

A User's Guide To

**BRILLIANT! TEST SCORING AND ITEM ANALYSIS**

February, 2005

Program Brilliant!: Test Scoring and Item Analysis .....	2
STUDENT TEST FILE.....	2
Test Key Attributes .....	3
Accessing Program Brilliant! .....	3
Program Brilliant! Execution .....	4
Printing the test results .....	4
Link to the Excellent!! Grades Management Templates for Microsoft Excel.....	5
STUDENT SCORE REPORT in student_scores.txt.....	5
Test Statistics Report.....	5
SAMPLE TEST STATISTICS REPORT .....	6
ITEM ANALYSIS REPORT .....	7
SAMPLE ITEM ANALYSIS.....	7
INTERPRETATION OF DIFF(T) VALUES .....	8
RELATION OF ITEM DIFFICULTY TO RELIABILITY IN AN ACHIEVEMENT TEST.....	8
ITEM DISCRIMINATION INDEX, (DISC).....	8
DISCRIMINATION INDEX INTERPRETATION .....	8
ACHIEVEMENT TEST INTERPRETATION.....	9
MASTERY TEST INTERPRETATION .....	9
DISTRIBUTION TABLES FOR DIFFICULTY AND DISCRIMINATION INDICES.....	9
SUMMARY .....	10
STUDENT RESPONSE RECORD.....	10
Test Scoring Enhancements .....	10
Survey items.....	10
Scoring either more than one exam or multiple exam versions.....	10
Unscored student exams.....	11
Assigning different point values to test questions .....	11
Partial credit answers .....	11
Subset Scores .....	11
APPENDIX A: FORMULAS USED IN PROGRAM BRILLIANT!.....	12
APPENDIX B: FORMAT OF THE STUDENT TEST FILE .....	13
FIVE-RESPONSE ANSWER FORMS (USING STANDARD-200 ITEM OPSCAN FORMS):.....	13
TEN-RESPONSE ANSWER FORMS (USING STANDARD-120 CONTROL FORMS):.....	13

## Program Brilliant!: Test Scoring and Item Analysis

Program Brilliant! will score selection type tests that have been read from OPSCAN answer forms into a file (the *student test file*). In addition to scoring a test and obtaining a test performance profile, program Brilliant! provides a means of test analysis to identify poor or erroneous items. The instructor can use the information provided to improve test items for future use.

This program provides for enhanced scoring to include:

- Multiple tests identified by the special code field may be scored together.
- Items may be assigned different point values for correct answers.
- More than one response for an item may be awarded points.
- Subsets of items may be defined and scored.

OCCS will score tests submitted to the computer operations staff. Complete an Optical Scanner usage form. You may request data files and program output to be emailed to your ODU email account.

### ***STUDENT TEST FILE***

The data required by Program Brilliant! is a file containing scanned student test forms and one or more answer keys. The filename of the student test file for 5 response (blue) OPSCAN forms is usually in the form BLxxxxx.dat where xxxxx is the time of day the answer forms were read. For 10 response (Green) forms the filename is GRxxxxx.dat. The key to error-free scanning is the instrument used to mark the answer forms. Number 2 pencils are acceptable, but an electronic scoring pencil is preferred.

- The fields read from the OPSCAN form are the NAME, STUDENT ID, SPECIAL CODE and ITEM RESPONSE fields. The KLM Special Code field is used on student forms to identify which key to use when scoring more than one test. The NOP field is not used on student forms for any scoring attributes, however, if the KLM field is active to identify key forms, then student entries in the entire KLMNOP field are displayed on the output reports.
- Best practice is for students to enter their ID numbers left justified; unused columns of the ID field should remain blank. No validation of the ID field occurs to prohibit scoring a test.
- Only one answer may be marked in each answer field (A thru E or A thru J) for a question. **An asterisk will be put in the output field for any answer field with more than one response marked and the question will be scored as incorrect.**
- Student answers left blank will be scored as incorrect in computing the *percent correct*. The *relative percent* score is based only on answered questions. A difference in *percent correct* and *relative percent* scores may indicate an incorrect marking instrument was used (pen ink will not be read by the scanner) or that a student did not answer all questions.
- The student test file may be created from the standard 200 questions, 5-response (BLUE) answer form or the 120 questions, 10-response (GREEN) answer form. The program looks for any answer on the key with a response of F through J to determine if a 10-response form is in use. **Scoring a test using 10 response forms without any keyed answer above the E distracter will have correct scores but incomplete item analysis if any students select an answer above E.**

## **Test Key Attributes**

- The *KLM Special Code field* may be used to indicate a **two-digit or three-digit test code** to identify a key to be used for scoring. If there is a **single key** submitted to OCCS for scoring, the special code field is ignored and all forms are scored with the single key. (For the faculty interactive version of Brilliant!, the user is prompted to use or ignore a single key with a test code.) If there are **multiple keys** read with the student tests, then the SPECIAL CODE KLM fields on student forms are used to identify the corresponding key used to score that form. You may use either the KL fields to identify a two digit test code (00 to 99) or the KLM field to identify a three digit test code (000 to 999). It is not good practice to mix two and three digit test codes in one packet of forms.
- The *NOP special code field* may be used on test keys to indicate item point values. Multiple test keys can be prepared for a test to award points for partial credit answers. To set point values for a key, set the N field is set to 9; the O field is set to 1, 2 or 3 since you can set up three point value keys for each test to be scored; the P field is set to the point value to be awarded to all keyed items on the form and can be set to 1 through 9.
- An additional use of the *NOP special code field* is to specify item subsets on the test. The item subsets created on this key form are scored separately and output in columns to the scores sheet and the score transfer file. To specify a subset form, set the NOP field to 411. Instead of marking an answer for each question, this form uses answers A-E to assign each item into one of five score subsets. If a ten response answer form is used, then you can specify up to 10 item subsets.
- Separate Item Analysis and Test Statistics reports are generated for each test when using KLM test codes to score multiple tests. The student Response Record and Score Reports are single reports consolidating all tests scored.

## **Accessing Program Brilliant!**

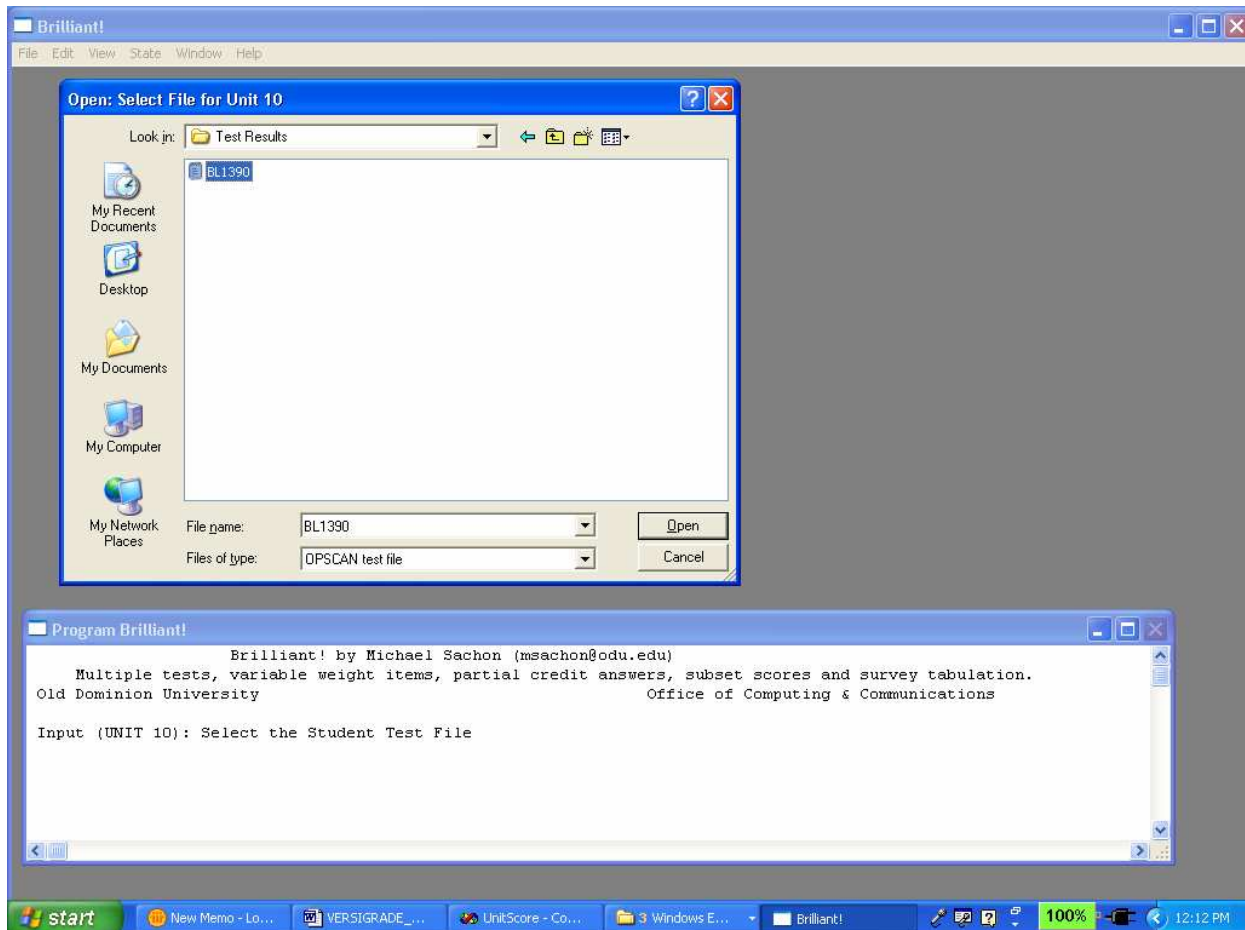
OCCS Operations Staff will scan forms and run the Brilliant! program. The data file will be emailed to the instructor's ODU email account. Program Brilliant! is accessible on the ODU LAN so that faculty may score tests at their desktop. This is a benefit if scoring multiple exams in one file because you may interactively correct problems with incorrect or invalid student test codes.

Program Brilliant! is available to faculty on the Violet network server. You can click on the *Start* button at the bottom left of your screen. Click on *Run* from the list of options to display the RUN Window. Click on *Browse* to display the Browse window. Select *Voll on Violet* from the list of network drives that are mounted. Click on the down arrow beside the *Look in:* box to display your network drives and select the Violet server. Double click on the *Data* folder, followed by the *Grades* folder, and finally the *Program Brilliant!* folder.

Double click on the *Program\_Brilliant!.exe* to start the program. While you may copy the executable to your hard drive, it is not recommended. If you execute Program Brilliant! directly from the network you are assured of running the latest version of the program. You can generate a shortcut on your desktop by right clicking on the filename and holding the button down while dragging the filename to your desktop.

## Program Brilliant! Execution

The program will initially prompt to identify which student test file to score. The default folder on first execution of the program will be the folder where the program resides. Traverse the network to your folder containing your student test file. It is recommended that you save opscan files and test results on your network file space (typically H: drive). The output files are created in the same folder as the location of the student test file. The path to this location is saved in a file on your C: drive so that on subsequent executions of the program, the default file folder will be the location of the last student test file scored.



## Printing the test results

Printed output from Program Brilliant! is two files, TEST\_SCORES.TXT and ITEM\_ANALYSIS.TXT. These files are saved in the disk folder containing the Student Test file. The files are overwritten with each subsequent run of the program. You may choose to send the output files directly to your current default printer or save them to disk.

You can change the default printer by clicking on *Start*, then click *Settings*, then click *Printers* from the Settings list box. A *Printers box* will display all available printers. Click an icon to select a printer, then click *File* on the menu bar, and click on *Set as Default* in the drop down list. A check mark beside *Set as Default* indicates the printer is already your default printer. You can set the default printer while Program Brilliant! is prompting for input..

## ***Link to the Excellent!! Grades Management Templates for Microsoft Excel***

The score transfer file (the ST File) is a program option to create a file containing student id, student name, and raw, absolute percentage, relative percentage and standard scores. The file may be formatted as either a comma delimited file or as an Excel SYLK file. The SYLK format provides column format definitions understood only by Microsoft Excel, while the comma-delimited file does not contain format information but can be imported into any spreadsheet application. Once loaded into Excel, the scores may be copy-and-pasted into a class grade book. OCCS provides a spreadsheet grade book solution, the Excellent!! Grades Management Templates for Microsoft Excel. Reference the documentation for more information.

### ***STUDENT SCORE REPORT in student\_scores.txt***

The Faculty STUDENT SCORE Report is a list of student scores sorted alphabetically by last name lists containing both the complete Student ID and a corresponding truncated Student ID along with the following scores:

- *ITEMS CORRECT* is the number of correct responses on the entire exam.
- *Percent Correct* is the percentage of correct responses for the entire exam.
- *RELATIVE PERCENT* is the percentage of correct responses of only those questions attempted by the student. The difference between the absolute score and the relative score is the percent of unanswered questions by the student.
- *STANDARD SCORE* is the student's score converted to a relative score based on a class mean of 500 and a class standard deviation of 100. For example, a standard score of 600 is one standard deviation above the mean.
- *PERCENTILE (%-tile)* is the percentage of the class scoring below the student's score plus one-half the percentage of the class with the same score. If KLM multiple key test scoring, then the percentile score is calculated within that test version only.

A similar list of scores without student names is output in sequential order by a truncated student number for posting. On this "post" report, there are *Answer Form* columns that include the number of keyed questions that were unanswered and the number of invalid responses caused when the student put more than one answer in an answer field (where the scanner outputs an asterisk to the student test file). If more than one test is being scored, then the standard and percentile scores are not printed to the post report to eliminate confusion when students with the same score on different tests would not have the same standard and percentile scores.

### ***Test Statistics Report***

A page of statistics on the test includes the *MEAN SCORE* (average), *MEDIAN* (midpoint in the distribution of scores), the *STANDARD DEVIATION* (a measure of dispersion) and the *HIGH* score and *LOW* score distribution end points.

These statistics are listed for the raw score (number of points), percentage score, relative percentage score, and the standard score. A score distribution report shows the frequency of each score and the cumulative frequency of scores in ascending order.

The Test Statistics Report lists the obtained statistics computed on the raw scores and the recommended statistics based on the number of items on the exam and the "chance" score. The *Reliability Coefficient* is a measure of the consistency of the test in measuring whatever the test



## SCORE DISTRIBUTION

Test Score			Cumulative		Test Score			Cumulative	
Raw	%	Freq	Freq	Percent	Raw	%	Freq	Freq	Percent
50	43.1	1	1	4.0	89	76.7	2	18	72.0
52	44.8	1	2	8.0	90	77.6	1	19	76.0
53	45.7	1	3	12.0	92	79.3	1	20	80.0
58	50.0	1	4	16.0	93	80.2	1	21	84.0
62	53.4	1	5	20.0	94	81.0	1	22	88.0
63	54.3	2	7	28.0	97	83.6	1	23	92.0
69	59.5	3	10	40.0	99	85.3	1	24	96.0
76	65.5	1	11	44.0	116	100.0	1	25	100.0
78	67.2	1	12	48.0					
79	68.1	1	13	52.0					
83	71.6	1	14	56.0					
87	75.0	1	15	60.0					
88	75.9	1	16	64.0					

### **ITEM ANALYSIS REPORT**

The ITEM ANALYSIS gives, for each item, a cross tabulation of each response for the item. The cross tabulation table for each item contains columns for groups of students based on performance on this exam. The rows correspond to each response and the cell values are the count of students selecting the response from each group. The columns correspond to the total class count, the class total percent correct (this is the difficulty index (DIF(T))), and the counts of responses in each student group of high scores, middle scores, and low. The high and low groups are comprised of (approximately) the top 27 percent of scores and the bottom 27 percent of scores. The column values contain the number of students in a group selecting each response, those omitting the item (OMIT) and those marking more than one response (denoted ERR and in the student test file as an asterisk). Responses which were not selected by at least one student are not output to the table unless it is a correct answer. Printing only foils that have student counts provides visual feedback on the non-performance of unused distracters. The last column is the item discrimination index, denoted DISC.

### **SAMPLE ITEM ANALYSIS**

	<b>17</b>	<b>COUNT</b>	<b>%</b>	<b>HIGH</b>	<b>MID</b>	<b>LOW</b>	<b>DISC</b>
<b>A</b>		15	20	3	7	5	-0.10
<b>C +</b>		51	68	17	28	6	0.55
<b>E</b>		9	12	0	0	9	-0.45

The interpretation of these computed values depends upon the classification of the test as either an achievement or a mastery test. An achievement test involves fairly complex tasks requiring the student to make a fine discrimination between the alternatives. Mastery tests examine the basic or minimal knowledge expected by all students to indicate understanding of the material. On mastery tests, most students would be expected to know the material being tested, so student scores would tend to be higher than on achievement tests. The Difficulty Index (column % for the correct answer) is the percentage of students correctly answering the item of all students examined (including those students omitting the question). The general interpretation is the higher the DIFF(T) value, then the easier the question. For foils, the % column shows the

percent of students selecting each foil. Foils not being selected by any students are not listed in the table.

#### INTERPRETATION OF DIFF(T) VALUES

<b>ITEM CLASSIFICATION INTERPRETATION</b>	<b>MASTERY</b>	<b>ACHIEVEMENT</b>
<i>SUGGESTED FOR MAXIMUM ITEM RELIABILITY</i>	ABOVE 70*	40-60
<i>SUGGESTS INADEQUATE INSTRUCTION AND/OR INAPPROPRIATE OBJECTIVES</i>	BELOW 60*	LOW VALUES: BELOW 30 HIGH VALUES: ABOVE 70

\* Depends on the Criterion Level established.

The table contains general guidelines for interpretation of the Difficulty index; however, the discrimination index should also be considered in determining reliability of an item.

#### RELATION OF ITEM DIFFICULTY TO RELIABILITY IN AN ACHIEVEMENT TEST

A test composed of items with a difficulty index of approximately 50 will be more reliable than a test with a linearly distributed range of difficulties. It will be much more reliable than a test composed of two groups of items with highly divergent item difficulties (i.e., only very hard and very easy items). Items either correctly or incorrectly answered by the entire class do not contribute to reliability. Items of mid-difficulty that are non-discriminating between the high and low scores groups do not contribute to test reliability.

#### ITEM DISCRIMINATION INDEX, (DISC)

The Item Discrimination index is defined as the difference in the proportion of correct responses between a group composed of the high 27 percent of scores on the test and a group composed of the low 27 percent of scores. Stated another way, the discrimination index is the difference in the item difficulty (as a decimal fraction instead of a percent) between the high scoring students and the low scoring students. The discrimination index is an indicator of how each item reflects achievement in the tested material. Positive values indicate that a greater proportion of the high achieving group answered an item correctly than in the low achieving group. The converse is true for negative discrimination values. An item with a negative discrimination value may be detrimental to a test. Negative discrimination may indicate the item is measuring the wrong objective or contains a fault that is misleading to the overall higher performing group.

#### DISCRIMINATION INDEX INTERPRETATION

<b>ITEM CLASSIFICATION DISC INTERPRETATION</b>	<b>MASTERY</b>	<b>ACHIEVEMENT</b>
GENERALLY ACCEPTABLE ITEM	.10 TO .39	.40 TO 1.00
CONDITIONALLY	ABOVE .40	.20 TO .39

ACCEPTABLE ITEM		
POOR ITEM (REVISE OR REJECT)	BELOW .10*	BELOW .19

The value of the discrimination index is dependent on the clarity in expressing the question, the appropriateness of the correct response, and the attractiveness of the wrong response (i.e. the distracter or foil) to the less knowledgeable student.

#### ACHIEVEMENT TEST INTERPRETATION

High discrimination index values are expected if the test is designed to measure achievement. Score discrimination between "better", "good" and "poor" performing students based on the total test performance is expected for each achievement item. Conditionally acceptable items, that is, items in which performance on the item does not correspond to performance on the total exam, may need to be reworked to provide clarity to the wrong answer choices that "fooled" the best scoring students.

#### MASTERY TEST INTERPRETATION

Having lower discrimination values for good mastery test items is reasonable since the test is attempting to show acceptable knowledge of the material vice a differential in achievement. Thus, a high discrimination value may indicate that the complexity of the item is beyond that intended for general mastery of the material.

#### DISTRIBUTION TABLES FOR DIFFICULTY AND DISCRIMINATION INDICES

Following the item analysis output are distribution tables for the Difficulty and Discrimination Index values in ascending order for each index. Listed for the Difficulty Index table are the index value and the list of questions having this difficulty index. Following each item number is the corresponding discrimination index for the item in parenthesis. Likewise, the Discrimination Index distribution table has the corresponding difficulty index in parenthesis following each item with a given discrimination index.

The item difficulty index distribution table lets you quickly determine the more difficult items (items with low difficulty index values). Further, you can look at which of these more difficult items have corresponding low or negative discrimination values indicating the students scoring highest on the exam performed no better or worse on the item than the lowest scoring students on the test. You can then look at the detailed item analysis for the item and the test item to determine what may have caused this effect.

The item discrimination index distribution table lists in ascending order discrimination index values from -1.0 to 1.0 and the corresponding items with each discrimination index value. Beneath each item number is the corresponding difficulty index for the item. You can use this table to identify non-discriminating items (both high and low score groups performed the same on the item) or negative-discriminating items (low score group performed better on an item). The combination of the two indices can be used to identify items which you may want to revise or can identify key errors.

## SUMMARY

Together, the difficulty and discrimination indices and the cross-tabulated table provide a means for the instructor to judge the quality of each item. The quality of the entire exam can be inferred from the distributions of the Difficulty and Discrimination indices. The distribution of values for most items should follow the guidelines given in the previous tables for the particular type of test.

## STUDENT RESPONSE RECORD

The STUDENT RESPONSE RECORD lists the student's responses to each item. In place of a correct answer, a plus sign (+) is listed. Otherwise, the student's incorrect response from the answer form is listed. A blank space is listed for items not answered and an asterisk is listed if a student marked the item with more than one response. One copy of the response record is sorted by last name, and another copy for posting is sorted by truncated student ID number.

### Test Scoring Enhancements

Program Brilliant! provides significant enhancements for test scoring.

- By completing the NOP special code fields on test key OPSCAN forms, answers can now be awarded different point values.
  - NOP special code keys allow more than one answer for a question to receive full or partial credit.
- To improve test security, the program supports processing of multiple tests using the KLM special code field to match each student test with its key.
  - For multiple test versions, a report listing student tests that can not be scored due to test code identification includes those student's scores using all available keys.

### *Survey items*

The Item Analysis Table will also provide counts for students responding to **unkeyed** items. This feature may be used for survey and it will also identify missing answer key item tabulation. The response record listing student names will list survey item responses as well as test items. The "post" response record, sorted by truncated student id numbers, contains only test responses. If Opscan forms are processed without a key, the item analysis table (with response counts and percentages) and the response record are output as survey tabulation.

### *Scoring either more than one exam or multiple exam versions*

The program will score multiple tests in one scanned file if each student test and corresponding answer key are identified on the Opscan form. A key for each exam is identified by completing the KL or KLM special code field. Acceptable codes are 000 to 999 if using the KLM fields, or 00 to 99 if using the 2 digit KL fields to identify keys. It is not recommended to mix two and three digit test codes in the same packet of exams. Students code their exam using the same KLM fields. When scoring multiple exams, each exam will have separate Test Statistics and Item Analysis reports. The Response Record and the Student Score Report are single reports encompassing all tests scored.

### ***Unscored student exams***

If a student's exam has a missing or incorrect test code, the student is listed in the response record and score reports as having an unscored exam. To facilitate obtaining test results for unscored students, the Test Analysis report will list all unscored students and their scores using all available keys. Once it is determined which exam was taken, the student scores may be copied from the test analysis file to a spreadsheet or gradebook. Request that operations email the Item Analysis and Score Report files.

### ***Assigning different point values to test questions***

Instead of all questions having a single of 1 point counting toward the score, the NOP special code fields can be used to specify the number of points (1 thru 9 points) that are awarded for each answer on the key. For the same exam, additional NOP key answer forms can assign other point values to item responses. Up to three point values can be assigned by putting 91x, 92y, and 93z in the NOP special code field on different Opscan key forms. If the same answer for an item is listed on more than one NOP answer field key form, that answer is awarded the sum of points from the two NOP keys *with a maximum allowable point value of 9 points*. The Item Analysis table will display the number of points awarded for each response. The Response Record will show the points awarded for each student's correct answers instead of a plus sign.

### **Partial credit answers**

The NOP key forms can be used assign credit for more than one answer for a question. Items can be constructed to allow partial credit for correct but incomplete answers to complex questions. For example, you can define a 911 key to assign 1 point for keyed items on that form, and a 923 key to assign 3 points to keyed items on that form. If question number one has a B marked on the 911 form and a C marked on the 923 form, then a student with answer B receives one point and another student with an answer C will receive 3 points for that question. It is important to note that only one response to a question will be read by the scanner. A student cannot answer B and C to this question as the scanner, detecting more than 1 response to the question, will put an asterisk in the output field and the question will be scored as an incorrect.

### ***Subset Scores***

A different type of NOP key form can be included to define up to 5 question subsets for a five response form, or 10 question subsets for a 10 response form. To create a subset definition key, enter 411 in the NOP special codes field. The subsets are defined by completing the key marking an answer letter to assign each item to a subset. That is, an A response for any item assigns that item to subset 1, a B response for an item assigns it to subset 2, etc. The subsets are scored with the subset scores output to the score reports and to the score transfer file (in Excel or comma delimited format). You may define NOP keys (variable credit or subset score definition) in conjunction with defining different KLM test code keys. Make sure the same KLM test code is correctly completed to link the different types of key information for each test.

**APPENDIX A: FORMULAS USED IN PROGRAM BRILLIANT!**

1. STANDARD SCORE (SS) =

$$\left( \frac{(\text{RAW SCORE}) - (\text{RAW SCORE MEAN})}{(\text{RAW SCORE STANDARD DEVIATION})} * 100 \right) + 500$$

2. STANDARD DEVIATION (SD) =

$$\frac{(\text{SUM OF SQUARES OF EACH SCORE}) - \frac{(\text{SQUARE OF SUM OF SCORES})}{(\text{NUMBER OF STUDENTS})}}{(\text{NUMBER OF STUDENTS})}$$

3. RELIABILITY (KUDOR-RICHARDSON 20) COEFFICIENT (R) =

$$\left( \frac{K}{K-1} \right) * \left( 1 - \frac{\text{SUM}(P*Q)}{S^2} \right)$$

K = NUMBER OF TEST ITEMS

P = PROPORTION OF CORRECT RESPONSES FOR AN ITEM

Q = PROPORTION OF INCORRECT RESPONSES FOR AN ITEM

SUM = SUMMATION OF P TIMES Q FOR ALL ITEMS

S<sup>2</sup> = VARIANCE OF THE RAW SCORE

4. STANDARD ERROR MEASUREMENT (SE) =

$$SO * \text{SQRT}(1-R)$$

SO = STANDARD DEVIATION OF SCORES

R = RELIABILITY COEFFICIENT

**APPENDIX B: FORMAT OF THE STUDENT TEST FILE**

The test records are stored in the student test file with the following format depending on the type of OPSCAN answer form used.

**FIVE-RESPONSE ANSWER FORMS (USING STANDARD-200 ITEM OPSCAN FORMS):**

BROOKS N	12345678	123	012340123401234012340....
20 CHARACTER NAME FIELD	10 CHARACTER STUDENT ID	6 CHARACTER TEST CODE	ITEMS RESPONSES 1..200 (200 CHARACTERS) A-E responses are coded as 0-4, respectively

**TEN-RESPONSE ANSWER FORMS (USING STANDARD-120 CONTROL FORMS):**

ALLEN W	123456789	123	01234567890123456789....
20 CHARACTER NAME FIELD	10 CHARACTER SSN	6 CHARACTER TEST CODE	ITEMS RESPONSES 1..120 (120 CHARACTERS) The A-J are coded 0-9