

## ASSESSMENT OF STUDENT PERFORMANCE IN STATISTICS

Carl J. Huberty, University of Georgia, USA

*It has been claimed that importance of assessment of student learning in universities has declined considerably over the past couple decades (based on resulting lack of variability in grading). The purpose of this paper is to illustrate an approach to assessment of student learning in an introductory statistical methods course. A three-type-assessment approach has, over a number of years, resulted in a fair bit of variability in reflected student learning. The approach and some results are presented following a discussion of some issues related to student assessment and grading.*

### INTRODUCTION

There have been many manuscripts written about the art, the skill, the science, the craft, etc., of teaching in general (e.g., Brookfield, 1990; Gullette, 1984; McKeachie, 1994) and of teaching statistical methods in particular (e.g., Cobb, 1993; Garfield, 1995; Hawkins, Jolliffe, and Glickman, 1992; Moore, 1995; Simpson, 1995). But, assessment of student performance is very seldom considered an important dimension of instruction in and of itself by many writers.

It is hypothesized that there are at least six reasons why student assessment is not taken too seriously by some instructors. First, it is conjectured that some instructors do not view student assessment as a learning process. Rather, they see it strictly as an evaluation process. Related to this, a second reason is that some instructors do not view student assessment as a means of providing feedback regarding instructional effectiveness. A third reason is that assessment is not easily carried out; it takes considerable effort on the part of the instructor; an effort in which some instructors prefer not to invest. A fourth potential reason is that some instructors may fear student anxiety which some students may exhibit if they know that *not* getting an “A” is a fairly strong possibility. Fifth, some instructors may feel that good student assessment is not rewarded (in terms of recognition or salary adjustments) to any great extent. A sixth reason why student assessment is not taken too seriously is that some instructors fear that unless a large majority of students are “awarded” an “A,” then student evaluation of the instructor will not be overwhelmingly positive.

In the second section of this paper, an example of a student assessment approach will be presented. This approach has been utilized by the author for over 20 years. The approach varies a little across course levels, but the general idea remains fairly constant.

## AN ASSESSMENT APPROACH

The learning environment context considered herein is the teaching of applied statistical methods to graduate level students in the behavioral sciences. A three-course sequence is being considered herein. This is a sequence in statistical methods from (1) introductory methods through (2) comparison of multiple means and proportions to (3) multiple regression analysis.

### *Class Activities*

Some of the class activities utilized include the usual lecturing and question-and-discussion sessions. Students are, on occasion, requested to formulate 3-4 questions about concepts covered in class discussions or in the textbook. These questions, prioritized by each student, are either discussed with the entire class or discussed in “learning groups.” Each learning group is comprised of 3-4 students -- I determine the initial group structure (by “mixing” student areas of study), and then “rotate” group membership for subsequent sessions. Class time is also spent reviewing “practice items” as well as administering quizzes and the test.

### *Student Assessments*

Three types of in-class assessments are utilized: four quizzes, a midterm test, and a final examination. All three of these types of assessments are of the short-answer variety. Some items (mostly in the quizzes) are multiple-choice items. [I would maintain that most of the multiple-choice items call for more than recall of facts and routine computations.] Other items in the quizzes, the test, and the examination are other types of short-answer items -- these items call for explanations, identifications (sometimes from computer printouts), and definitions.

All quizzes are 10-point assessments, and typically take the students about 20-25 minutes to complete. Class time is spent discussing each quiz item immediately after all students turn in their responses. Quiz items vary somewhat from section to section of a given course; changes are made because of content covered and item revision.

The midterm test length varies somewhat from course to course with the total number of points ranging from about 40 to about 70. In the most advanced of the three courses, one part of the midterm test is of the take-home variety. Items on the test for a

given course are not generally the same from section to section; some items are revised or replaced. In each of the three courses the total number of points on the test remains constant from section to section. About 110 minutes are allowed for test completion. Test results are reviewed during the next class meeting.

The final examination is administered after the last scheduled class session. As with the test, the total number of points for the examination varies somewhat across the three courses (from about 45 to about 75), but remains constant across sections of a given course. There is virtually no time limit for students to take the examination.

### *Other Student Activities*

In addition to the four quizzes, midterm test, and final examination, there are two other activities that are related to the assessment process. One is article critiques. Students are expected to locate published journal articles in which data analysis techniques covered in the given course are implemented. It is suggested that students locate articles in their area of study. If they have difficulty locating relevant articles, I have a number of references from previous classes from which current students may select an article. The critiques are fairly structured. Across the three courses, students locate two or three articles; different statistical techniques are used in the sets of critiques.

The second student activity is a research project. There are two choices regarding the type of project: (a) a substantive study using data analysis methods discussed in class meetings; or (b) a methodological study related to, but may be different from, methods studied during the term. [Very, very few students have opted for the second type.] Structure for the study, including reporting methods for the first type, is provided. Research project reports are turned in near the very end of the term.

### *Scoring and Grading*

#### Scoring

Student performances that yield numerical assessments are the only ones considered regarding student evaluation. Aspects such as class attendance, class participation, and homework are *not* considered in the evaluation process.

Of the four quizzes, each student's lowest score is dropped. For each student the three highest quiz scores are summed to yield a Q score; the maximum attainable Q score is 30. The Q scores for a current class are included in a distribution of scores with the three most recent classes for a given course. The arithmetic mean and standard deviation for the complete four-class distribution are used to transform a Q score for the current class to a unit standard score; this is labeled  $z_Q$ .

The midterm test score, T, for the current students is included with the T scores for the three most previous classes for a given course. The mean and standard deviation for the four classes are used to yield a  $z_T$  score for each student in the current class. Similarly, a  $z_E$  score is calculated for each student in the current class. Thus, three unit standard scores are determined for each student:  $z_Q$ ,  $z_T$ , and  $z_E$ .

Students are given rather complete descriptive numerical information for the current class after each quiz and the midterm test -- the three most recent classes are included for the test.

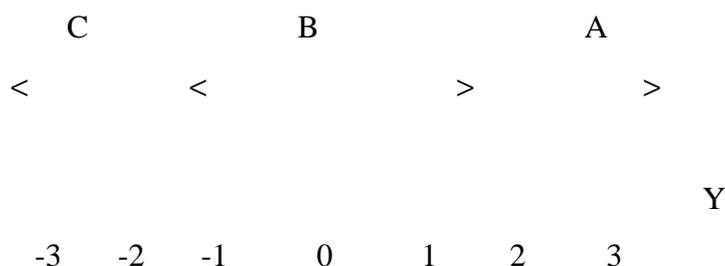
The three standard scores are combined to yield a composite score,

$$Y = 0.5z_Q + 1.0z_T + 1.5z_E .$$

For a number of years, students were asked to suggest the weights used to get a value of Y. Virtually all classes arrived at, and agreed with, the above weights.

It is after the first two terms in Y are calculated when students are informed as to how they "stand." Each student is given his/her score on the partial composite,  $0.5z_Q + 1.0z_T$ , along with a "guess" for their final grade -- a guess may be "A-B" or "B-C" or a single letter. Of course, the final grade depends upon the third term in Y,  $1.5z_E$ .

Each of the journal article critiques is "scored" with a maximum score of 10. Similarly, a maximum score of 20 is obtainable for the research project.



Grading

Earned grades for each course are basically determined by the Y score. The schematic above has been used.

Even though the critique scores and the project score are not explicitly considered in the above, they may be, and have been, taken into consideration for the student whose Y score falls in one of the two gaps. Of course, some judgment calls need to be made. There are three reasons why critique scores and the project score are not included in a composite: (a) very limited variability results; (b) the work involved is not that conducive to clear awarding of points; and (c) when two students collaborate, each is given the same number of points.

The above outlined scoring procedure has yielded some somewhat varied grade distributions over the three-course sequence taught by the current author over the past 25 years or so. These are given in Table 1. It is obvious, and expected, that the percent of "A" grades increases over the three course levels, while the percent of "C" grades decreases.

## DISCUSSION

Student assessment is not simple! It involves considerable effort on the part of the instructor, who is required to exercise judgments of various kinds at various times. But, student assessment is very important, for both the student and the instructor. Generally speaking, student assessment should reflect some variability in acquired student knowledge and skills within classes, at both the undergraduate and graduate levels of education. It is recognized that there will be some classes (e.g., advanced-level courses) in which there will be some limited variability.

Occasional assessment throughout an academic term has been found to encourage student learning, and has been very helpful to me as an instructor. Students are encouraged by the assessment to gain an understanding of concepts, and to appreciate some misunderstandings of concepts. Whether concepts need to be reviewed, instructional emphases redirected, and instructional pace needs to be changed are regularly learned by the instructor.

How student performance is assessed to reflect expected student variability may vary among areas of study. An example of how student performance is assessed in the area of statistical methods has been described above. No claim is made that the described

approach is applicable in all statistics courses. A certain degree of satisfaction has been achieved; improvements, however, are continually being considered and implemented.

Table 1. Grade distributions for the three courses

Course		A	B	C	D or F
1	#	134	249	124	10
	%	26	48	24	2
2	#	132	222	58	2
	%	32	54	14	.5
3	#	64	39	3	0
	%	60	37	3	0

REFERENCES

Brookfield, S. D. (1990). *The skillful teacher*. San Francisco: Jossey-Bass.

Cobb, G. W. (1993). Reconsidering statistics education: A National Science Foundation conference. *Journal of Statistics Education*, 1(1).

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63, 25-34.

Gullette, M. M. (Ed.) (1984). *The art and craft of teaching*. Cambridge, MA: Harvard University Press.

Hawkins, A., Jolliffe, F., and Glickman, L. (1992). *Teaching statistical concepts*. Harlow, Essex, England: Longman.

McKeachie, W. J. (1994). *Teaching tips*. Lexington, MA: Heath.

Moore, D. S. (1995). The craft of teaching. *Focus*, 15, 5-8.

Simpson, J. M. (1995). Teaching statistics to non-specialists. *Statistics in Medicine*, 14, 199-208.