Online Computer Scoring of Constructed-Response Questions

Richard Vigilante *
Jesuit Distance Education Network
District of Columbia, U.S.A.

Abstract

The time faculty spend teaching online courses can be considerable. The discussion/collaboration course format of New York University's Virtual College online program requires extensive, daily instructor discussion, monitoring, and feedback--an average of 28 hours per week for a typical six-week course. To reduce some of the 25 percent of online faculty time currently spent on grading written assignments and examinations, NYU and the Educational Testing Service (ETS) will develop automated scoring tools using natural language processing (NLP) technology to evaluate constructed-response questions in NYU online courses. Constructed-response questions and answers will largely come from electronic versions of the textbook package adopted for each online course. Testing of the scoring heuristics will be accomplished by applying the automated scoring process to questions completed by students in their courses. Computer-generated ratings for all questions will be compared with ratings provided by course.

Introduction

Properly designed and delivered, online courses provide students with the same level of dynamic, hands-on instruction that characterizes the best on-campus course, laboratory and faculty access. Since 1992, New York University's Virtual College program has been offering such a high-quality graduate program in information systems. During each of NYU's six-week online courses, students and faculty collaborate to analyze, design, and build case study information systems. Students are divided into groups of 4-5 participants to work concurrently on various phases of systems projects. Functioning as members of their virtual project teams, the students establish discussion guidelines, critique and edit each other's work, manage online workplace responsibilities, and at times run the asynchronous Lotus LearningSpace hosting software as if it were an online chat service (Aranda and Vigilante, 1995).

During a typical online course, students generate over 1,000 documents related specifically to the curriculum. This is an average of 50 discussions, analyses, questions, and assignments per student--a level of participation that would be rare in most on-campus courses over a similar time

period. Computer conferencing and electronic mail provide continuous faculty access to answer questions, evaluate assignments and examinations, and provide advisement.

The courses' collaborative, computer conferencing-based delivery modality provide a highly-effective instructional experience, but at the cost of operational efficiency.   The program's seminar format limits class size to around 20 students (to control redundant student participation), and requires extensive, daily instructor discussion, monitoring, and feedback.

Online Faculty Time Requirements

The time faculty spend on online courses can be considerable. Reducing faculty workloads through appropriate automation, while maintaining instructional quality, has been a goal of the Virtual College program. To quantify how online faculty spend their time, instructors are asked to maintain daily activity logs throughout each of their online courses. For each day that the course is in session, faculty indicate to the nearest quarter-hour the time devoted daily to the following course activities:

Maintain Course Syllabus. Update course announcements and modify syllabi information as necessary

Prepare Assignments and Study Questions. Preparing or revising individual and group assignments and study questions

Prepare Course Lectures. Preparing presentation material (lectures, topics, case studies) comparable in purpose to in-class lectures

Lead Discussions.   Creating and sustaining (both asynchronously and synchronously) student discussions and questions about the lectures, topics, etc.

Monitor Groups. Monitoring and responding as necessary to group discussions, activities and overall group or project progress

Answer Questions. Responding to student questions about individual or group assignments through individual e-mail, broadcast e-mail, announcements, threaded discussions, and message boards.

Grade Assignments. Grading written assignments and examination questions.

Maintain Online Resources. Maintaining and updating lesson links, lesson downloads, and resources. Adding, deleting and editing links to other Web sites. Uploading documents to server for student use, monitoring uploaded student documents, and advising students about downloaded documents and their use.

Provide Advisement. Responding to student questions about their academic programs and progress; responding to student inquiries about career or professional goals or requirements; and searching for and providing students with supplementary readings and related instructional materials

Other Activities. Responding to all other course-related activities, including hardware, software and network matters

The Virtual College discussion/collaboration course format necessitated significant faculty time commitments--an average of 28 hours per week (equivalent to 12 hours per week for a standard 14-week course). Table 1 shows the distribution of the 28 hours faculty spent on average each week delivering their online courses (Vigilante, 1999). Faculty time estimates included all information searching, analyzing and synthesizing, as well as the mechanics of keyboarding, table and graphics preparation.

TABLE 1. Faculty Time Spent on Telecourse Activities

Faculty Activity	Percent of Time
Preparing and Responding to Discussions	22%
Providing Additional Information	16%
Preparing Assignments	13%
Responding to Assignment Questions	15%
Grading Written Assignments and Exams	25%
Other	9%
TOTAL	**100%**

Automated Scoring Project

To reducing some of the 25 percent of online faculty time currently spent on grading written assignments and examinations, NYU has contracted with the Educational Testing Service (ETS) to adopt and utilize some of the automated scoring approaches developed at ETS over the past decade. These approaches have been operationalized in several content domains including computer science, architectural design, mathematics, and business (Burstein, Braden-Harder, Chodorow, Hua, Kaplan, Kukich, Lu, Nolan, Rock, and Wolff, 1998). Since February 1999, the ETS *e-rater* system has replaced one of the human graders for the essay part of the Graduate Management Admission Test.

During the project, ETS will develop automated scoring tools using natural language processing (NLP) technology to grade constructed-response questions in NYU online courses. These tools should help reduce a portion of online faculty time currently spent on reading and grading free-response questions and assignments. This project will test automated scoring in four online graduate courses--Information Technology, Database Management, Network Administration, and Information Security--to be offered during the fall 1999 and spring 2000 semesters. The constructed-response question types to be scored will be similar for all courses.

During the project, ETS staff will design knowledge-based (as opposed to *e-rater's* corpus-based) heuristics and algorithms for scoring the constructed-response questions for use in four graduate information systems courses--information technology, database management, network administration, and information security. Development of the heuristics will explore the use of both student-generated responses to questions and electronic versions of course textbooks and curriculum materials. Testing of the scoring heuristics will be accomplished by applying the automated scoring process to short-answer questions completed by students in their courses. Computer-generated ratings for all questions will be compared with ratings provided by course faculty (Leacock and Kukich, 1999).

The project will design and test a scoring model for constructed-response items that will initially produce a pass/fail rating. Later versions of the scoring model may produce multi-level ratings (e.g., excellent, good, fair, and poor). Project design issues will include:

- Develop a scoring prototype based on the criteria specified in a set of scoring guidelines developed by faculty and curriculum developers

- Use a content-based approach to determine a rating for a response

- Develop a prototype design that is generalizable to constructed-response type questions that have the same structure but different content

The online testing will be supported by the Assessment Manager feature of the LearningSpace course delivery software used by the Virtual College program. Assessment Manager enables the instructor to privately test, review, grade, survey, and give feedback on participant performance. Instructors can use the Assessment Manager's question bank repository to create quizzes, exams, surveys, or self-assessments.   Assessments completed by the students are returned back to the Assessment Manager for automatic grading, summarization and private review by the instructor. Graded assessments are then privately posted to the students.

The testing features of the Assessment Manager include: (1) ability to randomize questions so each student gets a unique assessment; (2) choice of true/false, multiple choice, and constructed-response question types; (3) ability to track assessment time and prevent students from taking multiple assessments; (4) ability to import questions from test banks; (5) ability to return corrected assessments to student portfolios; (6) ability to auto-grade and record assessments; and (7) ability to import and export assessment tools.

Sample Constructed-Response Question

To the maximum extent possible, constructed-response questions for the project will come from electronic sources associated with the textbook package adopted for each online course. Based on these criteria, ETS staff will design knowledge-based (as opposed to corpus-based) heuristics/algorithms for scoring specified short-answer questions.   Development of the heuristics will explore the use of both student-generated responses to questions and electronic versions of course textbooks and curriculum materials. Testing of the scoring heuristics will be accomplished by applying the automated scoring process to questions completed by students in

their courses. Computer-generated ratings for all questions will be compared with ratings provided by course faculty.

Constructed-response questions to be used in the initial Information Technology course will come from the Whitten and Bentley *Systems Analysis and Design Methods (4th Ed.)* textbook published by McGraw-Hill (Whitten and Bentley, 1998). The *Instructor's Guide* accompanying the textbook provides standard answers against which student responses will be measured. The following is a representative example of the question type to be used in the Information Technology course:

Question: Differentiate between upper-CASE and lower-CASE.

Answer: Upper-CASE describes tools that *automate or support* the "upper" or *earliest phases of systems development*--the *survey, study, definition,* and *logical design* phases. Lower-CASE describes tools that *automate or support* the "lower" or *later phases of systems development*--the *physical design, construction, implementation,* and *support* phases (Dittman, 1998).

A computer file version of the textbook will be used to create a thesaurus-like representation of the text's content. This will be used to recognize when a student uses a synonym of key words or phrases (shown in italics above) in the standard textbook answer. Some of these synonyms will be found in the section of the text that contains the answer; others will be found elsewhere in the text. Students can often use the vocabulary of a field without knowing the answer to a test question. Having the full text will enable development of a mechanism for detecting answers that, although framed within the general vocabulary of the course, do not address the question.

Conclusion

ETS will prepare a report analyzing the effectiveness of the computer-based scoring prototypes used in the two fall 1999 courses. The report will provide recommendations for necessary enhancements to the scoring prototypes to increase their reliability, to be incorporated into the two spring 2000 courses. A second and final report providing a summative evaluation of the overall project will be issued in mid-2000.

Continued work on this concept will ultimately develop methods and tools that will streamline and at least partially automate the process of designing content-based NLP scoring systems for free-response tasks. The objective will be to provide curriculum developers with the capability to define and input parameters for scoring system design so that new scoring applications can be designed with minimal involvement of research staff. An additional goal will be to develop a method to validate automated ratings of text that does not rely on correspondence with faculty ratings. Such an internal validation method will permit the automated rating procedure to evaluate whether or not the rating of a response meets specified requirements for reportability.

References

Aranda, R. and Vigilante, R. (1995, July). Re-engineering Management Training with Networked Multimedia. *Telecommunications*, 45-47.

Burstein, J., Braden-Harder, L., Chodorow, M., Hua, S., Kaplan, B., Kukich, K., Lu, C., Nolan, J., Rock, D., and Wolff, S. (1998). *Computer Analysis of Essay Content for Automated Score Prediction* (RR-98-15). Princeton, NJ: Educational Testing Service.

Dittman, K. (1998). *Instructor's Guide to accompany Systems Analysis and Design Methods* (4th Ed.). New York, NY: McGraw-Hill.

Leacock, C. and Kukich, K. (1999). Proposed ETS Automated Scoring Pilot Project with New York University. Princeton, NJ: Educational Testing Service.

Vigilante, R. (1999). Online Collaborating in the Virtual College. In D. G. Brown (Ed), *Vignettes from America's Most Wired Campuses.* Boston, MA: Anker.

Whitten, J. L. and Bentley, L. D. (1998). *Systems Analysis and Design Methods (4th Ed).* New York, NY: McGraw-Hill.

Dr. Richard Vigilante is Executive Director of the Jesuit Distance Education Network. He can be reached at One Dupont Circle, Suite 405, Washington, DC 20036. Email: vigilante@ajcunet.edu, Phone: (202) 862-9893, Fax: (202) 862-8523.