

11/1 (2013), 41–52

tmcs@math.klte.hu
http://tmcs.math.klte.hu

Teaching
Mathematics and
Computer Science

Problems of computer-aided assessment of mathematical knowledge

ISTVÁN VAJDA

Abstract. Although conventional written and oral exams are dominant in assessment nowadays, computer-aided assessment is developing dynamically. There are several assessment systems, but most of them evaluate only multiple choice questions and even the most sophisticated ones cannot follow the process of thinking of students in detail. Why is it? In this article I will analyse the difficulties of the implementation of assessment system focused primarily on mathematics questions and present some of my experience related to the eMax system, developed at Óbuda University.

Key words and phrases: computer aided assessment.

ZDM Subject Classification: D60.

1. Introduction

Shortly after the emergence of personal computers, computer-aided learning and teaching came more and more in focus of interest. While many of its enthusiastic supporters overestimated its role, others declined it strongly stressing its disadvantages. It is clear nowadays that multimedia and computers give a great deal of help in demonstration and motivation of students but cannot manage all the problems of learning and instruction.

Computer assisted assessment of student’s knowledge, abilities and skills is an important application within learning and teaching. Systems that apply multiple-choice tests developed rapidly in this domain, facilitating the work of assessing tests thus allowing frequent testing for teachers. However we cannot deny the

objection that this type of assessment focuses on encyclopaedic knowledge and it is less informative about thinking abilities and intelligence. To solve this problem, some of the teachers and researchers assemble more complex and sophisticated multiple-choice questions, while others try to implement new computer assessment systems, which are more appropriate to estimate deeper knowledge. However, these attempts have not been completely successful so far. In this article I will analyze the difficulties of the implementation of assessment systems while focusing primarily on mathematics questions and present some of my experience related to the eMax system, developed at Óbuda University.

2. Assessment of students' work

Tracing the development of student knowledge, skills and abilities is an important part of the process of teaching. The declared aims of the assessment are not only the scoring of students, but also the improvement of the efficiency of teaching. These two aims are simultaneously present in conventional teaching. Although there are different ways of assessment, they fulfil both aims of personal scoring and also give some motivation for better scores. The assessment usually does not help the students to recognize their lacking knowledge or how to improve their skills and abilities and change their learning methods for an improvement.

The modern pedagogical literature disassociates the two functions, it distinguishes formative and summative assessments. Summative assessment is applied usually at the end of a learning period, to determine at that particular point in time what students know or do not know. Its typical forms are exams or final tests. Contrarily formative assessment is part of the teaching process. It informs both the teacher and students about student understanding at a point when timely adjustments can be made.

While summative assessments are relatively rare – especially in higher education – it would be quite beneficial to use formative assessment more frequently so that to facilitate students' learning. This aim can be achieved only if the students and the teacher too receive a detailed feedback about the result of the assessment. If a student group is large the teacher cannot fulfil this task in oral form. If the students need longer time to solve the questions, which is quite common in mathematics for example, then the written form is adequate too. Since the teacher usually cannot cope with the large amount of correcting work needed, formative assessments will be used too rarely to make a sufficient effect. It would be a great advance, if at least some of the assessments could be computer-aided.

Another advantage of computer-aided assessment is that it can be utilized without the teacher too, so students can assess themselves solving their homework or making individual practise without the consequence of their fortuitous unsatisfactory performance. This opportunity already exists in some of the topics of mathematics [2], [13], [17]. Even using summative assessment, computer-based assessment has some advantages. The questions can be stored in a database so they can simply be reused. If the set of questions in the database is large enough, tests can even be generated automatically by the computer, while taking the criteria determined by the teacher into consideration [9]. Computer based assessment have additional advantages as well, as the computer can generate statistics, diagrams automatically, so the teacher can survey which tasks raise difficulties for the students and can compare the performance of different student groups.

Naturally computer aided assessment has disadvantages too and the elimination of that is usually not simple. In the following part of the paper I will focus on the problems in computer-assisted assessment of mathematics and on our solutions in the mathematics module of the eMax system.

3. The eMax system

Questions of computerized exam can be classified into two major groups, based on how students can enter the answer into the computer:

- passive questions
- active questions

While in case of the preceding type students have to choose their answer from a set of given answers (multiple choice, hot spot, sorting), they are expected to create their own answer in case of the latter (textual answers, giving formulae, creating graphs) [11], [7].

eMax is a computer assessment system developed at Budapest Tech (predecessor in title of Óbuda University). It provides several passive question types and two active ones, namely the short text and some special types of the calculation questions [7], [10]. The latters are provided by the mathematics module of the system, which was developed by our team. Calculation questions can be answered by an originally unknown quantity, calculated by using given data, which can be numbers, vectors or matrices. The mathematics module of the eMax system provides calculation questions mostly in topics of vector geometry and matrix algebra.

Considering the intelligence of evaluation, assessment systems can be classified into three groups:

- manual evaluation,
- semi-automatic evaluation,
- automatic evaluation.

Manual evaluation means that the exam system is able to take in the answers only, so exams can be paper free.

Semi-automatic evaluation means that system is able to assess most of the answers, but the problematic ones are checked by a human teacher too. Automatic evaluation is able to evaluate all the answers entirely.

The aim of the mathematics module was to achieve semiautomatic evaluation of active mathematics questions, and also to score incomplete solutions proportionally.

While designing the system and considering the aims of teaching and assessment, we had to make several decisions, such as:

- Which formats can the students use entering their solutions?
- How to store the entered solutions?
- What algorithms should the computer use for evaluation of the solutions?
- How to find the right parts of the solution if the student uses a bad syntax or the solution is not complete?
- How to score the solutions?

Subsequently we will discuss the above mentioned design aspects.

4. Entering solutions

It is a serious restriction for students to enter mathematical solutions into the computer, unlike to write on a paper or to take part in an oral exam. It is even more so, if computer-aided assessment is new for them, which can significantly degrade their achievement. Some of the systems avoid this problem by using only passive question types, this resolves other problems too – so students only have to choose the right answers from the ones given by the system. A good example for this type of systems is GRE [5].

Multiple choice questions are the most common type of passive questions, but opinions are very contradictory about them. One characteristic opinion is that they can be used widely and effectively, and with improving the underlying

methods, students’ knowledge will be assessed better and better [4]. On the contrary, the Mathematical Sciences Education Board (USA) pleads a seriously disapproving opinion.¹

Other systems require constructed response questions. In case of mathematical systems they usually allow to enter formulae, but mostly in one row only, like in case of calculators.² An example for this can be found on the homepages of Algebra Homework Help or WMI2 [2], [17]. Some systems provide entering formulae with an equation editor [1], [16].

The constructed-response questions are spreading first of all in systems that help the self control and practice. This type of questions require a more solid knowledge from students than multiple-choice questions, but entering of solutions can be slow which can be unfavourable for students if there is only a short time available. Even in case of constructed response questions it is common, that students have to give only the result of the question without the solution process.

The user interface for students of the eMax system provides an equation editor to enter the solutions of calculation questions. Students have to enter not only the result of the calculation, but a step by step solution, so that the answer can be evaluated even if it is not whole. Each rows must match to the following syntax:

`<new variable> = <formula> = <numerical value>`

Formulae can only contain the originally given data and the variables defined earlier. The note of the result and one of the new variables must be the same. The embedded syntax checking system helps the students to enter the solution correctly, yet firstly they had difficulties nevertheless.

5. Storing entered data

Multiple choice test systems store only the code of the solution, but in case of constructed response questions the system has to store the whole solution. If the answer is a short text, it can be stored into a string variable, if it is a formula

¹“Testing as we know it today arose because very efficient methods were found for assessing large numbers of people at low cost. A premium was placed on assessments that were easily administered and that made frugal use of resources. The constraints of efficiency meant that mathematics assessment tasks could not tap a student’s ability to estimate the answer to an arithmetic calculation, construct a geometric figure, use a calculator or ruler, or produce a complex deductive argument” [12].

²But the system often creates and shows the mathematical form too.

it has to be converted into a text first. Although there are many ways for this conversion, eMax as many other systems use MATHML [6], because it is a very precise way to describe math formulae.

A solution can include sequences of formulae or short texts, but these can be stored simply too.

6. Algorithms for evaluating mathematical solutions

Mathematical questions are classified in many ways [3]. The most common aspects of classification are:

- The mathematical topic of the question
- Whether the examined knowledge is a concept, a theorem, a proof or an algorithm.
- How deep knowledge is required in the given question?
- Is the question open-ended or not?

The first two of the above points do not require an explanation. Concerning the third aspect it can be stated, that those questions do not need a deep knowledge for which students only have to recognize the right answer. It is however harder, if they have to reproduce the correct answer or to accomplish a calculation or find another algorithm, but the hardest case is if they have to find connections or to prove a theorem.

A closed-ended question means that the solution is a definite answer, while in case of an open-ended question, there exist more correct answers [3].

The above aspects influence the method and the difficulty of the evaluation of students' solutions [8], [15]. If students have to recognize the right answer, then the system has to check only the entered code, but in case of constructed response questions the evaluating algorithm must be more complex even if students have to enter only the result. The result can have different forms even if it is definite. If the answer is an integer value, then evaluation is usually simple but also in this case the system has to recognize different forms of the answer, like instead of 1230 the student can also write $1.23 \cdot 10^3$. In case of a non-integer result, the system have to accept different fraction forms for example $\frac{1}{2}$, $\frac{2}{4}$, $\frac{3}{6}$, ... and 0.5. Students often have to determine an irrational number that can be given in algebraic form for example $\sqrt{2}$ and with an approximate value. It is important to decide how much difference – absolute or relative – can be tolerated by the system. The situation is more complicated if the solution is a mathematical formula that

has several equivalent forms, for example $\frac{1}{1-x^2} = -\frac{1}{(x-1)(x+1)}$. The evaluating algorithm has to recognize the right answer without reference to its form in the students' solution. To achieve this, the system can use some test data to check the solution or to transform the formula given by the student into a standard form. Computer algebra systems can cope with this problem.

If we want the system to follow the solution process, the implementation becomes far more difficult. But why do we need to check it? It is quite often, that the student makes an error, but still gives a right answer. For example $\log_2 xy = \log_2 x \cdot \log_2 y$ is a typical bad transformation, but if $x = y = 4$ the result of the calculation will be correct. Another problem is that students often make simple errors due to inadvertence, but their method of solution is principally correct. In this case the system underscores the solution if it takes only the result into account. So parts of the solution should be examined separately for a more accurate evaluation. Since a mathematical question usually has several solutions, and parts of the solutions can be different, the evaluation algorithm has to be aware of the possible solutions the students may give [8], [10].

The mathematics module of eMax system uses several algorithms to evaluate calculation questions. The most important of them is the test algorithm, the aim of which is to check if the principles of the solution are right or not, using appropriate test data. This algorithm decides also if the numerical calculation of the solution is correct or rather how many calculation errors are made [11]. You can see the outline of the algorithm in Figure 1.

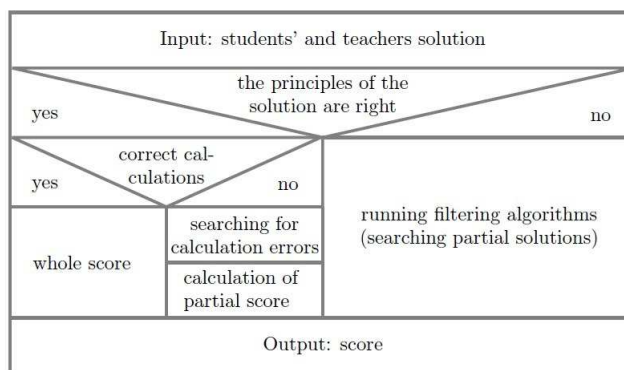


Figure 1. Test algorithm

7. Not complete and partially wrong solutions

The hardest task for the assessment system represents students' solutions that are not complete or have wrong parts. The deficiency is sometimes formal only, if the student for example does not write the solution entirely, but works out some missing parts of it in his head. In case of a conventional evaluation the teacher does not consider this as an error, because the students' thinking can be followed easily. But a computer based assessment system that looks for the logical connections between the parts of the answer rigorously, will find the same solution wrong. An example copied from a student's exam demonstrates this in Figure 2. This exam was written in the eMax system.

$$\begin{aligned} \underline{AB} &= \text{vector}(A) - \text{vector}(B) = \text{vector}(6, 4, 0) \\ \underline{CA} &= \text{vector}(A) - \text{vector}(C) = \text{vector}(-1, 5, 0) \\ \underline{CB} &= \text{vector}(B) - \text{vector}(C) = \text{vector}(-7, 1, 0) \\ BA &= \sqrt{6^2 + 4^2} = \sqrt{52} \\ CA &= \sqrt{1 + 5^2} = \sqrt{26} \\ CB &= \sqrt{7^2 + 1} = \sqrt{50} \\ k &= \sqrt{52} + \sqrt{26} + \sqrt{50} = 19,38 \end{aligned}$$

Figure 2. Imprecisely described solution

In this task the student has to determine the perimeter of a triangle. The vertices of the triangle are $A(2, 0, 5)$, $B(-4, 1, 0)$ and $C(3, 0, 0)$. In the example the student calculates first three side vectors using an unusual notation of \underline{BA} , \underline{CA} , \underline{CB} . Then he calculates the absolute values BA , CA and CB of these vectors, but they are independent from the side vectors in the system. In this case the student gives a correct solution, nevertheless he receives e.g. 10% of the full score. Unfortunately a computer assessment system cannot yet recognise the right steps of the solution if they are not deduced rigorously from the former steps because the formalization of it can be done in many ways and a numerical value – if exists – can be false due to a former computational error. Thus a not complete solution can only be evaluated if the steps of it are in a strict logical connection. The order of the steps is not determined because the mathematical thinking follows a treelike structure. In addition the solution can be top-down or bottom-up and these two can be mixed. So the logical connection can be recognised from the description of the steps and not from their order.

A further difficulty is that a question may have more than one correct solution, hence the system tries to decide at first which solution is chosen by the student. In most of the cases this is possible on the basis of typical parts of the solution, but it is not always straightforward because different solutions can overlap each other.

The mathematics module of the eMax system offers a partial solution for the above problems. We introduced the concept of key operation, which is a characteristic operation suitable to recognize which solution is applied by the student. Sometimes we cannot find an appropriate single operation, so we need to use a set of operations [10]. After deciding which example solution matches to the student’s answer, the system runs the test algorithm. If the test algorithm finds that the student’s solution is principally correct, then the system calculates the students score, otherwise it runs the filtering algorithms in order to search partial solutions. For the time being there are four filtering algorithms in the mathematics module of the eMax system, two of them are used to search for bottom-up solutions, while the other two search for the top-down ones [11].

8. Scoring solutions

Scoring of students’ solutions should serve first of all the fulfilment of pedagogic aims, but its realisation depends on the entering of the solutions and the evaluation algorithm.

In case of multiple-choice questions if only one of the given answers is correct, the only aspect is to decide on the scores of the right answer. If there can be more than one correct answer among the given possibilities or there may be no correct answers at all then students can receive a part of the scores if they mark some right answers but not all of them. Another question is how many points a student should get who choose all of the right answers but a bad answer too.

In case of short answer questions there are similar problems because if the correct answer is complex, students can give some parts of it correctly while other parts not. A good example for this is the short answer module of eMax [14].

In case of constructed response questions, assuming that the system declares the steps of the solution, then students’ solution can be scored as the sum of the points of the correctly answered parts of the solution. Another possibility is that the student receives only the sum of the scores of those parts of the solutions which are before the first wrongly answered part. This can be more suitable than the preceding one if the student made a principal error. If the student made only

a calculation error or another mistake through inadvertence that do not affect the subsequent steps of the solution then it is better to score these steps too. A further decision relates to the point if the student makes a calculation error, but works with the right relationship. Then the teacher has to decide whether or not to score the corresponding steps.

9. Test results of the mathematics module of the eMax system

The mathematics module of the eMax system provides only a small part of calculation questions at the moment, so we cannot use it to assess a real math exam. Due to the lack of subsidy it was not improved in the last years, so the last test results produced by the system are from 2008. The exam organized to check the system contained six questions in the subject of vector geometry and matrix algebra. The participants were 25 computer science and engineering students, who sent in 133 solutions. These solutions were evaluated by the system and a teacher too. If the difference of the scores given by the system and the teacher were not greater than 10%, then the evaluation of the system were considered right. As it can be seen in Table 1 the evaluation of the system proved to be right in 87% of the questions. 5% of the solutions were marked to manual evaluation and only in case of 7% was the difference between the scores given by the system and the teacher more than 10%. From this, in case of 2% the student made syntax errors while entering the solution, and in case of 5% the evaluation algorithm could not recognize perfectly the parts of the solution.

Table 1. Comparing scores based on system and manual evaluation

Question	$\leq 10\%$	Manual	Algorithmic problem	Bad usage
1	96%	0%	0%	4%
2	84%	4%	8%	4%
3	88%	8%	0%	4%
4	96%	0%	4%	0%
5	80%	8%	12%	0%
6	80%	12%	8%	0%
Together	87%	5%	5%	2%

10. Conclusion

Computer aided assessment of students' works is prevalent and is spreading more and more. Its advantages are the speed of the evaluation, reduction of teachers' work, keeping testing data and automatic demonstration of the results. Nevertheless, it is less sophisticated than human assessment, consequently students' deficiencies remain often hidden. There are several approaches to solve or at least to lessen this problem, and some computer assessment systems can already produce reliable scores if the questions are not too complex. Naturally, the more we want to take the creativity and the way of thinking of students into account, the more problems occur in the process of computer-aided assessment, so there is plenty of room for further research work in this field.

References

- [1] ALEKS, <http://www.aleks.com/>.
- [2] Algebra Homework Help – People's Math, <http://www.algebra.com/>.
- [3] A. Ambrus, *Bevezetés a matematika-didaktikába*, ELTE Budapest, 2004.
- [4] D. Bálya, *Az informatika kihívása a teszt-technológiában*, BME TIO, 1997.
- [5] GRE – Graduate Record Examinations, <http://www.ets.org/gre/>.
- [6] H. Hagen, MathML., <http://www.pragma-ade.com>.
- [7] A. György, V. Kos, B. Schmuck, D. Sima, S. Szöllősi and I. Vajda, Intelligens vizsgakiértékelő rendszer (előkészítő vizsgálódások), IX. Országos Neumann Kongresszus Győr, Széchenyi Egyetem, 2006.
- [8] A. György, S. Szénási and I. Vajda, Szemiautomatikus tudáskiértékelés a vektoralgebrai feladatok példáján, Informatika a felsőoktatásban 2008, Debreceni Egyetem, augusztus 2008.
- [9] A. György and I. Vajda, Intelligens tudásellenőrző kérdéssorok automatikus generálása, Informatika a felsőoktatásban 2005, Debreceni Egyetem, augusztus 2005.
- [10] A. György and I. Vajda, Intelligent Mathematics Assessment in eMax, IEEE Africon 2007, IEEE, 2007.
- [11] A. György and I. Vajda, The mathematics module of the emax system, 2-nd European Computing Conference (ECC'08), 2008.
- [12] Mathematical Sciences Education Board and National Research Council, *Measuring What Counts: A Conceptual Guide for Mathematics Assessment*, National Academies Press, Washington, 1993.
- [13] QuickMath, <http://www.quickmath.com/>.

- [14] D. Sima, Á. Miklós, B. Schmuck and S. Szöllösi, Rövid szöveges válaszok szemi automatikus kiértékelése, Informatika a felsőoktatásban 2008, Debreceni Egyetem, augusztus 2008.
- [15] I. Vajda and A. György, Electronic Assessment in Mathematics, *POLLACK PERIODICA An International Journal for Engineering and Information Sciences* **2** (2007), 203–214.
- [16] I. Vajda and A. György, The Mathematics Module of the eMax System, New Aspects on Computing Research, WSEAS, September 2008.
- [17] Welcome to WMI2!, <http://matek.hu/>.

ISTVÁN VAJDA
ÓBUDA UNIVERSITY
HUNGARY

E-mail: vajda.istvan@nik.uni-obuda.hu

(Received March, 2012)