Notes





# A Device To Detect Student Cheating

Robert G. Mogull, (E-mail: E-mail: mogullr@csus.edu), California State University

**Abstract**

*In this article, a method is offered by which to recognize, identify, verify and/or confirm that student "cheating" has occurred on multiple-choice formatted examinations. The approach rests upon the probability of more than one student incorrectly answering identical questions. The technique or "device" can be used either to screen or to confirm and it is seen to be easily applied in situations involving several students and for any number of exam questions and tests.*

## Goal Of Study

Copying between or among students during exams is a constant threat and reality. Yet, the identification of incidences of "cheating" is often difficult, especially when class sections are large and impersonal. Although it is likely that most "cheating" (like copying) occurs without the knowledge of the instructor (Cizak, 1999, pp. 13, 35), suspicions often arise after exams are graded, when several students are discovered to have missed the same questions. A pattern of identical or similar incorrect answers will appear among the student answer forms. The goal of this paper is to offer a method to recognize and/or to confirm that copying has occurred among students on multiple-choice formatted examinations, where recognition and confirmation are based upon probability.

In a recent article, Sotaridona and Meijer (2003) presented two indices to detect answer copying on multiple-choice tests. One index was developed by matching incorrect answers to a Poisson distribution, while the second index incorporated both correct as well as incorrect answers. The examples that were used were developed by simulations with various simulee sample sizes, number of test items and rates of answer copying.

This study differs from that of Sotaridona and Meijer (S-M) in several respects. First; in this paper, only incorrect answers are employed to detect copying. It will be seen that more than adequate sensitivity is attained with just incorrect answers. Second; where the S-M examples were based upon comparisons to a theoretical distribution and were verified through simulations, in this paper, an actual classroom experience of the author provides the example. Third; although the S-M paper is a valuable contribution, it is of limited use to those who wish to apply a detection device. Despite their interest and impressive development, the S-M indexes are simply too sophisticated (that is, too complex) for practical application. In this article, a detection method is presented with an eye on both simplicity and ease of application in real settings.

The scenario presented below reflects an actual experience of the author and, indeed, was the motivation and inspiration behind this paper. Class sections are large, with around 120 students per section. Four exams are administered during the semester, where each test consists of 25 multiple-choice questions and each question is provided with five answer options. Students fill out Scantron forms which are graded by machine. An Item Analysis form then indicates the number of Scantron test forms graded, the average number of questions answered correctly, and (very importantly) the frequency of incorrect responses per question. A suspicion of student "cheating" arose at the conclusion of the semester when two students were noted to have attained the same grades on exams.

The goal here is to determine the likelihood that two students, presumably working independently, would score the same on an exam and, more pertinently, would *miss identical test questions*. The objective is to calculate the probability of this phenomenon. If the chance of this occurrence is unusually small (to be discussed later), then the likelihood of collusion/collaboration/copying or "cheating" is high. Although an obvious pattern may exist, there is a need to quantify and distinguish it in numerical (that is, in statistical) terms. With what degree of certainty can it be claimed that collaboration between students occurred or that the two students did *not* arrive at their exam answers independently?

**The Details**

This part of the article will be presented in a format which allows interested readers to easily follow the steps in applying the methodology to their own unique experiences and classroom situations.

The technique focuses on the likelihood of students answering incorrectly the same multiple-choice questions. *Therefore, the first step is to calculate the probabilities of incorrectly answering specific* exam questions. In this example, two particular students were under suspicion, since their Scantron forms indicated not only the same grades on tests but also the same questions with incorrect responses. Column 1 in Table 1 below indicates that on the first examination, for example, *both* students missed questions numbered 2, 5, 6, 10 and 17. The Item Analysis form (not shown) reported that the frequencies of incorrect responses to *these specific questions* were 106, 181, 141, 121 and 79 respectively for the two class sections combined (see column 2 in Table 1). There were a total of 236 Scantron forms graded, so the proportions of all student papers that missed these questions were 106/236, 181/236 and so on. Therefore, the probabilities of incorrectly answering these five specific questions were approximately .45, .77 and so on (see column 3 in Table 1). Thus, column 3 indicates the proportion of all 236 student responses that were incorrect for the specified questions or the probability that a *single* randomly chosen student would miss each *specified* question.

The next step is to calculate the likelihood that *two* randomly chosen students, X *and* Y, would both miss each (same) specified question. This probability is the square of the probability of a single student missing each question. *Hence, column 4 in Table 1 indicates the probabilities* that two students would both answer incorrectly the specified questions *if they worked independently*. There is roughly a 20% chance that any two randomly selected students would incorrectly answer question number 2, roughly a 59% chance of question 5 and so on.

Finally for Examination 1: the overall probability that any two students would *both* miss questions 2, 5, 6, 10 and 17 is the product of the individual probabilities of any two students independently missing all five questions, or about (.202) (.588) (.357) (.263) (.112) = .001 = .1%. If all 14 significant digits of probability are reported, the more exact likelihood calculation is .00124770524027. Note the extremely small magnitude of this statistic. It indicates that two randomly chosen students, *if working independently*, would have only about a .1% likelihood of missing the same five specified questions from Exam 1.

Calculations of probabilities for Exams 2, 3 and 4 follow the same reasoning and procedure. For Exam 2, the likelihood of any two random students missing *all* nine same specified questions is .000000062450225501. For Exam 3, the probability is .000130301107292 for all five same specified questions and, for Exam 4, the chance is *.0000137480644603 that they would independently miss all seven same specified questions.*

Consequently, the relatively exact probability that any two students would miss *all* 26 same questions (over the four exams) is .0000000000000000000380540748206. This calculation is arrived at *either* by multiplying the respective overall probabilities over the four separate exams *or* by multiplying the individual probabilities from column 4 over all 26 specified questions.

Let us again interpret this result and place it in context of our goal. The above computation (which is taken to 31 digits to the right of the decimal) indicates the probability or likelihood that two randomly chosen students would *independently* answer incorrectly all 26 same specified questions (out of 100 total questions over four separate exams). The smaller this probability, the less likely and the more implausible it is that the two students under suspicion worked independently. That is, the smaller this probability, the greater the likelihood that they collaborated/colluded/worked jointly/shared/copied answers or "cheated" on the four tests. To express this as its complement and perhaps more dramatically, our level of confidence in concluding that they did indeed "cheat" is (1 - *.0000000000000000000380540748206*) = *.9999999999999999999619459251794*. Thus, the likelihood of "cheating" in this (actual) case is nearly an absolute/complete/total certainty.





**Table 1. Probabilities Of Incorrectly Answered Exam Questions**

| Question # | Wrong Frequency | Wrong Probability | Wrong Probability Squared |
|---|---|---|---|
| | | EXAM 1 N = 236 | |
| 2 | 106 | 0.449152542373 | 0.201738006320 |
| 5 | 181 | 0.766949152542 | 0.588211002585 |
| 6 | 141 | 0.597457627119 | 0.356955616203 |
| 10 | 121 | 0.512711864407 | 0.262873455904 |
| 17 | 79 | 0.334745762712 | 0.112054725654 |
| P (2 random students incorrectly answering *all* 5 questions) = .00124770524027 | | | |
| | | EXAM 2 N = 229 | |
| 4 | 51 | 0.222707423581 | 0.0495985965181 |
| 8 | 62 | 0.270742358079 | 0.0733014244582 |
| 10 | 122 | 0.532751091703 | 0.2838237257110 |
| 13 | 53 | 0.231441048035 | 0.0535649587155 |
| 16 | 72 | 0.314410480349 | 0.0988539501533 |
| 20 | 172 | 0.751091703057 | 0.5641387464010 |
| 23 | 84 | 0.366812227074 | 0.1345512099310 |
| 24 | 65 | 0.283842794760 | 0.0805667321372 |
| 25 | 99 | 0.432314410480 | 0.1868957495090 |
| P (2 random students incorrectly answering *all* 9 questions) = .000000062450225501 | | | |
| | | EXAM 3 N = 219 | |
| 4 | 71 | 0.324200913242 | 0.1051062321470 |
| 6 | 117 | 0.534246575342 | 0.2854194032650 |
| 16 | 62 | 0.283105022831 | 0.0801484539521 |
| 21 | 145 | 0.662100456621 | 0.4383770146580 |
| 22 | 77 | 0.351598173516 | 0.1236212756200 |
| P (2 random students incorrectly answering *all* 5 questions) = .000130301107292 | | | |
| | | EXAM 4 N = 226 | |
| 1 | 65 | 0.287610619469 | 0.0827198684313 |
| 8 | 62 | 0.274336283186 | 0.0752603962723 |
| 10 | 137 | 0.606194690265 | 0.3674720025050 |
| 11 | 122 | 0.539823008850 | 0.2914088808840 |
| 15 | 138 | 0.610619469027 | 0.3728561359550 |
| 18 | 78 | 0.345132743363 | 0.1191166105410 |
| 20 | 154 | 0.681415929204 | 0.4643276685730 |
| P (2 random students incorrectly answering *all* 7 questions) = .0000137480644603 | | | |

∴ P (2 random students incorrectly answering *all* 26 questions) = .0000000000000000000380540748206

Note that the probabilities of "cheating" have also been determined independently *for each separate test.* For example; the probability of two students independently missing the same five questions on Exam 1 was .00124770524027. Hence, the confidence level in concluding that the two students "cheated" on Exam 1 is (1 - .00124770524027) = *.99875229475973* or about *99.9%*. The likelihoods that the two students independently missed the same questions on *each* of the other three exams are even smaller and the levels of confidence in concluding that they "cheated" on those exams is, therefore, even greater.

## Further Comments

Below are several additional comments regarding both the technique presented and the conclusion.

- This article has demonstrated calculations where just two students were involved in "cheating" over four exams. The methodology can be easily modified to adjust for cases where more than two students are under suspicion. For example, if three students were involved, then the probability of three students missing the same exam question is the probability of a single random student missing the specified question to the third power. Thus, the current column 4 entries in the table would be replaced by the *cube* of column 3 entries. Further, if say two exams were administered instead of four then, in the final step, the overall sub-probabilities of just the two exams would be multiplied (instead of for the four exams demonstrated in the example). Of course, the greater the number of students involved and/or the more questions that are collectively missed, the smaller the probability that the students worked independently and, therefore, the greater the confidence level in concluding that "cheating" had occurred.
- The technique demonstrated may be used *either* as a general screening device or as a device to confirm. But, after the suspects are screened and identified and as a further step, Scantron forms then can be examined to reveal whether the particular *choices of incorrect answer options* were also identical. Thus, the instructor would examine the Scantron forms first for identical *grades*, then for identical missed *questions* and, finally, for identical choices of incorrect answer *options*. However, although this final step would further seal the fate (typically capital punishment) of the suspects in this example, it would be of very marginal mathematical or practical benefit to improve upon an already existing confidence level (in this actual case) of .99999999999999999996194592517 94.
- Related to this last point is the issue of how confident an instructor needs to be in order to confront and accuse the suspected students. There is no definitive answer to this question. In social science research which utilizes inferential statistics, usually a confidence level of at least 90% is employed. Therefore, as a rule of thumb, the user could employ this as a threshold for minimal confidence.
- The sensitivity of the technique or "device" depends, in part, upon the probability of the individual exam questions. For example; say an instructor wishes to employ the "device" yet obtain or screen for a minimal confidence level of 90%. To rephrase this; say that an instructor would not act unless or until s/he has a 90% confidence level. Also say that there is evidence of copying from just Exam 1 in the example. Then, the desired confidence level could be obtained from the first three questions alone, since (.202) (.588) (.357) is roughly .042, which yields a confidence level of (1 - .042) = .958. (Using just the first *two* specified exam questions would provide a confidence level of "only" .881.) Similarly, the desired minimal 90% level could be obtained on Exam 2 by examining the first specified question alone, on Exam 3 by examining just the first two questions, and on Exam 4 by examining the first question alone. The high sensitivity of the method/technique/tool/device therefore allows it to be employed *on each separate exam* in order to identify "cheating."
- As stated above, the sensitivity of this "device" depends partly upon the probability of students missing the *individual exam questions*. To be more specific; the "device" is more sensitive with questions that are *easier* to answer correctly than with questions that are difficult. To illustrate: on Exam 1, there was a 77% chance that a single random student would incorrectly answer question #5 and a 33% chance of missing question #17. And, the likelihood that two random students, if working independently, would *jointly* miss question #5 is 59% versus only 11% for question #17. Therefore, the confidence level in concluding that they collaborated on the more difficult question #5 is (1 - .588211002585) = .411788997415 or about 41% and on the easier question #17 is (1 - .112054725654) = .887945274346 or about 89%. Hence, it is more likely that two students *working independently* would miss a difficult question than an easier question. If they *both* miss an *easy* question, the chance of having worked independently (that is, of *not* "cheating") is smaller. Hence, the "device" attains relatively greater power with easier test questions.





## Summary

The purpose of this paper has been to offer a method either to identify or to confirm that "cheating" has occurred between students taking multiple-choice examinations. The technique rests upon calculating the likelihoods that more than one student would incorrectly answer identical exam questions. That is, the "device" determines how likely it is that students would miss the same questions *if they worked independently*. If the probability is small, it suggests/confirms/verifies that the suspected students did *not* work independently. The methodology of identification and/or verification of the likelihood of "cheating" is shown to be easily applied in situations with multiple students and any number of exam questions and tests.

*The author is grateful for helpful comments on this paper by Michael D. Mogull. Any errors, however, are the sole responsibility of the author.

## References

1. Cizek, G. J. (1999). *Cheating on Tests: How to Do It, Detect It, and Prevent It.*
2. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.
3. Sotaridona, L. S., & Meijer, R. R. (2003). "Two New Statistics to Detect Answer Copying." *Journal of Educational Measurement*, 40(1), 53-69.

## Notes