

## How well did your midterm or exam "work" ??

See also item #28 on <http://www.eos.ubc.ca/research/cwsei/resources.html>

### Top two reasons for taking a little time to analyze test results

1. Effective teaching involves measuring student thinking, then making adjustments to your teaching based on the data. Quizzes, tests, exams and assignments are the instruments; work students produce & exam results are the data.
2. Students must be able to make adjustments too. They need test results to help steer their efforts in appropriate directions. Teachers need good data in order to provide that all-important feedback.

**What do you NORMALLY do with marks?** Average them? Find the median? Standard deviation? Frequency distribution? Consider adjusting grades up? Down? Adjust to a "curve"? Do you scratch your head over why question X was answered poorly?

**Other ways results COULD be use:** Determine "reliability" of the test; Discriminate between top & bottom students; Between good vs bad questions or hard vs easy questions; Discover misconceptions or decide about dropping a question from scores. Here are some ideas about how.

**Using ScanTron results within EOS** (ScanTron is the machine that generates automatically graded multiple choice test results):

- Counts for each option show how often options were chosen for all questions. **For example**, options that are chosen by no one (circled red) are not contributing to the assessment.
- \* and \*\* point to poorly answered questions ... are these hard? "Bad" questions? Check them! **Check for:** Unclear options, misleading or vague questions, key misconceptions.

\* = q'ns missed by 50% or more students.  
\*\* = q'ns missed by 75% or more students.

Check for Answer Key Errors

Q	answer	#A	#B	#C	#D	#E	#total	#WRONG%
1)	B	14	86	13	13	7	135	47 34
...								
3)	C	22	5	60	9	40	136	76 55 *
4)	D	51	16	3	56	8	136	80 58 *
5)	B	13	103	1	16	3	136	33 24
6)	A	108	10	13	0	5	136	28 20
7)	D	6	5	35	63	27	136	73 53 *
...								
11)	D	27	32	4	32	41	136	104 76 **
12)	B	17	103	0	5	11	136	33 24
13)	D	47	23	6	42	16	136	94 69 *

### Online assessment options

- Using Vista's "Assessment Manager", the "View Reports" option provides statistics by question or student.
- Stats for each question: class avg; avg of top (and bottom) 25%, and a discriminator (see below<sup>1</sup>).
- Stats for each student: every choice of every student as a downloadable table (spreadsheet) or graphically.

### Classifying questions to find which were hard, easy, appropriate or inappropriate

This figure shows portions of a spreadsheet used to accumulate and sort answers, generate scores, and compare results question by question.

- **Red** avg. & **green** discrim<sup>1</sup>. suggests a difficult question.
- **Green** for both averages suggest "easier" questions (**redish** discriminator<sup>1</sup> means little).
- **Red** avg. & **red** discrim. suggests a "poor questions".
- **Green** avg. & **green** discrim. is impossible since 'discriminator' implies best students = worst students.
- **Negative** discriminator values mean more "poor" students got it right than "good" students.

Note each test or quiz should be considered individually.

<sup>1</sup>Discriminator =  $(avgT - avgB) / (avgT + avgB)$  where  $avgT$  &  $avgB$  are averages of Top and Bottom performing students respectively.

**Scores for each student on each question.**

student58	0	0	1.5	0	0	0	0	0.5	1	1	34
student59	1	2	4.5	0	1	0	0	0	1	0	32.5
student60	0	0	2	0	1	0	0	0	0	0	31.5
student61	1	2	2	0	0	0	0	0	0	0	31
student62	1	0	3	0	1	0	0.5	0	1	0	30.5
student63	1	2	2.5	1	0.5	0.5	0	1	1	1	30
student64	1	0	1.5	0	0	0	0	0	1	0	29
student65	0	0	2.5	0	0	0.5	0	1	1	0	21

**Analysis by question**

max	1	2	3	2	1	2	1	1	1	1
Question #s	B15	B16	B17	B18	B19	B20	B21	B22	B23	B24
Averages:										
top 50%	93	85	77	49	78	33	45	63	93	62
bot 50%	83	66	61	26	59	7	22	53	86	41
All	88	75	69	38	69	20	34	58	90	52
ScanTron						**	*			
discrimi	6	13	12	31	14	66	34	8	4	20
$d=(t-b)/(t+b)$	B15	B16	B17	B18	B19	B20	B21	B22	B23	B24

easy moderate hard (bad?) hard easy

### **What's the cost? How much time does it take to do this?**

ScanTron results and Vista reports are provided with no time invested (after learning how).

Spreadsheet analysis can take 15-30 minutes using a template. Also, after analyzing tests & questions a couple of times, comprehensive analysis of every test may become less important.

**To learn more** about using ScanTron, Vista, or Spreadsheets, ask an STLF to demonstrate, set it up, provide a template, or explain reasons, applications and techniques.

### **Returning students' work, and your analysis – Pros and Cons**


**Pros:** Closes the learning feedback loop, allowing students to learn from mistakes and adjust study habits & priorities. Also, students can learn to gauge their own performance better when they see question score analysis and/or feedback about specific efforts.

**Cons:** Time to generate feedback & analysis, although with practice this becomes routine. Also, some students need incentives to review their own work.

### **Options for *easier* generation of *effective* feedback**

- Show at least some of your analysis results, whether it's a simple frequency distribution or more detailed analysis. Then, in class, discuss only the one or two questions that caused significant difficulty.
- Clarify your actions (e.g. adjustments to scores) with carefully explained reasons.
- Incorporate "difficult" questions back into class – e.g. using clickers. For example, a difficult question might be deconstructed into a sequence of decisions to help students (and you) identify where they went wrong.
- Instead of commenting on individual tests or exercises, write feedback for the *whole class* while you do the marking. Then deliver *all* feedback to every student by email or PDF on the course Vista or web site.
- Show students the marking rubric *before* the test or exercise. Consider annotating it with "why" for the various rubric components.
- Add an assignment that forces students to review their work and corresponding feedback. For example, have them analyze one test question in terms of errors or improvements (maybe for bonus marks).

### **Precedent and other tools**

- Opensource JAVA software for test analytics: Basic descriptive stats, graphs, classical item analysis, factor analysis, and item response theory (ITR). Features at <http://www.itemanalysis.com/jmetrik-features.php>
- Microsoft Excel plug-in for test item analysis. Feature at <http://deltasigmasoft.com/index.html>  
\$0.00: # test items <= 25; # students <=500.  
\$30.0: (or \$150 for Dep't) no restrictions.
-  "item analysis" or "item analysis software".
- [http://en.wikipedia.org/wiki/Psychometric\\_software](http://en.wikipedia.org/wiki/Psychometric_software)

### **"Advanced" analysis**

Design and analysis of tests is a whole sub-discipline in educational psychology, called psychometrics. On this page we only touch on simple, small scale ideas for assessing effectiveness of your own individual questions or whole tests. There are many instruments, statistical methods, and theories about measuring performance. For example, large-scale, high-stakes assessments such as SATs and medical exams are often designed and analyzed using Item Response Theory (IRT). This basically involves modeling the response of individuals with a given ability to each item on a test. However, this degree of sophistication becomes useful primarily when many test items are used for each specific concept or skill, and when testing conditions (English abilities, time available, etc.) are uniform for all students. For more details, contact an STLF (below) or see for example [Wikipedia](#) or <http://luna.cas.usf.edu/~mbrannic/files/pmet/irt.htm>.

---

**Contact EOS-SEI:** To talk about your course(s) or teaching and learning in general, visit EOS-South 361, or contact Francis Jones ([fjones@eos.ubc.ca](mailto:fjones@eos.ubc.ca)) or Brett Gilley ([bgilley@eos.ubc.ca](mailto:bgilley@eos.ubc.ca)). See also <http://www.eos.ubc.ca/research/cwsei/>.