Cheating On Multiple Choice Exams Is Difficult To Assess Quantitatively

D.B. Marx and D.E. Longer

Abstract

Multiple choice exams have the advantage of being easily graded by computers so that students may obtain test results quickly. Many instructors with large classes rely heavily upon multiple choice exams to reduce the hours normally spent grading examinations. Multiple choice exams, if not administered properly, have the disadvantage of providing an easy avenue for students to cheat. Cheating incidents have reportedly increased from 23% of all students in 1941 to over 75% of all students in 1980. The instructor is faced with the problem of detecting cheating, and once detected, proving its existence. Two statistical methods that calculate the probability of cheating on multiple choice exams are evaluated. Both methods make assumptions which weaken their use under actual classroom testing situations. Based on these weaknesses the authors concluded that no satisfactory method exists for proving cheating through statistical analyses. The recommended approach for instructors is to design exams and classroom settings that discourage cheating.

Introduction

With the advent of the computer age, more professors are writing examinations which use the computer to score and tabulate grades. In addition, some exams are produced by computer from a 'pool' of questions. Since multiple choice questions lend themselves to both computer generation and grading of an exam, this type of examination has become more popular with professors in recent years. Large classes make it desirable, from the instructor's point of view, to have an exam which can be graded easily by computer. The ease of grading multiple choice exams is only one reason for their growing popularity among instructors. The advantage of having a large number of possible questions from which to choose in constructing an exam also allows many different possible examinations to be arranged. Moreover, the instructor will often allow a class to have access to possible exam questions in order to prepare for the exam which may motivate students to increase their learning.

When the size of a class gets extremely large, however, it becomes difficult for the instructor to know each student individually. Although there are usually some students who will try to cheat during an exam, the urge to do so may increase as class size increases. Studies support the idea that cheating depends upon other factors (Baird 1980) in addition to class size.

Marx and Longer are members of the Department of Agronomy, University of Arkansas Division of Agriculture, Fayetteville, AR 72701.

Factors that contribute to cheating include: seating arrangement (Vitro 1969; Vitro and Schoer 1972), curriculum (Schab 1972), and surveillance (Buchard 1970). The relative simplicity of noting one letter or number (answer) from a neighboring student's paper increases the ease and probably the incidence of cheating during a multiple choice exam. The reported incidence of cheating has increased from about 23% in 1941 (Drake 1941) to about 55% in 1955 (Stannard and Bowers 1970) to over 75% in 1980 (Baird 1980). The problem for the instructor becomes one of detecting cheating and once detected, proving its existence. With the increase of legal cases being brought against universities for various reasons, the instructor should try to be as exacting as possible before awarding a failing grade for cheating, since many students will appeal any accusation of copying or cheating to either a University court or state judicial system.

The senior author has been called as an expert witness in several incidents involving students who were accused of cheating. In all cases the statistical evidence was weighed heavily and was based primarily upon analyses of responses to multiple choice questions. Instructors need to know what statistics can and cannot do when they are faced with a potential cheating incident. The authors' objectives are to outline some of the techniques which are used to detect and prove cheating and comment on their strengths and weaknesses. In addition, some previously successful methods used by the authors will be suggested for dealing with cheating situations.

Discussion

Assume that a multiple choice exam with ten questions has been distributed to each student in a class with each question having four possible answers of which only one is correct. These conditions can be easily extended to a different number of questions or possible answers. The most common way to determine if a student has copied another student's paper is to compare their answers. This is usually done by taking those questions missed by the student thought to be copying (copier) and comparing with those missed by the student of whose paper was copied (copyee). A probability is usually calculated using the binomial distribution, or in some instances if the number of questions missed is large, the normal approximation to the binomial. In our hypothetical example, let us assume that the student thought to be copying missed four of the ten multiple choice questions and the copyee missed five questions, but three of the five missed by the copyee coincided with those missed by the copier. That is, the copier answered four questions

incorrectly. Of the four answered incorrectly, he had the same wrong answer as did the copyee on three of the questions, but the fourth question he answered incorrectly was answered correctly by the copyee. The fact that the copyee answered the fourth question correctly rather than missing it (but differently than did the copier) is important as seen in the second method of calculating probabilities. The easiest way to attribute a probability of the described situation is to assume no knowledge. In other words if the copyee were not cheating he would choose answers at random. This is an indefensible assumption but is sometimes made. With this assumption of no knowledge, the binomial probability uses all the questions and gives a probability of matching for any one question as one chance in four or .25 and a probability of not matching as .75. In our example the students matched on the four they both got correct as well as on three they both missed but answered the same. The binomial gives as a probability for situations like the above as:

$$\sum_{x=r}^{n} {}_nC_x \cdot p^x \cdot (1-p)^{n-x}$$

where nCx is the combinations of n things taken x at a time which can be rewritten as $n!/((n-x)!x!)$ and $x! = x \cdot (x-1) \cdot (x-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$; n is the number of questions on the exam; r is the number of questions answered identically by both students; p is the probability of matching answers on any one question. In our example this reduces to,

$$10!/(7! \cdot 3!) \cdot .25^7 \cdot .75^3$$
$$+ 10!/(8! \cdot 2!) \cdot .25^8 \cdot .75^2$$
$$+ 10!/(9! \cdot 1!) \cdot .25^9 \cdot .75^1$$
$$+ 10!/(10! \cdot 0!) \cdot .25^{10} \cdot .75^0$$

$$= .0030899 + .0003847 + .0000285 + .0000009$$
$$= .0035040$$

The interpretation of this number is that if the students in question had not copied, the chance or probability that their answers would have matched on as many or more of the questions on the exam as they actually did is about three and a half chances in 1000. Clearly the assumption of no knowledge is not valid for most exams. Some questions will be relatively easy and thus students will answer them correctly. Therefore the number of questions answered identically will be large, leading to a low probability of occurrence that the students would have independently answered the exams in this manner. The test therefore is biased towards saying cheating has occurred.

A second way of comparing papers is more complex but does not assume that the answers are all equally likely. We shall assume the same situation as in the first example but will elaborate on it somewhat by giving the actual answers as well as the percent of the remaining class who gave the same incorrect answer as did the copyee (Table 1).

The second method of comparing papers, used by the **National Board of Medical Examiners** involves an

Table 1. Exam answers for two students

Question	1	2	3	4	5	6	7	8	9	10
Copier	A	D	B	B	C	A	D	B	C	D
Copyee	B	D	B	C	C	A	D	B	C	B
Correct Answer	B	D	B	B	A	C	D	A	C	D
Percent[1]	50	60	70	50	40	50	90	40	80	70
Percent B[2]	*	*	*	*	50	40	*	60	*	*

[1]Percent of the class answering that question correctly.
[2]Percent of the class who missed that question but answered the same as did the copyee.
*Percent not needed for calculations

'agreement analysis' to determine if one student copied off another's exam. The number of questions answered incorrectly by both students (joint wrongs) is determined. In addition the number of these for which there were identical wrong answers given is tabulated. In the example in Table 1 there are three joint wrongs and all three were identical answers. A group of students not suspected of copying but who also missed the answers is selected and the percentage of that group that gave them the same wrong answer as that given by the copyee (the student from whom the answers were allegedly copied) is calculated. This is designated as percent B. These percentages are averaged for all joint wrongs $(].5 + .4 + .6] \div 3 = .5)$. The binomial distribution is used to calculate the probability for the actual or larger number of identical joint using equation 1 (sum for r = number of identical joint wrongs to r = n). In equation 1, n is the number of joint wrongs, r the number of identical joint wrongs, and p the averaged percentage for all joint wrongs. For our example this reduces to n = 3, r = 3, p = .5, and consequentially the probability for two students who did not copy of matching on all three joint wrongs is given as .125.

Clearly this method alleviates some of the objections of the first method of calculating the probability of the two exams being alike. However, there are still some problems with this method. The averaging of the joint wrongs will lead to a bias in that some questions may be quite likely to be answered in the same incorrect way like others might not. Averaging them loses this distinction. This analysis also does not take into account the questions which were answered correctly by the copyee but incorrectly by the suspected copier. Therefore this method leads to a probability of cheating which is too high. In addition, for the second method the exam should be quite long since a shorter exam would not yield enough joint wrongs to warrant this method of analysis. Another assumption inherent in both methods is that they are based on the ability of the instructor to assess that a student is copying off of only one other paper and that paper is known to the instructor.

From the above inherent weaknesses of the assumptions which have to be made in the statistical analyses to show cheating has occurred, it is the

opinion of the authors that statistical analyses should be kept to a minimum in trying to show cheating occurred. Instead, effort should be made to reduce the incidence of cheating. This can be done in a variety of ways. First a random, but assigned, seating arrangement can be given for each class on exam days. The seating arrangement would change for each subsequent exam. Secondly, proper monitoring of exams is important. Students should not be seated close to each other if at all possible. A more difficult, but very effective method of reducing copying, is to distribute two or more exams. Usually alternating rows in the classroom are given the same exam and thus the neighbor to the left and right of every student has a different exam. The exams to be alternated may even have the same questions but simply presented in a different order. Finally, if a student is strongly suspected of copying we would suggest giving differing exams with different questions such that if the student were copying he would miss almost all the answers he copied. However, if he were to do his own work the exam would not hinder his score. This solution makes much more work for the instructor, but the benefits to the students in fairness and protection to the instructor for the integrity of his grades makes the extra work worthwhile.

Summary

The authors conclude that it is almost impossible to prove someone copied based upon the results from two methods of statistical analysis of multiple choice answers. The assumptions in the analysis of independence, common probability of an incorrect answer, and various approximations such as the normal approximation to the binomial make a defense of the analysis difficult. Instructors should be aware of the weaknesses of these analyses and should not rely on them in order to show copying has occurred. Rather than trying to prove that a student copied, the instructor should try and minimize the chance of a student being able to copy. This can be done by giving two or more multiple choice exams so that neighboring students have different questions and answers. This is admittedly more work for the instructor but the advantages are easily outweighted by the reduced risk of copying. Computer generated questions can be easily rearranged thereby considerably reducing the additional work. A second method of reducing the level of copying is to administer essay exams. In this case the suspected copier can be more easily identified and indeed it is much harder to read a paragraph on a neighbors paper than a single written letter. A better method is to administer two separate exams constructed in such a way that if a student is copying he will fail the exam. However, if the student is not copying, his score will not be diminished by the two exam procedure.

Peer Review Comments

The main objective of the paper, the evaluation of these two methods, is a good one, However, one example is not an adequate basis on which to judge the methods as not useful. As such, the paper does not add much to our knowledge. It seems as if the author is setting up a straw man and then knocking him down.

Rather than attack the assumptions of the method at the outset, I believe it would be better to give the methods a fair hearing and then evaluate the assumptions. I suspect that the conclusion may be the same, but at least a fair evaluation was made.

Data for testing the methods could be obtained in a couple of ways. If the author has test answer data for students in computerized form, real student data could be used to show, using a relatively large number of cases, what proportion of students known not to have cheated would have been judged as cheaters by these techniques. One could also determine whether any discrepancies were related to score on the test or other factors.

If this sort of data is not available, then a large number of hypothetical cases must be developed. These cases should include a range of test scores and a range in numbers of common errors. Since a 10-question multiple choice exam does not effectively exploit the strengths of this exam type, I would suggest using at least 25 questions in the examples.

Either approach would, I believe, provide an adequate basis on which to judge the methods.

I commend the author on the very excellent style in which the paper is written.

Rebuttal

Dr. Marx, because he is a statistician, has been called upon to testify as an expert witness in several cheating trials. Much of what we discussed in the article is based upon fairly extensive reviews of the literature in preparation for his testimony. Our methods for assessing cheating or proving it are mathematical treatments that have been developed and used by others. We attempted to show their flaws from a mathematical point of view. As such, I think that the conclusions we presented were accurate and will stand. Conducting trial experiments would not be helpful. The flaws in these approaches were discovered after they had been used by others and reviewed by ourselves and others. The mathematical flaws in both techniques are verified without need of hypothetical work.

D.E. Longer

References

Baird, J.S., Jr. 1980. "Current trends in college cheating." *Psychology in the Schools.* 17—515-522.

Buchard, J.D. 1970. "A methodology for conducting an experimental analysis of cheating behavior." *Journal of Experimental Child Psychology,* 10:146-158.

Drake, C.A. 1941. "Why students cheat." *Journal of Higher Education,* 70:418-420.

National Board of Medical Examiners; 3930 Chestnut Street, Philadelphia, PA 19104. (unpublished).

Schab, F. 1972. "Cheating in high school: A comparison of behavior of students in the college prep and general curriculum." *Journal of Youth and Adolescence,* 1:251-256.

Stannard, C.I., and Bowers, W.J. 1970. "The college fraternity as an opportunity structure for meeting academic demands." *Social Problems,* 17:371-390.

Vitro, F.T. 1969. The effects of probability of test success, opportunity to cheat and test importance on the incidence of cheating. Ph.D. dissertation, Univ. of Iowa, Ames, IA.

Vitro, F.T. and Schoer, L.A. 1972. "The effects of probability of test success, test importance, and risk of detection on the incidence of cheating." *Journal of School Psychology,* 10:260-277.

Demographic Profile of Students Majoring In Animal Science

T.A. Mollett and E.K. Leslie

Introduction

W.B. Martin (1981) propounded the hypothesis that effective teaching involves combining teaching skills with human sensibilities so that both science and art contribute to the cognitive process of learning. The successful application of this hypothesis in the classroom or lecture hall requires that the teacher or lecturer be knowledgeable about the audience. Meeting this requirement allows the presentation of new material to be related or made relevant to the experiences or interests of the students. This is a challenge to agricultural educators when one considers that as many as 60 to 70% of today's agricultural students lack farm or other agricultural experiences (Hasslen, 1983). Moreover, it has been suggested that students who lack a farm background or a significant amount of farm experience are disadvantaged as students. Such students often encounter difficulties in the classroom that may carry over into sub-optimal job performance (Helsel and Hughes, 1984). Because of the challenge presented by the nontraditional student clientele currently pursuing baccalaureate programs in agriculture, it is imperative that we, as educators, re-evaluate our curricula to determine if our courses are meeting the needs of our students. However, a prerequisite to this evaluation process is the need to develop an accurate profile of the students to be served by the curriculum. Therefore, the objective of this study was to develop a demographic profile of those students entering the Animal Sciences program at the University of Missouri-Columbia (UMC).

Methods

These data were collected over eight consecutive semesters beginning Winter Semester of 1980 and continuing through the Fall Semester of 1983. Student responses were obtained by distributing the following questionnaire to freshman and sophomore animal science students enrolled in the entry level animal science course (Introduction to Animal Science).

Animal Science 11 Student Survey

Instructions to Student: This is an anonymous survey. The information collected in this survey will help us determine which subject material needs to be presented to you based on your background and interests. Select the most appropriate answer for each question and blacken in the appropriate circle on the answer sheet.

1. I am a   (a) female, (b) male.
2. I have lived most of my life in   (a) Missouri, (b) the central time zone excluding Missouri, (c) none of the above.
3. My expected occupation upon graduation is   (a) farming, (b) work in agricultural related fields, (c) go to professional school, (d) go to graduate school, (e) work in a field unrelated to agriculture.
4. My major area of emphasis is   (a) animal agriculture only, (b) animal agriculture/pre-professional (pre-vet, pre-med, etc.)
5. I was reared   (a) on a farm |200 acres, (b) on a farm {00 acres, (c) in a town with less than 10,000 people, (d) in a city of 10,000 to 50,000 people, (e) in a city of more than 50,000 people.
6. Of my family's income   (a) 0%, (b) 1-25%, (c) 26-50%, (d) 51-75%, (e) |75% of the income comes from agriculture.
7. On our family farm (if any)   (a) crops, (b) dairy, (c) beef, (d) swine, (e) other are the main source of income.
8. I have had   (a) 0, (b) 1, (c) 2-3, (d) 4-5, (e)|5 years 4-H and/or FFA experience.
9. I have had   (a) 0, (b) 1, (c) 2, (d) 3, (e) 4 years of high school vocational agriculture.
10. I have worked on a farm or ranch for   (a) 0, (b) 1-2, (c) 3-5, (d) 6-10, (e)|10 years.
11. I have had   (a) no, (b) very little, (c) some, (d) considerable, (e) extensive **experience** with beef cattle.
12. I have had   (a) no, (b) very little, (c) some, (d) considerable, (e) extensive experience with dairy cattle.
13. I have had   (a) no, (b) very little, (c) some, (d) considerable, (e) extensive experience with sheep.

Mollett was formerly assistant professor at the University of Missouri and is currently on the staff of the University of Maryland, Eastern Shore, Princess Anne, MD 21853. Leslie is in the Dept. of Extension Education at the University of Missouri, Columbia, MO 65211.