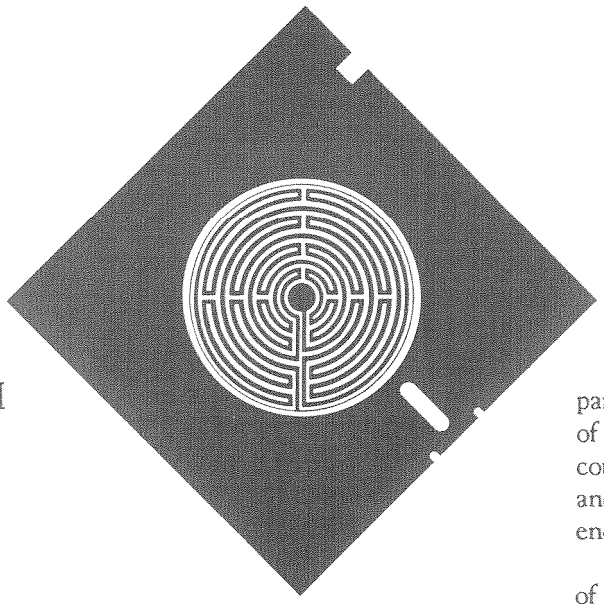
PROGRAM IN THE ACADEMIC COMPUTING LAB



USING THE PROGRAM TO ANALYZE TEST ITEMS

Diane L. Willey
Professor of Education

WHY TAKE THE TIME? Questions on a test should reflect important course content, so that performance on the test will be an indication of the students' actual learning. Using the item analysis program adds only about 10-15 minutes to your test scoring time for a class of 35-40 students. Yet, the information you obtain can be used to help you improve your tests, revise your instruction, and increase the accuracy of your grading decisions.

HOW DO I DO IT?

From the program options described in Steve Scherer's companion article, you should select "Option 1: Scan and analyze data" with a test key to obtain an item analysis of students' responses. Upon completion of the item analysis, there are three primary areas of the data to examine: 1) the patterns of student responses; 2) item difficulty level; and 3) item discrimination.

1. **Student Response Patterns.** The pattern of the students' responses allows you to check for keying errors and to examine the effectiveness of your answer options in multiple choice questions.

Check for miskeying. If most students have picked an incorrect answer option, e.g.:

Ques # 5	A	B	C	D	E
	3	31	[1]	2	0

then you have probably marked the wrong answer on the key (You hope!). If the item is not miskeyed, this response pattern may indicate that incorrect information was presented in the class, text, or other content source.

Check for guessing. If all answer options are selected with approximately equal frequency, the class as a whole probably has no idea of the correct answer and is guessing. This response pattern may also indicate that the item is assessing content which was not covered, the item is not written clearly enough for the students to understand what is being asked, or the item is too difficult for the class.

Check for ambiguity. If several students select one answer option with about the same frequency as the correct answer, then the item may have two correct answers. This pattern may also indicate that conflicting information was presented by different course content sources such as lecture and the text.

2. **Difficulty level.** The difficulty level represents the percentage of students who got the item correct.

$$p = \frac{\text{\# students selecting correct answer}}{\text{Total \# of students attempting the item}} \times 100$$

The greater the number of students getting the item correct, the easier the item. The "ideal" difficulty level for a particular item will depend upon the level of overall test difficulty you desire, the course content which the item is testing, and whether you are using norm-referenced grading.

Overall test difficulty. The average of the individual item difficulties gives you a difficulty level for the test as a whole. If you decide that your grades are too high (or too low) on a given test, you can adjust the overall test difficulty by substituting different items with lower (or higher) difficulty indices.

Course content. For course content which you consider very important, you may include test items which you expect all or most students to answer correctly. If too few students are answering these items correctly, you may choose to revise the items or your instruction. Typically, a mastery-focused test should have an average test difficulty of about 90 percent.

Norm-referenced grading. Norm-referenced grading involves ranking all students' final course point totals (or averages) and assigning a predetermined number of As, Bs, Cs, etc. If you use this grading procedure, then 50 percent is the ideal difficulty level for your test items. Students' scores will then be maximally spread out across the score scale, so that grade cutoffs may be more easily determined.

3. **Discrimination Index.** The discrimination value of each item describes how effectively an item separates the students who know the content well from those who do not. It is calculated by subtracting the proportion of students with the lowest total test scores who got the item correct from the proportion of students with the highest total test scores who got the item correct.

(continued on page 8)

Collection *(continued from page 6)*

having to feed the sheets through the scanner again

This program only allows one correct response for each question, so if you want to allow more than one correct response, you may need to eliminate that question from the analysis by weighting it with a value of zero.

Option 2: To create only a data file with no printing.

The data file, called C:/NCS3000/DATAFILE.TXT, contains one record for each sheet scanned. This record begins with the STUDENT ID followed by 100 sets of five binary digits (ones and zeros) which mimic the marks made for each question. For example, if an A and a D were marked on a single question, the result for that question on the data file would be 10010.

The data file is stored on a hard disk, so you don't need to worry about the size of the data file. However, when you get ready to transfer it to a floppy disk, you may need to convert the data to a more compact form using Option 4 in the menu to make sure it will fit onto your disk.

Option 3: To transfer the data file from the hard disk to a floppy.

If you have just used Option 1 or Option 2, the data file called C:\NCS3000\DATAFILE.TXTwas created to hold the images of the scanned sheets. This option can be used to copy that file to your own diskette. Since this file is in image form, it may be very large, if you scanned many sheets. A translated version of the data which replaces each set of five binary digits with one character may be more appropriate for your use. Option 4 allows you first to convert the raw data to the more compact form.

Although this option has been set to default to the data file created by the scanner, it can be used to transfer any file from one place to another. If the default is the file you want to copy, you need only press the Enter key when prompted for the file name.

After you have chosen the source file to be copied, you will be prompted for a destination file name. First make sure you have a diskette in the drive to which you wish to copy, and then include the drive designator in the path name for the destination file. For example, if you want to save the original data to a file by the name MYDATA on a diskette in drive A, use the name A:MYDATA when asked for the destination file.

Option 4: To convert groups of five binary digits to characters.

This option will compact the data by a factor of about five. On the original data file, each question response was represented by five binary digits. These binary digits are converted to unique characters, thereby replacing five characters with one character.

You are given a choice of how to form the conversion. You may choose to convert a blank and each single mark as a unique character and lump all multiply-marked answers into one value, or you may choose to have 32 different characters produced.

If only one mark per question is allowed, the conversion will produce an A, B, C, D, or E for exactly one mark; an @, if no marks were made; or an M, if multiple marks were entered. You can get a copy of the conversion into the 32 characters from Academic Computing Services.

Analysis *(continued from page 7)*

$$D = \frac{\text{(\# in top 25\% group who got item correct)}}{\text{Total \# in top group}} - \frac{\text{(\# in bottom 25\% group who got item correct)}}{\text{Total \# in bottom group}}$$

	Scored in top 25% of group	Scored in bottom 25% of group
Got item correct		
Got item wrong		

The discrimination value ranges from -1.00 to +1.00 and may be interpreted in the same way as a correlation coefficient.

A positive D-value means that students who had high total test scores got the item correct more often than students who had low total test scores. This is an acceptable item which should be retained.

A negative D-value means that students who scored the lowest on the test were more likely to get this item correct than students who scored the highest on the test. This pattern is the opposite of that desired! This item should be eliminated from the test scores for this group of students and revised for future use.

A zero D-value means that students who had high test scores and students who had low test scores got the item correct with about the same frequency. This is a weak item *or* an item with a high or low rather than moderate difficulty level. If all the students get the item correct, there is no "bottom 25 percent group" and thus no discrimination is possible. Similarly, with this program, if all of the students who get the item correct are in the top 25 percent group or in the middle 50 percent group, there is still no bottom 25 percent group, and consequently, discrimination will again be zero.

As the numerical value of D approaches 1.00, the item's discrimination between the top and bottom groups increases. A D-value of +1.00 is most desirable, since it indicates that all the students in the upper group got the item correct and all the students in the lower group got the item incorrect.

Using the information provided by the analysis program allows you to identify strengths and weaknesses in the machine-scored items on your examinations. Retaining strong items and revising weak items strengthens the data on which you are basing your grading decisions and your conclusions about what students are learning in your courses.