Adequate Testing and Evaluation of On-Line Learners

June 25, 2001

James R. Mallory
National Technical Institute for the Deaf

Abstract

One logistical and pedagogical challenge with online or remote teaching is properly evaluating the deaf or hearing learner's knowledge in a fair, secure and efficient manner. NTID has been teaching Distance Learning (DL) courses for the past six years and has implemented several different means for remotely evaluating students in C++ and Visual Basic computer programming classes. The three methods used were remote proctoring, online testing through multiple choice and fill in questions, and downloading and uploading answers to template type of questions using electronic conferencing software.

When implementing a testing method, factors such as security, ease and timeliness for data collection, ease of use by the students, turn around time and feedback to the students, flexibility of how the testing is done (limited to multiple choice or open-ended) and administration/implementation are but a few of the issues that one should consider. Each method has its strengths and weaknesses in these different areas.

In this session, findings of the three different testing methods will be presented, as well as the advantages and disadvantages of each from an administrative, teacher and student's perspective. This presentation will be followed by an interactive discussion with the audience regarding which testing systems are best for which content areas and learners.

Because of the problems and issues with these kinds of assessments, some faculty choose to use authentic assessments such as student projects, group projects, term papers, portfolios of student work, and so on.

Introduction

"In 1847 the first graded school was invented in America. The assumptions about the course of learning upon which the school was built were straightforward: students would be grouped by age, and each age level would be assigned to a grade. Age grading in our schools became the dominant organizational structure. It was further assumed that, since students were grouped by age, the content and aims for each grade should be the same for all children in that grade. Effective teaching was defined as the ability to enable all children in a grade to achieve the goals for that grade level." (Eisner) Many original goals and principles of instructional assessment have not changed since the first school was invented in America, although implementations, strategies and rapid changes in technology have created a vast array of new options in this field.

During the past 20 years, post secondary education has sprouted an additional format of remote or online asynchronous type of learning. As remote or online learning becomes more widely used, the issue of testing or evaluating the learner's knowledge also continues to grow, thereby increasing issues relating to the assessment of that knowledge. The commonly shared concerns by faculty with any type of remote testing seems to be:

1. How does an instructor know if that student is actually taking the test himself/herself and is not getting help or is not really someone else?

2. The teacher's discomfort with the technology.

3. Will the testing be limited in a conceptual way?

4. How will options be limited using this kind of testing?

5. How will the qualitative results of students be determined?

6. Will the way the questions are worded for an online test affect students' responses?

7. Will this technology be easy for the students to use?

8. How easy and how timely will data collection be?

9. How difficult will administration of this type of testing be?

Traditional Classroom Approaches

One NTID faculty member, Greg Emerton, summarized what the author has found to be typical faculty concerns using online testing in a traditional classroom. Greg opted not to use computer testing for some of the following reasons: "I do my exams in class to maximize control of the testing situation." G. R. Emerton (personal communication, May 22nd, 2001) This security issue is probably the most prominent concern in online testing. Another faculty member, Marc Marschark, states his opinion of online testing. "I was concerned that if I had students answer essay questions online, they might be getting assistance (my tests are 'open book' -- not 'open friend'). Of course, that's no different than would be the case for take-home exams...but I never use those." M.

E. Marschark (personal communication, May 22nd, 2001) Emerton also was concerned about the conceptual limitations of online testing. "My exams typically consist of a mixture of true/false and multiple choice questions and a number of short answer essays will allow the individual student to express what he or she knows about the topic. Reading across all of the answers for a given question also gives me a better sense of how the class is responding than scoring individual tests."

The responses of professors Emerton and Marschark are not uncommon. They are both master teachers and have many years experience teaching deaf and hard-of-hearing students at the post-secondary level. Their mindset is common among secondary and post-secondary instructors. Technology and the assessment vehicles described in this paper will address many of these issues, whether they are traditional on-campus or remote online students. Some of these issues apply to remote online students only. The evolution of traditional classroom teachers to online testing is slow. It is often more work and not worth it pedagogically for teachers in traditional classrooms to change their method to a more automated way of assessing students. Online or distance learning instructors are more enthusiastically adapting these models.

The Fundamentals of a Fair Online Assessment

The same fundamental principles apply to a proper online assessment that applies to a traditional on campus assessment. Items such as reliance on rote learning and other issues are still important, although these are outside the scope of this paper. Linda Suskie lists "Seven Steps to Fair Assessment:

1. Have clearly stated learning outcomes and share them with your students.

2. Match your assessment to what you teach and vice versa.

3. Use many different measures and many different *kinds* of measures.

4. Help students learn how to do the assessment task.

5. Engage and encourage your students.

6. Interpret assessment results appropriately.

7. Evaluate the outcomes of your assessments.*"* (Suskie)

Suskie's recommendations can all be accomplished with remote or online assessments. The myth that students cannot build a rapport with each other or with the teacher is not accurate from the author's experience. If the class size is reasonable, 15 to 20 students maximum, a relationship can be forged among classmates and with the teacher through desktop and online conferencing as well as email. (Mallory)

Security Issue Based on the Student Population

Before discussing issues with the assessment tool and the technology, one needs to first look at the demographics of the audience this type of online testing usually serves. In this paper, we

separate the student audience into three groups, high schools students, traditional college aged students and remote adult students that have spent at least a handful of years in the world of work. Rutgers University professor Donald McCabe surveyed 4,500 students at 25 high schools. "An astonishing 3 out of every 4 high school students admit to cheating on a test at least once in the past year." (Sohn) At the University of Virginia which grabbed headlines and triggered questions about the university's code of honor during a counter plagiarism campaign, it was found that 97% of the students did not cheat. (Gilgoff) The mature adult remote audience, which is already in the workforce, is far less prone to cheat than the younger audience described above.

Many of the security issues mentioned above have not been a problem with RIT's online learning courses. In fact, during the recent one year period (Winter 2000 quarter to Spring 2001 quarter), a total of 142 out of 589 online course instructors use proctoring, which means that only 24% of faculty teaching online courses use course proctors, which is the most secure of the three methods described in this paper. (Martin) Most of the courses in RIT's distance learning program use term papers, projects, and other alternative methods of testing to the traditional timed, closed book, multiple choice and true/false type exam. These alternative assessments are easier to administer than remote proctoring, and can be a more authentic form of assessment.

Security Issue with Online Testing

This topic is usually the biggest challenge with most instructors. In spite of the facts previously mentioned about the student population, many faculty don't seem to have a problem giving a take home exam but they are fearful of offering a test online. There are obviously trade offs with any kind of online system. The security issue should be focused more on the maturity and integrity of the students rather than on the testing method itself.

Remote Proctoring Testing Method

With this method, a remote proctor, (usually the student's supervisor, manager of an office area, library or other establishment) administers the exam and then collects it and submits this to the online learning office via FAX or US Mail. The exam is usually a traditional paper format where the student fills in the appropriate answers. The tests can be open book or closed book and a time limit can be established if appropriate. Upon receiving the test from the instructor, the online learning office then forwards the test to the proctor. The proctor has signed a contract with the online learning office agreeing to certain terms.

Remote proctoring seems to be the most secure of the three methods and allows the maximum flexibility for asking questions in a format that does not limit the teacher in a conceptual way. Instructors can design the tests in the manner they have been accustomed to with no limitations. The only thing to keep in mind is that students cannot ask for clarifications. Mistakes in the test cannot readily be clarified for the students, so tests should be as clear and as error free as possible.

One pitfall of this type of assessment could be the lack of integrity of a proctor. Another downside could be if remote students knew each other, took the test at different times and clued each other in on the content of the test. This could happen since this testing is asynchronous

(students take it at their convenience as arranged by the proctor). Experience with thousands of students at RIT has shown that none of these issues have been a problem.

Another pitfall from an instructor/student perspective is turnaround time. There is usually at least a one-week window between the time the instructor submits the test to the online learning office for dissemination to the time it is graded and returned to the student. Tests seem to trickle into the instructor's mailbox using this method. Grading has to be done manually one test as a time similar to traditional classroom exams. This slow turnaround time is often annoying to the remote student who is anxious for feedback.

Online Testing Through Multiple Choice and Fill in Questions

This is the most time efficient way of distributing, grading and providing feedback to the students but it is also the most limiting conceptually. Most of the work for this kind of testing is done up front developing the exam to fit the online format. An exam has to be designed in a format that fits properly on the computer monitor when viewed through a web browser. The author chose NTID's IDEA Tools designed by NTID developers to implement this, although there are several many other products on the market that would work fine for this task.

Multiple-choice types of questions are ideal for this type of system. (see Figure 1) There are also ways to incorporate fill-in types of questions, but they are much harder to grade and provide immediate student feedback using this method compared to multiple choice due to the automated nature of reporting the results. There are usually automatic grading instruments that will provide timely feedback to a student on how they did as soon as they complete the test or at a time designated by the instructor. Different weighting of questions is often difficult to achieve with these systems, so it is easiest to design an exam where each question is worth the same number of points.

Figure 1. Sample of Online Assessment Question for a Computer Programming Class

```
{
cout << "Please input your age. ";
cin >> age;
if (age >= 16)
   { cout << "You can drive!";
      cout << "if you are careful!"};
cout << "\nDriver program finished.";
getch();
}
```

○ if you are careful!
     Driver program finished.

◉ Driver program finished.

○ You can drive! if you are careful!
     Driver program finished.

○ no output would occur -> error in code

○ none of the above answers

Students enjoy the prompt feedback from this kind of assessment system, but this feedback may not incorporate a curve or a slight change that the instructor may have added after the test was posted. Again it is advisable to have the test as error free as possible once it is posted.

Another logistical item to pay attention to is that spacing sometimes changes once it is viewed on the web through a browser. The instructor could design a perfect, error free test in Microsoft Word, but once it is posted on the web and viewed from a browser it may look slightly different. Notice that in figure 1, the third choice with contains the wording "Driver program finished." Does not have a space in front of it. When testing students in a programming course, one space can often mean the difference between a correct or incorrect output in a computer program output.

Each question must also be limited in size so that the sample code and the output choices all fit on the monitor screen. If there are several multiple-choice answers for a particular section of programming code for example, the students need to be able to see the original code for each answer that is provided. In figure 1, for example, there are two lines of code that do not show at the top. The user would have to scroll up to see these, and would then not be able to see the last answer for this selection.

The online method is a great way to test students who are not able to take advantage of the proctor system, such as technicians who are out in the field, people who work swing shifts, or

independent consultants who are stuck in remote geographical areas. It is also a great way to combine and compare the results of on-campus students to remote students. During one quarter, traditional, on-campus, college aged deaf students were tested along with a group of 15 hearing Pittsburgh Telephone company technicians using the same online testing method. The deaf students took the test in their department's lab while being proctored by a lab assistant and the remote students took the test independently on their own. The results of the two groups were similar.

Downloading and Uploading Answers Using Electronic Conferencing Software

In this method, a test "template" is developed using a common software format such as Microsoft Word. This template is designed so that the students can simply type the answers to the questions directly into the template and then upload the exam to the conference drop box during the allotted time period.

This method is fairly secure for remote testing, because the electronic conference can detect which student downloaded the test, at exactly what time the test was downloaded and what time the test was submitted back to the conference. FirstClass™ was used for the electronic conference of choice.
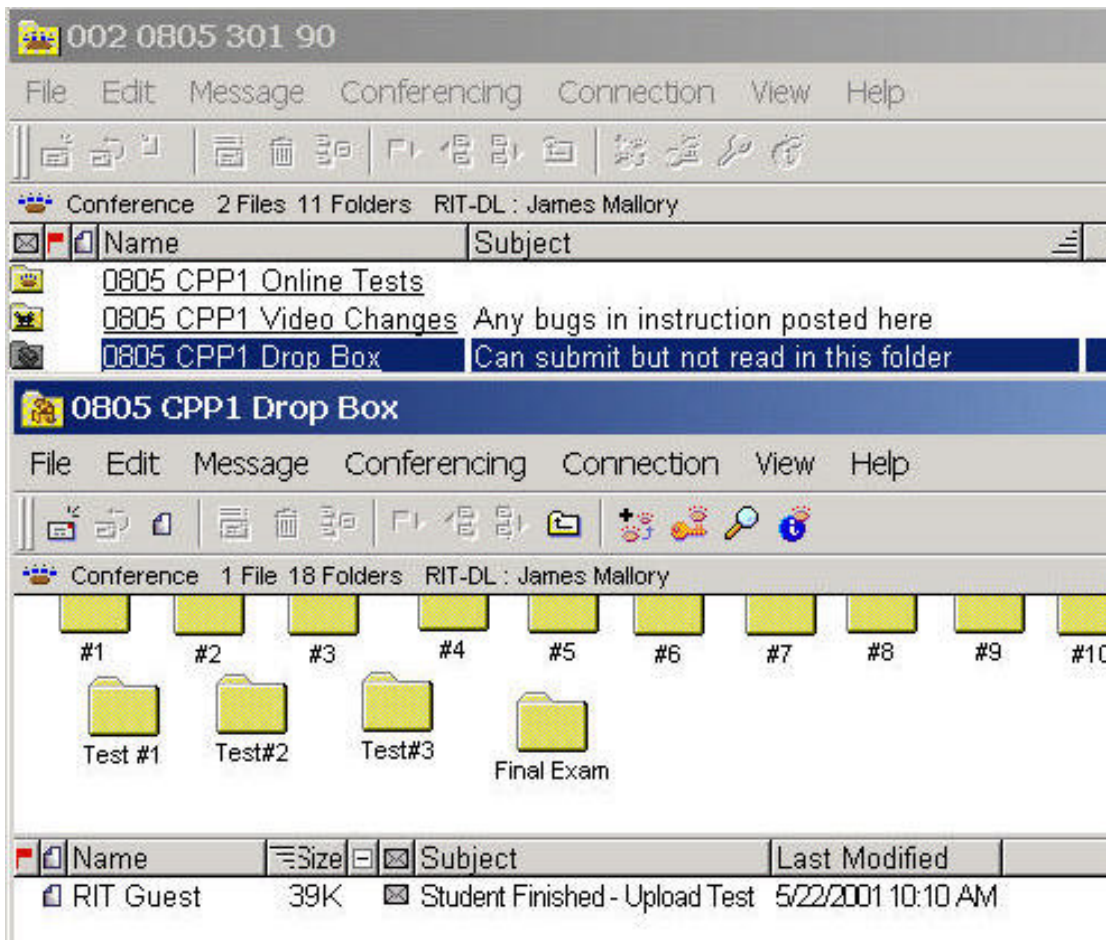
Figure 2. Example of a Posted Exam in FirstClass

Figure 2 shows the "History" feature from within FirstClass. This feature allows the instructor to track when the student downloads the test. In this example, the test was posted in the folder named "0805 CPP1 Online Tests". Figure 2 shows that the instructor, James Mallory, uploaded the test for all students in the class to be able to access at 6:47 AM on 5/22/2001. The student, "RIT Guest" saw that the test was available at 10:06 AM and downloaded the test at 10:06 AM and proceeded to fill in the answers. This student knew that he had a two-hour time limit to complete this test and upload it to the instructor's drop box.

Figure 3 shows that the RIT Guest student uploaded the completed test to the drop box at 10:10 AM on 5/22/2001. This completed test is actually an attachment to a message sent to the Drop Box. The drop box permissions are set so that only the instructor can read any of the information in the drop box.

Figure 3 Example of an exam submitted in a FirstClass Drop Box
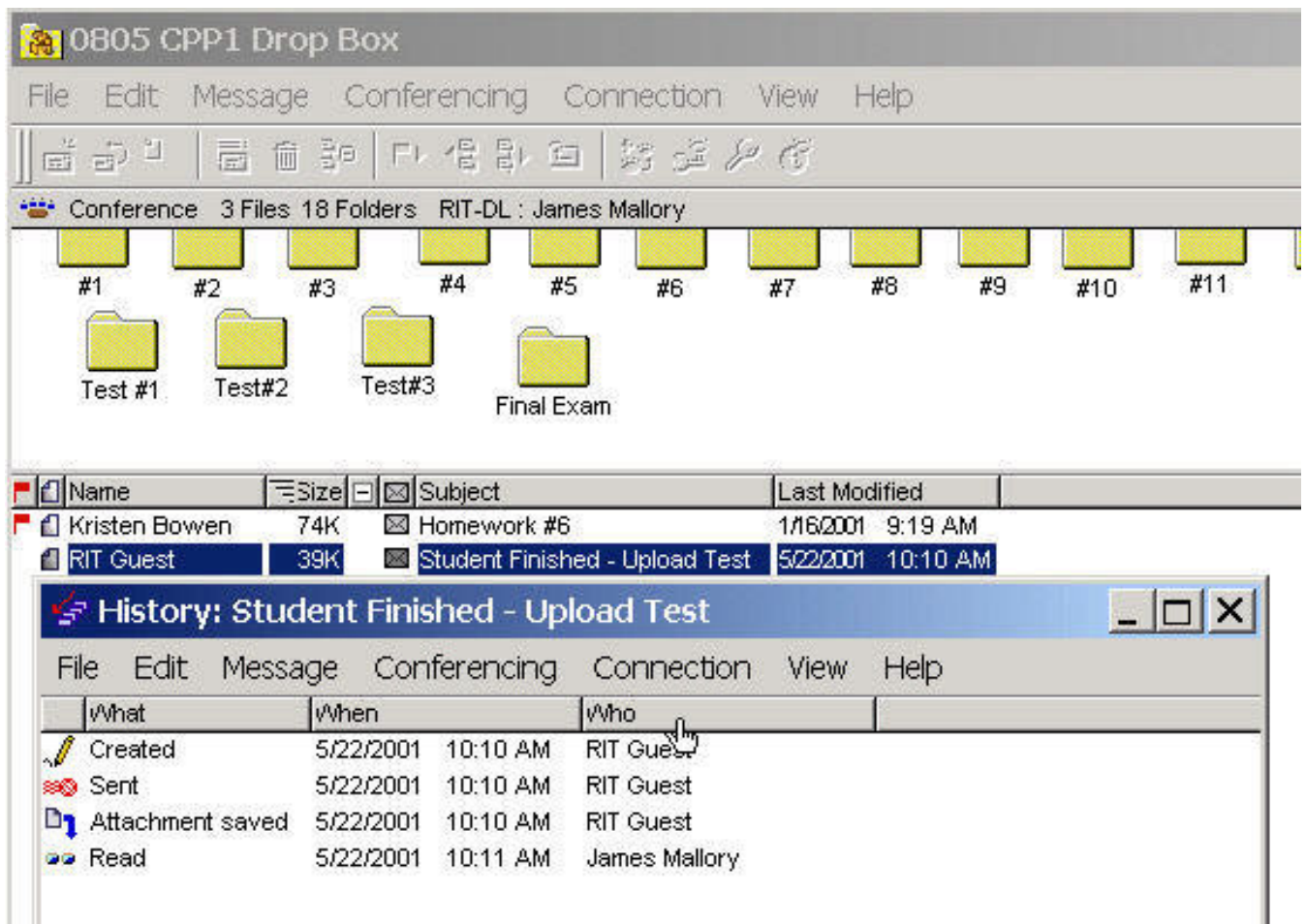
Summary

Online assessment has become the model of choice for most remote teaching faculty at RIT and elsewhere throughout the country. It is very slowly gaining acceptance into the traditional classroom, however, as a viable alternative to traditional teaching. There are always trade offs, of course, and there are still things that traditional campus test taking can do that remote testing cannot. For most purposes, however, online testing can accomplish many of the same desired results that traditional testing can. The application of these models depends to a large extent on the student population that it is serving and the comfort level of the instructor using this technology. A less mature audience or students with certain learning disabilities may not be a good fit for this type of assessment.

References

Eisner, E. W. (1999) The Uses and Limits of Performance Assessment. Phi Delta Kappan [Online]. Available: http://www.pdkintl.org/kappan/keis9905.htm

Gilgoff, Dan. (2001). Click on honorable college student. US News & World Report, Available: Vol. 130 Issue 20, p51

Mallory, J. R. & Laury, D. J. (2001). Using Desktop Video Conferencing with Deaf Students in Online Classrooms, Instructional Technology and Education of the Deaf Supporting Learners, [Online]. Available: http://www.rit.edu/~techsym/

Martin, M. J. (2001, May). Survey of online faculty. RIT Education Technology Center, Email: mailto:mjmetc@firstclass.rit.edu Message: Faculty Survey 992 to 003

Sohn, Emily. (2001) The Young and the Virtue less. US News & World Report, Available: Vol. 130 Issue 20, p51

Suskie, Linda. (2000) Fair Assessment Practices. AAHE Bulletin [Online]. Available: http://www.aahe.org/bulletin/may2.htm

Contact Information:

James R. Mallory
Department of Applied Computer Technology
National Technical Institute for the Deaf
Rochester Institute of Technology
Rochester, New York 14623-5604
Telephone (716) 475-2865
Fax: (716) 475-7101
E-mail: jrmnet@rit.edu