2004

Cheating in Online Student Assessment: Beyond Plagiarism

Rowe, Neil C.

Monterey, California. Naval Postgraduate School

On-Line Journal of Distance Learning Administration, Summer 2004
https://hdl.handle.net/10945/36015

Cheating in Online Student Assessment: Beyond Plagiarism

Neil C. Rowe

U.S. Naval Postgraduate School
Code CS/Rp, 833 Dyer Road
Monterey, CA 93943 USA
ncrowe@nps.navy.mil
(831) 656-2462

Abstract

Online student assessment features in many distance-learning programs. The prevention of plagiarism has been subject of much attention, but insufficient attention has been given to other problems of dishonesty in online assessment. We survey the types of problems that can occur and what can be done about them. In general, we believe educators are too unaware and/or deliberately oblivious to these problems, and most countermeasures proposed are insufficient.
This paper was published in the On-Line Journal of Distance Learning Administration, Summer 2004.

Introduction

When a student scores well for an online assessment, does that prove that they know the material? This question is becoming increasingly important as online distance-learning programs become popular. While traditional paper-and-pencil assessment of students can be done in distance-learning programs, it is appealing to think that technology can both teach material and assess learning. Traditional assessment also requires costs: the time of human proctors, care in control of the assessment materials before and after administration, and grading effort, all of which are simplified in online assessment. But can we trust the results?

Unfortunately, often we cannot. Everybody lies at one time or another (Ford, 1996), and cheating is common in education (Cizek, 1999; Lathrop and Foss, 2000; Dick et al, 2003). (Bushweller, 1999) cites disturbing statistics such as that 70% of American high school seniors admit to cheating on at least one test, and 95% of the students who said they cheated were never caught. (Dick et al, 2003) reports 12 studies of cheating, mostly with college students, in which an average of 75% of students reported cheating sometime during their college career. (Cizek, 1999) also reports that cheating increased significantly in the second half of the twentieth century, and that cheating increases with the age of the student at least through age 25, which has serious implications for distance learning with its often-older students. Cheating also tends to increase as the bandwidth (information per second) of the communications channel between assessor and assessee decreases; that is, people who feel more "distant" cheat more (George and Carlson, 1999; Burgoon et al, 2003). Online assessment has a narrower bandwidth than classroom assessment (instructors cannot watch students work, for instance) and is therefore encourages cheating.

Other reasons support online cheating too. From a practical standpoint, it is often easier to cheat online (since what or who the assessee brings to the assessment cannot be seen, and assessment can often be done at any time of day), which increases temptation. Many students are more comfortable with computers than their instructors are, and many know full well the potential of computers for cheating. In addition, students often have less commitment to the integrity of distance-learning programs than traditional programs because distance-learning programs often lack tradition, are often taken by people with pressures from other jobs, and many programs are new and not fully debugged. These reasons could justify cheating in the minds of some (Bell and Whaley, 1991).

Matters are exacerbated by the willful ignorance of many instructors and administrators of the possibilities for cheating in online assessment. Some attention has been paid to the increased ease of text plagiarism using the Internet (McMurtry, 2001; Heberling, 2002) but little to the problems of focused assessments using instruments such as multiple-choice and calculation questions, necessary in most science and engineering courses. Several studies of distance-learning assessment have ignored the cheating problem (Bull, 1999). (Kaczmarczyk, 2001) for instance reports that professors who do distance learning don't see cheating as a major problem ? but if 95% of the students cheating are not being caught as per Bushweller, professors are poor judges of cheating frequency. There is undoubtedly much cheating that is not caught, just as there are hundreds of times more break-ins by hackers to computer systems than those that are noticed. Since problems of information security are so common today, it is not surprising that problems of "educational security" could be common too.

Some writers have argued that cheating should not be much of a problem if a course is well designed. A combination of online and traditional paper-and-pencil testing, such as used at the Open University in Great Britain, may reduce problems of cheating but only a little. If a student does much better in online assessments than traditional examinations, they could be cheating online ? or they could be more nervous in traditional testing. Group projects can reduce cheating if students monitor one another, but group projects are not appropriate for many subjects and learning skills. Others have argued that assessment should be continuous so it is less cost-effective for students to cheat (Bork, 2001). This does require considerable work in setting up a course. It also gives students less opportunity to study and digest the material offline, a key feature of self-education, as computer screens are more difficult to navigate than the pages of a book, and will remain so for some time to come. It creates more of a climate of distrust, suggesting that students cannot be trusted to learn without constant testing. It is also logically impossible to simultaneously satisfy three important criteria for continuous assessment: That the assessments are of equal size, that the assessments test all material of the course to the same degree, and that each assessment tests some material covered before the previous assessment (see Appendix). Most seriously, continuous assessment almost inevitably overemphasizes a student's short-term memory, and the purpose of most education to cause long-term effects on a student. This is why final exams at the end of a course are so important.

Others have argued that it is insulting to students to even suggest they might cheat. While an explicit cheating policy, as well as the very act of testing, decreases the trust level of a testee because restrictions on people decrease their trust (Sztompka, 1999), assessment is central to education because the main purpose of an educational institution is to validate student knowledge (hence the importance of degrees and diplomas). If an institution claims to provide a service, they must prove to society that they do. Otherwise, their reputation will suffer -- reputation is very important in today's educational institutions -- and accreditation can be denied (Kaczmarczyk, 2001). So accurate assessment methods help insure the survival of educational institutions.

Some anecdotal evidence (Kaczmarczyk, 2001) suggests students cheat less in distance learning than with traditional instruction. This may be because new technologies typically first attract smarter and more motivated users with less reason to cheat. A good example is the Internet, which during the 1980s had virtually no incidents of vandalism, theft, and crime as its users were highly professional; then things changed dramatically in the 1990s with the appearance of a larger and broader class of users. As distance-learning technologies become more popular as we indeed wish, we will increasingly see a broader spectrum of students. We need to be prepared now to meet the increasing ethical challenges to come by enacting good policies. We consider now three of the most serious problems involving cheating in online assessment that have not been sufficiently considered previously.

Problem 1: Getting assessment answers in advance

A big problem with online assessments is that it is hard to ensure all students take them simultaneously (Olt, 2002). Otherwise, earlier students can supply answers to later students if some of the same questions are used: It is easy for the earlier students to take screen shots (copies of what is on the screen) under most operating systems, and otherwise, earlier students could just memorize the questions. Creating "windows of availability" for assessments as in WebCT and Blackboard helps a little but does not solve the problem unless the windows are on the order of minutes in width, not days. An interesting idea is to reward by a grading factor those answers that are the most atypical, but that will not work when there is only one correct or good answer.

If all-at-once assessment with a single test is not possible, assessment questions can be drawn from a large pool and each student given a random selection, as is possible in WebCT and Blackboard (Olt, 2002). But it is hard to grade students fairly when they get different questions since some students will get harder questions than others. A way to reduce unfairness is to ask many questions, but then assessments become long and tedious. A more serious problem with pools is that instructors systematically underestimate how large the pool must be to make negligible the overlap of questions between tests. This is related to the classic "birthday fallacy" where people systematically underestimate the likelihood that two people in a group of 20 have the same birthday -- it's actually around 50%. If M is the number of questions on a test and N is the number of questions in the pool, the expected number of questions in common between two randomly chosen test sets is approximately M*(M/N). Table 1 shows the expected overlap in questions for representative values of test size and pool size. If instead one wishes to give each student a different set of questions, the pool must be at least S times the size of the test where S is number of students, requiring even larger pool sizes. So it is difficult to prevent unfair advantage to later students when drawing questions from a pool, as most instructors will not have the patience to provide an adequately large pool.

Even with a large pool, a different serious danger is that students may be able to log in as the instructor and read the answer key themselves. Most assessment software is protected by short passwords -- in Blackboard these can be as few as eight characters, easy to guess with today's systematic "cracker" software. Even when students cannot guess the instructor's password, they can use "social engineering" methods that have been successfully used to scam even smart people into revealing their passwords, like "emergency" calls from alleged programming staff or "please change your password temporarily for system testing" requests (Mitnick, 2002). Since few instructors are security experts, they can easily fall for many of these scams.

Table 1: Example calculation of the overlap of two randomly chosen student tests.

Number of questions selected at random	Pool size	Average number of questions in common for two students
5	10	2.5
10	20	5.0
20	40	10.0
30	60	15.0
40	80	20.0
5	15	1.7
10	30	3.3
20	60	6.7
30	90	10.0
40	120	13.3
5	25	1.0

10	50	2.0
20	100	4.0
30	150	6.0
40	200	8.0

Even if students take an assessment simultaneously and the instructor's password is adequately protected, students can use "spyware" to electronically sneak a look at how other students are answering questions during an assessment or what the instructor is typing on their computer. Spyware is software that secretly sends messages about you to other people. It has become a problem on the Internet where some free utilities secretly install spyware to send information to advertisers about what sites you are visiting (Mintz, 2002). The software technology of spyware is not difficult, and students who steal test answers could sell them to other students. Students could also use software called "sniffers" (McClure et al, 2001) to decipher the message packets of a local-area network containing fellow students or the instructor and thereby read their answers or passwords. Students could also use a variety of hacker attack methods to gain server-administrator privileges on the course-server machine ("privilege escalation"), which is just as good as obtaining an instructor password, unless the machine is kept "patched" regularly with operating-system fixes. Students don't need to be software experts to do these things, just to download the spyware or sniffers from a Web site and follow a few simple installation instructions, just as how most hackers attacking computer systems don't understand their attack software. Installed spyware and sniffers can be recognized by careful computer forensics (Prosise and Mandia, 2003), but it requires some work.

Even without special software, students may be able to find answers by using computer-forensics tools themselves on computers used by other students or instructors. When a user logs off a computer, they leave in memory and on disk many records of what they have been working on, and it's not difficult for this information to be retrieved with built-in tools and free software. For instance, a student or instructor working on an assessment over the Web will leave the final version of the pages they downloaded, with their answers, in the cache of their Web browser. Even if the power is turned off, the cache will still remain on disk, and even if files are requested for deletion, operating systems often send them to a "recycle bin" before actually deleting them.

Problem 2: Unfair retaking or grade changing for assessments

Another serious problem with online assessment is that may be possible for students to retake an assessment multiple times until they are satisfied with their performance, even if that was not the intention of the instructor. WebCT and Blackboard use a "server" architecture for assessments where the answers and assessment software are stored on a central machine. If the server software is not properly designed, students can break their connection to the server during an assessment, then claim they lost power and test answers and need to start over, giving them extra time to consult collaborators or unauthorized reference materials. Students could also crash (stop) the server after the grading is done but before the grades have been recorded; crashing is not difficult with the many hacker tools currently available. Another trick is to change the system clock so the grading server thinks that a new test assessment is actually prior to an earlier assessment; many operating systems do not adequately control access to their system clock. Thorough testing must be done to ensure that these problems cannot occur, and WebCT and Blackboard have never provided the necessary test data. Password theft of the instructor's password as discussed above also permits a student to change previous grades, since instructional software must allow instructors to correct grading mistakes. Blackboard doesn't even bother to tell the instructor when they last logged in, a key clue to this kind of manipulation. Again, computer forensics can detect these unauthorized activities, but this is often not easy.

Problem 3: Unauthorized help during the assessment

Probably the most serious problem with online assessment is confirming that the student is in fact who they say they are. Since several distance-learning methods such as online discussion groups and email between students encourage collaboration, students have an excellent excuse of habit for unauthorized collaboration on assessments. A poor student could easily hire a good student to take their tests, or a team of good students, or could arrange "consultants" to contact during an assessment for the hard questions. Just because the student provides their password doesn't mean they are the ones answering the questions at a remote site. This issue of "authentication" has been subject of much research in computer security, but usually the problem is ensuring that a given person is present, not that they are alone, which requires different methods. Note that "high-tech" solutions such as infrared or electromagnetic monitoring of test-takers are not adequate for preventing unauthorized collaboration because communication can take many forms including aural, optical, and olfactory.

One approach is to include some traditional tests in any distance-learning program, as with the Open University, with proctors and the usual test security. But this costs money. Also, since proctors must have no stake in the outcome for the student (unlike supervisors or colleagues) because collusion is still possible, contractors may be necessary to proctor. For this reason and uniformity of test conditions, it will be hard to be fair to a student who is the only student within a large geographical area, as can be the case with distance learning in the U.S. military. It is also possible for one student to impersonate another, so each student will need an identification card and it will need to be checked at the assessment site. Such tests are an imposition on the students and will need to be minimized in number because of their logistics. Hence much is riding on the outcome of these traditional assessments (since a bad score should surely override great scores on online assessments in which we are not sure who is taking the test); students will be under pressure, some students perform unfairly poorly under pressure, and this is a good incentive to cheat.

Countermeasures

(Cizek, 1999) provides a good overview of methods for recognizing, responding to, and preventing cheating in traditional paper-and-pencil assessments, and many of his insights apply to online cheating. Since one can observe little of the test-taker online, statistical methods are often the only hope to detect cheating. Most distance-learning assessments are multiple-choice, true/false, and matching questions since they are much easier to grade automatically than short-answer and essay questions. Then the number of identical incorrect answers between two students remains a good clue, provided the assessment is not too easy, and can be given confidence intervals for a hypothesis of cheating. (While several sophisticated techniques can establish confidence intervals, the distribution of the number of answers in common between noncheating pairs of students should usually be close to a normal (Gaussian) distribution.) Software, including some we have written, is available to compute these automatically for every pair of students in a class. However, noting a similarity between answers does not establish guilt ? one student may be an innocent victim. So suspicion of copying does tell us who to penalize.

Note that the number of correct answers on an assessment and the consistency of a student's marks between assessments are not reliable clues for cheating, as extreme values can reflect honest good students or honest bad students, and inconsistency can reflect honest students having a bad day. Thus cheating by getting the answers from the instructor's password or computer is very difficult to detect by statistics, unless follow-up discussions with the student or deliberate entrapment is used.

As for preventing cheating, Cizek suggests several ideas which we can consider for online assessment:

- *Define cheating and encourage honesty.* This is also a good idea in distance learning: Students could be asked to read and sign a policy statement like an honor code or integrity policy at the beginning of the course. However, academic-integrity policies (Lathrop and Foss, 2000) are less effective with the typically older students encountered in distance learning than with younger students, as the ethical principles of older students are more difficult to change.
- *Know the assessment takers.* Statistics on student performance can suggest unexpectedly good performance, but this information is not especially reliable and the instructor should be cautious in using it. Informal discussion with students after unexpectedly good performance on an assessment can often reveal the student's

true level of knowledge.

- *Understand what students face.* This is always good advice in instruction.
- *Maintain assessment security.* Instructor passwords should be hard to guess, and this should be checked by software. Instructors should keep duplicate print copies of grades to guard against changes. Assessment documents should not be stored as files on instructor machines, but only on a server machine, and the server software must be kept up-to-date to minimize hacker attacks to obtain instructor or administrator privileges. The server should have intrusion-detection software to catch attacks before they happen and should do auditing to reveal attacks. The server should also have effective physical security to prevent events like theft of disks. To keep track of all this, the server site must have a designated Security Manager.
- *Proctor the assessment.* Proctors not personally related to the student are important when students use computers to do the assessment. Proctors can ensure that students take the assessment at a designated time, without collaborators, and without unauthorized materials. But if computers are used, proctors can't see everything stored on them, which may include unauthorized materials. And such proctoring can still be victimized by both "low-tech" cheating such crib notes and "high-tech" cheating like handheld devices communicating with collaborators outside the test room.
- *Control the assessment situation.* Prohibit all handheld devices (calculators, personal organizers, pagers, cell phones, headphones, etc.) since all can be used to store and transmit information, possibly from outside the assessment room (Lathrop and Foss, 2000). If computers are used for the assessment, communication should be made as difficult as possible between them and the rest of the Internet. Disable most networking capabilities on the machines, including wireless ones. Close all ports (critically ports 21 (file-transfer), 23 (remote-login), 25 (mail), and 443 (secure Internet connections)) except for the HTTP Internet port 80 which the grading server can use, and this port should be restricted to only connect to the server and not any other machines. All printer connections should also be disabled (so students cannot print screen shots) and removable-disk drives removed (so students cannot copy to or from their own storage media). Do not allow access by students to the testing computers except during the period of the test to prevent students from storing answers on it. Even with all these steps, proctoring is still important since there are so many ways to cheat electronically.
- *Make the assessment a learning experience.* Overly difficult or overly easy tests tend to encourage cheating because the student doesn't see the point, so instructors should avoid them.
- *Use constructed-response test formats.* A chief advantage of online assessments is automatic grading, for which constructed-response questions cannot be used much because the instructor would need to specify too many possible correct answers. One way to get different questions for each student is by changing numeric parameters in the question (Heron and Pain, 2003). But that requires a programmed calculation to obtain the right answer in each case, something not possible with most instructional software, is only possible for certain types of course material, and does not prove that the student is alone.
- *Use varied test formats.* Drawing questions randomly for each student from a pool is one example, with the disadvantages cited earlier. It helps further to reorder multiple-choice answers randomly if possible. Questions also could be randomly parameterized as just mentioned.
- *Avoid situations that encourage cheating.* This means avoiding take-home tests, unproctored tests, and student grading of tests in distance learning.
- *Plan for the unexpected.* There should be a plan for when technology problems prevent an assessment being taken at a given time. As with conventional testing, a different test should be used for late test-taking. This includes when a computer or network is reported to have crashed midway through an assessment, since such a crash could have been caused deliberately or students may simply be lying about it.
- *Entrapment.* Though Cizek does not mention it, a useful way to catch the advance stealing of tests and answers is to plant fake tests in possibly accessible places, like on the grading server, while keeping the true test offline until test time. Then if a student uses answers from the fake test, we know our security precautions

have been faulty and can take measures. This works best when the fake test looks as close as possible to the true test, as with slightly reworded questions or the multiple-choice questions with reordered answers. It also helps if assessments can be encrypted and restored quickly when needed. Fake tests are analogous to "honeypots", computers on the Internet with no legitimate purpose other than to attract attackers for study purposes.

Based on the above, we make the following recommendations for online assessment:

- Human-proctored traditional paper-and-pencil tests with traditional security procedures should be used for major assessments in distance learning.
- If manual grading is too burdensome, human-proctored tests taken at a computer are a second-best choice provided that the computer's software and networking capabilities are tightly restricted as described above.
- If students take the same assessment at different times, it is critical to draw questions randomly from a large pool and reorder them (and answers to multiple-choice questions) randomly.
- We should automatically and routinely compare answers given by students on assessments. When similarities beyond those due to chance are observed, especially for incorrect answers, it is usually best to just ask students to take a different assessment covering the same material since it is hard to prove guilt. Retaking should be done in a more secure manner than the original test, as for instance with essay questions instead of multiple-choice.
- Countermeasures for cheating should be a consideration in purchase of distance-learning management software. The burden is on the vendor to prove that their grading servers and client machines cannot be easily broken into and crashed. Assessment software must permit random test construction. Unless vendors do these things, their software should not be used.

Conclusion

In summary, online assessment raises serious security issues. Many methods of cheating are facilitated, some quite new, and it is inevitable that cheating will increasingly be automated and distributed as software packages. While there are countermeasures, most are unsatisfactory in some way. For these reasons, online assessment in distance-learning programs should be done with caution until more progress is made on the technical development of countermeasures. Certainly, practice quizzes can continue to be done online, and tests with essay and short-answer questions can be done online if plagiarism safeguards are used and instructors have the time to grade them, but traditional one-location one-time face-to-face testing for much of the student's grade will need to be the assessment norm for distance learning in the foreseeable future.

References

Bell, J. B., & Whaley, B. (1991). *Cheating and deception*. New York: Transaction Publishing.

Bork, A. (2001). What is needed for effective teaching on the Internet? *Educational Technology and Society,* 4(3), 139-143.

Bull, J. (1999). Computer-assisted assessment: impact on higher education institutions. *Educational Technology and Society*, 2(3), 123-126.

Burgoon, J., Stoner, M., Bonito, J., & Dunbar, N. (2003, January). Trust and deception in mediated communication. 36th Hawaii Intl. Conf. on Systems Sciences, 44a.

Bushweller, K. (1999, April). Generation of cheaters. *The American School Board Journal*, 186(4), pp. 24-32.

Cizek, G. J. (1999). *Cheating on tests: how to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum.

Dick, M., Sheard, J., Bareiss, C., Carter, J., Joyce, D., Harding, T., & Laxer, C. (2003, June). Addressing student cheating: definitions and solutions. *ACM SIGCSE Bulletin*, 35(2), 172-184.

Ford, C. (1996). *Lies! Lies!! Lies!!! The psychology of deceit*. Washington, DC: The American Psychiatric Press.

George, J., & Carlson, J. (1999, January). Group support systems and deceptive communication. 32nd Hawaii Intl.

Conf. on Systems Sciences, 1038.
Heberling, M. (2002). Maintaining academic integrity in on-line education. *Online Journal of Distance Learning Administration*, 5(2).
Kaczmarczyk, L. (2001). Accreditation and student assessment in distance education: Why we all need to pay attention. Proc. 6th Conf. on Innovation and Technology in Computer Science Education, Canterbury, UK, 113-116.
Lathrop, A., & Foss, K. (2000). *Student cheating and plagiarism in the Internet era: a wake-up call*. Englewood, CO: Libraries Unlimited.
McClure, S., Scambray, J., & Kurtz, G. (2001). *Hacking exposed: network security secrets and solutions, third edition*. New York: McGraw-Hill Osborne Media.
McMurtry, K. (2001, November). E-Cheating: combating a 21st century challenge. *T.H.E. Journal,* 29(4), 36-41.
Mintz, A. P. (ed.) (2002). *Web of deception: misinformation on the Internet*. New York: CyberAge Books.
Mitnick, K. (2002). *The art of deception*. New York: CyberAge Books.
Olt, M. (2002, Fall). Ethics and distance education: strategies for minimizing academic dishonesty in online assessment. *Online Journal of Distance Learning Administration*, 5(3).
Pain, D., & Le Heron, J. (2003). WebCT and online assessment: the best thing since SOAP? *Educational Technology & Society*, 6(2), 62-71.
Prosise, C., & Mandia, K. (2003). *Incident response and computer forensics, second edition*. Emeryville, CA: McGraw-Hill/Osborne.
Sztompka, P. (1999). *Trust*. London: Cambridge University Press.

Appendix: Proof of the logical impossibility of perfect continuous assessment

Suppose we have N assessments in a course, separating the course into N even segments. Suppose T total points in the course determine the grade. Then if the assessments are to be the same size, to provide a fair burden for student studying, each must have T/N points. If we want the assessments as a whole to test each part of the course equally, the total number of points allocated to testing material in each segment of the course should also be T/N. Consider the first segment of the course. Its assessment can only test material covered in that first segment, so all T/N of its points must be devoted to material of the first segment. But that then leaves no points left for later assessments to cover this material and still provide evenness of coverage. Hence each segment of the course can only test the material since the last assessment if we are to satisfy our criteria. But that means only short-term memory of the students is being tested unless we have just a single assessment, a final exam.

Acknowledgements: Thanks to Peter Denning, David Floodeen, and Joseph LoPiccolo for helpful suggestions for this paper.