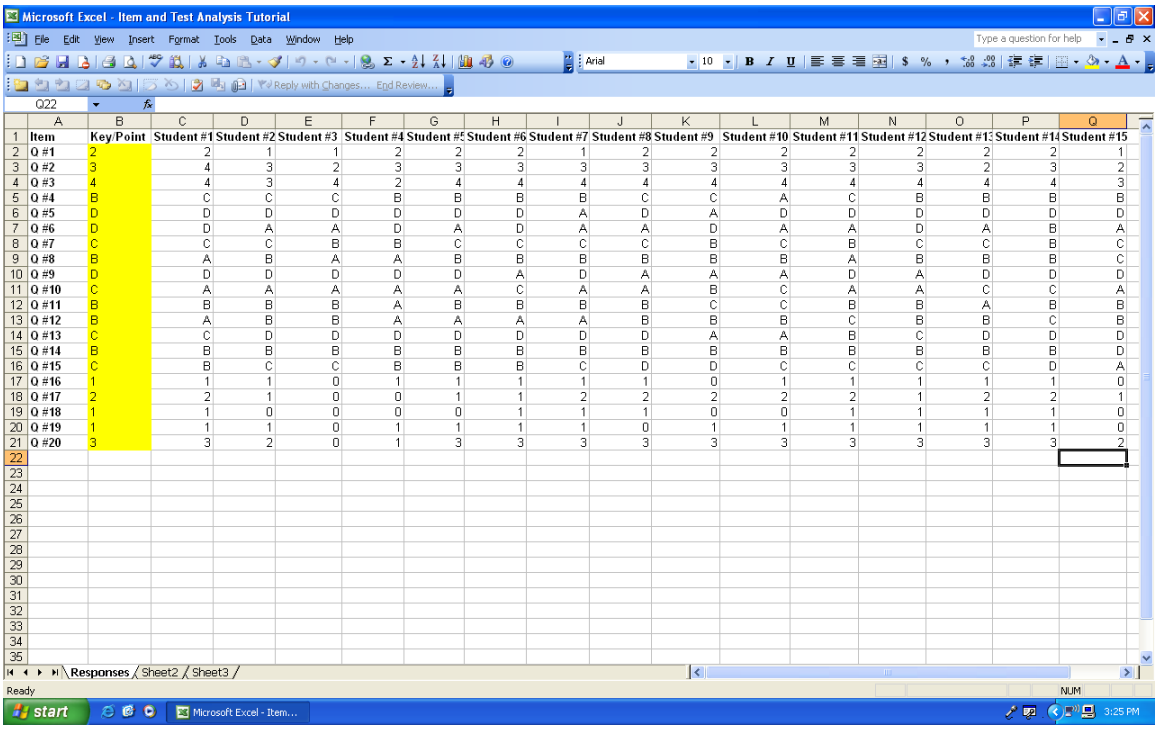
Using MS Excel to Analyze Data: A Tutorial

Various data analysis tools are available and some of them are free. Because using data to improve assessment and instruction primarily involves descriptive and simple inferential statistical analysis, general purpose computer packages such as MS Excel are sufficient. The advantage of using spreadsheet programs such as MS Excel is that it is available virtually anywhere (at school, at home) and on any computer platform (PC and Mac computers).

The following tutorial illustrates step by step procedures of analyzing assessment data using the MS Excel spreadsheet program. In this tutorial, we assume that the test has 20 questions, among which 15 are in the multiple choice type and 5 are in the constructed response type. We also assume that the test was given to a class of 15 students.

1. Entering student responses:
Start the MS Excel spreadsheet program. Double-click at Sheet1 at the left-hand corner of the sliding bar at the bottom, type in "Response" to rename Sheet1. We will designate rows for individual items, and columns for individual students. In row 1 starting with cell A1, we type in the following labels for columns: "Item", "Key/Point", "Student#1", "Student#2", … "Student #15". For column A starting with cell A2, we type in the following labels for rows: "Q #1", "Q#2", "Q#3", … "Q #20". Format the column and row labels as you like by using different fonts, sizes, colors, etc. The column width may also be increased by dragging the cell division lines forward. The data table is now set and ready for entering student responses. First, enter the Keys to multiple choice questions and maximal points for constructed response questions. Then enter students' responses (A, B, C, D or 1, 2, 3, 4) for multiple choice questions and points earned for constructed response questions. You have now created your data table, and ready to conduct item and test analyses. Save your Excel book by choosing Save As from the pull-down menu of File.



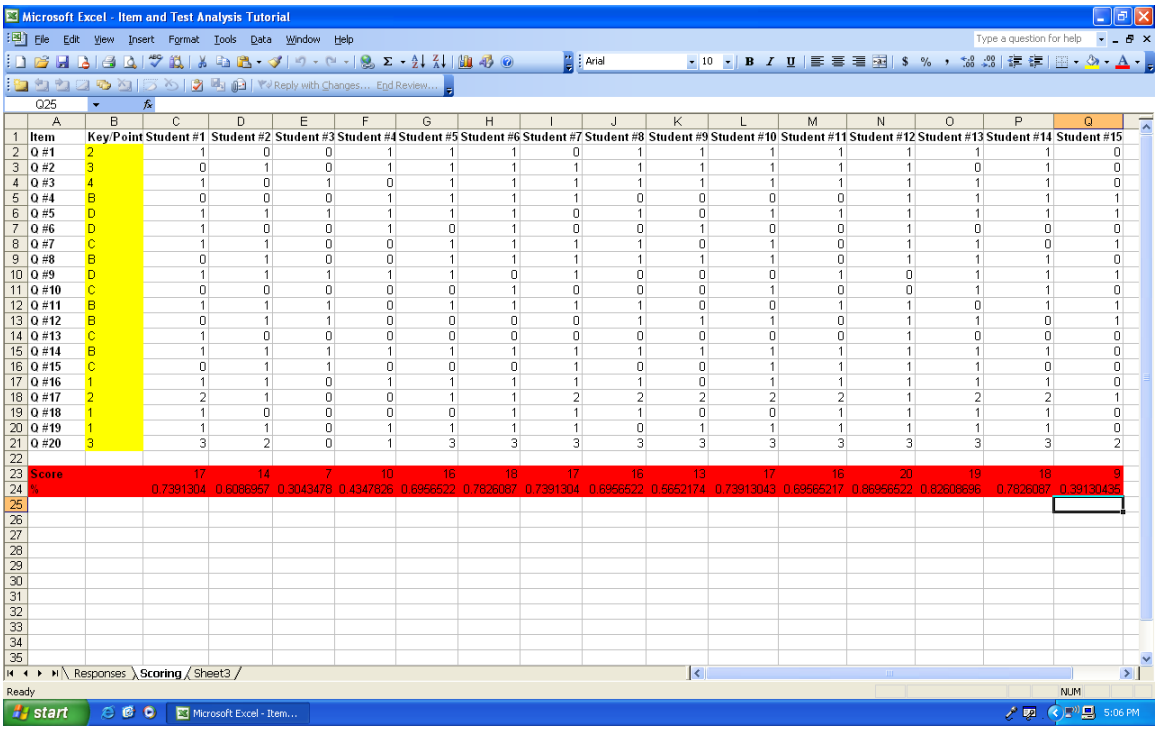
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Item	Key/Point	Student #1	Student #2	Student #3	Student #4	Student #5	Student #6	Student #7	Student #8	Student #9	Student #10	Student #11	Student #12	Student #13	Student #14	Student #15
2	Q #1	2	2	1	1	2	2	2	1	2	2	2	2	2	2	2	1
3	Q #2	3	4	3	2	3	3	3	3	3	3	3	3	3	2	3	2
4	Q #3	4	4	3	4	2	4	4	4	4	4	4	4	4	4	4	3
5	Q #4	B	C	C	C	B	B	B	B	C	C	A	C	B	B	B	B
6	Q #5	D	D	D	D	D	D	D	A	D	A	D	D	D	D	D	D
7	Q #6	D	D	A	A	D	A	D	A	A	D	A	A	D	A	B	A
8	Q #7	C	C	C	B	B	C	C	C	C	B	C	B	C	C	B	C
9	Q #8	B	A	B	A	A	B	B	B	B	B	B	A	B	B	B	C
10	Q #9	D	D	D	D	D	D	A	D	A	A	A	D	A	D	D	D
11	Q #10	C	A	A	A	A	A	C	A	A	B	C	A	A	C	C	A
12	Q #11	B	B	B	B	A	B	B	B	B	C	C	B	B	A	B	B
13	Q #12	B	A	B	B	A	A	A	A	B	B	B	C	B	B	C	B
14	Q #13	C	C	D	D	D	D	D	D	D	A	A	B	C	D	D	D
15	Q #14	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	D
16	Q #15	C	B	C	C	B	B	B	C	D	D	C	C	C	C	D	A
17	Q #16	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0
18	Q #17	2	2	1	0	0	1	1	2	2	2	2	2	1	2	2	1
19	Q #18	1	1	0	0	0	0	1	1	1	0	0	1	1	1	1	0
20	Q #19	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0
21	Q #20	3	3	2	0	1	3	3	3	3	3	3	3	3	3	3	2

Figure 4 Sample Student Response Worksheet

2. Scoring student responses: First, we will copy the Responses worksheet and paste to “Sheet2”. Click at cell A1 and keep the mouse pressed, drag the mouse to the last cell Q21 and then release the mouse, select Copy from the pull-down menu of Edit, click “Sheet2” at the bottom of the slide bar to open another worksheet, then select Paste from the pull-down menu of Edit. Rename “Sheet2” as “Scoring”. Start with cell C2 by click at it, type in the formula dialogue box “=IF(Responses!C2=Responses!$B2,1,0)” and then press Enter, cell C2 is now scored. Click at cell C2, place your cursor at the lower right-hand corner of cell C2 until your cursor changes to a cross, press down your mouse and while keeping it pressed, drag the mouse forward to cell Q2, release your mouse. Without clicking, place your cursor at the lower right hand corner of cell Q2 unit it changes to a cross, press down your mouse and while keeping it pressed, drag the mouse downward to cell Q16 (the response to the last multiple-choice question by the last student), release your mouse. Now all students’ responses to multiple-choice questions have been scored. The above mouse dragging and releasing process is called AutoFill, a very efficient procedure in MS Excel. No scoring is necessary for constructed response questions, because students’ earned points on those questions have already been entered. You have now completed scoring all students’ responses.

The final analysis of scoring is to calculate each student’s total score on the test. Create two new rows labeled “Score” and “%”. Click at cell C23, click at the summation sign ∑ on the tool bar, use your mouse to select cells C2 to C21 then press Enter. The total score for student #1 is now calculated. Use AutoFill to calculate the total scores for other students. You have now calculated all students’ scores. Next, click at cell C24, type in the formula dialogue box “=C23/23” and press Enter. The denominator is the maximum points of the test, in this example it is 23 (i.e. each multiple choice question is worth 1 point, plus the points for constructed response questions).

Once again, use AutoFill to complete calculation of percentage scores for other students. You now have calculated all students' scores in percentage.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Item	Key/Point	Student #1	Student #2	Student #3	Student #4	Student #5	Student #6	Student #7	Student #8	Student #9	Student #10	Student #11	Student #12	Student #13	Student #14	Student #15
2	Q #1	2	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0
3	Q #2	3	0	1	0	1	1	1	1	1	1	1	1	1	0	1	0
4	Q #3	4	1	0	1	0	1	1	1	1	1	1	1	1	1	1	0
5	Q #4	B	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
6	Q #5	D	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1
7	Q #6	D	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0
8	Q #7	C	1	1	0	0	1	1	1	1	0	1	0	1	1	0	1
9	Q #8	B	0	1	0	0	1	1	1	1	1	1	0	1	1	1	0
10	Q #9	D	1	1	1	1	1	0	1	0	0	0	1	0	1	1	1
11	Q #10	C	0	0	0	0	0	1	0	0	0	1	0	0	1	1	0
12	Q #11	B	1	1	1	0	1	1	1	1	0	0	1	1	0	1	1
13	Q #12	B	0	1	1	0	0	0	0	1	1	1	0	1	1	0	1
14	Q #13	C	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
15	Q #14	B	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
16	Q #15	C	0	1	1	0	0	0	1	0	0	1	1	1	1	0	0
17	Q #16	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0
18	Q #17	2	2	1	0	0	1	1	2	2	2	2	2	1	2	2	1
19	Q #18	1	1	0	0	0	0	1	1	1	0	0	1	1	1	1	0
20	Q #19	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0
21	Q #20	3	3	2	0	1	3	3	3	3	3	3	3	3	3	3	2
22																	
23	Score		17	14	7	10	16	18	17	16	13	17	16	20	19	18	9
24	%		0.7391304	0.6086957	0.3043478	0.4347826	0.6956522	0.7826087	0.7391304	0.6956522	0.5652174	0.73913043	0.69565217	0.86956522	0.82608696	0.7826087	0.39130435

Figure 5 Sample Scoring Worksheet

3. Conducting item analysis – item difficulty and discrimination:
We will now calculate item difficulty and discrimination for each of the items on the test. First use your mouse to select cells A1 to A21 and then choose Copy from the pull-down menu of Edit, click at "Sheet3" at the bottom of the sliding bar to open the worksheet, then choose Paste from the pull-down menu of Edit. Rename "Sheet3" as "Item Property". Now create three new columns by typing in "Points", "Difficulty" and "Discrimination" in cells B1, C1 and D1. Type in maximum points for each of the items in column "Points". Click at cell C2 to begin calculating item difficulty. Type "=SUM(Scoring!C2:Q2)/(15*B2)" in the formula dialog box and press Enter (15 is the total number of students), the difficulty for item 1 is calculated. Use AutoFill to calculate the item difficulties for other items. Now click at cell D2 to begin calculating item discrimination. Type in the formula dialogue box "=CORREL(Scoring!C2:Q2,Scoring!C$23:Q$23)" and press Enter. The item discrimination for item 1 is now calculated. Use AutoFill to calculate item discriminations for the rest of items.

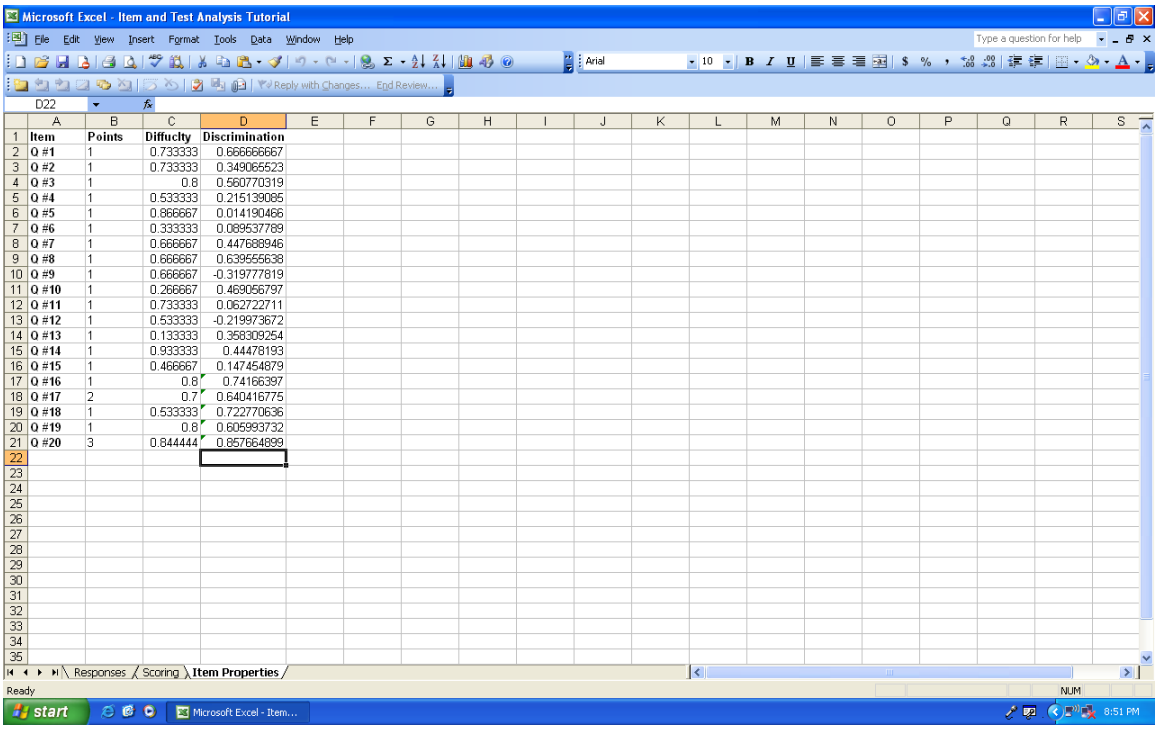
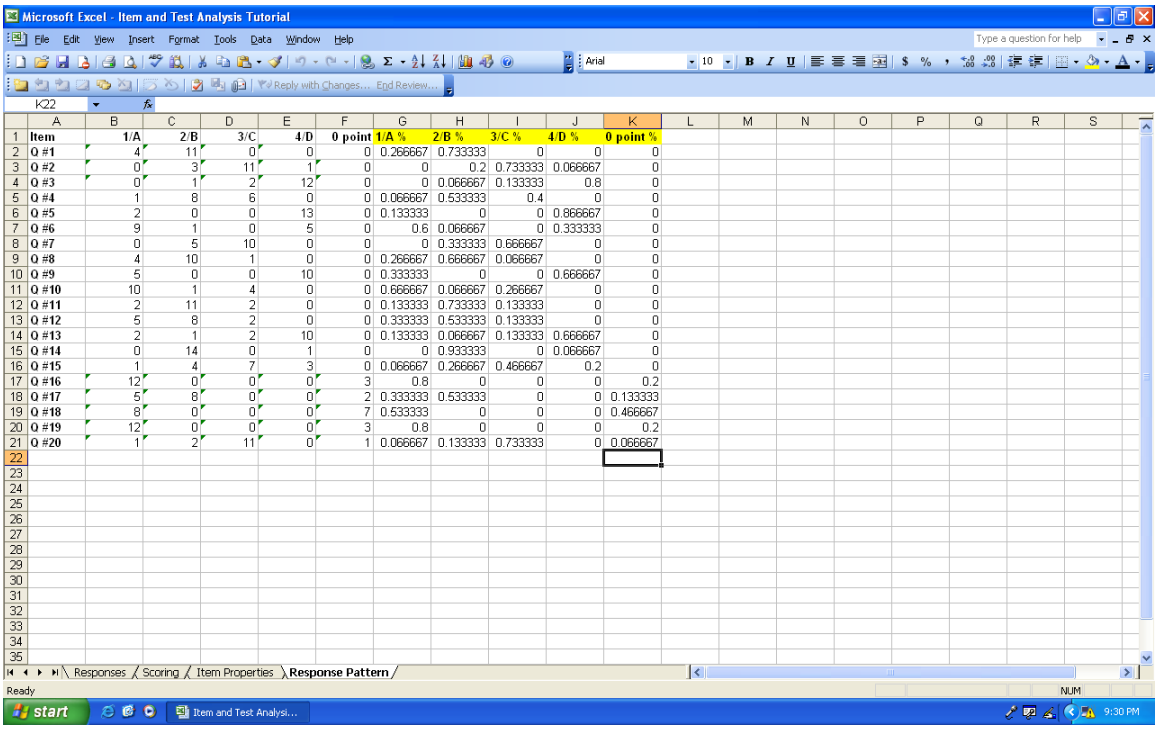


Figure 6 Sample Item Analysis Worksheet

4. Conducting item analysis – item response patterns:

While the Item Properties worksheet remains open, select and copy cells A1 to A21, choose Worksheet from the Insert pull-down menu, and then choose Paste from the Edit pull-down menu. Double-click at “Sheet3” and rename it as “Response Pattern”. Drag Response Pattern worksheet to after the worksheet of Item Properties. Create the following columns by typing in “1/A”, “2/B”, “3/C”, “4/D”, “1/A%”, “2/B%”, “3/C%”, “4/D%”. Click at cell B2, type in the formula dialogue box “=COUNTIF(Responses!C2:Q2,1)+COUNTIF(Responses!C2:Q2,"A")” and press Enter. The number of students who chose 1 or A is now counted. Use AutoFill to count the frequencies for the rest of items. Now do the same for responses “2/B”, “3/C”, and “4/D”. Please note that for the column of “0 point”, you only need to type in “=COUNTIF(Responses!C2:Q2,0)” in cell F2. To convert the frequencies into percentages, start with cell G2, type in the formula dialogue box “=C2/15” and press Enter. Use AutoFill to complete calculations for the rest of cells.

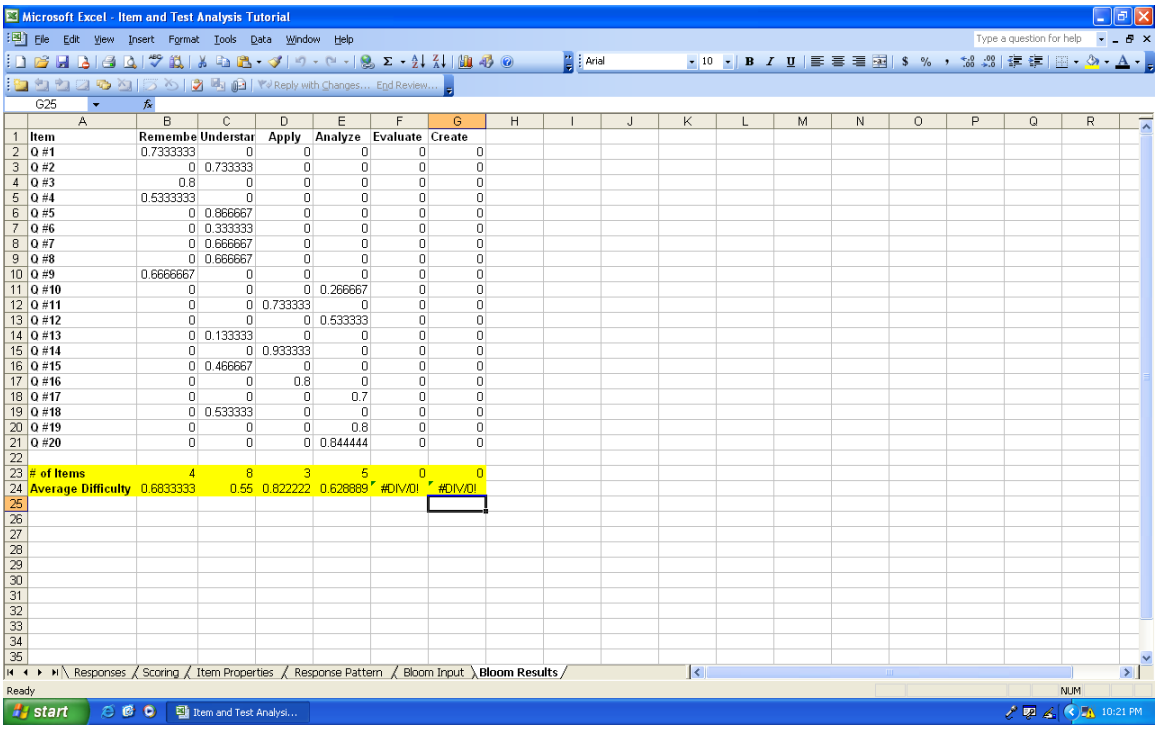


Item	1/A	2/B	3/C	4/D	0 point	1/A %	2/B %	3/C %	4/D %	0 point %
Q #1	4	11	0	0	0	0.266667	0.733333	0	0	0
Q #2	0	3	11	1	0	0	0.2	0.733333	0.066667	0
Q #3	0	1	2	12	0	0	0.066667	0.133333	0.8	0
Q #4	1	8	6	0	0	0.066667	0.533333	0.4	0	0
Q #5	2	0	0	13	0	0.133333	0	0	0.866667	0
Q #6	9	1	0	5	0	0.6	0.066667	0	0.333333	0
Q #7	0	5	10	0	0	0	0.333333	0.666667	0	0
Q #8	4	10	1	0	0	0.266667	0.666667	0.066667	0	0
Q #9	5	0	0	10	0	0.333333	0	0	0.666667	0
Q #10	10	1	4	0	0	0.666667	0.066667	0.266667	0	0
Q #11	2	11	2	0	0	0.133333	0.733333	0.133333	0	0
Q #12	5	8	2	0	0	0.333333	0.533333	0.133333	0	0
Q #13	2	1	2	10	0	0.133333	0.066667	0.133333	0.666667	0
Q #14	0	14	0	1	0	0	0.933333	0	0.066667	0
Q #15	1	4	7	3	0	0.066667	0.266667	0.466667	0.2	0
Q #16	12	0	0	0	3	0.8	0	0	0	0.2
Q #17	5	8	0	0	2	0.333333	0.533333	0	0	0.133333
Q #18	8	0	0	0	7	0.533333	0	0	0	0.466667
Q #19	12	0	0	0	3	0.8	0	0	0	0.2
Q #20	1	2	11	0	1	0.066667	0.133333	0.733333	0	0.066667

Figure 7 Sample Item Response Pattern Worksheet

5. Conducting test analysis – student performance by groups of items:
We will assume that we would like to calculate students' performance by Bloom's cognitive levels. First, we will create a new worksheet labeled "Bloom Input". The worksheet will have the following columns: Item, Remember, Understand, Apply, Analyze, Evaluate, and Create, and the following rows: Item #1, Item #2, … Item #20 (the item labels may be copied from a previous worksheet). Now go over each item in the test, classify it into one highest Bloom's cognitive level by typing "1" for the highest cognitive level and "0" for the rest of cognitive levels. Now select and copy the entire worksheet, insert a new worksheet and paste. Re-label the new worksheet as "Bloom Results". Click at cell B2, type in the formula dialogue box "='Bloom Input'!B2*'Item Properties'!$C2" and press Enter. Use AutoFill to complete calculations of difficulty for the rest of cognitive levels and items.

Now calculate the average difficulty for each of the Bloom's taxonomy levels. Create two new rows labeled "# of Items" and "Average Difficulty" at cells A23 and A24. Click at cell A23, type in the formula dialogue box "=COUNTIF(B2:B21,">0")" and press Enter. The number of items at Remember level is calculated. Use AutoFill to count the numbers of items at the other cognitive levels. Click at cell B24, type in the formula dialogue box "=SUM(B2:B21)/B23" and press Enter. The average difficulty for Remember is calculated. Use AutoFill to calculate the average difficulties for the other cognitive levels.



	A	B	C	D	E	F	G
1	Item	Remembe	Understar	Apply	Analyze	Evaluate	Create
2	Q #1	0.7333333	0	0	0	0	0
3	Q #2	0	0.733333	0	0	0	0
4	Q #3	0.8	0	0	0	0	0
5	Q #4	0.5333333	0	0	0	0	0
6	Q #5	0	0.866667	0	0	0	0
7	Q #6	0	0.333333	0	0	0	0
8	Q #7	0	0.666667	0	0	0	0
9	Q #8	0	0.666667	0	0	0	0
10	Q #9	0.6666667	0	0	0	0	0
11	Q #10	0	0	0	0.266667	0	0
12	Q #11	0	0	0.733333	0	0	0
13	Q #12	0	0	0	0.533333	0	0
14	Q #13	0	0.133333	0	0	0	0
15	Q #14	0	0	0.933333	0	0	0
16	Q #15	0	0.466667	0	0	0	0
17	Q #16	0	0	0.8	0	0	0
18	Q #17	0	0	0	0.7	0	0
19	Q #18	0	0.533333	0	0	0	0
20	Q #19	0	0	0	0.8	0	0
21	Q #20	0	0	0	0.844444	0	0
22							
23	# of Items	4	8	3	5	0	0
24	Average Difficulty	0.6833333	0.55	0.822222	0.628889	#DIV/0!	#DIV/0!

Figure 8 Sample Test Analysis Worksheet

6. Conducting test analysis – test validity:

Calculating correlation coefficient is typical in validation. The analysis requires two sets of scores: one is from the test under validation, and another is from a different test that is considered to be credible thus the criterion. Fist, create a new worksheet and name it "Validity". Create four rows labeled as "Student", "Test Score", "Criterion Score", and "Pearson Correlation Coefficient". Type in student IDs (or just serial #), their test scores, and criterion scores. Click at cell B5, type in the formula dialogue box "=CORREL(B2:P2,B3:P3)" and press Enter. The correlation coefficient is now calculated. Correlation coefficients range from -1 to +1, with negative values indicating negative relationships and positive values indicating positive relationships. The bigger the absolute value is, the stronger the relationship. Because correlation may be a result of a random effect, the obtained correlation coefficient needs to be compared to a critical value to decide if the correlation coefficient is likely caused by chance. Therefore, a statistical table of critical values of correlation coefficients needs to be consulted. The following Table are commonly used critical values at 95% confidence level.

pairs of scores	critical value
3	0.9969
4	0.95
5	0.8783
6	0.8114
7	0.7545
8	0.7067
9	0.6664
10	0.6319

11	0.6012
12	0.576
13	0.5529
14	0.5324
15	0.5139
16	0.4973
17	0.4821
18	0.4683
19	0.4555
20	0.4438
21	0.4329
22	0.4227
27	0.3809
32	0.3494
37	0.3246
42	0.3044
47	0.2875
52	0.2732
62	0.25
72	0.2319
82	0.2172
92	0.205
102	0.1946

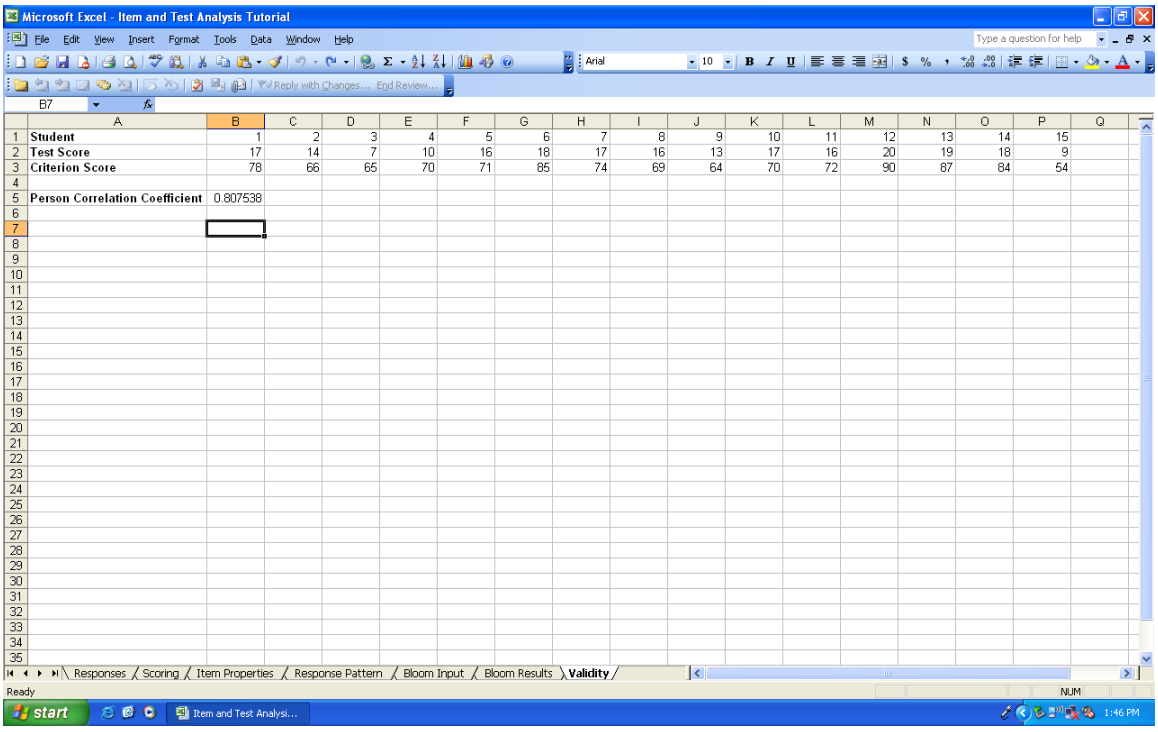


Figure 9 Sample Validity Analysis Worksheet

7. Conducting test analysis – reliability:

Open the Scoring worksheet created before by clicking at the label at the bottom of the sliding bar. Select and copy the worksheet, then insert a new worksheet by choosing Worksheet from the pull down menu Insert , then choose Paste from the pull-down menu Edit. Re-label the new worksheet as "Reliability" and drag it to the end of the worksheet list on the sliding bar at the bottom. Now create a new column by typing in "Var" at cell S1 for variances. Click at cell S2, type in the formula dialogue box "=VAR(C2:Q2)" and then press Enter. The variance for item 1 is now calculated. Use AutoFill to calculate the variances of other items and of total test scores. Ignore the "divided by zero" error message in cell S22 (because the cell is blank). Now type in any blank cell "Cronbach Alpha=", click at a cell on its right, type in the formula dialogue "=(20/19)*(1-SUM(S2:S21)/S23)" and press Enter. The Cronbach's alpha is now calculated.

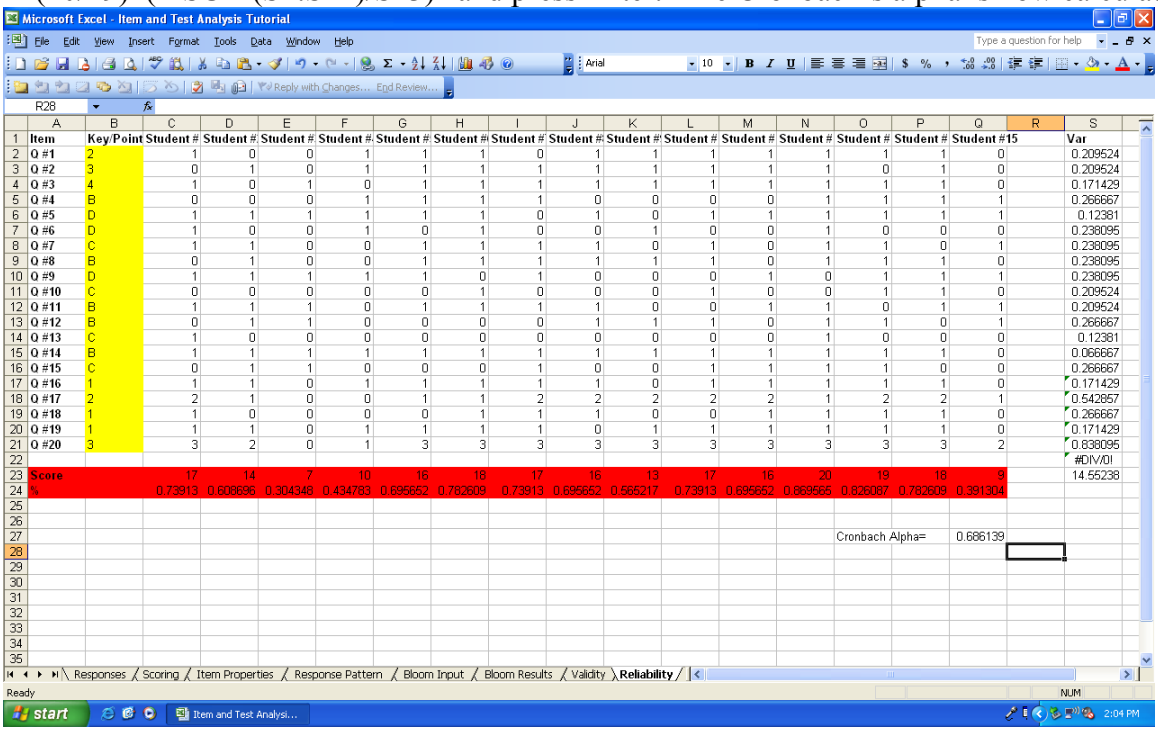


Figure 10 Sample Reliability Analysis Worksheet

8. Conducting test analysis – comparing student group performances:
The last test analysis, which may also be a part of validation for consequence validity, is to compare student group performances. In order to compare students' group performances and test statistically if the difference between two groups is significant, i.e. not likely due to chance, we need to enter students' test scores by groups. The grouping may be done by any criteria, such as gender, race, course session, etc. First create a new worksheet and label it as "Bias". Next, create two new rows: Group 1 Students' Scores, Group 2 Students' Scores. Enter students' scores in the two rows.

Now we will calculate the descriptive statistics for the two groups of scores. Create the following rows: "Group 1 Mean", "Group 2 Mean", "Group 1 SD", "Group 2 SD", "Group 1 Maximum", "Group 2 Maximum", "Group 1 Minimum", "Group 2 Minimum", "Group 1 Range", and "Group 2 Range". Click at cell B7, type in the formula dialogue box "=AVERAGE(B4:J4)" and press Enter. The average for group 1 students is now calculated. Use AutoFill to compute the average for group 2 students. Follow the same procedure to calculate the standard deviation

(formula: STDEVA), maximum (formula: MAX), and minimum (formula: MIN). To calculate the range, type in the formula "=B13-B16".

Finally, to test if the difference between the means of the two groups is statistically significant, we will calculate t-test statistics. A t-test gives the probability at which the difference may happen by chance. Normally, a 0.05 criterion is used as cut-off. If the calculated t-test probability is equal or greater than 0.05, then we claim that the difference may happen by chance, meaning that the difference is not statistically significant; if smaller than 0.05, we claim that the difference is not likely to happen by chance, meaning that the difference is statistically significant. Create a new row labeled as "t-test=", click at the cell right after, type in the formula dialogue box "=TTEST(B4:J4,B5:G5,2,2)" and press Enter. The probability is now calculated. This calculated probability is greater than 0.05, therefore we can claim that, although there is a difference between the two group means, the difference is not statistically significant, i.e. the difference may be caused by chance.

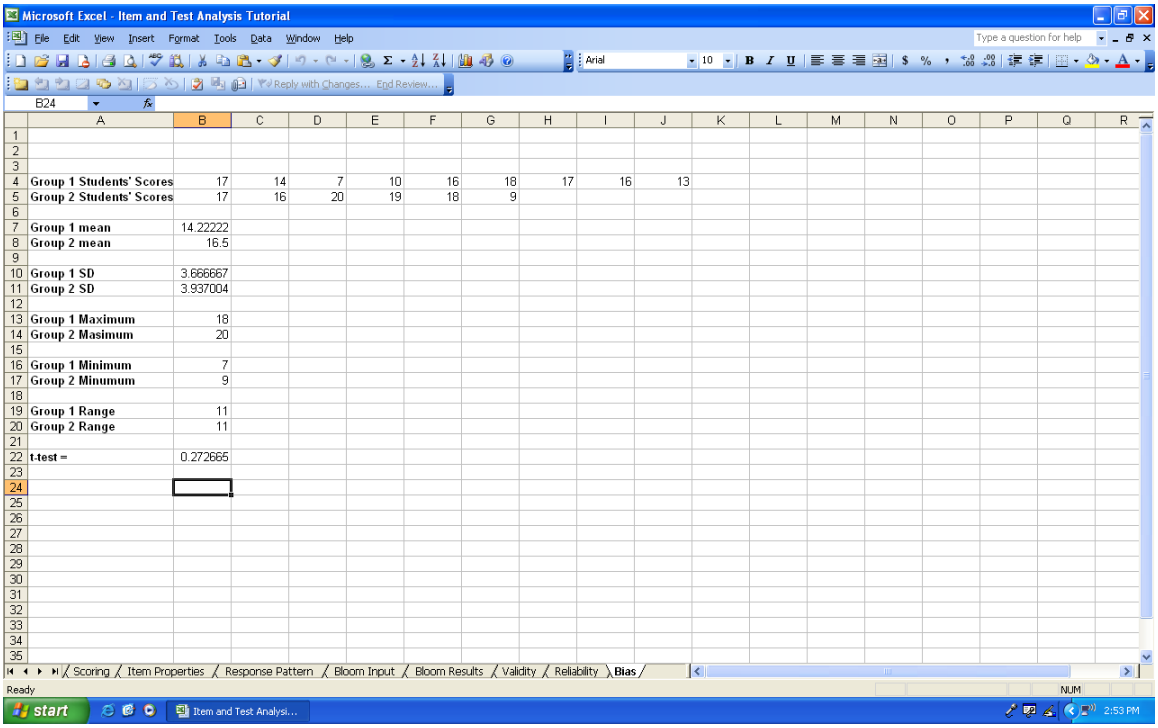


Figure 11 Sample Bias Analysis Worksheet