A Note on the Probability of Misclassifying Students on the Basis of a Multiple Choice Examination

WALTER H. CARTER, JR.

Department of Biometry,
Medical College of Virginia,
Richmond, 23219

Introduction

As college enrollments continue to increase, the amount of time spent by faculty members grading test papers is also increasing. Often this extra time is taken at the expense of research and other teaching duties. In an effort to regain this time, many educators are giving multiple choice examinations and using the computer to: obtain the number of questions answered correctly, standardize the scores, rank the students, analyze the test to determine if it actually does discriminate among the students, and obtain various other parameters which are of interest to the teacher. After obtaining this information from the computer, the decision still has to be made as to which students failed to exhibit a satisfactory amount of knowledge of the subject matter on the test. This usually results in the students being assigned to groups such as excellent, pass, fail, or A, B, C, D, F. However, since a test can only sample a student's knowledge of the subject matter, misclassifications will occur, eg, a student placed in the pass group really belongs in the fail group or vice versa.

Since testing is somewhat analogous to random sampling, statistical theories find an application in the general problem of classifying students on the basis of examination performances. Here a method of calculating the probability of misclassification of students on multiple choice examinations will be discussed and applied to a test. The probability of misclassification is the probability that, in a comparison between two students, the student who gave the correct answer to fewer questions, say Z_1, actually knew the correct answers to as many or more questions than the student who gave the correct answer to Z_2 questions ($Z_2 > Z_1$). Such an event can occur because students can guess the correct answer to a question for which they do not know the answer.

* This investigation supported in part by Public Health Service Research Grant RR 0016-08 from the National Institutes of Health. I thank L. Kornhaber, Department of Biometry, Medical College of Virginia, for computational assistance in preparation of the tables.

Krutchkoff (1967) has defined the separation level of grades as the probability that a student with a higher grade actually knew the answers to more questions than the student with a lower grade. This is very similar to the probability of misclassification defined above. However, to arrive at an expression for the separation level of grades, it was necessary to make two rather restrictive assumptions:

1. Partial knowledge plays no role in a student's guess at the answer to a question for which he does not know the correct response.
2. The class of students taking the examination is homogeneous.

The first restriction is clearly too restrictive, as illustrated by a hypothetical example. Consider a student who does not know the correct answer to a given question which contains four possible answers. As a result of his partial knowledge of the subject matter, this student is able to eliminate as incorrect two of the possible answers. Hence, the student is now able to guess the correct answer with probability ½ instead of ¼. The second assumption is not a realistic one and is unnecessary in the derivation of the probability of misclassification. For a mathematical derivation of the probability of misclassification, see Carter (1971).

The probability of misclassification is based on each student's partial knowledge of the subject

MCV QUARTERLY 7(1): 14-16, 1971

TABLE 1*

	9(5, 4)	9(7, 2)‡	9(7, 2)‡
8(6, 2)†	0.839	0.889	0.332
8(3, 5)	0.410	0.478	0.047
8(4, 4)	0.448	0.516	0.063
8(5, 3)	0.410	0.478	0.047
8(7, 1)	0.477	0.545	0.077

Krutchkoff's separation level = 1 − PMC(8, 9) = 0.558
PMC(8, 9) = 0.442, where PMC is the probability of misclassification.

* The entries in Tables 1, 2, and 3 are the PMC's, ie, in the i, j position we have tabulated the probability that student i actually knew as many or more correct answers than student j.

† 8(6, 2) denotes a student who answered 8 questions correctly, 6 on one half of the test and 2 on the other half.

‡ These two students have different guessing distributions.

matter. Since we have seen how guessing can be affected by partial knowledge, it is reasonable to consider the probability of guessing correctly the answer to a question as a random variable. This is contrary to what is usually done. The probability of guessing the correct answer to a question is customarily taken to be $1/r$, where r is the number of possible answers to a question. In this paper, this probability is considered to be a random variable possibly taking on a different value for each individual taking the test. It is assumed that this variable follows a Beta destribution with unknown parameters which must be estimated.

As a result of the effect of partial knowledge on guessing, and since guessing only occurs on questions for which the answer is unknown, it seems appropriate to include on the examination several questions chosen such that the students would not be expected to know the answer. However, these questions should be chosen in such a manner as to allow a student's partial knowledge to aid in arriving at the answer. The parameters of this Beta distribution are then estimated for each student from his performance on these questions using a method due to Weiler (1965).

Application

The methods developed here were applied to an examination given by the Biometry Department to 127 first year medical students at the Medical College of Virginia in September 1968. The test was given to determine the mathematical and statistical backgrounds of these students. The students with high grades were to be assigned to a more advanced course in statistics than those with lower grades. It was decided that those students who correctly answered nine or more questions were to be placed in the advanced course and those who answered fewer than nine questions in the elementary course. Since there were five students who answered eight questions correctly and three who answered nine questions correctly, it was of interest to calculate the probability of misclassification for each pair of students with these two scores.

To estimate the parameters of the underlying Beta distribution, the examination was randomly divided into two parts such that on each part there was an approximately equal number of questions designed to measure a student's partial knowledge. The probabilities of misclassification were then calculated for the students (Table 1).

Since students will generally perform differently on the set or sets of questions designed to measure their partial knowledge of the subject matter, it is possible, by using the method just discussed, to calculate the probability of misclassification for students who have correctly answered the same number of questions. This is useful in that we can now rank

TABLE 2
The Probabilities used to Rank Students who Correctly Answered 8 Questions

	8(6, 2)	8(3, 5)	8(4, 4)	8(5, 3)	8(7, 1)
8(6, 2)		0.935	0.915	0.935	0.897
8(3, 5)			0.561	0.606	0.529
8(4, 4)				0.638	0.563
8(5, 3)					0.529

The assumptions made in the derivations of Krutchkoff's separation level will not permit the calculation of a separation level for students who answered correctly the same number of questions.

students who have the same raw score by calculating the probability that one such student knew the answer to more questions than another student with the same number of correct answers. This probability has been calculated for the students who correctly answered eight and nine questions and the results appear in Tables 2 and 3, respectively.

Conclusion

The students' ability to guess correctly questions on a multiple choice examination creates a problem in the determination of the number of questions the students actually knew. Therefore, it is not unlikely that mistakes will be made when grades are assigned. In this paper, a method for calculating the probability of misclassification of students as a result of a multiple choice test in which the assumption of a uniform guessing distribution is relaxed has been discussed and illustrated. To do this it is necessary to include on the examination several questions, related to the subject matter on which the students are being examined, for which the students will have to guess the answer.

Another problem which frequently occurs in the evaluation of students' performances is the assignment of meaningful class ranks to students who correctly answer the same number of questions. However, since students usually will perform differently on the set of "guessing" questions, it was shown that it is possible to calculate the probability that one student knew the answer to more questions than another student who correctly answered the same number of questions.

TABLE 3

The Probabilities used to Rank Students who Correctly Answered 9 Questions

	9(5, 4)	9(7, 2)	9(7, 2)
9(5, 4)		0.636	0.123
9(7, 2)			0.082

The author recognizes the need for placing confidence intervals on the estimated probabilities of misclassification. However, to calculate such quantities, a knowledge of the distribution of these probability estimates is necessary. This is not known and mathematics needed to arrive at this distribution would be extremely complicated.

References

Carter WH: The probability of misclassification of students on multiple choice examinations. The Journal of Educational and Psychological Measurement, December 1971, in press.

Krutchkoff RG: The separation level of grades on a multiple choice examination. The Journal of Experimental Education 36: 63, 1967

Weiler H: The use of incomplete Beta functions for prior distributions in binomial sampling. Technometrics 7: 335, 1965