3710 Mid-Term, Spring 2011
Reflection Time

It is now time to determine if I am satisfied with how the test functioned in the context of my class goals.

What was the goal of the exam?

My goal was to use the test *to distinguish between those who know the material best and those who do not know it as well*. The "material" was the knowledge, understandings and applications related to Chapters 1-8.

Thus, the test needs to be difficult (but not unreasonably difficult).

It also needs to fair.

The steps I use for test reflection:

- First, I determine internal reliability.

- Second, I reflect on validity (Table of Specifications, face validity)

- Third, I perform item analysis.

- Fourth, I determine final test scores for "grade book".

Item Analysis: Let's Begin -

32 who people took the test are used for analysis.
Below are the number right by item number:

01.____25_____	02.____25_____
03.____25_____	04.____18_____
05.____30_____	06.____27_____
07.____24_____	08.____29_____
09.____25_____	10.____24_____
11.____26_____	12.____18_____
13.____23_____	14.____25_____
15.____21_____	16.____23_____
17.____9_____	18.____19_____
19.____19_____	20.____18_____
21.____26_____	22.____16_____
23.____27_____	24.____9_____
25.____27_____	26.____29_____
27.____8_____	28.____7_____
29.____13_____	30.____19_____
31.____16_____	32.____16_____
33.____13_____	34.____24_____
35.____32_____	36.____25_____
37.____27_____	38.____17_____
39.____26_____	40.____24_____
41.____20_____	42.____31_____
43.____17_____	44.____9_____
45.____28_____	46.____28_____
47.____21_____	48.____16_____
49.____25_____	50.____17_____

Item Difficulty Index: I will set acceptable item difficulty (aka "p-value") at around 75%; that is, if 75 percent of the class gets the answer correct, then the item is not too difficult. For this test, this standard translates into 24 or more people getting the item right. Of the 50 test items, using this criterion, 25 were not too difficult and I eliminate them from further analysis. These 25 are fair. However, now I will look at the most difficult items (the remaining 25) to analyze further.

Below are the 25 items that might be too difficult.

13.___23___	43.___17___	28.___7___
15.___21___	47.___21___	30.___19___
17.___9___	04.___18___	32.___16___
19.___19___	12.___18___	38.___17___
27.___8___	16.___23___	44.___9___
29.___13___	18.___19___	48.___16___
31.___16___	20.___18___	50.___17___
33.___13___	22.___16___	
41.___20___	24.___9___	

I will analyze them further using a process called item discrimination or the item discrimination index.

Item Discrimination Index: The theory: a difficult item is good . . . as long as it differentiates between those who know and those who do not. So, let's examine the 25 items to see if they "discriminate" properly.

One does this by comparing the best tests to the worst tests. I decide to do this by comparing the top quarter to the bottom quarter. I will take the top 25% (8 tests), compare them to the bottom 25% (8 tests), and compute the discrimation index (aka D). The concept is to compare the percentage correct in the top group to percentage correct in bottom group. I decide to set my minimum D value at +.20. See next table.

TABLE

Item Number followed by total correct answers	# Right in High Grp	# Right in Low Grp	% in High	% in Low	"D Value" High%-Low%
13.____23_____	7	3	.87	.37	**+.50**
15.____21_____	7	4	.87	.50	**+.37**
17._____9_____	4	2	.50	.25	**+.25**
19.____19_____	7	3	.87	.37	**+.50**
27.____8_____	7	2	.87	.25	**+.62**
29.____13_____	6	0	.75	.00	**+.75**
31.____16_____	8	4	1.00	.50	**+.50**
33.____13_____	5	2	.62	.25	**+.37**
41.____20_____	7	3	.87	.37	**+.50**
43.____17_____	6	6	.75	.75	**.00**
47.____21_____	8	4	1.00	.50	**+.50**
04.____18_____	5	4	.62	.50	**+.12**
12.____18_____	6	3	.75	.37	**+.38**
16.____23_____	8	4	1.00	.50	**+.50**
18.____19_____	6	4	.75	.50	**+.25**
20.____18_____	6	3	.75	.37	**+.38**
22.____16_____	6	4	.75	.50	**.+25**
24._____9_____	3	3	.37	.37	**.00**
28._____7_____	3	1	.37	.12	**+.25**
30.____19_____	8	3	1.00	.37	**+.63**
32.____16_____	5	5	.62	.62	**.00**
38.____17_____	4	3	.50	.37	**+.13**
44._____9_____	4	0	.50	.00	**+.50**
48.____16_____	5	1	.62	.12	**+.50**
50.____17_____	6	2	.75	.25	**+.50**

My analysis shows that I have five (5) items that are not functioning as I intended. They are poor test items. They could have been poorly written, asked for information students were unable to attain (from class discussion, reading, etc.), or the result of misteaching on my part.

I can use this analysis to improve my future teaching and my future item writing.

But, what should I do about these findings vis-à-vis the students who took this exam? Clearly they are unfair to the current class.

I decided to award <u>each student</u> 2.5 points (.50 point added back in for each poor item) to their otherwise final point total since I "own" some amount of overall "error" in students' scores.

I decide to do this for <u>all students</u>, whether the student missed these five items or not.