An Innovative Testing System
for Improving Testing Quality

Yossi Raanan
School of Business Administration
College of Management Academic Studies (Israel)
email: yraanan@colman.ac.il

Abstract

Testing has been part of higher education from its inception. It has taken on many forms: written tests, verbal tests, papers submitted for grades, oral presentations, and more. The most prevalent is the written test, for a variety of reasons. These tests are administered using two major methods - the "open" questions method, where the students have to reply, in writing, to a question submitted to them, and the multiple choice method, where the students have to select the correct answer to the posed question from a set of a few options given to them. The pros and cons of both methods are fairly well known, but there is little disagreement that "open" questions cause difficulties in assessments and naturally they take longer to check. This inherent delay between taking the test and receiving the results causes major difficulties in the management of learning and certainly leads to consequences that are non-quality.

The written testing methods will be analyzed and their benefits and drawbacks presented. Then, an innovative system will be presented that minimizes the drawbacks and produces higher quality results for all - students, faculty and administration. In addition, it improves student satisfaction and produces outcomes rapidly - much more in tune with today's learners.

Keywords

17th Toulon-Verona International Conference
Excellence in Services
Conference Proceedings ISBN 9788890432743

307

Liverpool John Moores University
Liverpool (England)
August 28-29, 2014

1. Introduction

Testing has been a part of the academic experience, in one form or another, for centuries. It is certainly the most commonly used tool, among all tools at the faculty's disposal, for determining students' performance. It serves as part of the assessment process for individual students, for course, for programs and even for entire institutions (see Suskie, 2008 for a comprehensive discussion of these issues). As such, the students put a considerable effort into them, and spend a lot of time preparing for them. However, that amount of time is constrained by a number of factors, many of them are outside the student's influence. Most notable among these constraints is the testing order at the end of a given semester or other study period. If the student is required to take more than one test, the tests' schedule becomes a major constraint as usually students study for tests one after the other and not in parallel. Therefore, the time between the conclusion of test A and the beginning of its successor, test B, is an upper bound on the preparation time for test B. The determination of the testing schedule is an administrative prerogative, in itself constrained by many factors: total time allotted to the testing period; number of tests to be administered in that period; duration of tests; calendar issues (like Sabbath and Holidays and other non-testing days); faculty availability; and more. Students' convenience and wishes are usually not the top priority. This creates a variety of issues, beginning with dissatisfaction of the students (an increasingly important parameter in today's competitive higher education marketplace), continuing with demands for rescheduling that, if and when they are approved, cause considerable friction and noise in the system, and so on.

In Israel, one solution to these problems is the regular offering of two consecutive testing periods, both covering all tests needed for the semester. The rules are that every student is entitled to take the test twice (a different form is used during the second exam period, of course) - regardless of the results he or she got in the first test. Some use it for improving their grade, some use it for load-balancing, spreading their exams over a longer period of time and gaining better control over the process. (Those that retake an exam a second time are given the second grade as their exam grade - even if it's lower than the first; those who fail the first test have, of course, nothing to lose.)

This system is expensive, labor intensive, cumbersome and extremely wasteful of resources. In addition, it paralyzes the campus for a long period of time. The students, too, have a lot of issues with it, not the least of which is its total lack of flexibility. The "invention" of the "second chance" tests does improve averages - but at a very heavy cost to the institution and to themselves, as will be described later.

Last but not least, the mass education system that we now have in higher education makes that system, which worked reasonably well with small numbers of students, almost impossible with the ever-growing numbers. A new system is direly needed. In this paper a new system is proposed that eliminates most of the drawbacks of the existing system and substitutes instead an easier, smoother method of testing - that is also in tune with today's computer and communications technologies that our students are engulfed by.

The current examination system is described in section 2 below and analyzed in section 3. The new system is presented in section 4 and analyzed in section 5. The pros and cons of the new system are discussed in section 6, and the implementation status is reported in section 7.

2. Academic testing in Israel

2.1 Scheduling regular examinations

In Israel, much like everywhere, most academic teaching involves testing. For the most part, courses have a final examination. In some, that examination is the sole determinant of the students' grades while in other it is one of a number of inputs that are combined to generate the final grade. Examinations decrease in frequency as students progress in their studies (for example, seminars essentially never use them to assess the quality of work). Masters level programs have fewer examinations, relatively, than bachelors, but overall examinations can be found in the vast majority of both undergraduate and graduate courses. Given that an undergraduate student takes between 5 and 10 courses per semester, on average, examinations become a major concern for the students.

Having to take 10 final course examinations in one semester requires a considerable time period. If we allow 2-3 days of intensive study period prior to each examination the required timespan is almost one month, and frequently more for many reasons.

That, in itself, may not be a problem although, for the students, the actual dates of individual examinations are critical. Naturally, the planning of specific dates for hundreds of examinations is an arduous task requiring skilled personnel, computers and a lot of coordination among a variety of stakeholders, most of them administrative. In order to take the students' point of view - being, of course, a major stakeholder, representatives from the Students' Union, are consulted before the schedule becomes final. They bring into the discussion what is referred to, these days, as "Wisdom of crowds" - in this case the students' crowd that determines both the order and the spacing of the various examinations. These inputs are considered, favorably, and given priority, but they cannot all be accepted. Other constraints that have to be satisfied include:
- Quantity of students taking the test;
- Seats availability for tests (which, as is well known, is one quarter to one third of the nominal number of seats in a given classroom or halL0;
- The duration of the test (with the allowances for handicapped students);
- Special requirements like using computers during the test;

The resulting schedule, although based on best efforts to meet students' requirements and wishes, cannot satisfy everyone. There are, naturally, many personal considerations and constraints, and even if the schedule is optimal for the "average student", it may be severely inadequate for a particular student for personal reasons.

2.2 An additional test date

In addition to the regular, normal, test date for each course requiring a written examination, the custom in Israel is (in most programs) to schedule a second, additional, test date. This additional test has the following purposes:
- To enable students to get a better grade if they think that the one they received on the first examination does not reflect their knowledge correctly;
- To assuage the fear that their semester \ year \ degree \ career may be doomed because they had a singular bad day;
- To set up a known test date in case the first one was missed for reasons of personal health or family emergency - without having to wait for the next time the test is given, which is at least a semester away, sometimes a year or even more;
- To enable students that failed the test to get a passing grade so that all dependent courses will not be delayed, or to avoid delaying completion of studies on time;

With time, however, this "second Chance" test date became part of the students' "examinations strategy". That is, they practice some form of risk management, usually reasoning along the following lines:

i.   Some examinations are hard to pass, while other are relatively easy to pass.
ii. If the two examination periods are considered as one long period, it is possible to have longer stretches of contiguous learning days.
iii. If the easier examinations are postponed to the "second chance" period, it frees up more days for studying for the hard ones.
iv. The final schedule is this:
a. Postpone the "easy" examinations to the "second chance" period, forfeiting the "first chance". This will give more contiguous learning days for the "hard" examinations.
b. Take the "easy" examinations during the "second chance" examination period.
c. If the grade in any of the "hard" examinations is not up to the desired level (whatever that is for a given student), retake it during the "second chance" period.

2.3 Another complication

Once the two examination periods became established and effectively the norm in higher education in Israel, students began to regard it as their absolute right. That is, in their opinion, they are entitled to have the two examination periods and take the tests in any manner they see fit. As a matter of fact, since 2007 that has become the law for all undergraduate studies (Knesset, 2007).

However, that was not the end of legislative interference in the examination procedures in higher education. A year later, an amendment decreed that in addition to the two periods of examinations, students that miss a test date or are absent from a significant portion of the semester because they were called for reserve military service, are entitled to a third test date, in proximity to their return (Knesset, 2008). Indeed, that was the practice prior to the legislation, but the law made mandatory rather than voluntary.

Since students will be students, there is a continuous encroachment of the barriers standing between them and the coveted "third test date". Reasons such as illness, family crises, etc., are frequently considered valid enough so as to allow the student to take that "third test", particularly if one is already scheduled for the "right" reason.

3. Problems with current system

There are a lot of problems with the current system. The main problems are given below.

3.1 Appropriation of the campus

In this system, all students that have to take a particular examination take it on a preset date and time. All of them are all given the same test questionnaire (some variations exist but by and large this is true for the vast majority of examinations). Therefore, in order to make cheating more difficult, the students are seated so that there are at least two empty seats between any two. This reduces the capacity of the halls and classes to about a third of its normal value, thus requiring three time as many seats as for teaching the same size class. Consequently, many rooms are required for the tests, essentially taking over the entire campus capacity during the first examination period. (The second examination period will be discussed below.) It basically precludes all other educational activity during that period as all available space is required for testing. The more accurate term would be "hijacking" the campus.

3.2 Labor intensive operations

Testing thousands of students, in scores of rooms, halls and classrooms is a very labor-intensive operation. It starts with the long, tedious, error-prone planning effort in which everyone takes a part: central administration (space allocations), faculty (determining tests' durations and structure); department administrations (coordination and specific requirements); students' representative (expressing the students' point of view and preferences); human resources (recruiting and training the temporary proctors); office of the dean of students (for addressing students' special needs).

After the lengthy planning process, the administration of the actual examinations requires hordes of proctors that work many hours. Usually they are at least two per small room, more in larger halls and auditoriums, and on top of that corridors are also monitored by additional proctors. Finally, it wouldn't be right without a layer of supervisors and managers….

The final phase is that of checking the examinations. This is done by faculty (usually junior faculty) and, since the results have to be published in a reasonable time, this task is divided among a fairly large number of readers. This, in turn, creates another problem:

3.3 Inconsistent checking of the examinations

When examinations are tested and scored by more than one person, inconsistencies may invariably arise. This is true, of course, for questions that are open, that call for summaries or for assessments of facts and theories and, in general, for questions that have a non-unique solution or answer. Since examinations are critical to the students, this problem is quite serious and generates extremes of dissatisfaction, of feelings of unfairness, of resentment and in rare cases in accusations of intended bias or discrimination. All these reactions are damaging for the institution. In addition, the feeling that a test was incorrectly graded causes an appeal for re-evaluation to be submitted - thus generating additional work (the appeal process and the re-evaluation and detailed reporting), thus exacerbating an already difficult problem.

3.4 Interference with regular studies

Undoubtedly the most serious problem with the current system is its overlap with the teaching of the second semester. Since the examinations are all scheduled twice, and the testing break between the first and the second semesters can only accommodate one, the first one, the other - second examination session - runs concurrently with the second semester. This is done in order to enable the students to continue with courses that have prerequisites that, having failed them, the students would not be allowed to take the following course, thus delaying his graduation by one to two semesters. This, in turn, creates the following disturbances to the regular teaching process:
- The students are absent from the second semester classes - either because they are studying for the exam or because they are actually taking it;
- Class roster is not stable since students that fail the second time must drop out of courses depending on the failed examination as a prerequisite; This make it difficult to assign group work and, in general, to teach well;
- Once the students pass the second examination of the prerequisite course, and are allowed to turn their probationary status into a regular one, they are usually far behind the rest of the class - in comprehension, in knowledge, in assignments turned in and in reading; They begin to request special allowances and they are also a group that is behind the rest of the class, thus affecting the progress of everyone in the class;

All these problem described above are the result of the second examination period running, as planned, parallel to the second semester. When a student takes a "third test" that, naturally, is taken outside (i. e. later) than the "second examination period", they become even more

difficult to resolve and more severe. A problem that is unique to this "third test" is that of scheduling it. While the "second tests" are planned, in advance, with the necessary resources allocated in good time, the "third tests" are essentially completely unknown - not every course has students that qualify for this privilege; if there are students that have this privilege - their number is not known ahead of time - so no planning is possible until such time when a student comes forward and claims the right to be given the test. When scheduling is completed, the other issues - writing a special examination, finding a suitable room, assigning a proctor and so on - just begin and they require manual manipulation in an otherwise semi-automatic system, with all the ramifications such an intervention entails.

4. New testing system

The new testing system would change the point of view of the testing mechanism - and scheduling responsibility. Instead of the institution shouldering all the scheduling responsibility, why not shift at least part of it to the students themselves? After all, the institution is responsible for writing the tests and administering them in an orderly, fair and timely manner - but scheduling has nothing to do with these features. It is just a technicality that is the result of the institution's inability to offer the examination at the student's will.

As it turns out, there is a way to do just that - allow the students to take the examinations at will, whenever they choose. Naturally, this new system will require some lateral thinking, but once we get past the traditional hurdles, it will provide a remedy to most, if not all, of the problems that plague the current system and it will also have additional benefits, outlined below.

The idea behind the proposed testing system is to do the following:
- Create a relatively large question bank for each course \ topic;
- Assemble, in real time, a test comprising of questions randomly drawn from that bank;
- Have the student take that test on a computer;
- Let the students choose their examination's schedules using an Internet-based booking system, much like people buy ticket to show and cinemas on line;

The process works as follows:
  i. The teachers compose a large number of questions for the examination in the courses they teach. These questions pass the quality assurance testing that is normally used for test items and are then deposited in a central repository, with the relevant metadata (like course name, author, date composed, date approved, date loaded into the repository, etc.). Multiple-choice questions will be uploaded with the correct answer identified.
  ii. The students that have to take the examination in that course \ topic log in to a secure website, identify themselves using user IDs and passwords, and are then presented with the list of examinations they still have to take this semester \ year \ program.
  iii. They choose one and are then presented with a list of dates and the time slots available for testing on each of these dates. The students can then choose the best timeslot that meets their own, particular schedule.
  iv. On the examination's date, the students report to the computer laboratory at the start of their allotted test time. They are identified by the proctors there, and are assigned to a computer. Once they log in the examination system will present them with a list of remaining examinations they still need to take at that point in time.
  v. After the choice is made, the computerized system will draw a random collection of questions (according to the parameters set by the teacher in advance) and scramble their order. That test will then be presented to the student, who shall either choose the right answer from a list - in case of a multiple-choice question - or write the long-form

answer in the allocated space using the keyboard. It should be noted that for best performance, the database has to contain considerably more question than are required for an individual examination, but the system works even if the quantities are identical.

vi. Multiple-choice questions are checked and graded as the student marks the answer. So, for an examination that contains only this type of question, the grade can be divulged to the student as soon as the "submit" button is pushed. Questions that need to be read in order to be graded are automatically sent, by email, to the designated teacher or teaching assistant for grading. The grade is then added to the system and, together with the multiple-choice questions, if there are any in a given examination, they make up the complete grade.

5. Does the new testing system solve the "old" problems?

The short answer is "yes", and while there are, naturally, some issues with the new system, it does provide a good solution to the old problems. Let's take a look at the "old" problems and at the solution facilitated by the new approach.

- Appropriation of the campus for testing purposes: this is the direct result of the need to physically separate adjacent students, as they are taking the same test, and of the fact that once a room is allocated for a test - it is unusable for any other purpose during that time - even if only a single student is being tested there. With the new system, students can occupy adjacent computer laboratory seats - and the risk of them cheating is practically zero as, most likely, they are taking tests in totally different subjects. In addition, even if they are taking the test in the same subject, the random choice of questions from the database and the random order of the questions in the questionnaires and the reordering of the answers in the multiple-choice questions completely eliminate any risk of cheating.
- Labor intensive operations - the new system requires a lot less proctors, since many fewer rooms will be needed; it requires almost no administrative personnel to handle scheduling, photocopying and distribution of form; and it requires quite fewer graders since many of the questions will be checked automatically by the computerized system.
- Inconsistent checking of the examinations - a direct result of many human assessors - will be drastically reduced due to the increased use of multiple-choice questions.
- Interference with regular studies - particularly during the second semester - will be minimized since the determination of when to take the "second chance" test may be left up to the student and need not necessarily be postponed to the beginning of the second semester. In principle, given an available spot and provided the student got the result of the first examination on line, at the end of the test, it's possible to take the "second chance" test the following day.

6. Pros and Cons of New Idea

The positive sides of the new system are fairly obvious. Chief among them are:
- Reduced costs
- Less grading errors as a result of increased use of multiple-choice questions
- Campus not put on hold for long weeks
- Better instruction during the second semester due to the ability of scheduling both the "first chance" and the "second chance" examination period during the examination break.

These were all discussed above and were, in effect, the main incentive for developing this new approach. But, as frequently happens, there are additional benefits that arise from the use of this new system that only add to its appeal.

- Better quality examinations and questions - since the computerized system will retain all answers given by the students to the test items, the accumulated data will enable precise determination of the difficulty level of the questions. Clearly, a question that 100% answered correctly is too easy while a question that no one answered correctly is too difficult. This data will be kept not only on a per question basis but also on per answer supplied with the question, to see which one is too obvious or perhaps too misleading. Continuous monitoring of these statistics will enable the teachers of each course to fine-tune the questions and make them precise, error-free and suitable for their purpose.
- Like all computerized tests, issues of illegible handwriting will become a thing of the past. In the same vain, some handicapped students will benefit greatly from the move to computerized tests.
- Examination averages may be controlled, if desired. Since for each question in the database the information kept will include the percentage of right (or wrong) answers, the expected grade of each question can be accurately computed. Therefore, since the questions are independent, it will be easy to compile a test with a predetermined average grade.
- Student satisfaction will be drastically increased - giving them control over their schedules and allowing them to plan their examinations period as they see fit is an obvious improvement over the current situation. In addition, the tests will become fairer and more comparable across semesters, across years and across teachers, thus improving the very important sense of fairness and lack of arbitrariness in testing and grading.

The negative sides of this new system are:
- It is not suitable for all subject matters that the institution teaches. Clearly, with current technology, any examination that requires drawing cannot be given with it. In addition, if the examination calls for writing any but the simplest mathematical expressions, it too cannot be done this way.
- It may impose limits on typed answers that may be too prohibitive for some teachers (although there is a solution for that as well), and it is not suitable (yet) for examinations that require reading very long texts.
- It may not be suitable for all handicapped students.

Overall, it seems that the pluses outweigh the minuses by a wide margin. It is worthwhile noting that many licensing examinations worldwide are given using similar methods and they overcome most, if not all, of these difficulties.

7. Implementation Status

Part of this system have already been tested, within a limited pilot project, with extremely good results for all - teachers, students and administrators. Student satisfaction was very high, and they keep asking when the system will be fully operational.

Of course, there are big differences between a limited pilot project and a fully operational system. In a fully operational system all details must be taken care of and all angles covered. Therefore, the full system is progressing at a slower pace. Currently, we are concluding an in-depth system analysis phase of the new system - and discovering many small but important details that must be addressed. We are hoping to have a test run of the system by the end of

the summer semester, and to implement the system gradually during the next academic year, in a limited fashion. The rate of adoption will be determined at that point, but it is reasonable to expect that pressures from students not given access to the new system (since adoption is planned by faculties) will accelerate it. Hopefully, some preliminary results will be available next year.

References

Knesst (2007), *knesset.gov.il/privatelaw/data/17/3/145_3_3.rtf*
Knesst (2008), *www.knesset.gov.il/privatelaw/data/17/3/322_3_1.rtf*
Suskie, Linda (2008), Assessing Student Learning: a Common Sense Guide, 2nd Edition, John Wiley and Sons.