FORMAL ASSESSMENT OF AN INNOVATIVE WEB-BASED TOOL DESIGNED TO IMPROVE STUDENT PERFORMANCE IN STATISTICS

J. A. González
Technical University of Catalonia (UPC), Spain
Lluis Jover
University of Barcelona (UB), Spain
Erik Cobo and Pilar Muñoz
Technical University of Catalonia (UPC), Spain
jose.a.gonzalez@upc.edu

We have developed a web-based tool, called e-status*, that is able to generate individually different statistical or mathematical problems and to correct the students´ answers. The tool is well appreciated by the students since it is available anywhere and anytime. This ability allows the weaker students to practice the concepts as needed, without obstructing the progress of the more advanced students. Although the use of Information and Communication Technologies (ICTs) is widely extended in undergraduate education, there are few studies evaluating the effectiveness of learning methods based on ICTs. In this work, the authors propose a blinded randomized trial to assess the* e-status *effects on improving average exam rating on dentistry students. The data results will be available by February 2006.*

INTRODUCTION

Our aim is to formally evaluate the effect of using *e-status*, a web-based tool designed to improve learning, in a statistics introductory course. The tool is complementary to classical learning materials used: practical sessions on the computer lab and lecture-based instruction.

Computer Assisted Instruction (CAI) is a field in constant growth, especially since the Internet rise. One of the first papers on this subject comes from Suppes (1965), but nowadays we can find tens of thousands of web sites containing topics related to CAI. In general, these web sites deal with developments on this methodology and/or discussion on its advantages and challenges

Although its success has been sometimes questioned, see Chumley-Jones *et al*. (2002), Internet has demonstrated to have great potential in topics related to education. Li *et al*. (2003) pointed out: "*First, web-based instruction presents information in a non-linear style, allowing students to explore new information via browsing and cross-referencing activities. Second, web-based teaching supports active learning processes emphasized by constructivist theory. Third, web-based education is enhanced understanding through improved visualization and finally, the convenience, it could be used any time, at any place*".

In spite of the high number of published works emphasizing the CAI educational advantages in teaching introductory statistics course (see Basturk, 2005; Prosser, 2000, and the references therein), there are few formally assessing its effects on improving the quality of learning, because experimental methods are hard to apply in an educational environment, and conclusions must be carefully drawn, see Holt and Oliver (2002).

Our approach in this paper aims at evaluating how *e-status* improves the statistical abilities of students. The experimental assessment has been carried out in a biostatistics introductory course in a dentistry school. The course has 35 hours of classroom teaching: 15 one-hour sessions to introduce theoretical subjects, and 10 two-hours practical sessions in lab. Topics covered are: descriptive statistics, elemental inferential procedures, an introduction to probability and sampling, observer agreement with qualitative measurements, interval estimation, basics of sample size determination and parameter comparison between two populations.

This course has several desirable features to run experimental evaluation: 1) although it is a compulsory topic, the students show low interest on biostatistics, and the use of a web-based tool may motivate them; 2) only one teacher is involved in the lecture sessions, ensuring this way that all the students receive the same contents; 3) the course is spread throughout October and November, and December is reserved to other clinical activities and preparation for the final exam (after Christmas holydays). Problem solving with *e-status* is mainly done in December, not necessarily in the classroom, and we think that it favours the independence of the students work needed to employ standard statistical methods.

The following section briefly introduces an outline of *e-status*. In Section 3, we explain the experimental evaluation proposed. Finally, Section 4 gives some data about previous experience with the tool and emphasizes some implications of the design.

OUTLINE OF THE TOOL

e-status, González and Muñoz (2005), is a web-based software that aims to:

- offer a powerful, but simple, tool for problem creation and solving,
- give feedback to the students,
- provide the teacher with reliable data about students' performance.

We think of a problem as a scheme, and the exercises seen by the student are particular instances from that scheme. A problem is composed by a wording, a number of questions, an analytical model to find the answers, some data and maybe some hints. The data changes on any new exercise, so there is only one solution for each one, which is different from other exercises.

The main strength of *e-status* is its ability, by means of this model, to compute correct answers for given questions, even if the problem has random data. Therefore, the students can see immediately which questions were answered correctly and which were not. The model is a set of arithmetic and logic expressions, written according to some syntax similar to that of programming languages like C or Java. These expressions are converted to code executable by the computer, obtaining the parameters associated with the problem: specifically, initial values and answers.

Teachers edit the problems, let them available to the students and look into the data collected from the exercises solved. Thus, a teacher can easily control whether the students are working at the suitable course pace and achieving the established academic goals, or detect alarming situations, that could be interpreted as unequal activity among the students, insufficient progress, or even deficiencies in a problem that make it hard to understand. Teachers can use this information as a basis to decide changes in the didactic structure of the course. Depending on the situation, the action decided could be immediately applied (for instance, inserting a revision session), or delayed until next year (for instance, modifying the assessment procedure).

Both teachers and students have to authenticate themselves to log in the system. This step certifies reliable identification of the student, necessary to display rate of success, average mark and other performance indicators of their effort. Students can look up their history records at any time in order to check if they are fulfilling the course objectives.

The tool promotes the repetition as a basic practice for learning. Each problem is a chance to reflect on what the student has learned. The students can pick a problem and solve it as many times as the teacher allows. Then, they may stop and move on to another problem if they feel that progress is being made.

AN EXPERIMENTAL EVALUATION

To formally assess the expected improvement in the student's statistical abilities due to *e-status*, we have prepared two different sets of six practical exercises. Each set has been designed to cover the instrumental abilities involved in a different section of the course. The accumulative nature of the statistical concepts precludes, in our view, from designing exercises that try to cover not overlapping parts of the course contents. Both sets are assuming to cover one half of the course and topics of similar difficulty. The first set (hereafter A) includes procedures related to descriptive statistics and graphical representation for qualitative and quantitative data, agreement, inference about one proportion and comparison of means and proportions between two populations. The second set (hereafter B) includes exercises about probabilities under the Normal curve, proportion and mean interval estimation, assessment of sample size, inference about one mean and about one proportion, and goodness of fit chi-square.

Official curriculum states that the final score of the biostatistics course will be calculated by averaging: 1) a test exam of multiple choices with 30 questions (45%); 2) a written examination on a practical exercise (45%); and 3) compulsory web-based exercises (10%). Scoring of test exam and web-based exercises is done automatically by corresponding software tools. The written practical exercise is revised by only one blinded corrector, in order to avoid inter-observer variability. During the second week of November a document with all the practical information and instructions needed to use the *e-status* tool was provided to the students in the electronic board of the course, and the personal user identification and password needed to log in the application was sent by e-mail to each student. At the same time, a special classroom session was organized to show the use of the application in practice.

• Participants: all students enrolled in the biostatistics course 2005-06 at the dentistry school of the University of Barcelona (UB) were included in the study (n=121). Most of them were new students (n=101) but 20 were students that were enrolled in previous years but did not pass the previous final exam (the *old* ones, so to speak).

• Intervention**:** During the last two weeks of November and all December students could solve the exercises corresponding to set A or B depending on the group they were randomly assigned. The student could work on a particular problem (with new data) as many times as they desire, and their score (by averaging over the repetitions) is updated.

• Random allocation**:** Participants were randomly assigned either to set A or B using random numbers generated with *Excel* by an independent researcher. Two blocks were defined: lab group, and new/old students.

• Outcome**:** In the practice part of the final exam we have defined three scores (S_A, S_B and S_{both}), each one based on a subset of questions directly related to instrumental abilities covered by exercise set A and not by B (S_A); by set B and not by A (S_B); and by both sets A and B (S_{both}). Differences between scores S_A and S_B in students randomized to receive either intervention A or B were our primary outcome. Moreover, we have also defined subset scores in the test exam in order to explore possible differences related to exercises assigned.

• Statistical model**:** Let μ represent the overall mean in the exam; τ_t the fixed effect of *e-status* intervention t=a, b; π_j the fixed "period" effect of part j=A, B; α_i the random effect of student i; and ε_{ij} the random effect of measuring performance of set j=A, B in student i. If we call Y_{tji} to the exam result of set j in student i with intervention t:

$Y_{tji} = \mu + \tau_t + \pi_j + \alpha_i + \varepsilon_{ij}$

And D_{ia} (D_{ib}) to the difference in student i receiving intervention a (b) between performances in sets S_A and S_B:

$D_{ai} = Y_{aAi} - Y_{aBi} = (\mu + \tau_a + \pi_A + \alpha_i + \varepsilon_{iA}) - (\mu + \pi_B + \alpha_i + \varepsilon_{iB}) = \tau_a + \pi_A - \pi_B + \varepsilon_{iA} - \varepsilon_{iB}$

$D_{bi} = Y_{bAi} - Y_{bBi} = (\mu + \pi_A + \alpha_i + \varepsilon_{iA}) - (\mu + \tau_b + \pi_B + \alpha_i + \varepsilon_{iB}) = -\tau_b + \pi_A - \pi_B + \varepsilon_{iA} - \varepsilon_{iB}$

With expectations and variances equal to:

$E(D_{ai}) = \tau_a + \pi_A - \pi_B$

$E(D_{bi}) = -\tau_b + \pi_A - \pi_B$

$V(D_{ti}) = V(\varepsilon_{iA} - \varepsilon_{iB}) = 2\sigma^2_\varepsilon$

Being the expectation and variance of the averaged difference ($\underline{D}$) between both groups with sizes n, equal to:

$E(\underline{D}_{ai} - \underline{D}_{bi}) = \tau_a - \tau_b$

$V(\underline{D}_{ai} - \underline{D}_{bi}) = 2\sigma^2_\varepsilon/n + 2\sigma^2_\varepsilon/n = 4\sigma^2_\varepsilon/n$

That means that 51 students per group provide 80% power to highlight an effect equal to 0.7 times the intra-subject variability (one sided, α=0.05). Therefore, the number of students enrolled in the course (121) seems enough to detect a relevant effect, preserving power against deviations from these assumptions.

DISCUSSION

We have started an experimental evaluation of the learning effect of an ICT tool. The data won't be available until the middle of January but, eight days before the end of the term, 78 students (37 from A and 41 from B) have already used the tool, and 925 exercises have been solved (we expect these figures to increase significantly by the end of December). It must be emphasized that our experiment preserves the rights of participants, as all of them have access to the tool, though each student can only solve problems from their own assignment group. Every student can benefit from this design, if a relevant effect is associated with the use of the tool.

Our previous experience with *e-status* is only observational, mainly based on computer science students. The tool has been well accepted, as shown by the following figures: last year, 231 students enrolled had 22 problems of Probability and Statistics available in *e-status*. 169 students used the application, and solved nearly 5000 exercises (on average, 2.5 repetitions per problem). Although not formally tested yet, we have noticed a relationship between use of the tool and exam score, and we expect to prove that it is not only a matter of motivation on the part of the student.

A survey of opinion was carried out to know how the students valuate the tool and its usefulness, and we found out: a) it has a good level of acceptance; b) they had the feeling that feedback provided is beneficial, and c) it increases the expectation of good results in the final exam.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the Agència de Gestió d'Ajuts Universitaris i de Recerca (Generalitat de Catalunya) for partially supporting this research under grant 2005 ECTS 00017. We also wish to thank the Institut de Ciències de l'Educació (UPC) for their financial support.

REFERENCES

Basturk, R. (2005). The effectiveness of computer-assisted instruction in teaching introductory statistics. *Educational Technology and Society*, 8(2), 170-178.

Chumley-Jones H. S., Dobbie A. and Alford C .L. (2002). Web-based learning: Sound educational method or hype? A review of the evaluation literature. *Academic Medicine*, 77(10).

González, J. A. and Muñoz, P. (accepted in 2005). *e-status*, an automatic web-based problem generator: Applications to statistics. *Computer Applications in Engineering Education*, http://www3.interscience.wiley.com/cgi-bin/jhome/38664.

Holt, R. D. and Oliver, M. (2002). Evaluating web-based learning modules during an MSc programme in dental public health: a case study. *British Dental Journal*, 193(5).

Li Y., LeBoeuf E. J., Basu P. K. and Turner IV L. H. (2003). Development of a web-based mass transfer processes laboratory: System development and implementation. *Computer Applications in Engineering Education*, 11(1), 67-74, http://www3.interscience.wiley.com/cgi-bin/jissue/104531600.

Prosser, M. (2000). Evaluating the new technologies: A student learning focused perspective. *Proceedings of Evaluating the New Teaching Technologies Workshop*, http://science.uniserve.edu.au/pubs/procs/wshop5/prosser.html.

Suppes, P. (1965). Computer-based mathematics instruction. *Bulletin of the International Study Group for Mathematics Learning*, 3, 7-22.