Mathematics education and assessment

Heleen Verhage & Jan de Lange
Freudenthal Institute, The Netherlands.
Amesa conference, 1-5 July 1996.

Introduction

In a changing society changes in education will also take place. Mathematics education is changing rapidly in a number of countries as is represented in documents which are produced to formulate the principles, aims and goals of the learning of mathematics at school.

In the Netherlands, for instance, general goals of mathematics education are to:

– become an intelligent citizen (mathematical literacy);
– prepare for the workplace and for future education;
– understand mathematics as a discipline.

Of course, besides general goals like these, content related goals are formulated as well. Curriculum development projects are needed to translate these formal documents into practice. Preferably this is done in an innovative way. In several countries, like the Netherlands, Denmark and Australia, this process of renewal started some fifteen years ago and it is still going on. In other countries, like the USA and more recently South Africa, the renewal started some years later. In fact the renewal of mathematics education is a continuous process, and it will never be finished.

In many countries there is a tendency to put more emphasis on the relation between mathematics and the daily world around us. Mathematics is no longer seen as a discipline for the happy few but should as well be useful for ordinary people in daily life. This point of view has many implications for the selection of curriculum content and for the teaching and learning process. Briefly, meaningful mathematics has to be learned in a meaningful way. The NCTM, for example, formulated *'mathematical power'* as a central goal for mathematics education as follows:

– Application of knowledge to solve problems within mathematics and in other disciplines.
– Use of mathematical language to communicate ideas
– Ability to reason and analyse.
– Knowledge and understanding of mathematical concepts and procedures.
– Positive disposition towards mathematics.

In the Netherlands in the eighties the so-called Realistic Mathematics Education (RME) was introduced. By this is meant, among other things, that the mathematics taught must be derived from the reality around us and also be applicable to this reality. Realistic situations (indicated by contextual problems) are thus used as source and as area of application for mathematics education.

Another important characteristic of RME is concerned with how mathematics can be learned. Hans Freudenthal has stated that mathematics is *"a human activity and that one learns it by doing it"*. Students should, as it were, rediscover bits of mathematics and construct it themselves. They are stimulated to use their own strategies in this process. Different strategies, sometimes reflecting different levels, can be provoked and used productively in the learning process.

The problem of assessment

Very soon after the introduction of the realistic approach in the Netherlands, the mathematics community realised that assessment became much more complicated than it had been in the past. Everyone agrees on the fact that a fair test should reflect what has been learned before.

If the use of mathematics in context situations is a goal of mathematics education, then context situations, preferably new ones, should be part of the assessment. This sounds trivial, but it means that the design of, for instance, a test will require much more labour than it has in the past. Teachers have to make their own collection of contextual problems. They have to develop an eye for appropriate situations and material and collect everything that may ever be useful in the mathematics lessons.

If new contexts as a source for testing are available, then there is still the problem how to design a fair and balanced test. The most interesting questions related to context problems, are quite often very open. Does this mean that very open, or even vague, questions can be part of a test? If so, the next problem looms up.

It appears that the answers of the students to very open questions can be very different and even rather unpredictable. So how to deal with the answers of the students and how to grade such a test?

In summary, we have the following problems:

- How to find original context situations as a source for testing?
- How to design a test that reflects what has been learned?
- How to judge student work and grade the test?

The nature of these problems is rather practical. But the second problem also has a more fundamental aspect. To make this clear, in the next paragraph the different goals of mathematics education will be analysed.

Three different levels

If we analyse the goals of mathematics education, different levels can be distinguished. A possible categorization of them is described by Jan de Lange (1994). In this article he distinguishes three levels which refer to lower, middle and higher goals. Because the assessment has to reflect education, these categories can be used both for the goals of mathematics education in general and for the assessment. De Lange (1994) represents the levels of understanding in the form of a pyramid as shown in Figure 1.

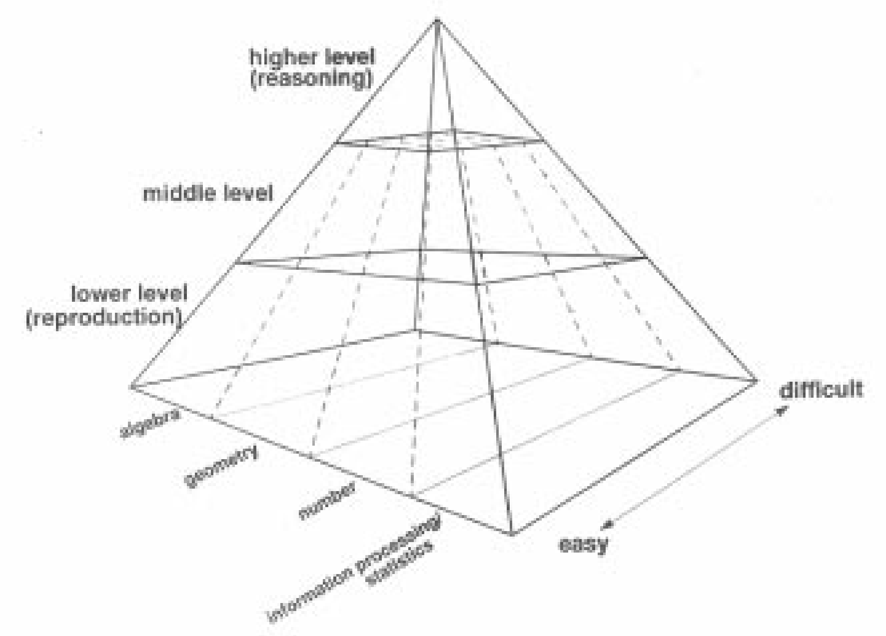


Figure 1

The lower level

This level concerns the knowledge of objects, definitions, technical skills and standard algoritms. Some typical examples are:

- adding (easy) fractions;
- solving a linear equation with one variable;
- measuring an angle using a compass;
- computing the mean of a given set of data.

But also straight forward *'real life problems'* can be at the lower level. For example:

> We drove our car for 170 kilometer and used 14 litres of gasoline. How many kilometers did we get per litre of gasoline?

According to De Lange’s categorization most of traditional school mathematics and traditional tests are at the lower level. One might think that a question at the lower level will be more easy than a question at one of the other two levels. But this need not be the case. A question at the lower level can be a difficult one. The difference is that it does not demand much insight; it can be solved by using routine skills or even by rote learning. An example of a difficult but meaningless question is the following:

> Which number is 75% of:

$$\frac{\sin^2 30^0 - \left(\frac{1}{2}\right)^2 \cdot (0.8)^{-1} + \sqrt{2.25}}{\frac{11}{20} + \left(\frac{2}{3}\right)^2 \cdot (\cos 60^0 + \tan 45^0)^2}$$

> Answer: $\frac{345}{496}$

The second level

The second level can be characterized by having students relate two or more concepts or procedures. Making connections, integration and problem solving are terms often used to describe this level. Also problems that offer different strategies for solving, or offer more than one approach to solve, are at this level. An example:

> You have driven your car 2/3 of the distance you want to cover and your tank is 1/4 full. Do you have a problem?

Children’s strategies for solving this problem are discussed and illustrated in Streefland (1994) as shown in Figure 2:

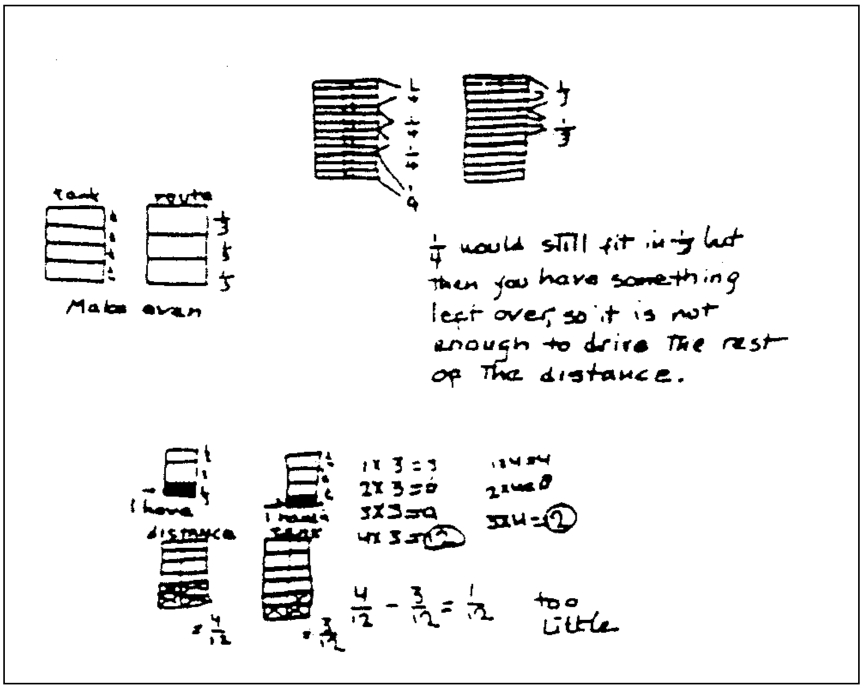


Figure 2

If this car-fuel problem is compared with the first one, it is clear that the second is much richer. To solve the first problem the children take their calculator (if they have one!) divide 170 by 14 and that's it. To solve the second problem requires much more, because they have to find their own strategy for interpreting and comparing the numbers in relation to the context situation.

The second example is taken from a final examination in the Netherlands (VBO/MAVO D, 1995-II) (see Figure 3).

Houses of cards

On the picture you see a house of cards. 15 cards were needed to make this house.

a. How many cards are needed for a four storey house of cards?

Rob plays a game with his father and mother. Whom can make the highest house of cards? The house of Rob has one more storey than mother's house. Mother has one more storey than father. Mother used 17 more cards than father.

b. How many more cards than mother did Rob use? Explain your answer.

A formula for cardhouses: $K = 1{,}5V^2 + 0{,}5V$
K is total number of cards, V is number of storeys

c. Someone says that he can build a 18 storey house. How many cards are needed?

d. You have 208 cards. How many storeys can you reach? Explain your answer.

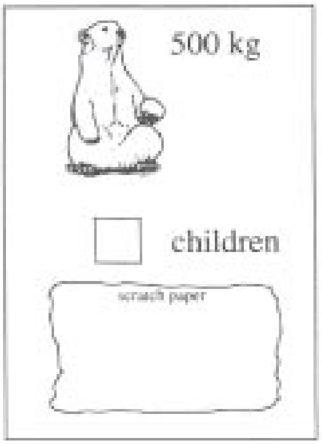
Figure 3

For these questions careful reading and some good reasoning are needed. There is quite a lot of information to read and students have to make decisions about their selection of strategies.s

The third level

The highest level has to do with complex matters like mathematical thinking and reasoning, communication, critical attitude, communication, creativity, interpretation, reflection, generalization and mathematizing. Students' own constructions are a major component of this level. A famous example from primary school is given in Figure 4 (Van den Heuvel-Panhuizen, 1996).

Test Item - Polar Bear



Instructions to be read aloud:
A polar bear weighs 500 kilograms.
How many children together weigh as much as one polar bear? (...)

Figure 4

Only the weight of the bear is given. This test item refers to the children's knowledge of measures, because they have to decide themselves on the weight of a child. So students’ own productions are needed to solve this problem, two examples of which are shown in Figure 5.

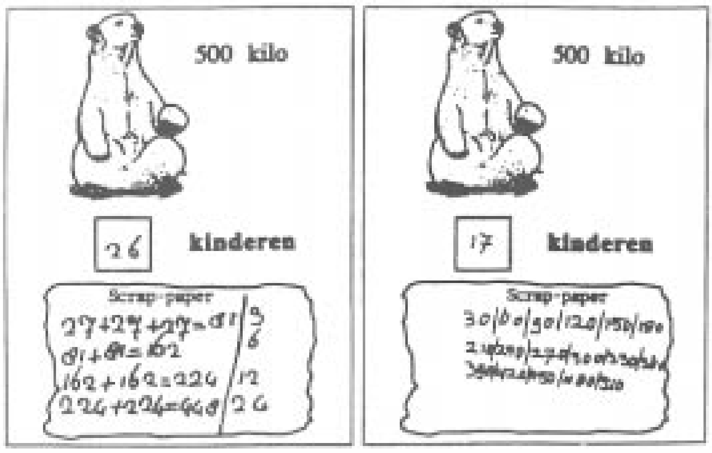


Figure 5

So far all the examples are taken from tests which took place locally or even nationwide in the Netherlands. After many years of experience it appears that it is possible to design tests with a reasonable number of questions at the lower and second levels. Good test items at the highest level, however, are hardly found. To assess the higher level thinking skills, different test formats, like project work, are probably needed.

One of the tasks given to the students in a project on ‘*independent learning'* was to find out whether all eggs have the same shape or not. A data set with the length and width of a number of eggs was given (students can use the computer to analyse these data). It appeared that the relation between length and width of the eggs can be approximated by a linear equation. Students own productions play a major role here. In fact the process is more important than the product in this instance. Tasks like this need a special way of organisation. It is nearly impossible to let students do such a project during one lesson of fifty minutes.

Assessment principles

The following definition of assessment is given by Gardner (1992):

“The obtaining of information about the skills and potentials of individuals, with dual goals of providing useful feedback to the individuals and helpful data to the surrounding community.”

Of course learning this definition by heart as such is an activity of the lower level!

De Lange (1987) describes five important principles underlying the ideas on assessment in realistic mathematics education, the first principle of which is that *‘the first and main purpose of assessment is to improve learning and teaching’*. Although this is not a new idea, it is easily underestimated in the teaching-learning process. Assessment is all too frequently seen as an end-of-chapter or end-of-course activity whose primary purpose is to serve as a basis for assigning grades. A properly designed assessment task should not only motivate students by providing them with short-term goals towards which they work, but also provide them with feedback concerning the learning process.

The second principle is that *‘methods of assessment should be such that they enable the student to demonstrate what they know rather than what they do not know*’. Very often testing is meant to find out what the students do not know. This is a rather negative approach and it does not give the students a fair chance to show what they do know. One result may be that the student loses confidence.

The third principle is that *‘assessment should operationalize all goals of mathematics education’*. So not only the lower level skills should be tested, but also those of the middle and higher levels. Especially the higher order thinking skills are difficult to assess in a straight forward type of test. For that reason a variety of effective assessment tools is needed.

The fourth principle is that *‘the quality of assessment tasks is not in the first place determined by its accessibility to objective scoring’*. Too often the possibility for objective scoring is the dominant factor when the quality of a test is evaluated. The danger of this mechanism is that a test consists only items which are easy to score, that is, at the lower level. Although it may be difficult to score more complex tasks, experience shows that such tasks may have some advantages as well. If more complex tasks are given to the students, it is no longer sufficient for them to give a single answer. The thinking process of the student will become visible and that will give the teacher more insight in what the student is actually doing.

The fifth principle is that *‘the assessment tools should be practical’*. If the previous principles are taken serious, different formats of testing are needed. There is a need for balanced packages of assessment tools, with all the elements of fair testing in it. The designers of such packages have to keep in mind what the practical and physical limitations of schools are. Whatever assessment packages are designed, we should bear in mind the demands made by assessment on the teacher.

In their so-called *‘Standards’*, the American NCTM has formulated aspects of assessment which will have increasing and decreasing emphases in the near future. (See Table 1).

Increased Emphasis	*Decreased Emphasis*
Evaluation of what students know and how they think	Evaluation of what students do not know
Assessment as an integral part of teaching and learning	Assessment as simply correct answers on a test to assign grades
Focus on a broad range of tasks and holistic view	Focus on a large number of isolated skills
Problem situations that require application of several ideas	Use of exercises or word problems that require only one or two skills
Use of multiple techniques, including written, oral, and demonstration	Use of only written tests
Use of calculators, computers, and manipulatives	Exclusion of calculators, computers, and manipulatives from all tests

Table 1:

Different formats

Different formats of tasks can be used to assess different goals. Which format is chosen, depends on the goals to be assessed. The following indicates what formats are available. These formats are again taken from De Lange (1987).

Multiple choice

The multiple choice format is widely used, although it is difficult to design properly. If used on a large scale, it is a cheap format, because the easy way of scoring that can even be computerised. A typical test item from a Dutch national examination in the past is shown in Figure 6.

2 Solve: $\frac{6x+9}{3} = 8$

The solution lies between
A –3 and 0
B 0 and 2
C 2 and 4
D 4 and 6
E 6 and 9

3 Given the function f: x - $\frac{1}{2}(x-3)$. What is x when the result is 5.

The answer is
A 1
B $3\frac{1}{4}$
C $5\frac{1}{2}$
D 7
E 13
F 16

Figure 6

Negative aspects of this format are that, in general, only lower level goals are tested and this format has meagre learning effects. The most fundamental problem with the multiple choice format is the characteristic of '*negative testing*'. Even if the correct answer is selected, one can never know if a student really does understand the problem. A correct answer can always have been a good guess.

Open questions

The category of open questions is a very broad one. To get more insight into the difference between several types of open questions, this category can be refined into sub-categories.

(Closed) open questions

In case of open questions no answer categories are given in advance. Still such questions can be rather closed. We call them (closed) open questions if the question is very straight forward and there is only one correct answer. The so-called short-answer questions are of this type. This type of question is widely used, easy to design and very practical. (Closed) open questions are meant to test lower goals and do not have a learning effect. Compared with the multiple choice format, students have more chances to come up with an answer, although we still have negative testing here.

(Open) open questions
By this we mean open questions where the process of getting the answer, which can still be one formula or one number, is more complicated. Lower level skills are no longer sufficient to answer the question because some complex reasoning is required. The students can solve the problem in their own way and they can be asked to give a (short) explanation of their reasoning. These questions are more difficult to design and are less frequently used. Middle and lower level goals are tested in a positive way. (Open) open questions can have positive learning effects and are still reasonable practical.

Kulm (1994) gives some examples to show how rather closed questions can be reformulated to make them more open. Here is one of his examples:

Textbook question

Find the perimeter of a rectangle with a length of 8 meters and a width of 17 meters.

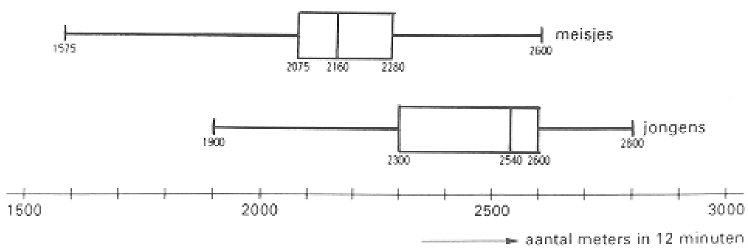
Alternative item

Susan wants to make a fence for her dog in the shape of a rectangle. She has 50 meters of fence. What are some sizes of rectangles that she could make? Which shape would be the best?

Extended-Response Open Questions
If we have an open question where the explanation becomes more important than the final answer, or, to put it in another way, the process becomes more important than the product, we speak about *'extended-response open questions'*. Such questions are rarely found in official examinations, because they are difficult to design and mark. On the other hand such questions are needed, because they have a positive learning effect and give us an opportunity to test higher level goals.

Properly constructed open questions, with a variety of short, long and extended responses, offer possibilities for assessment at a level that is above the lowest level.
The example in Figure 7 (taken from the W12-16 Experimental National Examination, 1995) illustrates these different types of open questions very clearly.

Coopertest
The Coopertest is a test to measure the condition of people. The distance you can run in 12 minutes is measured. 120 girls and 120 boys participate in a Coopertest. You can see the results in the following boxplots:



a. How many girls ran more than 2075 metres in 12 minutes?

Four girls were slower than the slowest boy.

b. Calculate what percentage of all participants ran between 1900 and 2600 metres in 12 minutes. Write down your calculation.

Compare the results of the 60 fastest girls with the 60 slowest boys.

c. Is it possible to conclude from these boxplots whom are on the average the fastest, these 60 girls or these 60 boys? Explain your answer.

A boxplot is a way to visualize data by means of only five numbers (the lowest and the highest score, the first and the third quartile and the median). Boxplots are very useful for comparing different groups. In the Netherlands the boxplot is part of the curriculum, so the students are familiar with a graphical representation like that shown in the illustration. Before you read further, I invite you to analyse the above questions. How would you classify them?

If you analyse the questions, you can see that they are of different nature. The first question is an example of a (closed) open question. It is very staight forward, and meant to test the students' knowledge of how to read a boxplot, which is a basic skill at the lower level.

The second question is more complicated and several steps are needed to answer it, combined with insight in what a boxplot really is. This is an example of a (open) open question. There is still only one possible answer (if we forget about technical details of interpolation), but serious reasoning is needed here.

With the third question the examiners try to test higher other thinking skills. This question goes in the direction of an extended open response question, where the reasoning is more important than the answer itself. The official model answer to this question even shows that several answers are accepted. Given the fact that the Coopertest problem is taken from a national examination, this third question is really remarkable.

Other formats

Besides the closed and more or less open questions, other formats are possible, like essays, oral tasks and project work. These formats have in common that they cannot be done during a fifty minute lesson. They all need their own way of organisation. Especially to assess the higher level goals, we need these other formats.

A well known other format is the statistical survey. Students design their own questionnaires, collect data by means of interviewing people, process their data by hand or by means of a computer, get the information they wanted, write a report and draw conclusions. Of course this can be done by means of group work.

In the Netherlands schools have some freedom about the way they organize their school examinations. Some schools use this freedom to trial new ideas and non-traditional formats. They are very creative in designing different activities for their school exams. A few years ago the concept of Integrated Mathematical Activities (IMA) was introduced for junior secondary education and at some schools IMA is part of the school exam. An example of an IMA is from Tilly Kayser, teacher at the Scholengemeenschap Lelystad (see Figure 8).

Translation:

> For this investigation you need green fingers. You need two flower pots and some seeds of two different plants. Let the seeds grow and measure the length of the two plants every day during one month. Draw four different graphs to illustrate the growth of the plants.

Another example is the so-called A-lympiad. You probably know about the Mathematics Olympiad, an internationally known contest for pre-university students. The A-lympiad is also a mathematics contest, but it is completely different from the Olympiad. With the A-lympiad three or four students work together for one day on one problem. The problem is not a pure mathematical one, but it is taken from real life. The A-lympiad has a history of six years now and the number of participating schools is increasing every year.

Last year the problem was to design a schedule for the elevators in an office building (De Haan, 1996). One of the tasks was to write recommendations for the director of the office and another was to write a letter to inform the employees. The complete task, including a simulator of the elevators in the building, can be found on the world wide web home page of the Freudenthal Institute: http:\\www.fi.uu.nl.

* GWA = Geïntregeerde Wiskundige Activiteit

Voor deze opdracht zul je een beetje groene vingers moeten hebben.
Heb je dat helemaal niet, vraag dan om hulp van iemand bij jou thuis.

Je krijgt mee naar huis:

2 bloempotjes
aarde en
een aantal zaadjes.

Je vult de potjes met aarde en stopt de zaadjes net onder de grond.
Alsjeblieft niet te diep, want dan lukt het niet !
De aarde maak je vochtig (pas op: niet te nat !)
De aarde moet je de volgende weken vochtig houden
Eén potje zet je in het licht op een vensterbank en het andere potje zet je in het donker (bijv. in een kast).

Zodra de plantjes opkomen ga je beginnen met de uitwerking van de opdracht. Iedere dag noteer je hoelang de plantjes zijn en hoeveel blaadjes er aan zitten.
De tabellen zien er zo uit:

plantje **1**		
	lengte	aantal blaadjes
1e dag	....	....
2e dag	....	....
27e dag	....	....
28e dag	....	....

plantje **2**		
	lengte	aantal blaadjes
1e dag	....	....
2e dag	....	....
27e dag	....	....
28e dag	....	....

Dan maak je bij deze tabellen 4 **bijbehorende grafieken.**

Vraag: Wat voor soorten groei herken je in de tabellen en/of grafieken.
Verklaar je antwoord !
Lever dit alles, inclusief de twee plantjes, in vóór 14 maart 1996

Tilly Kayser, 24/1/96

The A-lympiad is a good example of a task were middle and higher level goals are also tested. The A-lympiad shows how schools can be stimulated to experiment with different types of tasks. Some schools even use the A-lympiad as a part of their school examinations. Clearly these schools have overcome the problem of scoring. They have developed different categories which they use to mark the work of the students. Although there is much experience now with the A-lympiad, every year it again takes much time to design a proper task. For this reason one cannot expect that individual schools or individual teachers will ever design such tasks on a large scale.

Conclusion

We have argued that both the lower, middle and higher level learning goals have to be assessed during education. It appears that it is not a simple task to design tests that fulfil this condition. During the design of assessment tasks, one has to be very clear which goals are being operationalized and which contexts and formats are chosen. In fact a balanced assessment package which represents the range of mathematics that we aim for students to be able to do is needed. Especially the design of tasks which go beyond the lower level goals and assess the middle and higher level goals, is very time consuming. For this reason it is possibly more effective to have (examples of) tests designed on the level of province, or even nationwide, instead of on the level of the individual school. To realise a new culture of assessment, all participants (teachers, test designers, parents) have to develop a new attitude towards testing. Teacher training with special attention for assessment is needed to enrich both teachers' knowledge and practical skills on assessment.

Although assessment is not an educational goal in itself, it appears to be so important that it can be used as an instrument to realize changes in mathematics education in general. It is the task of the community of mathematics educators to take care that these changes go in the desired direction.

This article has previously been published in Pythagoras, Journal of the Association of Mathematics Education of South Africa, 42, April 1997, pp. 14-20.

Literature

CEVO (1996). *Syllabus voor het centraal examen Wiskunde.* Cito, Arnhem, The Netherlands.

Heuvel-Panhuizen, M. van den (1996). *Assessment and Realistic Mathematics Education.* Dissertation. Freudenthal institute, Utrecht, The Netherlands.

Gravemeijer, K., M. van den Heuvel-Panhuizen & L. Streefland (1990). *Contexts Free Productions Tests and Geometry in Realistic Mathematics Education.* OW&OC, Utrecht, The Netherlands.

Kemme, S.L. & M. Wijers (1996). *Als jonge honden op een bak met lever.* In: Nieuwe Wiskrant 16(3). Freudenthal institute, Utrecht, The Netherlands.

Lange, J. de (1987). *Mathematics, Insight and Meaning.* OW&OC, Utrecht, The Netherlands.

Lange, J. de (1994). *Assessment: No Change without Problems.* In: T.A. Romberg (Ed) Reform in School Mathematics and Authentic Assessment. SUNY Press, Albany NY, pp 87-172

Wijers, M. & K. Hoogland (1995). *Vijf jaar Wiskunde A-lympiade.* Freudenthal institute, Utrecht, The Netherlands.