

## Test Item Analysis using the GradeMaster Scanners

The goal of test item analysis is two-fold: to identify poorly written questions and to identify mistakes in the answer key. Both problems cause teachers to assign grades that do not accurately reflect what a student knows. You can analyze tests quickly and easily with the data produced by the GradeMaster scanners. As with all statistics, these methods rely on large samples. The smaller the class, the less helpful the results will be.

### ***Difficulty***

---

The GradeMaster scanners tell you what percentage of students answered each question correctly. Compare the measured difficulty to the expected difficulty. A question that is much easier or harder than expected indicates a problem, perhaps a badly worded question or a mistake on the answer key (or a topic you forgot to cover).

Best practice is to keep difficulties in the 20-80% range. Easy questions reassure the teacher that “everyone got it” but they provide little help in sorting out who deserves an A or a C. Similarly, questions that everyone misses tell the teacher nothing about which student knows what.

### ***Distractors***

---

- Distractors are the wrong answers on a multiple choice question. Writing them is one of the more difficult parts of constructing a quality test.
- Multiple choice questions are susceptible to guessing. If you provide four choices, a student making a random guess will answer 25% of the questions correctly, on average, in spite of the fact that he may know 0% percent of the material. His score is artificially high.
- This assumes that all four answers are equally plausible. If only two answers are plausible, students will guess 50% of them right, greatly inflating their score.
- With the GradeMaster results, you can identify at a glance distractors that no one guesses. You should consider rewriting them.

### ***Discrimination***

---

Consider the following scenario: You have a true-false question. By design it is of moderate difficulty; 50% of the class usually gets it wrong. One day you accidentally mark your answer key with the wrong answer. Half of the class misses the question, as expected, so the difficulty gives you no clue that something is wrong.

Your only hope of catching this mistake is to notice that the good students tended to miss the question and the weak students tended to get it right. *Any* question with that result is suspect. Discrimination attempts to measure that correlation.

The GradeMaster scanners calculate two discrimination scores. Check the literature for full details about either.

1. Point bi-serial. This classic score correlates scores on a particular question with the students’ scores on the overall test. The score ranges from -1 to +1, with a -1 indicating a bad correlation and a +1 indicating a good correlation. Questions that score close to or below zero should be checked for problems.
2. Discrimination index. This index uses just the highest and lowest scoring students (often the top and bottom 27%) to calculate discrimination. Scores close to or below zero indicate the question is suspect.
3. Note that extremely easy or hard questions inherently do not provide good discrimination; nearly everyone gets them right or wrong. The discrimination scores for these questions are not meaningful.