ASSESSING STUDENT LEARNING IN FIRST-YEAR QUANTITATIVE COURSES AT BABSON COLLEGE: EXPERIMENTAL DESIGN

RYBOLT, William H. & MCKENZIE, Jr, John D.
Babson College
USA

From 2001-2006 we used a number of approaches to assess how well our first-year students learn statistics and mathematics when introduced to different teaching methods. The topics introduced in their two courses include those found in a standard applied probability and statistics course. For example, descriptive linear regression. Most of these assessments have been based upon analyses of opinions and examination results from the students. This paper reports on designing experiments to determine whether electronic quizzes enhance student learning. A second paper presents the implementation of these experiments and a preliminary analysis of the data from these experiments.

INTRODUCTION

During the past five years we have explored a number of techniques for assessing different methods of teaching in our first-year quantitative courses at Babson College. The emphasis in the first course is on calculus, linear programming, and mathematics of finance, while the emphasis in the second course is probability and statistics. This paper reports on our most recent effort to assess student learning using statistics. The basic premises are that the technique used should be built into the course and not regarded by either the instructor or students as an add-on, that it should be frequent, and that it should focus on fundamental skills and knowledge and not try to deceive the students.

Our principal assessment design uses instructors teaching two different sections of the same course. One instructor gives 10-minute quizzes at the end of each class based upon the material covered in that class. Another instructor gives the quizzes once a week based upon the material covered the previous week. Each of the paired sections is given a different type of quiz, either an electronic version hosted in EDU, or a paper version. The electronic quizzes are graded and returned immediately, while the paper quizzes are graded and returned in the next class. Because most of the electronic quizzes are generated with random data, each student receives a different version of the quiz. After a student takes a quiz once by himself or herself, he or she is encouraged to give or receive help from his or her fellow students. Thus the quizzes generate interactions among the students and cause the learning to become a cooperative experience. Students are allowed to take the electronic quizzes from one to three times.

During the semester we employed four different schemes for grading: best of three attempts, average of three attempts, best of three attempts with a one-point penalty for each retake, and single attempt. On each of the two midsemester exams and the final exam, we have several questions that are associated with a particular quiz type. By comparing the performance on a particular exam question with the type of quiz our goal is to determine if there is a measurable benefit to using a particular type of quiz. If different types of quizzes yield similar results, then we employ other criteria to determine the type we want to use for assessment in the future: time factors, convenience, or preference on the part of either the instructor and/or the students. At the present we only have the data for the fall semester of 2006. We will not have data for the spring semester until May of 2007. Determining the optimum type of quiz is only the beginning of our assessment plans. In the future we plan to investigate other course-related issues such as PowerPoint versus non-PowerPoint lectures and pre-class versus post-class quizzes using the same design.

We have made several exploratory attempts at using electronic quizzes to begin to answer these questions. Each attempt has generated issues that need to be addressed before we undertake more complete studies in these areas, but that is the function of this exploratory study. The next paper focuses on implementation issues associated with this type of assessment effort and a preliminary analysis of the generated data.

Babson College only offers undergraduate and master's degrees in business. It is located in Babson Park, Massachusetts, just outside of Boston. There is an undergraduate international population of approximately 1700 students. Roughly 40 percent of the students are female. When undergraduate students come to Babson, they are issued a laptop for their use during their years at Babson. The laptop has Windows XP and Office 2003 as the standard software. In their first semester, most students take an introduction to quantitative methods course. Excel 2003 is the most common software package used in this course. In addition to optimization and financial topics, this course involves some statistics. During their spring semester of 2007 most students take an introductory applied probability and statistics course which uses Minitab. The maximum size of each section of these classes is 30 students.

The classrooms in which we teach are networked and have one or two digital projection units. A few classrooms have both wireless and wired Internet access, but most have only one type of access. Students are encouraged to bring their laptops to class and have access to the Internet during class. They are able to download data sets, upload files, and access electronic quizzes during class.

ELECTRONIC QUIZZES

In the late nineties we began to experiment with web-based electronic quizzes using Java script. Five years ago one of us received a grant from the Davis Educational Foundation to develop electronic material and techniques to enhance and assess student learning in our introduction to quantitative methods course. One of the techniques that we investigated for both enhancing and assessing student learning was the use of electronic quizzes. These quizzes were also used for homework and for review. Our quizzes were implemented in a package called EDU. EDU is very similar to a package developed by Maplesoft, Maple TA. Our early efforts were devoted primarily to developing quizzes that we thought were reflective of the subject material in that course. Since then we have expand these quizzes to include the first-year introductory applied probability and statistics course.

Our quizzes are designed to take approximately 10 minutes. They are used both in and out of class. Most quiz questions are randomly generated so that a student can take a quiz multiple times. Each time they receive a slightly different quiz. Because students receive quizzes that are slightly different from one another, the answers to each quiz are slightly different. It is not possible to copy casually the answers from another student.

One of the benefits of this type of quiz is that we can encourage students to help each other with the concepts behind the quiz without being tempted to copy answers. The purpose of this experiment was to explore alternative methods for using the quizzes and assess objectively the impact of alternative approaches. Initial work was done under the auspices of an earlier Davis Educational Foundation grant received in 2004.

When assessing the merits of the electronic quizzes, we had three goals in mind. First, we wanted to demonstrate that the use of such quizzes did not cause any hindrance to learning. A frequent criticism of assessment techniques is that they take time away from learning and are counterproductive. Second, our goal was to create electronic quizzes that that would take no more time than their paper equivalent while being more convenient. Third, we hoped to demonstrate that these quizzes, while enabling assessment of the learning process also enhance the learning process. This enhancement potentially comes from several sources: active participation, immediate feedback with the correct answers, and repetition.

Our quizzes are used in a variety of ways by different instructors at Babson College. Some prefer daily quizzes, others use them only occasionally, and finally some not at all. Some instructors prefer to give paper quizzes lasting 20 to 30 minutes; others prefer electronic quizzes lasting five or 10 minutes. The goal of our work in the fall semester of 2006 was to evaluate and to measure the effectiveness of the use of these quizzes in our two first-year quantitative courses.

Specifically, we wanted to determine whether there is a significant difference in student learning between using electronic and paper quizzes. We also wanted to explore the impact of quiz repetition when the students are required to take the quizzes more than once. Finally, we wanted to explore the consequences of different quiz grading schemes.

We wanted to investigate the impact of the following grading schemes: best of three attempts, average of three attempts, best of three attempts with a one-point penalty for each retake, and single attempt. Hence the purpose of our experiments was to employ quizzes in different manners to measure the impact of these alternative methods on student learning.

EXPERIMENTAL DESIGN

When we designed our experiments for the fall semester of 2006, we explored several alternatives. Our original plan was to obtain demographic data from admissions for prior classes, construct a model to predict the student performance in our first-year quantitative courses, and randomly assign the students to three treatments. There would be three treatments: the students in one class would take no quizzes, in the second class students would take only electronic quizzes, and the in the third class students would take only paper quizzes. This plan was not implemented for several reasons. We could not get timely access to the demographic data and were not able to influence the assignment of students to specific sections.

Our actual design was based upon a crossover design. We divided the subject content of the each course into blocks. The aim was to construct each block so that the student competence in one block was different from that in other blocks. For example, in the probability and statistics course one block was probability and discrete probability distributions; a second block was normal distributions. After separating the material in the courses into blocks, we were able to assign different treatments to each class in each block by employing a different type of quiz for each section.

To measure the effectiveness of a quiz on learning a given block of material, we used performance on exam questions. Each course had two midsemester exams and one final exam. Each exam had roughly 10 questions, each worth approximately 10 points. For each exam, we included two to four questions on different blocks of material. The plan was to match the performance on each of the selected questions with the corresponding quiz. The goal was to identify types of quizzes or grading schemes that yielded better performance on the corresponding exam question.

We used a crossover design so that each class would receive the same number of each type of treatment. For example, students in one section would take an electronic quiz on discrete probability distributions, and the next time that section would take a paper quiz on normal distributions. An alternative design would be to select one section and give no electronic quizzes and select another section and give only electronic quizzes. We decided that this was impractical because of potential student complaints from those taking and those not taking the electronic quizzes. There was also concern about the impact on end-of-the-semester student opinion surveys which have a major impact on the perception of faculty teaching by the administration.

Our primary design used paired sections. We had hoped to have multiple sections with different instructors, but we only had four sections taught by two instructors. One instructor taught two sections of the introductory applied probability and statistics course on the same days. Unfortunately these sections were in very different rooms and at different times. The other instructor taught two sections of the introduction to quantitative methods course. These were held in the same room during adjacent time slots. When any given block of material was being covered, students in one section would take electronic quizzes and the students in the other section would take paper quizzes. When the course went to the next block of material the quiz types would be reversed. The quiz types were balanced so that at the end of the semester students were exposed to equal numbers of both types of quizzes. Using different types of quizzes was the basis of one hypothesis we investigated. Another hypothesis we investigated was the impact of the frequency of quizzes between each pair of sections.

We also had a single section of the introduction to quantitative methods course. In this course we explored the impact of the different grading schemes, mentioned above, on student performance.

Our primary goal was to evaluate alternative ways of administering the quizzes. First, we hoped to determine whether there was a quantitative difference in learning associated with the different quiz types. The idea was to have questions on an exam that were associated with a given

block of material. We wanted to compare the performance of students on that block of material who had taken electronic quizzes with those who had not taken electronic quizzes.

There was a need to compensate through the use of covariates for possible differences of ability between the students in each of the two sections. Unfortunately although the design seems valid, we were concerned that the effect we were trying to measure might not be detected due to the constant switching. That is, students might take two different types of quizzes during the same week. It might take them a while to become familiar with the different types of quizzes.

In addition to using the grades on the quizzes and three exams and covariates to compare the effectiveness of different quiz techniques, we used results from three additional surveys. At the beginning of each course, we had the students take a data collection survey in which we obtained demographic data about the students. We solicited data about such items as SAT scores and gender. At the end of the course, we administrated two surveys to solicit their opinions on various aspects of the course. One survey focused on their general experience with and their attitude towards the course. The second survey asked about their experience with and attitudes toward the electronic quizzes. We also gathered data on how many times each student visited Babson's Math Resource Center to obtain extra help.

The data that we collected were as organized as follows. The survey data were collected on a class website and entered into an Excel spreadsheet. All data for each student were placed on one line of a spreadsheet. For each student, we had their responses to the three surveys, their score on each quiz question, an index variable indicating the type of each quiz they took, and their performance on each of the questions from midsemester exam 1, midsemester exam 2, and the final exam. All quiz and exam questions that were part of the experiment were graded on a 10-point scale. Because we have more data than we could report in this paper, we selected a subset of the data to analyze for the purpose of investigating the effectiveness of the electronic quizzes.

ELECTRONIC VERSUS PAPER QUIZ EXPERIMENTS

We began each semester with one or two practice electronic quizzes to familiarize students with that environment. In the introductory applied probability and statistics course, we focused on the following question. Is there a significant different in the effectiveness on learning between electronic versus paper quizzes? We had two sections taught by the same instructor. These sections were taught twice a week on Tuesdays and Thursdays. In this course the quizzes were given at the end of each class. The quizzes were designed to take roughly 10 minutes. Each quiz was based upon the material covered in that class. The quizzes were given during the last 10 minutes of each class. Because there is a 30 minute break between classes, students were able to take an extra 10 to 20 minutes on each quiz if they wished. The first time a student took a quiz, they took it by himself or herself. If they did not understand the material, they were able to seek help before retaking the quiz. They were required to retake the quiz two more times before 8 a.m. the next morning.

After the students took the quiz the first time, they were encouraged to interact with other students. Because only one of the three attempts was individual, we though it best to use the average of the three tries as the score on the quiz. The purpose of the interaction was to enable students to learn from each other. Because the questions were randomly generated, it was not possible to copy casually another student's answer. It is necessary to work the answers to each quiz individually. The idea is to transform learning into a more cooperative experience. Thus, if a student needs help with the quiz, he or she could recruit help from a classmate, or if a student understands the material then he or she can volunteer help to a classmate. The hope was that by making the quizzes more of a cooperative experience, either inside or outside of the classroom, this would enhance student learning on the quizzes. By encouraging repetition on the quizzes it was hoped that students would better learn and retain the material.

The paper quizzes were given once at the end of each class and collected and returned the following class. There was no opportunity to retake the paper quizzes. The experiment was designed in such a manner that each person took the same number of paper and electronic quizzes. The correct answers to the electronic quizzes were immediately available after a student submitted his answers. It was expected that students would compare their answers with the correct answers

and learn what they did wrong. Thus they could correct their mistakes before they took the quiz a second and third time.

Two midsemester exams and a final exam were given in this course. Each quiz was graded on a 10-point scale. Each exam had 10 questions, each one graded on a 10-point scale. The design was to pair certain questions on each exam with the quiz question designed to help learn that material. Thus we could analyze the relationship between the performance on a particular exam question and the student performance on the quiz question. The goal was to determine which best facilitated learning, electronic quizzes or paper quizzes. Could we demonstrate a difference in performance for a given block of material between those who had taken electronic quizzes and those who had taken a paper quiz?

We had two sections of the introduction to quantitative methods course taught by the same instructor. The goal of the experiment in the two paired sections was the same as in the two probability and statistics sections but using different types of material and less frequent quizzes. Once a week when a quiz was given, it was administered it was given at the end of class and the students in one section took an electronic quiz two more times before 8:00 a.m. the next day. The quizzes were designed to require five to 10 minutes, but students were given 10 minutes during class and up to 20 minutes after class to complete. Again the course was divided into blocks. On each block one section was given an electronic quiz and the other and a paper quiz. The major difference to the other paired section experiment was that the quizzes were not given daily. Instead they were given once a week during non-exam weeks.

To analyze the results, selected questions on the final exam and midsemester exams were paired with the question on the quizzes that most closely covered the same topic. Again the performance on a selected exam question was calculated. When students took quizzes only once a week, did students who had taken an electronic quiz on that topic perform differently than those who took a paper quiz?

In the other section of the introduction to quantitative methods course, we gave daily quizzes. But this time the quizzes were divided into groups with different grading algorithms. Four schemes were used to determine the quiz score. Take a quiz: multiple times with a one-point penalty for each take, three times and chose best score, three times to get the average, and only once and count that score. Again the analysis was similar, but the hypothesis investigated was the impact of the grading scheme on performance. Could one show that one grading scheme was better than others?

There were a number of different lessons learned from administering the quizzes based upon this experimental design. These are discussed in the next paper.

This work has been supported by a grant from the Davis Educational Foundation.

REFERENCES

Aieta, J., & Rybolt W. (1997). Using the Web for Courseware and Assessment. Paper presented at 10th Annual International Conference on Technology in Collegiate Mathematics. Chicago, IL: Addison-Wesley & Prentice Hall.

Aieta, J., & Rybolt, W. (1998). Enhancing a Statistics Course by Using the Internet to Collect Student Responses from Customized Surveys: Experience, Lessons, and Recommendations. Paper presented at 11th Annual International Conference on Technology in Collegiate Mathematics New Orleans, LA: Addison-Wesley & Prentice Hall.

Barrett, H. (2003). Differentiating Electronic Portfolios and Online Assessment Management Systems. Paper presented at AERA Assessment Conference. Chicago, IL: American Educational Research Association.

Blackmore, J. (1996). Pedagogy: Learning Styles. East Perth, Western Australia: P. L. Duffy Resource Centre, Trinity College.

Clark, D. (2004). *Enhancing the IMS Q & TI Specification by adding support for dynamically generated parameterized quizzes*. Rockhampton, Queensland Australia, Central Queensland University.

Epstein, M., & others (2002). Immediate Feedback Assessment Technique Promotes Learning and Corrects Inaccurate First Responses. In *The Psychological Record 2002*. Vol. 52, (pp. 187-201). Carbondale, IL: Southern Illinois University.

Sanchis, G. (2001). Using Web Forms for Online Assessment. In *Mathematics and Computer Education*. Spring 2001, (pp. 105-113). Old Bethpage, NY: Mathematics and Computer Education.