WEB-BASED ASSESSMENT AND EXAMINATION SYSTEM – FROM EXPERIMENT TO PRACTICE

Jerzy Rutkowski, Katarzyna Moscinska, Piotr Jantos
Institute of Electronics
Faculty of Automatic Control, Electronics and Computer Science
Silesian University of Technology
Akademicka 16, 44-100 Gliwice,
Poland
jerzy.rutkowski@polsl.pl, kmoscinska@polsl.pl, pjantos@polsl.pl,

ABSTRACT

The Circuit Theory course at Silesian University of Technology, Gliwice, Poland has been recently converted into blended model and various forms of web-based formative and summative assessment have been implemented [1], [2]. In the paper the development of close-ended Computer Assisted exam has been presented. The new repository has been created based on the results of previous exams. The detailed statistical analysis of the exam results has been performed with respect to students’ outcomes as well as particular questions performance. We present the results of students’ survey on web-based assessment and the comments of the staff with respect to new idea of examination. The paper enlightens the numerous advantages of Computer Assisted exam and provides some tips for teachers, especially in case of technical university courses.

KEY WORDS
Web-based education, faculty development, e-learning

1. Introduction to e–education: tension among objectives

Education has become an important part of economic activity, from 25% in 2000 to a predicted 60% in 2020 [3], and application of Computers and Advanced Technology in Education (CATE) seems to be the most promising, rapidly developing form of knowledge delivery. When implementing Information and Communication Technology (ICT) to Web Based Education (WBE), three main objectives have to be taken into account [4]:

1. quality,
2. access,
3. costs.

and there is a strong tension among them, as depicted in Fig. 1. From academic teachers’ good practice point of view enhancement of quality is the key objective. Then, in teaching quality improvement process, development of a reasonable formative assessment program, and next, development of a summative assessment – the final examination, seem to be of the major importance.

![Figure 1. Tension among e-learning objectives](image-url)

The last CATE 2006 Conference has confirmed this opinion, as two awarded papers (best paper award [5] and best non-professional software award [6]) have focused on this element of education. Recently, at many other, not directly education-oriented conferences, the problem of Computer Assisted Assessment (CAA) program development is undertaken [7]. In recent years the number of students increased quite dramatically. Not surprisingly, with such increase and traditional method of student knowledge assessment, the overall quality of student intake dropped significantly. With class size as large as 100 or more, the traditional, not automated marking of courseworks amounts to hundred of hours [8], and then, redevelopment of the assessment program from traditional to Computer Assisted (CA) model seems inevitable. The general description of the CAA program is presented in Section 2. A new, information channel-based approach to e-exam development has been proposed and verified experimentally on Circuit Theory exam, as described in [1] and briefly repeated in Section 3. In the academic year 2006/2007 the newly developed e-exam
has been preceded by numerous formative assessments [2] and then, practically verified, as the only binding form of examination, with a pool of 175 students of Electrical Engineering. The results and findings are presented in Section 4.

2. Computer Assisted Assessment – General description

In general, a reasonable assessment plan should include evaluation of both problem-solving skills and conceptual understanding. In engineering courses, such as Circuit Theory, skill in problem-solving is very important, however it should result from conceptual understanding not problem-solving mechanics of similar numerical problems. There are two possible forms of examination tests:

1. open-ended,
2. close-ended.

Both forms have advantages and disadvantages, and it is generally accepted that the close-ended form is more effective and reasonable, especially for large classes - for the automated system, practically only the close-ended form can be used. This form limits the student to the set of possible alternatives offered. Then, term multi-choice test is commonly used. Of course, for both forms development of reliable and complete task (question) bank, covering all topics, is crucial. Such database has to be continuously updated, badly defined questions have to be deleted or modified before re-use, new questions have to be added. Both formative assessments and the final exam should use only the database problems, the exam problems should be similar to the tutorial and formative test problems. All questions can be classified into three categories, subject to a form of answer presentation:

1. numerical,
2. symbolic – formula/equation,
3. graphic – figure/diagram/plot.

Distribution of these forms over the test/exam should be possibly uniform. Some questions may probe multiple concepts, however most of the questions should be single-concept questions. They have the advantage that they provide a clear measure of what the student does and does not understand. If the student gives a wrong answer to numerical question, then it does not necessarily mean that he/she does not understand the problem. In numerical problems, students are often making “silly” mistakes in calculations, mainly due to carelessness. However, students must learn that in engineering calculations the accuracy of the final result is as important as the correctness of the process they follow to obtain it [7]. For that reason, such questions have to be included and 1/3 seems to be a reasonable proportion. Seven exemplary questions are presented in Fig.2.

Q4. Find $R_{12}$ and $E_1$.

- A. $R_{12} = R_1 + R_3; E_1 = E$
- B. $R_{12} = R_3; E_1 = E - \frac{R_1}{R_1 + R_3}$
- C. $R_{12} = R_3; E_1 = E$
- D. Other

Q8. For the given voltage waveform of the RC integrator input, plot the output voltage. Assume $RC = \tau$.

- A. $u_1$
- B. $u_2$
- C. $u_2$
- D. Other

Q9. What reactance of capacitive character should be connected in series with $j100 \ \Omega$ coil such that at $U=200V$ supply, coil voltage drops by 50%, i.e. down to 100 $V$?

- A. $X=-100\Omega$
- B. $X=-300\Omega$
- C. $X=-200\Omega$
- D. Other
Q11. The parallel RLC circuit, $R=100\,\text{k}\Omega$, $L=1\,\text{H}$, $C=1\,\mu\text{F}$ is connected to $I=1\,\text{mA}$ source. Calculate element currents at the resonant frequency.

A. $I_R=0$, $I_L=1\,\text{mA}$, $I_C=1\,\text{mA}$  
B. $I_R=1\,\text{mA}$, $I_L=0.1\,\text{A}$, $I_C=0.1\,\text{A}$  
C. $I_R=1\,\text{mA}$, $I_L=0$, $I_C=0$  
D. Other

Q14. A balanced delta load with phase impedance of $10\exp(j30)$ is connected to $230\,\text{V}$ lines. Determine the total real power supplied.

A. $P=13.7\,\text{kW}$; B. $P=15.9\,\text{kW}$;  
C. $P=41.2\,\text{kW}$; D. Other

Q17. Given the voltage divider: $R_1=10\,\Omega$, $R_2=20\,\Omega$, find sensitivity of $U_2$ with respect to $R_2$.

A. $-1/90$  
B. $1/30$  
C. $1/90$  
D. Other

Q20. What should be the value of $R$ so that current $I=20\,\text{mA}$?

A. $R=3\,\Omega$; B. $R=12/8\,\Omega$; C. $R=2\,\Omega$; D. Other

Fig.2 Seven exemplary exam questions

3. E-exam description

Following the experiment findings [1], both formative and summative tests consisted of 20 multiple-choice questions. This number is the optimum for a single-session exam, as confirmed by many authors. The questions emphasize conceptual understanding over problem solving mechanics. If the students understand the concept, they are able to select the correct answer with little or no computation, within 3 to 5 minutes. After adding 20 minutes for answer revision, the total exam time is 20x5+20=120 min. Single-concept questions are used, obviously some of them are linked, but they are evaluated independently. Some questions test understanding of basic problems, some other test understanding of complex and more difficult to comprehend problems, however no weight is introduced. A question may require either “forward” or “reverse” reasoning skills. The latter ones require students to process the information in a different fashion than typically presented, e.g. given the output and system they must reason backward to find the input or given the input and outputs reason backward to find its parameter or structure. Classroom experience suggests that students who do not fully understand a concept may correctly answer a “forward”-reasoning question while they fail when answering a “reverse”-reasoning version of the same question [9]. All questions of Fig 2, except 9 and 20, are the “forward”-reasoning ones. In the applied e-exam, “forward”-reasoning questions prevail, what seems to be reasonable. The e-exam had been carried out in computer rooms, integrated with the installed information system, the Moodle LMS (e-learning platform) in its standard version. All students have had the same set of questions, however presented in a randomized order with randomized order of answers to individual questions. Four choices per question have been proposed, one of them is “other answer” and this answer is the correct one for 4 questions, 25% of all questions. After the e-exam, prior to marking, two parameters of the multiple-choice are set:

1. Pass/Fail threshold – $Th$  
2. Penalty point weight – $Pen$.

These parameters are adjusted based on the information channel concept, as described in [1]. In the experiment performed, two methods of examination have been compared: the traditional open-ended and e-exam. The following values of parameters have been designated and applied: $Th=7.5$; $Pen=0.25$. These parameters have been repeated during the regular e-exam, performed this winter for the same course of Circuit Theory. All parameters of this exam are:

- number of students: $N=175$,  
- number of questions: $Q=20$,  
- number of choices per question: $Ch=4$,  
- total time: $T=120\,\text{min}$,  
- Pass/Fail threshold: $Th=7.5$,  
- Penalty point weight: $Pen=0.25$,  
- No answer weight: $No=0$,  
- Correct answer weight: $Cor=1$.

As can be seen, the maximum total score is, twenty and the threshold has been set at 37.7% of this score.

4. E-exam results and findings

The following results have been obtained:

- number of “Passes”: $Pass=79$ (45%),  
- maximum score reached: $S_{\text{max}}=17.75$ (88.75%),  
- average score: $S_{\text{av}}=7.25$ (36.25%),  
- minimum score: $S_{\text{min}}=-1.25$,  
- maximum score with no penalties: $S_{\text{max, np}}=18$ (90%),  
- average score with no penalties: $S_{\text{av, np}}=8.85$ (44.25%),  
- minimum score with no penalties: $S_{\text{min, np}}=1$!

The automated form of examination allows its CA statistical analysis. For each question, its

1. “easiness”,  
2. “discrimination”,  
3. “percent choosing”
can be calculated [10]. Then, all questions can be evaluated, classified into:

1. “good”,
2. “fairly good”,
3. “not too bad”,
4. “bad”.

A question is considered “good”, if it is gives good “discrimination” - is answered well by top students, not well by bottom (below average score) ones, its “easiness” is close to 50% and has good “percent choosing” – distribution between five possible answers (three distractors+correct+no answer) is reasonably uniform. A question is considered “fairly good”, if its “easiness” is close to 50%, it gives good “discrimination” but its “percent choosing” not to diverse. A question is considered “not too good”, if it gives good level of “easiness (close to 50%) but non-uniform distribution of distractors and low “discrimination”. “Bad” question fails all three criteria. Fig.3 presents five histograms. The first two histograms illustrate distribution of total points (scores) $S$

$$S = n\text{Cor} - m\text{Pen}$$

where

- $n$ - number of correct answers,
- $m$ - number of wrong answers,

and wrong answers $m$ over the whole population of $N=175$ students. As can be observed, in both cases the distribution is of the normal shape, with mean values at 7.25 and 7 respectively.

The next histograms illustrate evaluation of individual questions. Fig. 3c presents percentage share of correct/wrong/no answer per each question. As can be seen, questions 6,11,14 and 17 (see Fig.2) have been correctly answered by less than 1/3 of all students and then, could be classified as “difficult” – with low “easiness”, in contrary to questions 1,2,20 which could be classified “easy”. Fig. 3d illustrates evaluation of distractors, percentage difference between correct and wrong answers per each question. As can be seen, for seven questions number of wrong answers is greater than number of correct answers, for question 11 this difference reaches nearly 30%!

The last Fig.3e evaluates “discrimination”, illustrates percentage difference between correct answers given by top students and correct answers given by bottom students. As can be seen, for five questions 1,6,12,14 and 17 this difference is below 30%. These questions have been deleted and scores re-calculated. The following results have been obtained: $S_{max}=13.75$ (91.7%), $S_{av}=5.82$ (38%).
Further analysis of the e-exam results, not presented here in details, divided students into three classes, instead of two: “High score”, “Average score”, “Fail”. Based on these analyses, some general findings can be drawn:

- “Average score” prevail, 80% of students gained less than 50% of the maximum score of 20,
- high percentage of incorrect answers – mean at 7,
- students from “Fail” class give “wrong answer” more often than “no answer”, due to wild-guessing rather than problem solving strategy applied,
- deletion of “bad” questions 1,6,12,14 and 17, did not change the e-exam scores significantly.

Generally, histograms illustrating “easiness” and “discrimination” of individual questions are especially valuable, as they allow to evaluate the effectiveness of teaching methodology and then give hints for corrections. However, sometimes a question is evaluated as “difficult”, not due to a failure in the teaching process but rather due to students preferences in problem learning. Students prefer typical problems with simple algebra involved and disregard those that require fluency in more complex calculations, such as differentiation or integration. Moreover, typically, percentage of wrong answers is much greater for problems delivered during the last lecture! Question 14 checks understanding of a very typical and easy to learn problem, however explained during the last lecture. The other “difficult” question 17 checks not only the CT knowledge but also requires an elementary skill in differentiation of function. Moreover, a question can be found “difficult” (low percentage of correct answers) as the result of good selection of distractor(s). The “most distractive” question 11 is a very typical one with little calculations involved.

Both the students and the staff have been questioned to express their opinion about the newly developed form of examination. For the students, the close-ended form of questioning has been used, with three possible answers to each question: *Yes*, *No*, *Difficult to say* (*?*). The results are collected in Table 1. When answering the first question, students have compared the CA form of CT exam with the traditional form of all other exams that they have passes so far (around 10 exams). The percentages have been calculated separately for students that have passed and students that have failed, but no significant difference with overall percentages has been observed.

### Table 1. E-exam – students’ feedback

<table>
<thead>
<tr>
<th>Question</th>
<th>No</th>
<th>?</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I prefer the CA form over the trad one</td>
<td>8</td>
<td>5</td>
<td>87</td>
</tr>
<tr>
<td>2. CA form is easier to pass</td>
<td>20</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>3. My score is adequate to my knowledge</td>
<td>10</td>
<td>16</td>
<td>74</td>
</tr>
<tr>
<td>4. I’ve been wild-guessing</td>
<td>94</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>5. I’ve been solving problems from the recommended textbook [11]</td>
<td>47</td>
<td>0</td>
<td>53</td>
</tr>
<tr>
<td>6. I’ve been solving problems of formative tests</td>
<td>18</td>
<td>34</td>
<td>26</td>
</tr>
<tr>
<td>7. Most of e-exam problems were easy to solve</td>
<td>26</td>
<td>24</td>
<td>50</td>
</tr>
<tr>
<td>8. E-exam problems were similar to those of formative tests</td>
<td>23</td>
<td>26</td>
<td>51</td>
</tr>
<tr>
<td>9. E-exam time was sufficient</td>
<td>27</td>
<td>4</td>
<td>69</td>
</tr>
<tr>
<td>10. Problems of re-sits were similar to those of the 1st term problems</td>
<td>27</td>
<td>4</td>
<td>69</td>
</tr>
<tr>
<td>11. I’ve spent a significant time to improve my knowledge before the re-sit</td>
<td>27</td>
<td>4</td>
<td>69</td>
</tr>
</tbody>
</table>

As can be seen, students praise the effectiveness of the online formative assessments and rate them highly as an efficient component of their learning. Majority of students have found the e-exam objective, its problems similar to those of tutorials and formative tests, the duration-time sufficient. Percentage of wild-guessing has been declared low, however it has not been confirmed by the exam statistical analysis - for seven questions number of wrong answers is greater than number of correct answers! There was no distinction between two possible wild-guessing strategies: systematic and occasional. The 6% that have declared *Yes*, probably have applied the wild-guessing strategy to all questions; many others that decided to take a wild-guess occasionally have declared *No*.

For the staff, the open-ended form of questioning has used, eight members of the CT team have been asked to outline positives and negatives of e-exam, as compared with the old traditional one. The consensus has been achieved when evaluating positives of e-exam:

- easy to check,
- more objective,
- cheating reduced practically to zero.

When outlining the negatives, the opinion was not so unanimous. Only 50% expressed absolutely positive attitude to the CA form of examination, other 25% expressed positive attitude however recommended some serious modifications. These recommendations are as follows:

- introduction of questions with only one numerical answer, practically elimination of multiple-choice form of numerical questions,
- increase of distractors from three to four (total of five choices), with “other answer” choice maintained,
- increase of the penalty point weight, from 0.25 to 0.5 or higher, even if this leads to significantly worse results – students have to be convinced that taking a wild-gues does not pay,
- providing questions with randomized parameters and/or structure of a circuit.

These recommendations, except the last one, will be taken into account when developing the next examination, summer this year. The CA question randomization is the new field of research and the first results have been already reported [7], [12]. To randomize question parameters and/or structure of a circuit, a specialized software has to be elaborated and incorporated into open-source LMS. Unfortunately this is a time consuming task and can not be accomplished without a financial and organizational support from the side of the University authorities. Up till now, teachers can not expect gratification for their involvement in course redevelopment, what never helps rapid action, regardless their strong emotional commitment. Moreover, teachers are heavily loaded by traditional learning and organizational duties, research activities, and therefore not much time is left for ICT incorporation. These are the main factors hindering rapid development of course/assessment program restructuring to CA form [13]. Nevertheless, it is our great belief that course/assessment restructuring will proceed dynamically and within next year the CT course will be redeveloped into a blended model and CAA system will be fully operational and ready to use by other teams.

5. Conclusion

Various forms of computer-based or web-based assessment are necessary in modern teaching. The implementation of Information and Communication Technology (ICT) enables to increase the quality of learning while preserving costs. There are numerous advantages of Computer Assisted close-ended exam, including reduction of workload, objectivity of marking and elimination of cheating. Both students and staff are satisfied with the new idea of examination. Permanent verification, modification and expansion of repository should be carried out, mostly with respect to data randomization.

References