International Journal of Mathematical Education in Science & Technology
Vol. xx, No. xx, xxx 2006, 1–24

Mathematics on the threshold

André Heck and Leendert van Gastel
Universiteit van Amsterdam, AMSTEL Institute,
Kruislaan 404, 1098 SM Amsterdam, The Netherlands.
Email: heck@science.uva.nl
(Received 31 December 2005)

Lowering the dropout rate of incoming mathematics and science students, and enhancing the provision of mathematics support for freshmen are two important aims of the University of Amsterdam. The approach recently adopted to support first year students is to set up a diagnostic pretest and posttest and use these tests to identify students being at risk of failing their mathematics courses and other modules in the first year. Follow up procedures are implemented and computer algebra based assessment and practice of mathematics skills play an important role in it. In this paper we describe this approach and its success.

1 The mathematics problem at university entry

In the last few years, a wave of consternation went through mathematics, science, economy, and engineering departments at the Dutch Universities. The mathematical abilities of incoming students dropped significantly and the freshmen had many problems in making the transition from school to university mathematics. Due to this reason, many departments, faculties and universities organized an initial assessment of basic mathematical knowledge and skills of incoming students. Because up to now the gap between school and university mathematics was always rather easily overcome, these diagnostic tests at university entry were a new phenomenon in the Netherlands. At one of the technical universities, only four percent of freshmen passed the initial assessment in 2005. The alarm was sounded in national newspapers and a letter of advice was sent to the Minister of Education.

1.1 *The international perspective*

The concern over the transition from school to university mathematics is not new and not restricted to the Netherlands, but can be heard in many countries around the world, and especially in the United Kingdom. The London Mathematics Society made already ten years ago disturbing claims about the mathematical preparedness of new undergraduates in the report *Tackling the Mathematics Problem* [1]. The report identified three keys areas seen to be lacking in incoming students:

(i) Students enrolling on courses making heavy mathematical demands are hampered by a serious lack of essential technical facility — in particular, a lack of fluency and reliability in numerical and algebraic manipulation and simplification (para 4a);

(ii) Compared with students in the early 1980s, there is a marked decline in students' analytical powers when faced with simple two-step or multi-step problems (para 4b);

(iii) Most students entering higher education no longer understand that mathematics is a precise discipline in which exact, reliable calculation, logical exposition and proof play essential roles; yet it is these features which make mathematics important (para 4c).

These are actually the same main problems that educators at Dutch universities perceive more recently with the incoming students. The Mathematics Working Group of the European Society for Engineering Education (SEFI) noted in the report *Mathematics for the European Engineer; a Curriculum for the twenty-first Century* [2] a similar trend across many European countries and it proposed a new engineering mathematics curriculum, partly to deal with the declining standard of mathematical knowledge and skills of new entrants to engineering degree courses. Two conclusions in the report *Mathematics in the University Education of Engineers* [3] were that it is time to reconsider what kind of engineering mathematics is needed and when, and to think about what pedagogical approaches in the computer era can best deliver the mathematical needs of students. The Engineering Council recommended in the report *Measuring the Mathematics Problems* [4] the use of diagnostic tests to measure the mathematical competence of students upon arrival at university and that prompt and effective support should be available to students whose mathematical background is found wanting by the tests. One of their findings was that diagnostic tests play an important part in

- identifying students at risk of failing because of their mathematical deficiencies,
- targeting remedial help,
- designing programmes and modules that take account of general levels of mathematical attainments, and
- removing unrealistic staff expectations.

Another finding in this report worth mentioning here is that the decline in basic mathematical skills is not the fault of the teachers in secondary schools.

Compared to their predecessors, they have to deliver a different curriculum, under very different and difficult circumstances, to quite different cohorts of students. Hoyles, Newman, and Noss [5] noted that mismatches between school and university mathematics in the United Kingdom might be influenced by the changing profile of entrants to mathematical subjects in higher education. Nardi [6] suggested that the gap between school and university mathematics could be characterized simply as a jump from empirical to abstract mathematics, from the informal to the formal. She noted that incoming students have to learn a new way of thinking and operating mathematically at university, but that they often assume that it is merely an extension of school mathematics. Therefore they are not prepared for the rigour and precision of university mathematics and do not easily adopt formalization. Whatever the reasons may be, many university educators all over the world notice nowadays a lack of adequate mathematical knowledge and computational proficiency of their beginning students and they find that many a student has problems regarding mathematical maturity.

1.2 *Recent educational reforms in the Netherlands*

Coming back to the Dutch situation, in order to understand the background of the consternation over the lack of mathematical preparedness of incoming students, it helps to know that in the past years the secondary education in the Netherlands has undergone several changes. These changes have a large impact on the mathematical abilities of the freshmen that are to pass the threshold of the university.

First of all, in 1998 the Dutch Ministry of Education introduced a new concept for education in the upper level of secondary education, the so-called 'Studiehuis' (study house), which emphasizes inquiry skills and self-responsible learning. The role of the teacher has changed into that of a coach, supporting the learning activities. In this new concept, projects and practical work are more important. In most cases, problems are framed in a rich context. The introduction of contexts in education could rather easily be implemented because at the same time a new examination programme was implemented, in which students are required to choose from four fixed combinations of subjects. These so-called profiles are: Nature & Technology, Nature & Health, Culture & Economy, and Culture & Society.

The pupils are trained to extract the mathematical content in a context. Meanwhile, this has lead to a reduction of the mathematics contents in the nationwide standard curriculum and to a reduction of time that students can spend to familiarize themselves with new concepts and to become proficient with mathematical techniques up to routine level. The goal of the study house is that students learn to learn, learn to gather information, analyze it, gain results and present these.

These two changes in Dutch education, study house and profiles, have often been under debate in the last couple of years and at this moment new plans are proposed by the Ministry of Education to change the examination programme in 2007 and to reposition the study house. The most important issues are a change from breadth to depth in learning, i.e., a shift to fewer subjects being treated deeper, and more attention to transfer of 'traditional contents' via teachers teaching subject matters instead of self-learning by students.

The next reform is specific to mathematics education and it concerns the introduction of the graphing calculator and the chart of standard mathematical formulas. Secondary school students typically start to work with the graphing calculator at upper level (age 15 years), and henceforth they seem to use the graphing calculator for almost all their problem solving. Students are also no longer requested to memorize mathematical facts and formulas, but instead they are allowed to use a formula chart, even at the national exam. As a result students forget many mathematical concepts and techniques learned before, such as calculating with fractions by pencil and paper and manipulating simple formulas. Students tend to become dependent on the graphing calculator for numerical calculations and to use without question the formula chart for symbolic manipulations.

Certainly, these reforms have their merits. The new students are open-minded and are quite smart in solving contextualized problems. The graphing calculator has certainly stimulated a more exploratory, active approach to problem solving. However, the drawbacks from the perspective of mathematics, science, and engineering education at university level, become visible as well. The shift of focus from depth to breadth in mathematics education involves the danger of students having a rather disjointed body of knowledge and skills. Certainly, the mathematical levels of incoming students are much more heterogeneous than before. Moreover, the students are not used to the more abstract level at university and their mathematic knowledge is rather fragile with regard to the nature of rigour in formal mathematics. Both reasoning in the pure, non-contextual mathematical setting, and formula manipulation are put under stress due to these reforms. In many of the courses, both the mathematics courses and all those courses that rely on mathematical knowledge and skills learned, students lack sound mathematical understanding and cannot cope well with the academic pace. There also exists a worry about the students' lack of stamina in learning and doing mathematics. As a result of all, the self-confidence of many a student suffers and there is a greater risk of dropping out.

1.3 *An initiative*

It is our strong belief that mathematical knowledge and skills at routine level can only be obtained by intensive practising. J. van de Craats and J. Bosch put this aptly in the preface of their recently published Dutch exercise book for basic mathematics [7], which we use in our remedial teaching activities (translation by the authors):

> 'Just like with any other skill, whether it is footballing, playing the piano, or learning a foreign language, there is also only one way of mastering mathematics: a lot of practising. For football you must train, playing the piano requires a fair amount of study, and learning a foreign language involves memorizing words. Without basic techniques you cannot get anywhere; in case of mathematics things are not different.'

At the University of Amsterdam, an approach is being developed to cope with this mathematics problem at the threshold of the university. The approach is based on the trust that incoming students are in principle quite capable of overcoming the initial difficulties and can attain the desired level in a rather short term. Because of the heterogeneity, it is very important that both students and teachers are informed about the individual level and progress of each student. So initial and continuous assessments are an important part of the approach. For mastering the mathematical abilities, many exercise questions are envisaged together with a condensed mathematics text. A trajectory will be designed that is based on these assessments and the mastery learning of the mathematical abilities by doing many exercises.

An important requirement for the approach is that not much staff will be available in the long term. There are no structural funds available for this goal. So the challenge is to design a trajectory that involves only a modest staff deployment. For this reason, a pilot is undertaken with automated assessment in mathematics using the commercial package Maple T.A. [8]. This system uses the computer algebra package Maple [9] to generate and mark thousands of mathematical problem exercises from generic templates and it can deliver these exercises in various forms to students, ranging from diagnostic assessments, self-tests, and practice sessions to placement tests and summative assessments.

The development of this approach is part of two projects with other higher education institutions in the Netherlands. The project Web-spijkeren [10] is funded by the SURF foundation and has as a goal to develop didactic scenarios and assessment instrumentation for freshmen with heterogeneous mathematics level. Partners are faculties of economy of the University of Maastricht and the Erasmus University of Rotterdam. The project MathMatch of the Digital University [11] develops course materials and a database with mathematical questions for mathematics, science, and technology studies. Partners are the University of Twente, the University of Amsterdam, the Vrije Universiteit

Amsterdam, and the Saxion University Enschede.

Our work is still in its early stages. This paper describes the approach we have taken in 2005, the instrumentation used, the experiences of the first ten weeks, the results of the first try-out, and future developments.

2 Supporting freshmen to pass the mathematics gap

We haven taken in 2005 the following four-step approach with all hundred students entering the study of chemistry, physics, astronomy, mathematics, and bio-exact sciences.

Step 1. Immediately upon arrival of the freshmen at the Faculty of Science, on their second day at university, we take a one-hour diagnostic test. The test has been designed in collaboration with university teachers of first year mathematics and a teacher trainer has commented on it.

We explain to the students that the test is not meant as an entry exam, and that the test will probably be too difficult to answer all questions correctly. Indeed, the mathematical level is the one that they are expected to reach after four weeks of (remedial) study, and the test may contain mathematical subjects that not every student has already met in mathematics courses. It depends on the profile chosen at secondary school which topics they have mastered. So, limits, vectors, and systems of equations will be new for some of the students.

Lack of mathematical skills at routine level, maybe because of the long period between their final secondary school exam and entrance at the university, is another reason that we expect that students will not have enough time to complete all tasks in time. This is why we suggest the students to continue with the next question of the diagnostic test whenever no immediate idea comes into the head when reading an exercise. Because it was the first time that students are asked to do a digital mathematics test, because it was our first time to use Maple T.A. on such a large scale, and because we wanted to get as much as possible insight in the students' mathematical performance, we gave them the printout of the digital test as scrap paper and asked them to make each exercise with pencil and paper, to type their answer in the digital test form, and finally to hand in the scrap paper.

All student make the same test, that can be found in the appendix. We report the results of the students in the form of teacher's interpretation: We expect that students who score more than fifteen points (out of twenty-five) only have to dot the i's and cross the t's. A score between ten and fifteen points means that the student simply has to brush up his or her mathematical knowledge. A score between six and ten points indicates holes in mathematical knowledge that can still be repaired in the first four weeks when students practise the basics of mathematics. A student who scores less than six points is advised to talk with his or her student tutor to make a work plan because there may be serious mathematical deficiencies.

Step 2. During the first four weeks of the Calculus 1 course that all students take, they have weekly a session of two hours to practise basic mathematics. The students make exercises in which they do basic algebraic manipulations, do simply calculus work such as computing derivatives of functions and limits of sequences, and learn again about properties of elementary functions and how to use them. In this set-up, we hope that students can brush up their mathematical knowledge and skills, and that they can reach the mathematical level that is needed for making a good start with their mathematics or science study.

Step 3. In the fifth week of the study year we let the students take the second diagnostic mathematics test. The test is taken in digital format only because then we can make it by randomization of each question similar to the first test. Through this pretest-posttest design, students and staff can see the progress made in the meantime. The test is held in the middle between the start of courses and the first intermediate examinations. Students can practise in the weeks before as much as they want with instances of this diagnostic test that are made available through the virtual learning environment Blackboard. We present the students' results in the second diagnostic test in the form of a study advice: a student who scores less than ten points is supposed to participate in the remedial mathematics courses given by two tutors. A student who scores between ten and fifteen points writes down a personal study plan that describes in concrete terms what the student is going to do to reach the required mathematics level. A student who scores fifteen points or more only has to take care that the present mathematics level is maintained.

Step 4. Two tutors, who are third-year mathematics students, take care of remedial teaching of mathematics. Furthermore, freshmen work in small self-help groups on mathematics problems under guidance of a student tutor. The hope is that the freshmen can still pass the first calculus exam by the special attention given to them. A recently published exercise book [7], of which most chapters are electronically available, is used for this purpose.

3 Computer algebra based diagnostic testing

In this section we give a short overview of the computer algebra based assessment system Maple T.A. for Blackboard [8]. Maple T.A. for Blackboard can be shortly described as a Blackboard-integrated system for creating tests, assignments and exercises, allowing automatic assessment of student responses and creation of model solutions via the computer algebra package Maple [9]. The integration with Blackboard concerns the user administration (authentication and grade book).

While some systems for computer algebra based testing and assessment already exist or are under development (for example, AiM [12] and Stack [13]), the main reasons for us to choose the commercial system Maple T.A. for Blackboard are that it

- is developed far enough to have a rich set of question types and assignment types;
- is based on two components, viz., Maple [9] and EDU Campus [14], which have been proved successful in educational practice;
- can be used with large groups of students without a high load on the computer environment;
- is integrated with the virtual learning environment Blackboard, currently in use at our university;
- can be used by students and instructors without much difficulty after a short introduction;
- offers licensed use of a familiar computer algebra system in an assessment environment.

Below, we will concentrate on the possibility of creating randomized mathematical free-response questions from a single question template. A detailed and more complete account of Maple T.A. in the context of diagnostic testing has been presented in [15].

In addition to the use of existing question banks, an instructor can create a bank of questions that is well-suited for assignments or tests in his or her particular course. There are several ways to author questions, but we prefer the LaTeX authoring mode because it supports advanced question types, it allows off-line editing, and it supports re-use of many course materials already written in this typesetting language. It lowers the learning curve of many a university teacher who wants to use the assessment system. One only has to convert via a web-facility the LaTeX code into the internally used EDU format.

Many of the question types available in Maple T.A. will be familiar to users of existing computer-aided assessment systems. Common examples of closed question types are multiple choice, multiple selection, true/false, ordering, clickable image, and matching questions. For these types answers are predefined, and can be marked automatically without difficulty. Common examples of open question types are fill-in-the-blanks (text or numerical value), essay, and graphical sketch problems. For these free-response questions it is more difficult to mark automatically.

But specific to Maple T.A. is the availability of mathematical free-response question types: (restricted) formula, multi-formula, and Maple-graded questions. The latter type is the one that we will discuss more deeply in this paper because, we believe that free-response questions are indispensable in evaluating mathematical knowledge and skills. It is a common misbelief to think that only standard mathematical questions leading to a clear and unique answer can be automatically assessed and that higher mathematical skills are out of reach in this approach. We refer to [16] for examples in which a computer algebra system is used to verify mathematical properties of given answers (out of an infinite number of possible, correct answers) and to assess advanced mathematical skills.

We will use the following typical example to discuss issues of randomization of questions and of marking and commenting on students' responses. It is the LaTeX code of a question in the posttest.

```
\begin{question}{Maple}
\name{removal of brackets} \type{formula}
\qutext{Remove the brackets and simplify: \var{MMLpoly}.}
\maple*{
  expr := $RESPONSE;
  evalb(simplify(expr-($res))=0) and
  type(expr, expanded) and
  evalb(StringTools:-CountCharacterOccurrences("$RESPONSE", "-")=1);
}
\code{
  $i=rint(5); $v1=switch($i,a,p,r,u,x); $v2=switch($i,b,q,s,v,y);
  $e=range(2,5); $n=range(2,5);
  $poly=maple("(($v1)^($e)-($n)*($v2))^2");
  $res=maple("expand($poly)");
  $MMLpoly=maple("printf(MathML:-ExportPresentation($poly))");
  $MMLres=maple("printf(MathML:-ExportPresentation($res))");
}
\comment{
  The correct answer is \var{MMLres}.\newline You can get this answer
  by removing brackets and collecting terms systematically.
}
\end{question}
```

The above code listing illustrates that Maple is not only used to check the student's response, but also to calculate the answer itself, to provide feedback, and last but not least to create randomised questions. In other words, the teacher does not calculate the correct answer, mark a student's answer or provide feedback: these are done by the computer algebra system.

In the first two lines of the `\code` part, several random variables are defined to create the polynomial $({v_1}^e - n\,v_2)^2$, where $e \in \{2,3,4,5\}$, $n \in \{2,3,4,5\}$, and $(v_1, v_2) \in \{(a,b),(p,q),(r,s),(u,v),(x,y)\}$. This means that we have in fact created here eighty exercises of similar type. This randomization of questions makes it possible to generate many tests with questions of similar types, which is ideal for pretest and posttest designs. But also in self-assessment it is very useful that students can redo questions of similar type until they have reached the requested level of understanding. In the last four lines of the `\code` part, Maple is used to compute the polynomial (`$poly`), the correct answer (`$res`) and MathML expressions (`$MMLpoly`, `$MMLres`) for pretty display in the question and the feedback, using standard mathematical notation. The dollar variables defined in the `\code` part can be used in other parts of the question. For example, when the question gets instantiated, then the `\var{MMLpoly}` in the question text (the argument of `\qutext`) is replaced by the MathML code stored in the variable `$MMLpoly`.

The `maple*` code specifies how the Maple system will verify automatically whether the answer given by a student to a free-response mathematical question is algebraically equivalent to the correct answer as predefined by the author. Like in many other Maple-graded questions, we subtract the author's answer from the student's answer, simplify the intermediate expression, and test whether it is zero. The code is here a little bit more complicated because we also request that the student's answer is in expanded form and has been simplified, i.e., contains no terms that could have been collected.

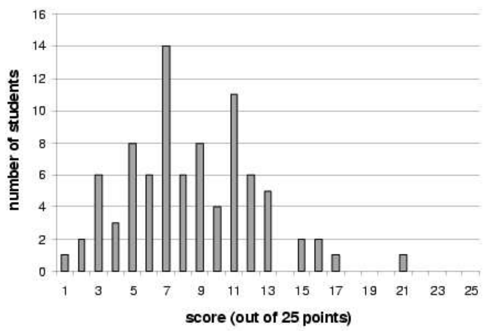
The feedback to the student is in our sample question just the correct answer, together with a short description of how the problem can be solved; a model solution has not been provided in this example. There exist a few options in Maple T.A. to give feedback that intelligently depends on the answer given by the student. For example, multiple choice and multiple selection questions allow a different comment for each option.

4 Our experiences with diagnostic testing of freshmen

In this section we discuss the mathematical performance of the freshmen in the pretest and posttest taken in the first weeks of the study year. We also look at common misconceptions in mathematics found amongst beginning students and at their progress made in the first four weeks. We address the psychometric qualities of the assessments and their predictive value with regard to identifying students at risk of failing course because of mathematical deficiencies. Finally we discuss the students' opinions about the diagnostic testing.

4.1 *Analysis of students' results*

The overall scores of the freshmen at in the pretest and posttest are shown in the two histograms of figure 1. The progress of individual students between pretest and posttest is represented in the scatter diagram of figure 2: each dot represents the score of a student at the pretest and posttest.



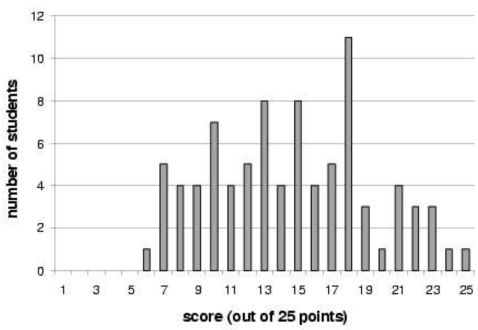


Figure 1. Histogram of students' result in pretest (left) and posttest (right).

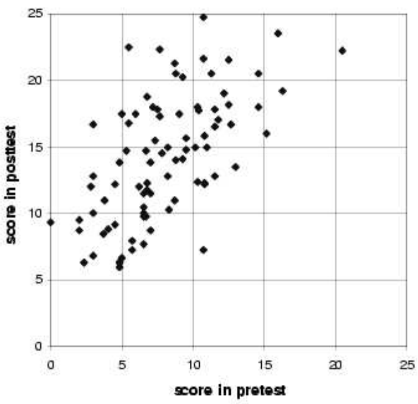


Figure 2. Scatter diagram of students' scores in pretest and posttest.

Needless to say that these low scores of beginning students worry university educators. Even though some students make remarkable progress in four weeks of remedial training, still only one-third of the student population entering the exact sciences programme of the university has reached a desirable mathematical level. The students' difficulties are of different kinds; we focus here on the basic and necessary mathematical techniques and computational skills. Some common mistakes made by students at the pretest are investigated in more detail below by looking closely at the pencil-and-paper work handed in.

4.2 *Computational mistakes*

We noticed that the incoming students had underdeveloped computational skills, even on simple calculations with fractions. Twenty-five percent of the students could not solve the first exercise without a calculator.

Question 1. *Simplify as much as possible:* $\dfrac{1}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}}$.

In regard to this question of simplifying a fraction we noticed the following three conceptual problems:

(i) Seventeen students failing on this question did not seem to know how to go on when they had added the three fractions in the denominator. They wrote down answers that are in itself correct such as $\frac{1}{\frac{13}{12}}$ and $\frac{1}{1\frac{1}{12}}$. We can only conclude that these students had forgotten to compute with fractions at routine level, maybe because the last three years at school they could use the calculator for this purpose. The second answer suggests that some of the students seemingly had been trained at school to separate the integral part of a rational number from the fractional part, but this notation is inconvenient for further computations.

(ii) Five students showed the misconception of linear reasoning [17] and gave the answer 9 by using the formula $\frac{1}{a+b} = \frac{1}{a} + \frac{1}{b}$. Such improper use of linear reasoning, leading to formulas such as $(a+b)^2 = a^2 + b^2$ and $a/b = \ln(a)/\ln(b)$, is found more and more in students' mathematical work. Although students can easily convince themselves that these formulas cannot be correct by substituting numbers, they seem not to be aware of it.

(iii) One student gave the answer $\frac{1 \cdot 24}{2+3+4}$. This can be seen as a case of overgeneralizing a mathematical formula that is in itself correct in two unknowns to the erroneous case of three or more unknowns. Indeed

$$\frac{1}{1/a + 1/b} = \frac{1 \cdot ab}{a+b},$$

but this does not generalize to

$$\frac{1}{1/a + 1/b + 1/c} = \frac{1 \cdot abc}{a+b+c}.$$

Two-third of the students failed on the second exercise of the pretest:

Question 2. *Simplify as much as possible:* $\left(\frac{6}{5}\right)^{-3} \times \left(\frac{3}{10}\right)^{2}$.

With regard to this exercise of calculating a product of powers of fractions we noticed the following two mistakes:

(i) Improper application of rules for multiplication of powers. The biggest problem is that students often mix the rules for multiplication of powers with variable bases and exponents. The following answers illustrate this:

$$\left(\frac{6}{5}\right)^{-3} \times \left(\frac{3}{10}\right)^{2} = \left(\frac{12}{10}\right)^{-3} \times \left(\frac{3}{10}\right)^{2} = \left(\frac{36}{100}\right)^{-1},$$

$$\left(\frac{6}{5}\right)^{-3} \times \left(\frac{3}{10}\right)^{2} = \left(\frac{18}{50}\right)^{-1} = \left(\frac{9}{25}\right)^{-1},$$

$$\left(\frac{6}{5}\right)^{-3} \times \left(\frac{3}{10}\right)^{2} = \left(\frac{12}{10}\right)^{-3} \times \left(\frac{3}{10}\right)^{2} = \left(\frac{15}{10}\right)^{-1} = \left(1\frac{1}{2}\right)^{-1}.$$

(ii) Misunderstanding of mathematical notation of powers. Some examples:

- Confusion with scientific notation: $\left(\frac{6}{5}\right)^{-3} = \frac{6}{5} \times 10^{-3}$.
- Interpreting a negative integral exponent as repeatedly taking reciprocal values: $\left(\frac{6}{5}\right)^{-3} = \frac{1}{\frac{1}{\frac{1}{5/6}}} = \frac{1}{\frac{1}{6/5}} = \frac{1}{5/6} = 6/5$.
- The power of a fraction equals the fraction obtained by multiplying numerator and denominator with the exponent: $\left(\frac{6}{5}\right)^{-3} \times \left(\frac{3}{10}\right)^{2} = \frac{18}{15}$.
- A negative power of a fraction equals the fraction obtained by taking positive powers of numerator and denominator, multiplied by minus one: $\left(\frac{6}{5}\right)^{-3} = -\left(\frac{6}{5}\right)^{3} = -\frac{216}{125}$.

4.3 *Algebraic mistakes*

In view of the many difficulties that freshmen already had with pencil-and-paper computations with fractions it will not be a great surprise that algebraic manipulations revealed an even greater variety of misconceptions. Occasionally, students got correct answers using incorrect symbolic manipulations. An example was the following question:

Question 8. *Simplify as much as possible:* $\dfrac{9r^2 - 4s^2}{3r + 2s}$.

This exercise could not be routinely done by many students because the special product rule $a^2 - b^2 = (a - b)(a + b)$ was not generally known or internalized. Another reason could be that students missed the 'Gestalt view' to notice this rule [19]. This means that they did not have the 'global substitution principle' at their disposal that would make them consider here $3r$ and $2s$ as entities. However, many university teachers wrongly assume that their students know this factorization rule by heart and can apply it routinely, even in cases where it is present in disguised form such as in the formula $\dfrac{4^x-1}{2^x+1}$ (Question 19 in the test). The main reason that thirty students failed on Question 8 seemed that the special product rule is not present on the chart of standard mathematical formulas that students use and that little or no attention has been paid to it in the Dutch curriculum; all depends on whether mathematics teachers at school found it worthwhile to discuss and practice this rule thoroughly. Nevertheless, the algebraic complexity of this rule is simple enough and the mathematical background of the students should be sufficient to figure out what to do in this case.

Students made cancelling errors. For instance, four students reasoned as follows:

> 'Division of $9r^2$ by $3r$ gives $3r$ and from this I must subtract the division of $4s^2$ by $2s$, which gives $3r-2s$. Thus, $\dfrac{9r^2-4s^2}{3r+2s}=\dfrac{9r^2}{3r}-\dfrac{4s^2}{2s}=3r-2s$.'

Implicitly, we see a kind of 'rule' $\dfrac{a^2-b^2}{a+b}=\dfrac{a^2}{a}-\dfrac{b^2}{b}$. Unfortunately, this step did lead to the correct final result. This example shows clearly that assessment in which only final results of work are asked and nothing of the process that led to the answers is recorded runs the risk of not identifying some of the misconceptions or alternative concepts that students may have. Other cancelling mistakes of students are:

- Cancelling terms based on the 'rule' $\dfrac{a^2-b^2}{a+b}=\dfrac{a-b}{1+1}$ gives $\dfrac{9r^2-4s^2}{3r+2s}=\dfrac{\cancel{9r^2}\,3r-\cancel{4s^2}\,2s}{\cancel{3r}\,1+\cancel{2s}\,1}=\dfrac{3r-2s}{1+1}$.
- Cancelling symbols by applying the previous rule only for the symbols: $\dfrac{9r^2-4s^2}{3r+2s}=\dfrac{9r^{\cancel{2}}-4s^{\cancel{2}}}{3\cancel{r}+2\cancel{s}}=\dfrac{9r-4s}{5}$.

Other mistakes found in students' answers to Question 8 were:

- Improper use of linear reasoning: Application of the 'rule' $\dfrac{1}{a+b}=\dfrac{1}{a}+\dfrac{1}{b}$ explains the mistake $\dfrac{9r^2-4s^2}{3r+2s}=3r-\dfrac{4s^2}{3r}+\dfrac{9r^2}{2s}-2s$. But maybe it is just a matter of systematic cancelling errors because we noticed this kind of errors many times in students' work.
- Improper use of common denominators $\dfrac{a-b}{c+d}=\dfrac{a}{cd}-\dfrac{b}{cd}$ gives $\dfrac{9r^2-4s^2}{3r+2s}=\dfrac{3r}{2s}-\dfrac{2s}{3r}$.

4.4 *Other common misconceptions*

Other common mistakes and difficulties found in the students' work were:

- The balance misconception [18], i.e., $A(x) = B(x) \Rightarrow op[A(x)] = op[B(x)]$ for every operation *op*. For example in answers to Question 4, in combination with the related misconception of linear reasoning, appears the statement $\dfrac{1}{f} = \dfrac{1}{a} + \dfrac{1}{b} \Rightarrow f = a + b$.
- Missing the 'Gestalt view' to notice for example in Question 18 that the equation $e^{2x} - 2e^{x} - 3 = 0$ may be considered as a quadratic equation in e^{x}. Probably this is related to the next difficulty.
- Problems with recognizing and checking quickly and confidently equivalent algebraic expressions. Not knowing the rule $a^2 - b^2 = (a - b)(a + b)$ or not knowing how to apply this rule in concrete situations belongs to this category of problems, in our point of view. Students also seemed to be too dependent on the chart of standard mathematical formulas.
- Lack of 'symbol sense' [19] and difficulties with keeping track of the overall problem-solving process while executing elementary algebraic procedures that are part of it. For example, when asked in Question 21 to compute the derivative of $\dfrac{x^2 - 4}{x - 2}$, every student immediately started to apply the quotient rule for derivation and did not consider simplification of the given formula in the first place in order to avoid cumbersome algebraic simplification.
- Application of clear mathematical rules, for example rules for removing of brackets, worked well in simple cases, but not when the complexity was increased, for example by increasing the number of terms or the number of variables. This explains the difference in students' performance regarding Question 6 and 7.
- Unfamiliarity with the compact algebraic notation with its specific conventions and symbols. For example, many students confused the notation of inverse function $f^{-1}(x)$ with that of reciprocal value $\frac{1}{f(x)}$. Two-third of the students could not determine $\sum_{k=-10}^{10} k$ (Question 9), either because they did not understand the summation symbol or because the negative lower limit confused them. Students did not known that one can define a function in terms of another one, like $g(x) = f(2x - 1)$ in Question 22.
- Overgeneralization of mathematical rules. For example, the rule $\ln(x)' = 1/x$ on the formula chart was often used by students in Question 20 to compute $\ln(1 - t^2)'$ as $1/(1 - t^2)$.
- The obstacle of reversing a mental process ([20]). When given properties of a function, like in Question 16, many students had difficulties in constructing a good example. We expect that they would not have had so many difficulties

with the reverse problem, viz., verifying that a given function satisfies certain properties.

- Lack of sufficient mathematical knowledge, well-trained skills, and effective attitudes for solving nonroutine exercise problems ([21]). Only one student was able to give a correct answer to Question 15, in which the correct response can be found by adding extra terms in the expression.

At this point it is good to mention that many students manage to overcome the problems initially faced when entering the university, despite their weak foundation in mathematics and their view on algebra and other mathematical topics as a menagerie of disconnected subjects and techniques at that moment. It is more an issue that we want to lower dropout rates by helping students to make in the first study year the necessary steps from superfluous knowledge of mathematics to deep understanding of what mathematics comprises at university level.

4.5 *Psychometric analysis*

We believe that a psychometric analysis is important because the students' population has new characteristics, the technology is new, the didactic scenario is new, the tests have been newly constructed, and because the topics comprehend quite a large part of the secondary school mathematics. So in this section, we address the psychometric qualities of the assessments and therefore we are concerned about questions like: 'Does it measure what it should measure?', 'Is the test about the right content?', 'Are the questions well posed?', 'Are the questions too difficult or too easy?', 'Do the questions discriminate between the higher and lower mathematical abilities?', 'Are the outcomes significant?', and 'What is the predictive value for other first year courses?'.

First of all, to ensure the content validity of the assessment, the questions have been reviewed by experts, both in the field of the mathematics of secondary education, as in the mathematics of the first year of university. Their general opinion was that the content was valid. The reviewers remarked that they would expect that the freshmen would score at least ten points out of twenty-five. The opinion of the teachers is that at the end of the trajectory, a score of two-third of the questions is desirable. A calibration of this norm has not yet taken place; for instance this can done by asking second-year students to take the test.

The item difficulty index (*p-value*) measures the proportion of examinees who answered the item correctly. In fact, the p-value is higher if more students answer the questions correctly, so it should rather be called the item easiness index. In principle, a very high or low p-value suggests that the question is not useful as a measurement instrument.

The item discrimination index (*Rit*-value) is a measure of how well an item is able to distinguish between examinees who are knowledgeable and those who are not. Usually, this is measured by the correlation between the student's performance on the given item (correct or incorrect) and the student's score on the overall test.

The psychometric measure of reliability regards the internal consistency of the test. It can be measured by *Cronbach's alpha*, which quantifies the extent in which patterns emerge in the answers by the students.

The pretest had the same form for all students, so a straightforward statistical analysis is possible. However, in the posttest the questions were randomly picked from a pool. Unfortunately, Maple T.A. did not give useful data on a per item basis, so a precise psychometric analysis was only possible for the pretest.

Table 1. Psychometric data for the pretest.

Topic	*Question*	*p-value*	*Rit-value*
Computing with numbers	1	.7733	.458
	2	.3058	.547
Substitution	3	.7170	.472
Reciprocity	4	.2907	.500
Computing with symbols	5	.1100	.327
Removing Brackets	6	.7713	.283
	7	.6570	.377
Rational Expressions	8	.3023	.248
Series and Limits	9	.3372	.354
	10	.3140	.025
Linear Equations and Vectors	11	.3344	.434
	12	.4302	.459
	13	.3934	.343
Functions	14	.1530	.507
	15	.0174	.256
Elementary Functions	16	.2384	.515
	17	.4343	.551
	18	.1279	.445
	19	.0930	.301
Differentiation	20	.1987	.561
	21	.2411	.432
	22	.1163	.283
	23	.4070	.261
Integration	24	.1879	.408
	25	.1395	.190
Cronbach's alpha		.7508	

The data collection was restricted to those students that did both the pretest and the posttest. The number of students involved in the test was 85. In table 1 we present the psychometric data for the pretest. The questions can be found in the appendix. The overall conclusion is that reliability is sound and that most questions have good difficulty and discrimination indices. Because the context of the tests is the mastery of mathematics abilities, a high p-value at the beginning of the trajectory means that the ability is already present among many of the students. This is the case for Questions 1, 3 and 6, but it should be noted that these are very basic questions, and a much higher p-value is desirable. A low p-value in this stage does signal the students that the particular ability is desirable, meanwhile there is work to be done (Questions 5, 14, 15, 18, 19, 20, 22, 24, 25). For instance, Question 15 about limits proved to be so difficult as this topic is hardly taught in secondary education. In the posttest, the score on this topic was much higher (about 0.5). Question 19 was about a special product within a function definition. Special products are not yet recognized by incoming students. Similarly, a high p-value at the end of the trajectory states that the ability is present among the students (Question 6, 7, 8). A low p-value in this stage means that although the ability may be desirable, the trajectory did not lead to its mastery (Question 5, 16, 17). So there is work for the teachers is to be done.

A low discrimination index (e.g., Question 10) means that is does not help to discern between the good and the bad performers, so the item is for a pretest only useful to signal the students that this is something they should know. In this case, the correct answer was the number zero, which is a logical candidate for a wild guess or a simple trial by substituting numbers. Probably, a better test will come out if the item is replaced by a similar one whose answer is not so likely to be guessed. The Questions 22 and 25 are multiple choice questions with rather low discrimination indices. The p-value happens to be below the guess ratio. Typically, for multiple choice questions, an approach is taken where the p-value is much larger that the guess ratio, because this will improve the discrimination index. So these two questions are candidates for improvement.

For the first 10 weeks period, we collected the mid-course data from the other courses that were followed by all students. It concerns a course in Calculus, an course "Symmetry and Pattern Formation in Nature" (SPIN) and a training course in the software system *Mathematica* [22]. The posttest correlated significantly with these courses, with a remarkable high correlation with the Calculus course (0.66 for Calculus, 0.37 for SPIN and 0.41 for *Mathematica*). We also calculated the correlation between the progress made from pretest to posttest, using the 'normalized gain' that is based on the formula

$$\textit{normalized gain} = \frac{\textit{posttest score} - \textit{pretest score}}{\textit{maximum possible score} - \textit{pretest score}}$$

and this also correlates quite significantly with the mid-term results of the Calculus course (0.49) and of the SPIN course (0.30). This indicates that substantial progress is a good indicator for success in other courses.

Figure 3 is the scatter diagram with the score of the posttest in the horizontal direction (max. 25 points) and the endterm score of the calculus course in the vertical direction (normalized to the interval $[0, 1]$). It is quite striking that most data points are above a diagonal, implying that a good score in the posttest, implies at least a similar endterm score in Calculus. The scatter diagram for the other courses show a similar pattern. The correlation coefficients between the posttest score and the endterm scores of Calculus and SPIN are 0.55 and 0.48, respectively. All this supports the opinion that a mastery of mathematical abilities is a *sufficient condition* for study success. However, a more longitudinal and extensive study is required for a more distinct statement.

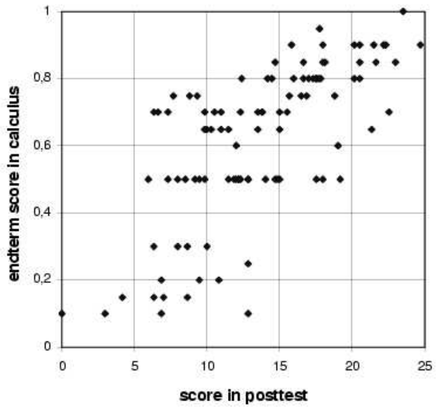


Figure 3. Scatter diagram of endterm Calculus scores versus posttest scores.

4.6 *Students' opinions about the diagnostic testing*

Just after the posttest, the students were asked to give their opinions by means of a survey. The response was sixty-five percent.

The students found that the goal of the diagnostic tests were sufficiently clear (90%). The chosen approach was appreciated by 86%. They confirmed that the tests helped to identify shortcomings in their mathematical abilities (69%), and also that the tests contributed to making progress with mathematics (68%). About 63% of the students found the tests stimulating to set to work. A minority (42%) said that they obtained more insight in the difference between school and university mathematics.

About the instrumentation, 61% of the students found that they could work well with the Maple T.A. software. This is remarkable, since there was no training beforehand. Students were permitted to use the graphing calculator and the standard formula chart. A small majority of the students (63%) thought that this influenced their score positively. It also means that a large number of students do not feel dependent on the graphing calculator and the formula chart, while teachers noticed during practice sessions that the students used them many times, even in case when it seems inappropriate.

At the end of the survey there was an open field for suggestions and remarks. About half of the students used this field. Nine students made it clear they preferred a test on paper instead of the computer; an important reason they mentioned is that they could only fill in the end result and no intermediate steps. Maybe they felt disadvantaged by not been given clearly the opportunity to earn partial credit (although we did this in fact at a manual check of the grading). Six students made a positive comment about the usefulness of the approach. Two students made a remark about the planning of the pretest as it was planned just before the very first week of the start of first courses. This happened to be the period of the university introduction with all kind of activities and festivities. Arguably, a pretest was not appreciated at that moment.

5 Future developments

A consolidation of the current approach is the first step to come. This means to inform and involve the staff to a higher extent, and a tighter integration with the mathematics courses of the first year.

There are many more possible locations in the curriculum where the technology can be deployed. In the course where *Mathematica* [22] is taught as a tool for mathematics, some classes now start with a small Maple T.A. test to refresh the relevant mathematical knowledge.

The effort of developing tests is considerable. Because of the complexity of the total system with several application interfaces involved, and the inherent complexity of mathematics, developing a question template can take up to an hour or so. So it makes sense to combine forces and to exchange materials and didactic approaches. The University of Amsterdam actively promotes collaboration and is now involved in the SURF-project Web-spijkeren [10] where didactic scenarios and assessment instrumentation are tried out, and in the DU-project MathMatch [11] where materials are developed. Recently, an initiative is started for a SURF Special Interest Group about the mathematical problems of the transition to higher education.

6 Conclusion

The main question to address in the conclusion is whether the approach described in this paper does solve the problem of the influx of freshmen with heterogeneous mathematics level in a staff-extensive way. As can be seen from the substantial progress of the students, the trajectory does lead to an overall considerable amelioration of the mathematical abilities.

On the whole, the students appreciated that they were confronted with the mathematics abilities as desired by the universities, and were informed about their own level. Although the first assessment was even before the first course day, the students did not feel uncomfortable with the fact that they were assessed. As can be seen from the questionnaire, they understood well the purpose of the testing. We believe that for this the repeated information to students, tutor students and staff is very important.

The teachers of the other courses for the freshmen appreciated the extra attention to the necessary mathematics abilities. The significant correlations with the other first year results support the relevance of the mathematical abilities for these courses. But a longer pilot period is necessary for conclusive results about this.

The Maple T.A. system is able to generate and assess questions from the templates in the database. So the whole trajectory can be based on a single test template: it is used for the entry assessment, the formative exercises and the diagnosis tests. This is a very strong advantage of the approach with a tremendous potential for saving time of staff. However, the current version of Maple T.A. showed some peculiarities in the submission process and in the automatic assessment by Maple of formula based items. These forced us to check manually the student results, so in this pilot the time efficiency requirement was not met. Nevertheless, we believe that once the peculiarities of the system have been removed, also the staff-extensive requirement comes close.

Acknowledgments

The authors would like to thank their colleagues Jan van de Craats, Peter de Paepe, and Lidy Wesker for their help in the development of the pretest and posttest, either by making fruitful comments on the exercises, or by suggesting or providing new exercises. We much appreciated the help of our colleague Wolter Kaper in setting up the Maple T.A. for Blackboard in short time and in running the tests. We also thank Louise Krmpotic and Robin Coe of Maplesoft Inc. for providing us, before official release, with the version 2.5 of Maple T.A. with Blackboard support and for helping us out with technical problems when needed.

Appendix A: First diagnostic test

(1) Simplify as much as possible: $\dfrac{1}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}}$.

(2) Simplify as much as possible: $\left(\dfrac{6}{5}\right)^{-3} \times \left(\dfrac{3}{10}\right)^{2}$.

(3) For kinetic energy E of a particle with mass m and velocity v holds: $E = \frac{1}{2}mv^2$. For the momentum p holds: $p = mv$. Express E in m and p, and simplify as much as possible.

(4) The Gaussian Lens Formula is: $\dfrac{1}{f} = \dfrac{1}{b} + \dfrac{1}{v}$. Write f as a quotient of polynomials in b and v.

(5) Simplify as much as possible: $\dfrac{a^2(8b)^{\frac{1}{4}}c^{\frac{1}{2}}}{\sqrt{abc^3\sqrt{2}}}$, assuming that a, b and c are positive numbers.

(6) Expand brackets and simplify: $(2a+3b)(3a-2b)$.

(7) Expand brackets and simplify: $(2x-y+z)^2 - z^2$.

(8) Simplify as much as possible: $\dfrac{9r^2-4s^2}{3r+2s}$.

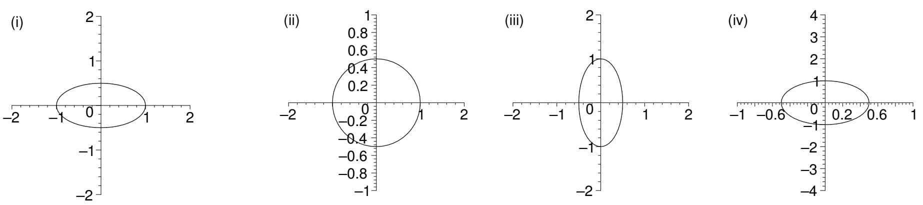
(9) Compute the exact value of the sum: $\displaystyle\sum_{k=-10}^{10} k$

(10) Given is a positive number a. Compute the exact value of the limit: $\displaystyle\lim_{n\to\infty} \frac{n^2-2^n}{an^2+e^n}$.

(11) Given is the system of two equations in two unknowns x and y and two constants a and b:
$\begin{cases} 2x+3y=a \\ 4x+6y=b \end{cases}$
Mark for each of the conclusions below whether it is true or false:
(i) For some values of a and b the system has no solutions.
(ii) For some values of a and b the system has exactly one solution.
(iii) For some values of a and b the system has an infinite number of solutions.

(12) Given is the point P with coordinates $(1, 2)$. We translate the coordinate system over the vector $\begin{pmatrix} -1 \\ 3 \end{pmatrix}$ and leave the point P at its current position. What are the coordinates of P in the new coordinate system?

(13) Which of the following graphs belong(s) to the curve described by the equation $x^2+4y^2=1$?



(14) For $x > -2$ is the function $f(x) = \sqrt{1+\dfrac{x}{2}}$. The formula of the inverse function is then: $f^{-1}(x) = \ldots$

(15) We assume that the standard limit $\displaystyle\lim_{x\to 0} \frac{\sin(x)}{x} = 1$ is known.
Given is a number $a \neq 0$. Compute the exact value of the limit: $\displaystyle\lim_{x\to 0} \frac{\sin(ax)}{\tan 2x}$.

(16) Determine the quadratic function $f(x)$ with stationary point $(1, -4)$ and a zero in $x = 2$.

(17) Compute the exact values of all solutions of the equation: $x^3 + 14x^2 = 72x$.

(18) Compute the exact values of all real solutions of the equation: $e^{2x} - 2e^x - 3 = 0$.

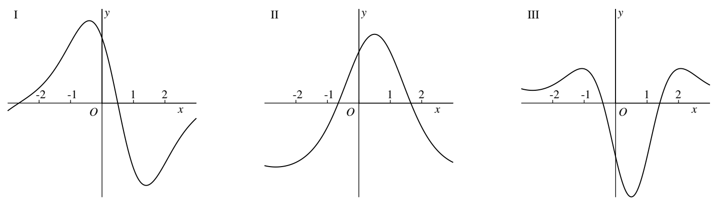
(19) Simplify as much as possible: $\dfrac{4^x - 1}{2^x - 1}$.

(20) Differentiate with respect to t and simplify your answer as much as possible: $\ln(1 - t^2) + t^2$.

(21) Differentiate with respect to x and simplify your answer as much as possible: $\dfrac{x^2 - 4}{x - 2}$.

(22) Given are a function $f(x)$ with $f'(1) = 3.8$ and the function $g(x) = f(2x - 1)$. What is $g'(1)$? Select one of the following responses.
◯ 1.9 ◯ 2.8 ◯ 3.8 ◯ 4.8 ◯ 7.6 ◯ For this you must know the formula of $f(x)$.

(23) Below are drawn in random order the graphs of the function $f(x)$, its derivative $f'(x)$ and the second derivative $f''(x)$. Identify them.



(24) Compute the exact value of $\displaystyle\int_1^2 (x^3 + \sqrt{x})\, dx$.

(25) Given is a function $f(t)$ with $\displaystyle\int_1^3 f(t)\, dt = 2.4$. What is $\displaystyle\int_3^5 f(t - 2)\, dt$?
Select one of the following responses.
◯ 0.4 ◯ 2.4 ◯ 4.4 ◯ For this you must know the formula of $f(t)$.

References

[1] London Mathematical Society, 1995, *Tackling the mathematics problem.* Available online at: `www.lms.ac.uk/policy` (accessed 8 December 2005).

[2] SEFI Mathematics Working Group, 2002, L. Mustoe and D. Lawson (Eds): *Mathematics for the European engineer — a curriculum for the twenty-first century.* Available online at: `http://learn.lboro.ac.uk/mwg/core/latest/sefimarch2002.pdf` (accessed 8 December 2005).

[3] Kent, P., and Noss, T., 2003, *Mathematics in the university education of engineers. A report to the Ove Arup Foundation.* Available online (accessed 8 December 2005) at: `www.theovearupfoundation.org/arupfoundation/pages/download25.pdf`

[4] Engineering Council, 2000, *Measuring the mathematics problem.* Available online at: `www.engc.org.uk/documents/Measuring_the_Maths_Problems.pdf` (accessed 8 December 2005).

[5] Hoyles, C., Newman, K., and Noss, R., 2001, Changing patterns of transition from school to university mathematics. *International Journal of Mathematics Education in Science and Technology*, **32**, 829–845.

[6] Nardi, E., 1996, *The novice mathematician's encounter with mathematical abstraction: tensions in concept-image construction and formalisation.* Dphil, University of Oxford. Available online at: `www.uea.ac.uk/~m011/thesis/` (accessed 8 December 2005).

[7] van de Craats, J., and Bosch, J., 2005, *Basisboek Wiskunde* [Basic Mathematics]. Pearson Education Benelux. Most parts of the book are online available at `www.science.uva.nl/~craats` (accessed 8 December 2005).

[8] Maple T.A.: Automated Online Testing & Assessment. Developer's website `www.maplesoft.com/products/mapleta` (accessed 8 December 2005).

[9] Maple: `www.maplesoft.com/products/maple` (accessed 8 December 2005).

[10] Web-spijkeren: SURF Project, information (in Dutch) online available at `www.web-spijkeren.nl` (accessed 8 December 2005).

[11] MathMatch: Project of the Digital University, information (in Dutch) online available at `www.du.nl/mathmatch` (accessed 8 December 2005).

[12] AiM: Assessment in Mathematics. `http://aiminfo.net` (accessed 8 December 2005).

[13] Sangwin, C.J., and Grove, M., 2006, STACK: addressing the needs of the "neglected learners". Conference proceedings of WebALT2006, Eindhoven, 5-6 January 2006.
[14] Brownstone's EDU Campus: `www.brownstone.net/products/ecampus.asp` (accessed 8 December 2005).
[15] Heck, A., and van Gastel, L., 2006. *Diagnostic testing with Maple T.A.* Conference proceedings of WebALT2006, Eindhoven, 5-6 January 2006.
[16] Sangwin, C.J, 2004, Assessing mathematics automatically using computer algebra and internet. *Teaching Mathematics and it Applications* **23**, 813–829.
[17] De Bock, D., van Dooren, W., Janssens, D., and Verschaffel, L., 2002, Improper use of linear reasoning: an in-depth study of the nature and irresistibility of secondary school students' errors. *Educational Studies in Mathematics*, **50**, 311–334.
[18] Bagni, G., 2000, "Simple" rules and general rules in some high school students' mistakes, *Journal für Mathematik-Didaktik*, **21**, 124–183.
[19] Arcavi, A., 1994, Symbol sense: informal sense-making in formal mathematics. For the Learning of Mathematics, **14**, 24–35.
[20] Norman, F.A., and Prichard, M.K., 1994, Cognitive obstacles to the learning of calculus: a Kruketskiian perspective. In: J.J. Kaput and E. Dubinski (Eds) *Research Issues in Undergraduate Mathematics Learning.* MAA Notes **33**, pp. 65–77.
[21] Selden, J., Selden, A., and Mason, A., 1994, Even good students can't solve nonroutine problems. In: J.J. Kaput and E. Dubinski (Eds) *Research Issues in Undergraduate Mathematics Learning.* MAA Notes **33**, pp. 19–26.
[22] *Mathematica*: `www.wolfram.com` (accessed 8 December 2005).