EXPERIENCES USING AN AUTOMATED TESTING AND LEARNING SYSTEM

Alyosha Pushak, Deb Carter, Teresa Wrzesniewski, Ramon Lawrence
University of British Columbia Okanagan
3333 University Way, Kelowna, BC, Canada
vysethelegend47@hotmail.com, deb.carter@ubc.ca, teresa.wrzesniewski@ubc.ca, ramon.lawrence@ubc.ca

ABSTRACT
Technology can improve teaching efficiency and delivery, but the most repetitive and time-consuming tasks of assignment, test preparation and marking are still mostly manual. Automated testing systems have the potential to dramatically reduce the amount of time spent on these repetitive tasks, freeing up instructor time for valuable student contact, and allowing for more uniformity on student and instructor achievement and evaluation. The challenge is determining if such systems are appropriate for a given course, the methods in which they will be used, and if the time and costs associated with their use are acceptable. In this paper, we describe our experiences on evaluating available automated testing systems, the reasons for selecting an in-house system, and initial results on its effectiveness. The results demonstrate that these systems can be very beneficial, especially for large science classes, and can be deployed at minimal cost. Insights for successful use are also presented.

KEY WORDS
automated testing, electronic assignments, evaluation, science, physics

1. Introduction

Technology has been deployed to assist education in a variety of ways including web-based learning and interaction, personal response systems ("clickers"), online communications, and electronic testing systems and resources. With the diversity of options and products available, it is challenging to determine which technologies are appropriate for certain teaching styles and courses, and what are the correct products and systems that provide the appropriate payback for student learning and institutional deployment.

In this work, we detail experiences in using an automated testing system for a large, introductory physics course. Unlike personal response systems, the system must be able to produce questions with randomized parameters, automatically mark questions that have numerical answers (not just multiple choice), and allow the instructor the ability to easily create questions and tests. These systems can replace traditional written assignments or practice questions that are marked by the instructor or teaching assistants. The motivation is that they provide a scalable mechanism for verifying that students are completing assignments, giving feedback to students while they are doing assignments, and efficiently assigning grades to students on work completed. To handle large class sizes, instructors often use workarounds to evaluate assignment and homework completion such as marking only a random subset of assignments, assigning fewer questions that are marked, or using significant teaching assistant time. Students expect individualized, personalized, on-demand, and real-time instruction and feedback.

An automated testing system allows every student to have their work evaluated and feedback provided while potentially saving significant time and costs that can be dedicated to student learning. This helps to realize the vision of student-centric, individualized learning where the instructor is no longer the evaluator (the system does that) but rather becomes the motivator and facilitator of student learning, which can occur at any time the student works with the system.

In this paper, we describe some of the automated testing systems evaluated, our reasons for building an in-house solution, and experiences in using the system. The contribution is a discussion on the benefits and issues with automated testing systems and recommendations for their successful introduction into the classroom.

2. Background

2.1 e-Learning

At its best, hybrid e-learning, a blended strategy of traditional classroom and electronic educational activities, provides an open system that respects learner diversity and motivates student achievement and engagement [1]. At its worst, hybrid e-learning provides a closed system that frustrates learners without prior domain knowledge or technological expertise and primes toxic events such as copying and submitting peers' answers when unable to successfully navigate an electronic system [12]. Thus, whenever adding new e-learning strategies such as an electronic assignment system, it is important to understand and ensure the learners' abilities, perceptions, and learning needs are met [9, 10, 11].

2.2 Automated Testing Systems

For this work, we define an automated testing system as a system capable of providing questions to learners and then automatically marking them. This general definition encompasses a range of technologies including electronic feedback systems ("clickers"), course content systems such as WebCT, Blackboard, and Moodle, and teaching and learning tools.

Most of these technologies are limited to *static* questions which do not change. Asking the same question again will be useless as the student knows the answer after the first time. There is no learning value in repetitive practice. The questions do not encourage or improve learning; they simply test if previous learning has occurred.

Our specific focus is on systems that can generate randomized versions of questions. A *question template* is a question that tests a specific concept that can be re-used by varying its contents. For instance, consider a physics question such as: "Calculate the distance travelled by a car in 2 hours that is travelling 60 km/h." There are versions of this question which can test the same concept by changing the numbers involved and the object described. A question template may be "Calculate the distance travelled by a x in y that is travelling z." The system can then use instructor supplied information on reasonable values of the variables x, y, z to generate an infinite number of questions of the same type. The system can also mark this question by having the instructor encode the solution as y*z while writing the question. Further, the instructor can encode common incorrect answers as well as explanations on why the solution is incorrect. Then when the student does the question and provides an incorrect answer, answer-specific feedback can be given to help direct the student to the correct answer or explain why their answer is incorrect.

The idea of question templates is not new. One of the first systems produced was WWWAssign built by Dr. Larry Martin for physics courses. Initially freely released, this system is no longer supported and has been replaced by the commercial product WebAssign [13]. WebAssign has content and questions from multiple textbook publishers.

There are other similar commercial systems most often affiliated with textbook publishers including Aplia [2] (mostly for social sciences), Mastering Physics [7], Gradiance [5] (computer science), and Carnege Learning (mathematics) [4]. The advantages of using these products are that the questions are completed and verified, there may be questions associated with the course textbook, and the web hosting is provided by the companies so no technical setup is required. The disadvantages are cost and inflexibility. The cost of these systems ranges from approximately $20 to $100 per student per course. Given the large enrolments for classes that typically use these systems, this is a cost most likely placed on the student rather than the institution. Although both parties see a benefit (students for learning, the institution for reduced costs), it is primarily supported by students, who are resistant to increased costs. Further, it is challenging to create new questions in these systems and their costs force instructors to use them significantly in a course. It is not cost-effective to use them for a few questions that require specific repetitive practice and learning or produce your own questions for use.

Another alternative is that most course management systems have built-in support for quizzes and tests including some such as WebCT Vista [3] and Moodle that have some support for randomized questions. If the institution has a course management system deployed, then simple questions can be done this way. We evaluated the support for randomized questions in WebCT Vista and found them reasonable for basic questions but not acceptable for more detailed questions.

Finally, there have been several university led efforts on building these systems and providing access. One such system is the Quest Service [8] at The University of Texas at Austin. This system has questions for many sciences and outside institutions can get access by requesting UT Austin accounts.

LON-CAPA [6] is an open source system built primarily at Michigan State University. It has hundreds of thousands of resources and questions that are shared by multiple institutions. Institutions are part of a sharing network for distributing questions and resources.

There has been some informal evidence that these systems have improved efficiency, especially for large first-year classes, and student learning and educational outcomes. Students enjoy the immediate response of the system and overall performance on tests and lectures have improved.

2.3 System Selection

Selecting an appropriate system depends on a very few key decisions. The most important decision is cost and who will bear that cost. If the instructor or institution deems that students should bear the costs of instructional systems similar to textbooks, then any one of the commercial products are suitable, especially if they have the textbook used in the course. This choice requires minimal instructor time and effort as all resources are prepared, and system details are managed. As long as students are actively engaged with the system throughout the semester, then students will generally appreciate the learning and grudgingly accept the additional costs.

If student costs are an issue (such as in our case), then the choices are much more limited: the Quest Service, LON-CAPA, or an in-house solution. The issue with requiring UT Austin accounts was a negative factor in using the Quest Service. We installed LON-CAPA and felt that it would be a suitable system for our needs. In the end, we decided to construct an in-house system for two reasons. First, part of the learning project was engaging computer science students building the system with users and instructors in physics. Constructing the system itself was a learning experience. Second, we wanted to construct a minimal platform to evaluate how long it takes to produce such a system, the challenges in system deployment, and also to allow for open question sharing and authoring. LON-CAPA has a restriction that questions cannot be released outside of the participating institutions (this makes sense so that students may not acquire the questions and answers). We would like the opportunity for anyone (including students) to have access to the questions. Some of the impacts of this decision are discussed in Section 4.2.

3. Technical System Details

The system constructed, called *AutoEd,* is designed to be easy to setup and use by both students and instructors. It differs from most of the commercial and open source products available in that it is written in Java and was designed from the beginning to support database storage of questions and answers. The database stores the questions, answers, and all user interactions. The web server code generates questions, marks student answers, and provides facilities for finding questions and building tests. Users and instructors interact with the system using any web browser. The system is built using open-source systems.

Students can:
1) Logon to the web site and see the courses with question content.
2) Practice questions on current course content with instructor-specified questions. The student will see an answer, an explanation, and marks when answers are submitted. A student can repeat questions on the same subject as desired.
3) Write tests or quizzes required by the instructor and receive grades.

Students get to practice the same questions that they will be tested on and get answers on demand.

The instructor can:
1) Provide students access to questions relevant to current content for practice.
2) Build tests and quizzes from questions and have the system mark them.
3) Perform reporting on student performance in the course and individual questions.

The big benefit for the instructor is reuse of resources. Generating a template once allows it to be reused in many situations by many people. It also allows comparisons to be made across courses and years. The system will store all student answers on every question.

The initial version of the system was constructed by an undergraduate honours student in computer science in approximately three months and continues to be improved. The short time for building a system shows that the technical challenge of system construction and deployment should not be the major cost in using a system and also illustrates how much money is appropriate to charge for using such a system.

The student view of the system is shown in Figure 1. Here the student is working through a physics assignment. Some questions have been completed and marked. The student can pick up an assignment at any time from where they left off.

From the instructor's perspective, they can see each student's progress on the assignment and overall student progress. Questions are built using a web interface. The questions are written in HTML with a special `<eqn>` tag used to indicate calculations to perform, random variables to generate, etc. Details on question authoring have been omitted for brevity.

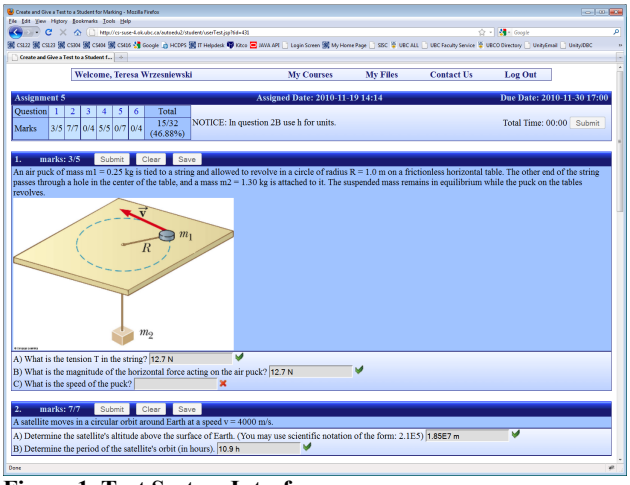

Figure 1. Test System Interface

4. Evaluation

4.1 Methodology

The electronic assignment system was made available to Physics 112 students in Fall 2010 for three assignments. Physics 112 is an introductory physics course mostly taken by students who are not interested in physics but require it for their degree. This course is already using clickers with great success [14]. The first two assignments in the course were traditional paper assignments. Assignments three and four were electronic using the *AutoEd* system. Students could decide to use paper or electronic submission for assignment number five. The questions for the electronic assignments were converted from traditional paper assignments. The only difference in marking is that a student could only get full marks for the question by having a correct answer in the electronic system. The system did not capture student work towards the answer which would be awarded part marks even if incorrect on a traditional paper assignment.

Two self-reporting paper surveys were distributed and collected on the first and final classes. The initial survey had 105 out of 111 distributed surveys completed (95.5% return) while the second survey had 76 out of 89 distributed surveys completed (85.4% return). The primary goal was to measure students' satisfaction and engagement with the *AutoEd* system. The final survey provided students an opportunity to provide comments.

4.2 Results

More than 90% of the students reported using the *AutoEd* for all three assignments and 92% used it for the last assignment where they had a choice. When asked which they preferred paper or electronic assignments, 70% chose electronic while an additional 15% chose both as equal. Over 72% of the students chose satisfactory or very satisfactory when asked about their experience using the *AutoEd* system while only 3% reported their experiences as very unsatisfactory. Almost half of the students (45%) reported that *AutoEd* helped their learning. Students also reported that *AutoEd* helped them complete their assignments (81%) and hand their assignments in on time (84%). Over 92% of the students thought the *AutoEd* system should be used in other Physics 112 classes and 87% of the students felt that it could be used in other classes. Suggested classes included chemistry, biology, mathematics, and English.

The most important benefits cited by students were the immediate feedback which aided learning, the randomization forced them to learn concepts rather than memorizing, and the system was easy to use as they could work anytime and come back where they left off.

Student feedback from the final survey provided some interesting insights. First, a common complaint was the issue of incorrect marking, especially with significant figures, for some answers. This issue was both a technical issue that was resolved during the course and an issue with proper question construction and validation. A more interesting comment (although given by only 5% of the students) was a request to block the ability to search Google or other Internet sites for the correct answer.

This is interesting at two levels. First, students always have the ability to search the Internet for answers whether the questions are given electronically or on paper. However, an electronic question makes it easier to just copy-and-paste the question into a search engine. The larger issue this raises is how well these systems can evaluate student learning as questions get released to the public. Even if questions are not released out of the system (such as in LON-CAPA) there is enough resources on the web to find answers. This may allow students to complete questions without learning just as with other forms of assignments. This should motivate instructors to treat this technology as a learning tool rather than simply an evaluation tool. The ability to cheat the system will increase as these systems are used more extensively.

4.3 Discussion

Overall, the pilot project of the *AutoEd* system was a full success in terms of student satisfaction and engagement. With any new technology, issues are to be expected. One of the key recommendations is to exhaustively test questions and answers. Unlike paper assignments where the answer key does not have to be perfect or can be developed during the assignment, an electronic question needs to be fully completed with answers and tested before it is first released. Several issues arose during the assignments because the questions were not tested and ready early enough in the process. Due to time constraints, the system did not encode many incorrect answers which would have greatly improved student feedback and learning.

The second insight is that question authoring does take time. In the evaluation, we took existing paper assignments produced by the instructor, modified them to add randomization, and had a computer science student implement them in the system. This worked quite well as assignments could be quickly added to the system without extensive instructor time.

There is a continual issue on how much time should be given for assignments. The automated testing system allows us to track when students first start an assignment and when they are completely finished. The chart in Figure 2 shows what instructors often suspect: students generally only start an assignment during the last day or two even for a two week assignment. In the chart, you

can see the day the assignment was assigned (November 3rd), and its due date at midnight on November 15th. A significant majority of the students did not start the assignment until the last two days and most completed it during the last day. The tendency to procrastinate and the ability to easily collect and distribute assignments electronically should encourage instructors to assign smaller assignments due in a few days rather than larger assignments. The barrier of collecting and grading assignments is removed with an automated testing system.

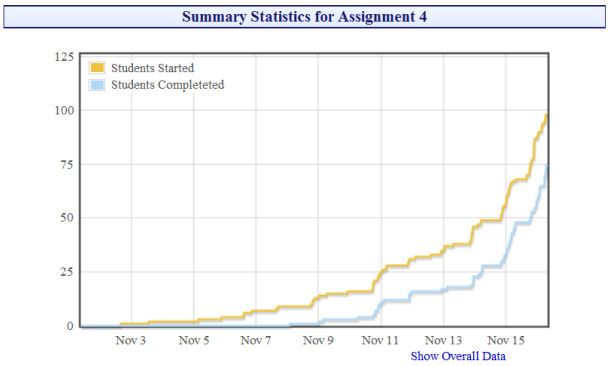

Figure 2. Student Assignment Progress

Finally, it is possible to use open-source systems or build an institution specific one with minimal effort. Although the upfront costs and time is higher, the payback in terms of student costs and instructor flexibility is significant. We are currently evaluating the system in several other courses and will release it as open source to the community.

5. Conclusion

Automated testing systems are another effective technology instructors can use to educate and engage students. Selecting between commercial and free products depends on the willingness for the institution and instructors to have students pay the substantial costs. We have demonstrated that it is possible to build a system at minimal cost and use it effectively in a large first year class. Students appreciate the immediate feedback, especially when feedback on how to arrive at the correct answer is provided. Two major issues must be considered when deploying these systems. First, they should be actively integrated into the instruction, not just used as evaluation systems. Second, instructors should be aware of the potential for students to easily find answers on the Internet and potentially adjust their grading accordingly. Our future work is to continue to improve the *AutoEd* system by producing an open source release for widespread distribution and producing an open repository of questions to be shared across institutions.

Acknowledgements

The authors appreciate contribution of PHYS 112 students who evaluated the system. We also acknowledge the support of UBC Okanagan who funded the research.

References

[1] Ahmed, H.M.S. (2010). Hybrid e-learning acceptance model: Learner perceptions. Decision Sciences Journal of Innovative Education, 8(2), 313-346.
[2] Aplia www.aplia.com
[3] Blackboard. http://www.blackboard.com
[4] Carnege Learning. http://www.carnegielearning.com/
[5] Gradiance. http://www.gradiance.com/
[6] LON-CAPA http://www.lon-capa.org/
[7] Mastering Physics. http://www.masteringphysics.com/
[8] Quest. https://quest.cns.utexas.edu/
[9] Pence, H. E. (2007). Preparing for the Web Generation. Journal of Educational Technology Systems, 35 (3), 347-356.
[10] Rovai, A. P. (2007). Facilitation online discussions effectively. Internet and Higher Education, 10, 77-88.
[11] Ruth, S. R., Sammons, M., & Poulin, L. (2007). E-learning at a crossroads: What price quality? Educause Quarterly, 30 (2), 33-39.
[12] Siemens, G. (2006). Knowing Knowledge (e-book). Retrieved Nov. 30, 2008 at www.knowingknowledge.com
[13] WebAssign. http://www.webassign.net
[14] Wrzesniewski, T. (2008). Use of PRS Clickers and Computer Animations in Large Introductory Physics Classes. Proceedings of 11th IASTED International Conference: "Computers and Advanced Technology in Education" (CATE 2008), 436-439.