet and on all student answer sheets.

-TEST I.D.

“Form” and “No.” are optional designations you may wish to use to identify this  
test or version.

. FORM

If a test has only one form, leave this area blank. If it has more than one form,  
mark in the column headed “FORM”, using any combination of two letters. For  
example, one form may be named AA and the other AB. A KEY SHEET must  
be prepared for each form and the student forms must be in separate piles. Each  
form of a given test will be graded and statistically analyzed separately.

FORM designations are used for sorting prior to scoring. The Key Sheet and all  
student answer sheets must be Marked with identical FORM designators.

NO. - (TEST NUMBER)

The Test Number is used only as the instructor's reference for his or her tests. If  
you wish to keep track of the number times you have given the test, mark the

test number (NO.) area. This number will be shown on the reports. If there is no test number, leave this area blank. This item is not used for sorting purposes, i.e., you may assign different test numbers to your students.

-DATE

The data you wish printed on your output report.

Shade in the appropriate circle below each number.

-USE - (TEST USE)

-OPTIONS

SCORING FORMULA

FORMULA A.        Score – Number of Correct Answers

FORMULA B        Score =  $R - W/4$   
                         -Number of Correct Answers – Number of Wrong Answers / 4

FORMULA C.        Score =  $R - W/2$   
                         =Number of Correct Answers – Number of Wrong Answer / 2

FORMULA D.        Score =  $R - W$   
                         = Number of Correct Answers – Number of Wrong Answers

SUBTESTS

EVEN IF THERE ARE NO SUBTESTS, ENTER THE NUMBER OF ITEMS (QUESTIONS) IN SUBTESTS AREA 1 (LOWER LEFT OF INSTRUCTOR FORM). If there are subtests, indicate the NUMBER OF QUESTIONS in each subtest in the SUBTESTS area. For example, suppose there are 3 subtests, the first contains 10 questions, the second contains 20 questions, and the third contains 30 questions; mark '010' in SUBTEST 1, '020' in SUBTEST 2, and '030' in SUBTEST 3,

respectively. In this case, SUBTEST 1 will consists of QUESTIONS 1 through 10, SUBTEST 2 will consist of QUESTIONS 11 through 30, and SUBTEST 3 will consist of QUESTIONS 31 through 60. In designing a test that contains subtests, make sure that questions are numbered consecutively. Shade in the appropriate circle below each number (1, 2, 3, or 4).

EXAMPLE 1. – NO SUBTEST

Only one formula is applied to the entire test. If Formula A is chosen, leave the Scoring Formula are Blank. If other formulas are chosen, note the following directions:

FORMULA B: Mark the circle under T, on the same line as R-W/4.

FORMULA C: Mark the circle under T, on the same line as R-W/2.

FORMULA D: Mark the circle under T, on the same line as R – W.

EXAMPLE 2 – SUBTESTS

Different formulas may be chosen for different subtests. If FORMULA is chosen for all subtests, leave the entire Scoring Formula Area Blank; otherwise, note the following directions:

FORMULA A: Leave blank the area under 1, 2, 3, or 4 for Subtest 1, 2, 3, or 4, respectively.

FORMULA B: Mark the circle under 1, 2, 3, or 4, on the same line as R – W/4, for Subtest 1, 2, 3, or 4, respectively.

SORT: (SN)

Test results are automatically listed with students' names in alphabetical order; however, if the area 'SN' is shaded, the test results will be listed according to numeric order of the student social security number. If "None" is shaded in the report, it will list students in the same order as submitted.

PUNCHED OUTPUT – Not currently available.

INDIVIDUAL FEEDBACK – Not currently available.

POSTING REPORT – A report with only the Social Security Numbers that identifies the examinees.

C O P I E S

Mark the circle 1, 2, 3, or 4 for one, two, three or four copies of test results as required. Mark "E" if each student requires her/his own report (Maximum = 30). If this area is left blank, ONE copy will be provided.

TEST OFFICE USE ONLY

Do not mark in this area.

-KEY ANSWER AREA

Mark a correct answer to each item. There can be only ONE correct answer. Do not skip any items. For example, if a test has 50 items, they should be numbered from 1 through 50. Items not marked on the key will NOT be scored or counted in any way.

If you have subtests, design the tests so that items are consecutively numbered. For example, if Subtest 1 contains 25 items and Subtest 2 contains 30 items, the scoring program will grade item 1 through 25 as part of Subtest 1 and Item 26 through 55 as part of Subtest 2.

DO NOT MARK ITEMS THAT ARE NOT TO BE SCORED. Items marked, even if they are beyond the range of the total items marked in Subtests, will be scored. (This feature allows the Subtest area to be considered optional if there are no Subtests.) Avoid any stray marks.

### III. GENERAL INSTRUCTIONS FOR USING STUDENT ANSWER SHEET

1. The student answer sheets have marking instructions on the back. Ask the students to read and follow them. Remind the students to use only a No. 2 pencil.
2. Make sure the students fill in the following items on the answer sheet.

- a. Student's name and initials
- b. Course sort number (IMPORTANT)
- c. Test I.D. – Form and No., if any.

If the Key Sheet has a form designation, make sure the students use the same designation; if there is no form, leave this area blank.

(Warning: Any student answer sheet with “Form” designation different from the “Form” designation on the key will not be scored.)

- d. Student Number – Social Security Number.

- e. Instruct the students to mark only in the required areas. Any stray pencil marks may be picked up as an intended answer or cause the NCS Machine to reject the answer sheet. Do not use the margins of the answer sheet as scratch paper.

#### IV. SUBMITTAL OF MARKED SHEETS TO TESTING CENTER FOR PROCESSING

1. Check ALL the sheets for accuracy and completeness in the following crucial areas:
  - a. Course Sort Number
  - b. Test Form Designation
  - c. Student Number
2. Stack the forms so that they are all oriented in the same way.
3. Return the sheets to the Testing Center.
4. Pick-up the results from the Testing Center. The receptionist will advise you of current turn-around times.

PART II  
STATISTICS

1. MEAN AND STANDARD DEVIATION

Two of the statistical figures provided by the Test Scoring System are the arithmetic mean, or simply mean, and standard deviation of scores. The formulas for computing the mean  $\bar{X}$  and standard deviation S are:

$$S = \sqrt{\frac{\sum_{i=1}^N X_i^2 - \frac{\left(\sum_{i=1}^N X_i\right)^2}{N}}{N - 1}}$$

Where  $\sum_{i=1}^N X_i$  = sum of scores

$\left(\sum_{i=1}^N X_i\right)^2$  = the square for the sum of all the scores

$\sum_{i=1}^N X_i^2$  = the sum of the squares of the scores

$N$  = the total number of scores

The mean of the scores provides the average of the scores in a given group, while the standard deviation provides the deviation of all the scores in the group from the group mean.

The more the scores in a group deviate from the mean, the larger the standard deviation will be.

## 2. Z – Scores ( Transformed Standard Scores )

A standard score is one of commonly used types of derived scores (derived scores are transformed raw scores). A standard score, denoted by  $z$ , is obtained by the formula:

$$z = \frac{X - \bar{X}}{S}$$

Where  $X$  = a raw score

$\bar{X}$  = mean of raw scores

$S$  = standard deviation of raw scores

A standard score indicates how far a given raw score departs from the mean, in terms of standard deviation units. From the formula, one observes that if a raw score is higher than the mean,  $z$  will be positive; if a raw score is lower than the mean,  $z$  will be negative;  $z$  is zero when the corresponding raw score is equal to the mean. Moreover, it can be shown that when any original set of raw scores are converted to  $z$ -scores, the mean and the standard deviation of the resulting distribution is 0 and 1 respectively.

Standard scores are particularly useful in comparing performance on two different tests of scales, when the distributions of raw scores have approximately the same shape. If both distributions were converted to standard scores, a student's performance on the two tests could be compared, because the mean and the standard deviation of each distribution of derived scores would be 0 and 1 respectively.

Since  $z$ -scores often assume negative and fractional values, it can be bothersome to use. In order to eliminate negative and fractional values,  $z$ -scores themselves are often transformed into derived scores with larger mean and standard deviation. Transformed standard scores are

denoted by Z. The transformed Standard Scores, provided by the Test Scoring System, has the mean of 50 and the standard deviation of 10, that is:

$$Z = 50 + 1/10 (X - \bar{X}) / S$$

Where X = a raw score

$$\bar{X} = \text{mean of raw scores}$$

Note: T-scores are Z-scores with the mean of 50 and the standard deviation of 10 that are derived from normally distributed raw scores.

Remarks: Since standard scores from two different distributions are comparable only if the distributions have the same shape, there are limitations in applying standard scores. Moreover, unless two raw score distributions are approximately normal, they rarely have the same shape.

Thus, standard scores are usually not used as derived scores unless the raw distribution is approximately normal.

### 3. KUDER-RICHARDSON COEFFICIENT

The formula for computing Kuder-Richardson coefficient  $r_{tt}$ , which is used in the Test Scoring System, is developed by Cyril Hoyt and is given by: ( See Computational Handbook of Statistics by Bruning & Kintz ).

$$\frac{\left( \frac{\sum_{k=1}^N P_x^2}{I} - \frac{\left( \sum_{k=1}^N P_x \right)^2}{NI} \right) - \left( \frac{\sum_{k=1}^N P_x}{k=1} - \frac{\sum_{k=1}^N P_x}{I} - \frac{\sum_{j=1}^I C_j^2}{N} + \frac{\left( \sum_{k=1}^N P_x \right)^2}{NI} \right) \frac{1}{(N-1)(I-1)}}{\left( \frac{\sum_{k=1}^N P_x^2}{I} - \frac{\left( \sum_{k=1}^N P_x \right)^2}{NI} \right) \frac{1}{N-1}}$$

Where  $P_k$  = the number of correct answers of the  $K^{\text{th}}$  student  
 $N$  = total number of students taking the test  
 $I$  = number of items in a given test  
 $G$  = the number of students who correctly answered the  $J^{\text{th}}$  item

The coefficient,  $R_{TT}$ , provides a measure of internal consistency of the items in a given test. A high reliability coefficient (.70 or higher) would mean that the items are highly consistent with each other; namely, the individual items in the test are producing similar patterns of responses in different people. Thus a high reliability coefficient would mean that the test items are homogeneous.

#### 4. POINT-BISERIAL CORRELATION COEFFICIENT

Intuitively, an item in a test should be considered “good” if the majority of high performers would answer it correctly. On the other hand, an item should be considered “bad” if the majority of low performers would answer it correctly. One technique to test this is to find a correlation of a continuous variable (whose values are the whole test scores) with a dichotomous variable which takes on the value 1 or 0 case the answer of a given item is correct or incorrect; and then interpreting the correlation according to some statistical method. We start by computing a point-biserial correlation coefficient  $r_{pbi}$  for each item. The  $r_{pbi}$  that corresponds to the  $i^{\text{th}}$  item in the test is given by the formula:

$$r_{pbi} = \frac{Y_1 - Y_0}{S_y} \sqrt{\frac{N_1 N_0}{(N-1) N}}$$

Where  $N_1$  = the number of students who produce correct answers for the  $i^{\text{th}}$  item in the test

$N_0$  = the number of students who produce an incorrect answer for the  $i^{\text{th}}$  item in the test

$N$  = the total number of students taking the test

$\bar{Y}_1$  = the mean criterion scores for those who produce a correct answer for the  $i^{\text{th}}$  item

$\bar{Y}_0$  = the mean criterion scores for those who produce an incorrect answer for the  $i^{\text{th}}$  item

$S_y$  = the standard deviation of the criterion scores

Note: The criterion score of a student is the number of his or her correct answers in the test.

From the formula for  $r_{pbi}$ , it is immediate that a “high”  $r_{pbi}$  corresponds to a “good” item and a “low”  $r_{pbi}$  corresponds to a “bad” item. In order to give an unambiguous meaning to “high” or “low”  $r_{pbi}$ , we utilize the formula

$$t = r_{pbi} \sqrt{\frac{N - 2}{1 - (r_{pbi})^2}}$$

Where  $N$  = the number of students taking the test. To give a meaningful discussion of this formula would be beyond the scope of this handbook, and we shall refer the interested readers to STATISTICAL INFERENCE by H.M. Walker and J. Lev. For our purpose, it suffices to say that it has a students’ distribution with  $N=2$  degrees of freedom and that the following discussion is based on a statistical table of statistic  $t$  (for example, see HANDBOOK OF STATISTICAL TABLES BY D.B. Owen). For

our purpose, we use a two-tailed test with .05 level of significance and let  $t_{.05}$  be the corresponding  $t$  value with  $N-2$  DEGREES OF FREEDOM. Now we can state the testing procedure. If  $t < t_{.05}$ , we say that  $r_{pbi}$  is “small” and that there is no evidence the  $i_{th}$  item of the test is measuring what the other items measure. If  $t > t_{.05}$ , we say that  $r_{pbi}$  is “large” and the  $i_{th}$  item is statistically consistent with the other items.

Note: Because of the limitation of computer primary storage space, the analysis of  $r_{pbi}$  will be printed out only for a class not exceeding 102 students. For larger classes, the  $t_{.05}$  value can be looked up from the statistical tables for statistic  $t$ .