

EVALUATING TESTS

L. P. Iarmak,
O. R. Gladchenko

To produce a good test is very important. Why is this so important? For one thing, good evaluation of our tests can help us measure student skills more accurately. It also shows that we are concerned about those we teach. For example, test analysis can help us remove weak items even before we record the results of the test. This way we don't penalize students, because of bad test questions. Students appreciate an extra effort like this, which shows that we are concerned about the quality of our exams. And a better feeling toward our tests can improve class attitude, motivation and even student performance.

Some insight comes almost intuitively. We feel good about a test if advanced students seem to score high and slower students tend to score low. Sometimes students provide helpful "feedback", mentioning bad questions, as well as questions on material not previously covered in class, and unfamiliar types of test questions.

Besides being on the right level and covering material that has been discussed in class, good tests are also **valid and reliable**. A valid test is one that in fact measures what it claims to be measuring. A listening test with written multiple-choice options may lack validity if the printed choices are so difficult to read that the exam actually measures reading comprehension as much as it does listening comprehension. It is least valid for students, who are much better at listening than at reading. Similarly, a reading test will lack validity, if success at the exam depends on information not provided in the passage.

A reliable test is one that produces essentially the same results consistently on different occasions when the conditions of the test remain the same. Teacher's grading often lacks consistency or "reliability" since so many matters are being evaluated simultaneously. In defining reliability one must refer to consistent results, **when the conditions of the test remain the same**.

For example, for consistent results, we would expect the same amount of time to be allowed on each test administration. When a listening test is being administered, we need to make sure that the room is equally free of disturbing noises on each occasion. If a guided oral interview were being administered on two occasions, reliability would probably be hampered if the teacher on the first occasion were warm and supportive and the teacher on the second occasion were abrupt and unfriendly.

In addition to validity and reliability, we should also be concerned about **the affect of** our test, particularly the extent to which our test causes undue anxiety. Negative affect can be caused by some far too difficult questions or by an unfamiliar examination task which has not been used in class. Where possible, one should utilize test forms that minimize the tension and stress.

Besides being concerned about these general matters of validity, reliability and affect, there are ways that we can improve our tests by taking time to evaluate individual items. While many teachers are too busy to evaluate each item in every test that we give, at least major class tests should be carefully evaluated.

Selection of appropriate language items is not enough by itself to ensure a good test. Each question needs to function properly; otherwise, it can weaken the exam. Fortunately, there are some rather simple statistical ways of checking individual items. This procedure is called « item analysis». It is most often used with multiple-choice questions. An item analysis tells us three things: how difficult each item is, whether or not the question tells the difference between high and low students, and which distractors are working as they should. An analysis like this is used with any important exam- for example review tests and tests given at the end of the semester.

Соціально-гуманітарні аспекти розвитку сучасного суспільства : матеріали Всеукраїнської наукової конференції викладачів, аспірантів, співробітників та студентів факультету іноземної філології та соціальних комунікацій, м. Суми, 19-20 квітня 2013 р. / Відп. за вип. В.В. Опанасюк. — Суми : СумДУ, 2013. — Ч.2. — С. 115-116.