MARJA VAN DEN HEUVEL-PANHUIZEN

IMPROVEMENT OF (DIDACTICAL) ASSESSMENT BY IMPROVEMENT OF PROBLEMS: AN ATTEMPT WITH RESPECT TO PERCENTAGE[1]

ABSTRACT. In recent attempts to reform assessment much attention is devoted to the formats of assessment and organizational aspects. Without good problems, however, it is impossible to improve assessment. This article focuses on the role of problems in the development of better assessment. An example of an attempt is given with respect to some key concepts and abilities on percentage. A series of problems for assessing these has been developed inspired by the ideas of Realistic Mathematics Education. The guiding principles were that problems should be meaningful and informative. The problems, which are rather different from the traditional kind have been tried out in two American grade 7 classes. The problems reveal a lot about the different levels of understanding of students. As such they can facilitate instructional decision making. Nevertheless it is clear that this is not the final stage of the search for better assessment problems. As a start to further developmental research, some recommendations will be given for the improvement of the problems.

INTRODUCTION

Assessment has always been a matter of great concern to anyone involved in education but recently interest in assessment has greatly increased. I believe there are two reasons for this. Firstly, there are changes in mathematics teaching. Teaching of mathematics is moving away from the mechanistic approach towards a way of teaching that is aimed at insight – at 'mathematical power' as it is called in the NCTM Standards – and its development by the student. These new ideas on the teaching of mathematics call for new forms of assessment (De Lange, 1987; NCTM, 1989; Romberg, Zarinnia, and Collis, 1990; Clarke, Clarke, and Lovitt, 1990). It has even been argued that innovation in education cannot succeed without new forms of assessment (Romberg, Zarinna, and Williams, 1989). Secondly, policy makers wish to have an overview of the 'output' of education. The NAEP in the USA, the National Curriculum assessment in the United Kingdom, and the PPON in the Netherlands, are examples of this.

A characteristic of recent attempts to change assessment at school level is that most attention is devoted to the changes in assessment techniques, including both the formats of assessment and its organizational aspects. There is a plea for alternatives to class administered written tests consist-

Educational Studies in Mathematics **27:** 341–372, 1994.
© 1994 *Kluwer Academic Publishers. Printed in Belgium.*

ing of multiple choice questions or short answer items, such as portfolio assessment, performance assessment, project assessment, group assessment, classroom observations, student self-assessment (see, for instance, NCTM, 1989). Another hot issue is the shift from standardized tests to forms of assessment for which the teachers themselves are responsible (see Graue and Smith, 1992).

Less attention, however, is paid to the problems, the tasks, used for assessment. Problems are the most crucial part of the assessment. Because the format cannot transform a bad question into a good one, any effort to achieve better forms of assessment can be forgotten without good problems. Although interest paid to this topic was not that big, some educators have paid attention to the kind of problems given to students. See, for instance, the work in the United Kingdom done by Bell, Burkhardt, and Swan (1992), in the United States by Stenmark (1992), and especially the work of the Australians Sullivan and Clarke (1987, 1991), whose plea for good questions does not focus on assessment specifically but rather on teaching in general. In the Netherlands Freudenthal (1973) and other advocates of Realistic Mathematics Education also stressed time and again that the problems we offer to students should always be meaningful (see, for instance, Treffers, 1987; Gravemeijer, 1982, 1990; Streefland, 1991). Later on this aspect was elaborated more explicitly for assessment problems (De Lange, 1987, in press; Van den Heuvel-Panhuizen, 1990; Van den Heuvel-Panhuizen and Gravemeijer, 1991; Streefland, 1992).

This article addresses the development of improved kinds of assessment with respect to the problems. It will be restricted to short task problems and their presentation. Apart from dealing with assessment of mathematics in general, special attention will be paid to a specific topic, namely percentage.

First of all, a general impression will be given of Realistic Mathematics Education and how within this approach attempts are made to improve assessment problems. Then the switch is made to percentage. After a brief consideration of some commonly used problems on percentage, some problems will be shown that have been developed in connection with a teaching unit on percentage based on realistic principles. Besides some background information about the development of the problems some data will be given on students' understanding and abilities evoked by those problems. Connected to this data some recommendations are given for improvement of the problems in question.

Finally, this introduction will not be complete without some terminological clarification. In accordance with the different levels of assessment and with all the different purposes it can serve, the term assessment certainly can have many different meanings. In this article assessment is considered as collecting information about knowledge, insight, and strategies of students that teachers need to have to their disposal in order to make instructional decisions. In other words, the focus here is on formative assessment to support the teaching-learning process. It is called didactical assessment. This focus on didactical assessment, however, does not imply a rejection of assessment for other purposes.

REALISTIC MATHEMATICS EDUCATION AND ASSESSMENT

Realistic Mathematics Education

Freudenthal and his colleagues from the former IOWO[2] were the leaders of a new approach to mathematics education that came into being in the Netherlands. It is called Realistic Mathematics Education (RME). This renewal started in the seventies and continues to the present. Characteristic of this new approach to mathematics education is the rejection of the mechanistic, procedure-focused way of teaching in which the learning content is atomized in meaningless small parts and where the students are offered fixed solving procedures to be trained by exercises, often to be done individually. Instead of conveying knowledge, skills, insight, and its applications, RME is aimed at the development of all these, by reconstruction. This means that RME has a more complex and meaningful conceptualization of teaching. The students are considered to be active participants in the teaching-learning process. Interaction is an essential feature of instruction. In this respect RME is quite in tune with other recent reform movements within mathematics education. RME was – and still is – not a development that stands alone. It has a lot in common with constructivist approaches.

The very feature of RME is, however, that not only the development of the student is seen as a guiding principle for mathematics education but also the mathematics itself, albeit in a special way; the subject matter is not derived from mathematics as a science, but by means of didactical phenomenology – as Freudenthal (1983) called it. The goal of phenomenological analysis is to find out what is worthwhile to teach and in what way. How mathematical concepts manifest themselves to students is investigated. Situations and contexts by means of which the students can constitute these mathematical concepts because of the fact that the contexts prompt them to do this are sought. Often these contexts are real world contexts but that is not always necessary. Essential is that they offer the opportunity to mathematize, that means that the situations can be organized with mathematical tools. Moreover, it is essential that the situations can be imagined by the students (see, for instance, Gravemeijer, 1982). Actually this latter aspect has given the new Dutch approach to mathematics education its name. It is called 'Realistic Mathematics Education' because offering students problem situations which they can imagine is one of the key aspects of this approach. In Dutch, the word 'to imagine' is called 'realiseren'. Although contexts can play an important role in stimulating and motivating the students, it should be clear that in RME the contexts also have a genuine didactical function.

Another consequence of the domain orientation of RME is that much attention is paid to models to bridge the gap between the informal, context-connected mathematics and the formal mathematics. These models form the continuous thread within a learning strand and the linking factor between learning strands. Often these models originate from the contexts that are used or have at least roots in them. See, for instance, the arrow language as it has grown out of the bus context (Van den Brink, 1984).

Assessment

In RME, teaching and assessment are strongly connected and integrated. This means that assessment plays a role in all stages of the teaching process. Moreover, it includes both looking backward and looking forward. Looking backward is done to investigate how the given instruction worked out. The purpose of looking forward is to look for stepping stones for further instruction. In this respect, there is a strong affinity with Vygotsky's zone of proximal development. By gathering all this information, assessment is guiding the teaching process and making instructional decisions. So, the better the assessment, the better the teaching.

Whatever specific purpose assessment has to serve, it is always aimed at revealing in one way or another what the students know and understand and how they are thinking. With respect to RME it means that assessment must take into account that the development of the student can proceed along different routes and, moreover, that there can be discontinuities in the learning process. As a consequence assessment must be both broad and open.

For gathering this information a variety of means can be employed but it is crucial that they must make the learning process and its stage of development visible. As a consequence classroom observations and individual interviews have a prominent place in RME. This does not mean, however, that the format and the organization of the assessment are the most important aspects of assessment. The main role is reserved for the problems used. Whatever format or organization has been chosen, at the very end – and again the domain orientation of RME manifests itself here

– one has to ask the students a question or to give them a problem. For this reason, the improvement of assessment in RME is especially sought (see also De Lange, in press) via the improvement of the problems that are used for assessment.

Requirements for assessment problems

What assessment problems are used depends on the educational goals that are pursued. In RME, that is based on the idea of mathematics as a human activity (Freudenthal, 1973) the main goal is that the students learn to do mathematics as an activity. This implies that they learn 'mathematics so as to be useful' (see Freudenthal, 1968). The students must be able to analyze and organize problem situations and apply mathematics flexibly in problem situations which are meaningful to them. Moreover, the problem situations should be rather new for them because this offers the students the opportunity for mathematizing. In other words, problem solving in RME does not mean carrying out a fixed procedure in typical situations. As a consequence the problems can be solved in different ways. In order to fit into RME, the assessment problems have two important requirements.

(i) Meaningful

First of all, the assessment problems – just as the problems used for teaching – should be meaningful. From the point of view of the students this means that the problems must be accessible, inviting, and worthwhile to solve. From the point of view of the subject matter, the problems need to reflect important goals. Something that is not worthwhile to learn is not worthwhile to assess either. Moreover, the problems should be correct mathematically. Further, it is important that the problems are not restricted to goals that can be assessed easily, but that they cover the entire mathematical area in width and in depth. That means that the assessment should cover all chapters of the subject matter and should include problems on each level: from basic skills to higher order reasoning (see also De Lange, in press).

(ii) Informative

The problems also must be informative to be adequate for assessment. From the point of view of the students this means that the problems should offer students the possibility to show of what they are capable instead of stressing only what they are not yet able to do. In order to achieve this it is necessary that the problems are accessible and moreover that they can be solved in different ways and on different levels. This means that the problems will make the learning process transparent to the teacher. They must provide her or him with a maximum of information on students' knowledge, insight and abilities, including their strategies. Again this means presenting open problems that can be solved in different ways and on different levels. Having the students make 'free productions' is one way of doing this. This implies that students are asked to think up problems by themselves instead of doing problems thought up by others. Free productions not only offer a cross-section of the understanding of the students of a class at a particular moment, but they can also give a longitudinal intersection of the learning path the students will follow globally (Treffers, 1993; see also Streefland, 1991). Apart from free productions also some less far reaching steps can be taken to make problems more informative (more can be found about this in Van den Heuvel-Panhuizen and Gravemeijer, 1991).

The role of contexts in assessment problems

In RME contexts play an important role in the construction of problems for assessment, as well as in regular teaching. In assessment problems they are important in several ways.

(i) Accessibility

The use of contexts, often accompanied by a pictorial presentation of the problem, enables the students to grasp the intention of problems immediately even without an extensive written or oral explanation. Therefore the problems have to be related to well-known everyday-life situations, or at least situations which can be imagined by the students. All this can contribute to the accessibility of the problems. Needless to say this is only true if the presentation and the wording of the instruction are clear.

(ii) Latitude and transparency

Compared with bare problems, contexts give the students more latitude to display what they know. The reason for this is that context problems offer the students more freedom for their solution. This contributes to the transparency of the assessment. If they are properly selected, contexts will allow a broad range of solution strategies – and sometimes also answers – on different levels. In a bare problem the operation to be carried out is more often than not fixed. Such problems only have the possibility to assess whether or not a student can carry out the procedures that have been taught. For this reason bare problems are not very well suited to *a priori* assessment to get indications for furhter instruction.

(iii) Offering strategies
As already noted, the contexts used in RME are not only meant to motivate the students, but also have a real didactical function. Again, if properly selected they can give the students a context-connected way of tackling the problem or they can lead them to a kind of model that can support the solving process.

Below an example is given of an assessment problem that fits the ideas of RME quite well. It will make more concrete what has been said about assessment problems. This example does not mean, however, that problems like the example are the only ones that are used in RME. It ought to be clear that in RME a wide variety of assessment problems is used (see De Lange, 1993). The given example represents only one of them: a short task problem.

The ice-bear problem as a paradigm

This problem brings us to grade 3 (aged 9 years). The students are already rather proficient in multiplication and division tables up to ten, and also have had some experience in doing multiplications and divisions with two-digit numbers. But written algorithms have not yet been addressed. Within RME this is an ideal moment for assessment which necessarily includes both looking backward and looking forward. The problem that has been developed for this is the following one:

> An ice-bear weighs 500 kilograms.
> How many children will weigh as much as one ice-bear?"
> Write your answer in the frame.
> If you like to use the scrap paper on the test sheet you are allowed to do so.

This instruction is given orally, together with the test sheet (see Figures 1 and 2). The scrap paper that is mentioned in the test instruction is drawn on the test sheet.

Strictly speaking (and neglecting for a moment that not all the data needed for solving are given) this problem can be considered as a division problem. It is, however, a division problem that can be solved in different ways, even by means of multiplication or addition (Figure 1), or in a ratio-like way (Figure 2), even if this is not taught in class thus far.

Students often encounter this kind of problem in television programs, in magazines and in shops. It arouses their natural curiosity. As such it is a meaningful problem. It is both worthwhile to solve and recognizable. Furthermore, it represents an important goal of mathematics education. Apart from being able to apply calculation procedures, students are required to solve real problems in RME, problems of which the solution procedure

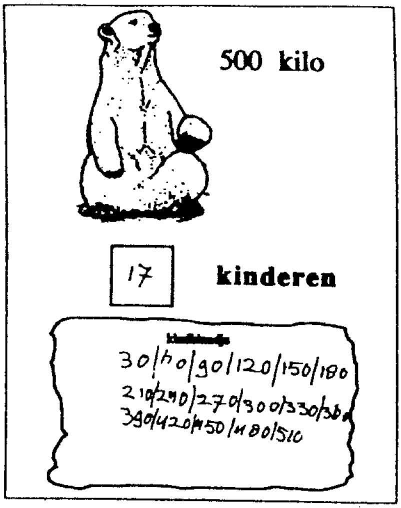

Fig. 1. Ice-bear problem with student work

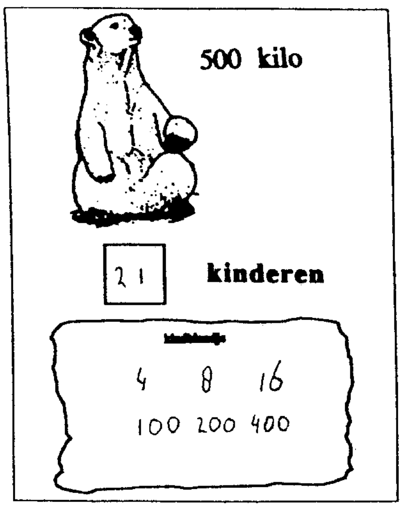

Fig. 2. Ice-bear problem with student work

is not known in advance, and where not all data are given. Often this kind of problem solving requires students' own contributions, like making assumptions about missing data. Therefore within RME measures that are both known and lived through are so important. Because of the absence of information about the weight of a child, the icebear problem becomes a real problem. The problem does not indicate what kind of procedure needs to be carried out, while it also calls for thinking of the weight of an average child first. Both offer the student room for own constructions which can be considered as the most important merit of problems like this one. On the one hand students are given the opportunity to show of what they are able, and on the other hand teachers are provided with information of how their students tackle problems, what knowledge of measures they have, what strategies they apply, and what models and notation schemes they use for supporting their thinking on the problem.

Apart from giving the students room for their own constructions, and the opportunity of applying their own combination of mathematical tools and insights, contexts can also indicate to students a solution strategy by offering them a context-connected way of tackling the problem. An example of such a context is given in Figure 3. The question that belongs to this problem is:

You have 47 beads. How many beads are left after you have made a chain of 43 beads?

The problem belongs to a paper-and-pencil test that was administered in grade 2 (aged 8 years) in November, 1993. 427 Dutch students did the test and 60% gave the correct answer to this problem. In the same paper-and-pencil test there was also an analogous problem without context (47-43=). This problem, however, was only done correctly by 38% of the students. Apparently the context elicits a kind of 'complementary subtraction' strategy that makes this problem attainable for more students. It is this power to reveal the informal strategies of the students that a context can have, that gives it its crucial role in assessment problems. Although there is already a lot of evidence that contexts can contribute to teaching and assessing mathematics, this does not mean, however, that everything is known about their use. Further research needs to be done in this respect.

The above mentioned ideas about assessment problems also play an important role in answering the question: How to assess percentage?

PERCENTAGE AS IT IS OFTEN TESTED

Before dealing with how to assess percentage, some common assessment problems on percentage will be considered. Is there really anything to

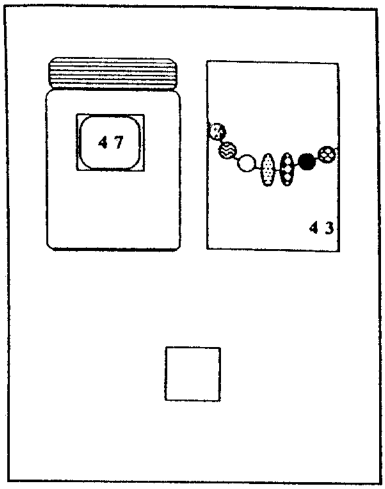

Fig. 3. 47–43 in context

improve? In Figure 4 a sample of problems is given that resulted from a small informal survey (in USA) of some mathematics achievement tests for grades 5 through 10. The survey contained two commercial tests and two school district tests. The examples have been slightly changed because they are not open to publication.

First of all, the multiple choice format of the problems is noticeable. This does not mean, however, that the problems should be rejected for that reason. Sometimes the multiple choice format can also be revealing (see Van den Heuvel-Panhuizen, 1993a). In this case where it is the only format that is used one might wonder whether this format will provide the information that is needed.

What information one needs depends on what one considers as valuable to learn. So the next question deals with the goals that are pursued. Many of these problems indicate that education is primarily focused on procedures and recall instead of getting a real understanding of percentage. Do students, for instance, understand percentage if they can convert a percentage into a decimal? And can one state without restriction that this question is mathematically correct? (see Davis, 1988). Without answering these and other questions, it can be claimed that problems like the ones in this sample have shortcomings both with respect to the meaningfulness of the goals they cover and to the degree of information they offer.

TABLE I

NAEP items on percentage and their scores

4th NAEP (grade 7) *		5th NAEP (grade 8) **	
Concept items	% correct	*Without calculator*	% correct
A. Express .9 as a percent	30	A. Change .35 to a percent	79
B. Express 8% as a decimal	30	B. Which of the following is true about 125% of 10?	55
Calculation items	% correct	C. A calculator sells for $9.99 in a certain state. The purchase price including tax is $10.69. To the nearest whole number percent, which of the following (three alternatives) is the best estimate of the sales tax in this state?	41
C. 4% of 75	32	*Calculator allowed*	% correct
D. 76% of 20 is greater than, less than, or equal to 20?	37	D. If the price of a can beans is raised from 50 cents to 60 cents, what is the percent increase in the price?	18
E. 30 is what percent of 60?	43		
F. 9 is what percent of 225?	20	E. Kate bought a book for $14.95, a record for $5.85, and a tape for $9.70. If the sales tax on these items is 6 percent and all 3 items are taxable, what is the total amount she must pay for the 3 items, including tax?	46
G. 12 is 15% of what number?	22		

* The items and scores are derived from Kouba, Brown, Carpenter, Lindquist, Silver, and Swafford (1988).
** The items and scores are a sample from the released 1990 NAEP items.

these factors with the goals and didactics of the curriculum – and this is particularly true for (i) and (ii) – taking them into account can turn out differently.

In RME where teaching is built on the informal knowledge of the students, the teaching of percentage could start with assessing what the students already know about percentage. It should be noted, however, that it is not correct to speak of 'start' here. Teaching percentage does not start

A fifty percent score is needed to pass the exam.
John missed 14 problems.

>>*What do you think? Can we congratulate John or not?*

Fig. 5. The exam problem

when the name percentage is mentioned for the first time, but it has its roots in all kinds of 'so-many-out-of-so-many' situations that have been dealt with long before, at least in everyday situations.

For assessing informal knowledge on percentage, a problem like the one shown in Figure 5 could be used. This problem is meant to evoke this knowledge. The development of this problem was inspired by a similar problem described in Burns and McLaughlin (1990), but somewhat changed. The Burns and McLaughlin problem does not include the question about congratulation, the students have to discuss the sense or nonsense of the statement instead.

The 'Per Sense' unit

The problem shown in Figure 5 is used in the first chapter of a teaching unit on percentage, called 'Per Sense'. This unit is developed for the 'Math in Context' project of the National Center for Research in Mathematical Sciences Education at the University of Wisconsin, Madison. This is an NSF funded Middle School Project. The aim of the project is to improve the mathematics curriculum of the middle school. The Freudenthal Institute of the University of Utrecht is involved in this project. It is their task to develop draft materials for it. The draft version of the 'Per Sense' unit was developed by Van den Heuvel-Panhuizen and Streefland, and was americanized by Meyer, Middleton, and Browne (in press). The unit is intended for grade 5. Its goal is to help students to make sense of percentage. The next part of this article will deal with (a part of) the assessment of this unit.

As stated, the unit starts with a chapter consisting of a series of problems for evoking informal knowledge on percentage. The problems are meant to be discussed in class. The experiences with these problems were very revealing (see Streefland and Van den Heuvel-Panhuizen, 1992). It turned out that if problems are offered to students with the opportunity to become involved, they can tell much about the students' thinking. In the case of the exam problem, John's success actually can be controlled by the students.

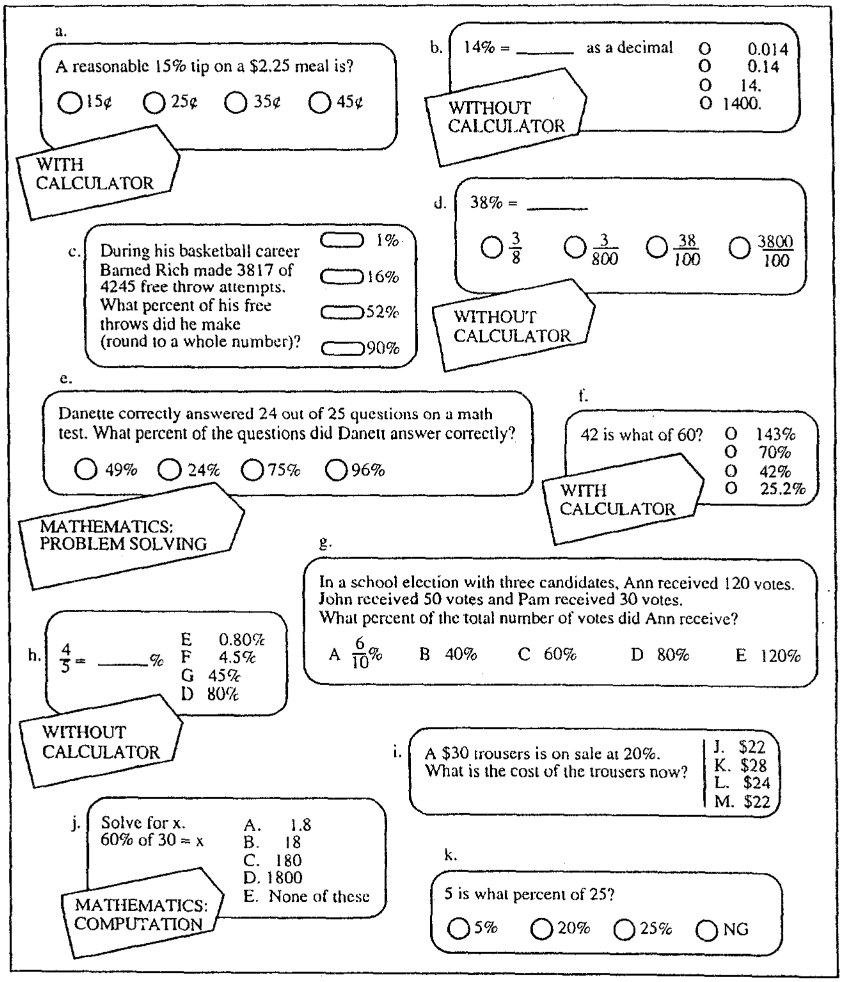

Fig. 4. Sample of problems taken from some mathematics achievement tests

This criticism does not mean that all problems in the sample fail completely. Some of them include good ideas and could easily be transformed into more meaningful and informative problems. Take for instance the first problem (no. a). By skipping the alternatives, abandoning the calculator, and changing the wording somewhat (for instance: 'What do you think is reasonable to leave as a tip after you had a nice $2.25 meal? A what percent tip is this?'), this problem could become a very good problem for assessment.

Another thing that should be kept in mind is that standardized tests are not the only tests being used in the classroom today. According to Grouws and Meier (1992) the most common form of testing in the classroom is the one in conjunction with course content. These tests may be teacher-made or produced by a publisher to accompany a textbook. They are generally more closely aligned with the curriculum and content being taught in the classroom than are standardized tests.

Nevertheless, the same kind of problems as shown in the sample are used in the NAEP studies. Table I shows items on percentage that are used in the last two studies and the percentage of the students that performed them correctly.

What can these problems reveal regarding the knowledge of students in the US about percentage? The data in the table reflect that the students are not very proficient at percentage. Converting a decimal into a percentage received actually the best score. According to Kouba, Brown, Carpenter, Lindquist, Silver, and Swafford (1988) this problem is labelled as a 'concept item'. Apart from wondering whether operationalizations like this one are correct and whether these problems represent important goals, some other imperative questions arise because of the results. Despite the fact that the 'correct' scores are very revealing, what do they really tell us about how to proceed with instruction? And, what would the results be if other problems, like the RME type of problems, had been used instead? The latter question will be answered to some extent in the next section.

ARE THERE OTHER POSSIBILITIES FOR ASSESSING PERCENTAGE?

How percentage is assessed depends on (i) what precisely has to be assessed (this means what kind of understanding of percentage and what kind of skills are assessed), (ii) the stage of the learning process the students have reached (have they just started or do they already have a lot of experience with percentages?), and (iii) what the purpose of the assessment is. As such, these factors have always to be acknowledged whatever approach is taken to mathematics education. Because of the strong relationship of

They can have him fail or pass the exam. This makes the problem so powerful. Like in the ice-bear problem the students are given room for their own constructions.

Apart from this first chapter, the unit contains two other parts for assessment: one at the end of each chapter and a final one at the end of the unit. The first one is meant for assessing what has been treated in the chapter. For this reason each chapter ends with a summary activity that can be used for assessment. The problems are strongly aligned with the content being taught and the contexts of the problems are often the same ones as used in the chapter.

For the end of unit assessment a test has been developed, called 'Per Sense Test' (Van den Heuvel-Panhuizen, 1992). Compared with the end of chapter assessment, this 'final' assessment has a broader scope. The goals to be assessed are at a more general level. They concern some key concepts and key abilities on percentage which are also endorsed by people who stick to other approaches of mathematics education. To make the difference between an end of chapter assessment and this more general assessment at the end of the unit more clear an example can be given of the unit-connected assessment. The assessment of the ability to use the percent bar, which is an important tool in the 'Per Sense' unit, is, for instance, done in an end of chapter assessment and not specifically in the final assessment. This, however, does not mean that the way in which the student solves the problems are not of interest. The name 'final' assessment should not be misunderstood. Its main purpose is to document the achievements of the students in order to make decisions about further instruction and this implies gathering information about their strategies.

The problems developed for this 'final' assessment are mostly appropriate to this article because of their general character. Some of the general goals that are assessed, and the problems that are used for it, and what they revealed are discussed next.

The goals to be assessed — a selection

The 'Per Sense Test' tries to cover some key concepts and key abilities on percentage. Here a selection will be presented including not only computational goals but also higher order goals concerning the understanding of percentage. This distinction is often omitted. Yet making this distinction is not the most difficult part of goal description. The most difficult is operationalizing both aspects, and determining what students — say at middle school level — need to know about percentage.

A key feature of percentage, that one has to understand in order to have insight, is that a percentage is a relation between two numbers or magnitudes that is expressed by means of a ratio. Grade 5 students do not necessarily have to explain this in this manner. On the contrary, it is even unwanted, unless one wishes schools to become temples of verbalism (Schoemaker, 1993). Instead of giving definitions, the insight into what percentages are should emerge in how the students are using them. This means that they have to show an awareness of the fact that percentages are always related to something and that they therefore cannot be compared without taking into account to what they refer (Fig. 6,1). The same is true for another consequence, namely, that if one changes the amount of reference the percentage changes too (Fig. 6,4). Another way in which their insight could emerge is when they show an awareness of the fact that a percentage remains the same if the ratio is the same (Fig. 6,2a).

With respect to the computational goals a great variety of types of computations with percentages can be mentioned such as being able to compute the part of a whole while the percentage is given (Fig. 2b). However, it is more important that students are able to use percentages in a situation in which they are needed, for instance when different parts of different wholes have to be compared (Fig. 6,3) than that they are able to carry out these computations in a rather isolated way.

FOUR PROBLEMS FROM THE 'PER SENSE' TEST AS AN EXAMPLE

Problems have been developed fitting to all the aforementioned goals (see Figure 6).

The birth of a problem has some similarities with the delivery of a baby. Sometimes it takes considerable time, and sometimes it is suddenly there, unexpectedly. So every problem has its own story. Take the four problems of the 'Per Sense Test'. Only some snapshots of their story will be given here.

Problem 1

To assess whether students understand that percentages are always related to something and that they therefore cannot be compared without taking into account to what they refer, the familiar situation of a sale was chosen. Two shops are having a sale. In the first shop one can get a discount of 25%, and in the other one the discount is 40%. Both shops have put up a big poster in the shop window. The manner in which the two shops advertise their discount could convey the suggestion that the two shops do not sell wares of the same quality. This cue is given on purpose to alert the students to consider to what the percentages might refer.

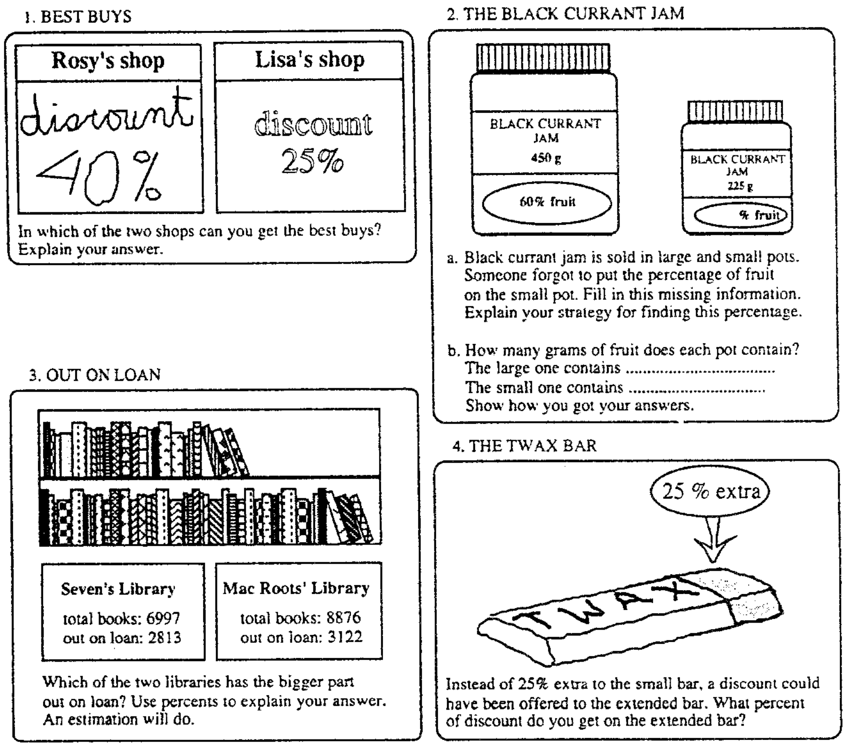

Fig. 6. Four problems from the 'Per Sense' test

Problem 2

In daily life, besides in sales, percentages are also used to describe how different kinds of substances are composed. In this case the percentages do not refer to a certain change that is happening or to one that has occurred already, but they describe a fixed situation. They indicate what part of the substances consists of this or that. In other words, they describe a proportion, a ratio. Because the composition of foodstuffs often is not hidden as in other substances, the context of foodstuffs was chosen to assess whether students understand that a percentage remains the same when only the absolute amounts or numbers of a mixture change and not their ratio. The problem is about quality jam. It contains 60% of fruit. The jam is sold in large and small jars. The question is whether the students are aware of the fact that the size of the jar does not influence the percentage of fruit. Or, in other words, is their understanding strong enough to withstand this visual distractor.

The same jam context is also used to assess whether the students can compute a part of a whole item when the percentage of that part is given. This is assessed by the question: How many grams of fruit are in 450 grams of jam?

Problem 3

Although the last part of problem 2 gives an indication whether students can do computations with percentages, it does not reveal whether they can apply percentages when a problem situation calls for it. The 'books loaned out' problem is meant to assess this. Students must compare different parts of different wholes. The library context has been chosen because of the bookshelves that can be found there. The shape of the bookshelves can be seen as a kind of bar which can help the students in finding (or estimating) the percentages out on loan in both libraries by first marking the numbers of books out on loan and then converting these to percentages. Actually the choice of this context has to do with the percent bar that has been introduced in the unit as a mathematical tool.

Problem 4

In order to assess whether students understand that a percentage changes if one changes the amount of reference, again a situation has been chosen from everyday life outside school. A common advertising gimmick is giving parts of something for free. A small candy bar plus one fourth of its length for free costs as much as a large candy bar minus one fifth of the price. Hence, the shaded part refers to two different percentages, depending on the chosen amount of reference that is either 25% or 20%. In this problem the pictorial presentation plays an even more important role than in the previous one. On purpose the bar is given a name of four letters. This gives namely structure to the bar and as such can help the students to organize the problem.

THE TRY-OUT OF THE 'PER SENSE TEST'

Because there were no grade 5 classes available at the moment that the 'Per Sense' unit had to be tried out, a switch was made to grade 7. Another reason for this switch was the assumption that the unit in its final draft state was rather difficult for the grade 5 level, which is not very surprising if the NAEP results are taken into consideration. The unit was tried out in three grade 7 classes of a school near Madison, Wisconsin, in May 1992. It took just over three weeks. One of these grade 7 classes was a special class with low attainers. This class did the test too, but the scores are not included in the results discussed here. The two other classes can be classified as regular classes. In total they consist of 39 students. The test was administered by the teacher of the two classes. The students worked individually on the test. Although the students did the 'Per Sense' unit before the test was administered the results cannot be considered as an output of this unit. The conditions under which the unit was tried out were less than ideal. The unit was not tried out in the way it was intended because there was no advance teacher training and the teacher guide was not completed on time. The try-out, however, is not of specific importance to the aim of the present study, which is to look for ways of improving assessment problems. As far as the test is concerned, it was the first time that a test like this was administered in these classes. The students were used to tests consisting of problems as shown in Figure 4.

The students' work was evaluated by both the teacher and the author of this article. The analysis done by the latter will be discussed in the next section. The teacher only marked the students' work. It would exceed the limits of this article to discuss both evaluations.

WHAT DO THE 'PER SENSE' PROBLEMS REVEAL

The first thing that struck in analysing the students' work was that offering rich problems to students results in getting rich answers. This means that simple marking becomes a thing of the past and that giving students credits for 'the' correct answers becomes a hard job. Correctness turns out to appear in several manifestations. Elicited by the problems, students give responses which do not allow confining to the strict criterion of correctness. This would do no justice to the richness of the answers. Instead of the criterion of correctness it is better to ask the question 'is the answer reasonable?'. In a way this implies taking the point of view of the students, and wondering what they might have meant by their answer, or what their reasoning might have been. By doing justice to the students' own way of thinking, the assessment does not only become more fair, but it also contributes to the enhancement of the information offered by the problems.

A difficulty in applying the criterion of reasonableness is that the cut-off point (what is reasonable and what is not) often is not easily determined. Which decision to make depends on many factors. What matters is the message that is conveyed by the response of the student. Sometimes an incorrect answer shows that a student has insight, and sometimes question marks must be put to a correct answer. Furthermore, one must be aware of the fact that a new, unexpected interpretation of the problem can be elicited while analyzing the students' work. In the case where the work is analyzed by the class teacher specific difficulties of the students might also be taken into account.

Another difficulty in analyzing students' work is the degree of precision. How many categories are needed and are categories as such sufficient? On categories alone one cannot build instruction. They are not sufficiently informative. This means that the categories are only meant as a cue that indicates a certain aspect of the response but which cannot substitute for the real answer.

Apart from the number of categories decisions have to be made as well about the nature of the categories. First of all, they should be connected to the problems and not be too general. Moreover, they should be aimed at how to continue the instruction. Although the categories can be distinguished in advance to a certain degree, it is the students' work that eventually determines what categories are distinguished.

With respect to the work of the 39 grade 7 students that was analyzed (see the Tables II through V) both the categories and the cut-off point were determined a posteriori. The dotted line in the tables indicates a possible cut off between reasonable and unreasonable answers. To make the categories more informative they are illustrated with examples of students' responses.

Apart from the results some conclusions will be given concerning the indications for instruction that can be derived from the answers of the students. However, the reader should not expect a complete teachers' guide. Only a few indications will be mentioned. Finally, some indications for improvement of the problems are given.

The analysis of problem 1 — Best buys

Results

The analysis of the responses (see Table II) shows that at least half the students (20 out of 39) understood that one cannot compare percentages without taking into account to what they refer. The majority of this group solved the problem by explicitly indicating that the answer depends on the list price. Three students did this more indirectly. They took as an example an item that has the same list price in both shops and they did some calculating. Two different students proceeded in the same way but made the wrong conclusion at the end which raised the problem how to evaluate this. Although 'the' correct answer was not given, it is clear that these students knew that a percentage is related to something. Because the latter was assessed, rather than correct close reading and carrying out tasks

TABLE II

The responses to problem 1, Best buys

BEST BUYS		
Answering categories	N	*Examples*
a. Taking into account the original price	15	- "It depends on the original price of the objects they are selling" - "Both, I mean how much does the item cost, nobody knows" - "Lisa's, because if you buy something that's already been used, you will have to fix it up or ..."
b. Taking the same price as an example	3	- "Rosy's, if something at both stores was $30.75. At Rosy's it would be $12.33, at Lisa's it would be $28.95"
c. Taking the same price as an example: wrong conclusion	2	- "Lisa's, because for example a shirt cost $50; 40%=$20 and 25%=$12.50; with Lisa's deal you're paying less"
d. Comparing the percentages absolutely	18	- "Rosy's, 40% is better than 25%" - "Rosy's, because it is closer to one hundred percent, so there would be more off"
e. No answer	1	

precisely, it was valued as reasonable. Finally, half the students compared the two percentages absolutely.

Indications for instruction

The responses not only show different levels of understanding but provide at the same time a global view of the learning path to be followed. This can be very helpful for planning further instruction. Moreover, different levels in class can also be exploited in instruction in a very concrete way. It offers the opportunity to discuss the levels in the students' own language which is often far more convincing than the adult talk of the teacher. By means of these discussions the students can learn from each other even when they have to learn different things. It is clear that the students who compared the two percentages absolutely need most help. They must learn that one cannot treat a percentage as an absolute number. To deal with this matter, the first step could be to present them a 'best buys' problem in which the percentage of discount in the two shops is the same. The students who made the wrong conclusion at the end need to learn how to check one's reasoning by wondering whether it's outcome is in accordance with the expected outcome. The students who took an item as example could be prompted to explain it in a more general way by asking them why they took an item with the same price as an example. Finally, the students who explained it in a general way, could be asked to find out when an item will be cheaper in the 25% discount shop than in the 40% discount shop.

Indications for improvement of the problem

Apart from indications for further instruction the analysis of the students' work also yielded some indications for improvement of the problem. What is the degree of certainty that this problem offers valid information about the level of understanding of the students? Do the students who compared the percentages absolutely really lack understanding of the relativity of percentages? To be certain about this, the problem could be improved by extending it with an additional question, one that will serve as a safety net. In this case this extra question could be: "Is there any possibility that your best buy could be at Lisa's? If yes, give an example." By means of this 'safety net question' one can pick out those students who understand the relativity of percentages but who still need extra help in expressing this.

Another remark that could be made concerns the clarity of the problem. What is meant by a best buy? Does it refer to the cheapest price or the biggest amount of discount? That students could be confused by this is shown by the ones who made the wrong conclusion at the end. It is possible that their mistake was made because they switched their point of view while reasoning.

The analysis of problem 2 — The black currant jam

Results

Only 10% of the students (4 out of 39) knew that the percentage of fruit is the same in the two jars (see Table III). With respect to the second question about 40% of the students (16 out of 39) gave a reasonable answer. Of those whose answers were reasonable and whose strategies were obvious almost half used a rather informal indirect strategy. In contrast to them there was also a large group who did computations which make no real

TABLE III

The responses to problem 2a and 2b, Black currant jam

THE BLACK CURRANT JAM (a)			
Answering categories		N	*Examples*
a.	Correct answer (60%) explanation indicates insight in 'same ratio, same percentage'	4	- "They are the same, except one is at a smaller scale, both pots contain 6/10 of fruits" - "Big got 60%, little got 60%"
b.	Correct answer (60%) without reasonable explanation	1	- "Guessed"
c.	Incorrect answer (30%) 'halving g-halving %' explanation	25	- "You look at the bigger bottle and its half g, so you take half of the %" - "225 is 1/2 of 450, so 30% is 1/2 of 60%" - "450 divide by 2 is 225, so you divide 60 by 2 and you get 30"
d.	Incorrect answer (30%) other or no (3) explanation	5	- 450÷60=7.5 and 225÷7.5=30
e.	Incorrect answer (others)	3	- 25%: "Both divided by 9, got my percent" - 22%: "First subtracted the grams to get the difference (450-225=245), then 22 by 10, and put a decimal, is 22%"
f.	No answer	1	

sense. It seems as if they were trying to recall and re-use the computational procedure instead of using common sense.

Indications for instruction

First of all the students need to get more experience in expressing a ratio (a 'so many out of so many situation') by means of a percentage. This could be done by offering them problems in which they can investigate what consequence all kind of changes will have.

TABLE III Cont.

THE BLACK CURRANT JAM (b)			
a.	Correct answer (270 g) based on a standard strategy	5	- 450×.60 = 270.00 - "0/450 = 60/100; 270/450 = 60/100"
b.	Correct answer (270 g) based on an informal strategy	1	- "10% of 450 is 45; 45×6 ("from 60%") = 270"
c.	Correct answer (270 g) no information about strategy	3	
d.	Reasonable answer based on a standard strategy	2	- 275: 450×0.60 = 275.00
e.	Reasonable answer based on an informal strategy	4	- 263: Bar, approximated 60% by repeated halving - 200: "450÷2=225 and you have to take a little more away to make 60%"
f.	Reasonable answer no information about strategy	1	- 250
g.	Incorrect answer (450 g) not able to work with percentages or no information about strategy (1)	12	- "It says on the bottle" - Bar divided in parts of 15% - 450÷225 = 2; "1/2 of 60% is 30%"
h.	Incorrect answer (others) not able to work with percentages or no information about strategy (3)	10	- 390: 450−60 = 390 - 7.5: 60÷450 = 7.5 - 13.3 or 13: "60/450×?/100"
i.	No answer	1	

As far as the computational part of the problem is concerned, the students who got stuck in senseless calculating could acquaint themselves with the informal strategies of their classmates.

Indications for improvement of the problem

That most of the students thought that the percentage of fruit is not the same in the two jars does not mean necessarily that they had no insight at all into the relational aspect of percentage. It might also be the case that their understanding was still unstable and that they could not yet withstand the visual distractor of the problem presentation. To check whether the latter is the case or not, this problem could also be improved by a safety net question. The question that could be asked next is: "Let's take a look at the taste of the jam in the two jars. Will they taste the same or not? Explain why you think so."

The analysis of problem 3 — Out on loan

Results

Although estimation problems are not very common in school about one third of students solved this problem by means of working with rounded off numbers (see Table IV). In total some 60% of the students (23 out of 39) arrived at the correct answer and another 20% who failed to get this did at least show that they were capable of working with percentages to some degree. Only 20% of the students failed entirely. Again, most of these students did impressive calculations which probably did not make sense to them.

Indications for instruction

Further instruction should pay attention to estimation by means of working with rounded off numbers. This would be valuable in particular for those students who fell into senseless juggling with numbers. As in the previous problem, the work of some of their classmates, put on a transparency, can be used to show them how this estimation can be supported by means of a number line or a bar.

Indications for improvement of the problem

It sounds rather paradoxical, but the fact that almost every student made use of percentages to compare the two libraries is a clear sign that this problem asks for improvement. It is important to remember that the purpose of the problem was to assess whether the students are able to use percentages in a situation in which they are needed. Actually the present problem fails to do this. The students are asked to use percentages instead of applying them spontaneously. Moreover, the problem can be solved without percentages.

TABLE IV

The responses to problem 3, Out on loan

OUT ON LOAN		
Answering categories	N	*Examples*
a. Correct answer (Seven's) estimation with rounded off numbers	12	- 7000, 3000, 45%; 9000, 3000, 33%
b. Correct answer (Seven's) estimation by means of bar or line	2	- 7000÷3000=1/2, 50%; and 9000÷3000=1/3, 30% - Number line, repeated halving to find the percentage
c. Correct answer (Seven's) other strategies, or no information about strategy (2)	9	- "S has 40% out and R has - 34% out" - The numbers on loan have been doubled - The numbers not on loan have been computed - "Because R has 8876 total and S has 6997. When S loans out books it hardly has any left. R still has a big selection"
d. Incorrect answer (both (3) or Mac Roots'(3)), but strategy indicates able to work with percentages to a certain degree	8	- "They both have the same amount out on loan" - Number line both 40% out on loan - "R has about 35% of his books out on loan" - "He has about 33%, S has 30%" - The percentage not on loan has been computed
e. Incorrect answer (Mac Root's) not able to work with percenates, or no information about strategy (2)	7	- 6997 + 2813 = 9810; 8876 + 3122 = 11998; "98% for S and 119% for R; R has a greater % - 6997 - 2813 = 4184; 8877 - 3122 = 5754; "4184/5759 = ?/100; R has a bigger percent, about 73%"
f. No answer	1	

The analysis of problem 4 — The Twax bar

This problem was the most difficult one of the series. Only nearly 20% of the students (7 out of 39) succeeded in solving this problem (see Table V). More often than not the answers of these students contained excellent explanations like the one marked (*) in the table. To recognize the quality of such answers, however, requires that one should not be distracted by the clumsy wording of the students, and that one wonders what the students might have meant by their responses. Another interesting thing was that some students really used the picture to solve the problem. One of them did some additional drawing in the given picture (see Figure 7), others drew a bar or a number line by themselves to find the discount.

The responses showed a great variety of incorrect answers. The most common mistake was the idea that the percentage would stay the same. Four students did not give an answer at all.

Indications for instruction

In this case the students' worksheets contained a nice example of how this kind of assessment can suggest instructional ideas (see Figure 7). Moreover, it even looks as if the teacher, while grading already adopted this same strategy as one of her students. The same visual explanation given by the student (Figure 7) was used by the teacher to explain the problem to some students who gave the wrong answer (see Figure 8). The bold handwriting in this figure is the teacher's. Whether it really happened in this way cannot be stated for sure because it was not possible to find out in what order the teacher graded the students' work. Her 'new' explanation which was only given to part of the students suggests that it happened in the way suggested.

Indications for improvement of the problem

Among the students who were unable to solve the problem there were at least three who did not understand the question and there were even more who did not understand what the problem was about. This was probably due to the wording of the problem and its presentation. A possible way to improve the problem could be to use two advertisements: one in which a four-part bar is extended with 25%, and another in which a discount is given to a five-part bar. Whether this will be a real improvement is still under investigation.

TABLE V

The responses to problem 4, The Twax bar

THE TWAX BAR			
Answering categories		N	*Examples*
a.	Correct answer (20%) taking into account the change of the amount of reference	7	- Marks and words in the picture - Number line, 4 parts and 1 part, each "25"; at the end "total" and "each 20%" - "Without the 25% extra, each part (bar) is divided into 4th. But if you have 1/4 more, it becomes so that each part into 5th's. So if you divide 100 by 5, you get 20. So 20% is the discount" (*)
b.	Incorrect answer (25%) not taking into account the change of the amount of reference	14	- "Because 25% extra is just like 25% discount" - 100+25=125 and 125-25=100 - "I guessed" - No strategy described
c.	Incorrect answer (125%) no understanding of the problem	2	- "100%=the total candy bar + your 25% = 125%, which equals the total of the new candy bar"
d.	Incorrect answer (others)	12	- 15%: "Added" - 15%: Bar, 5 parts, 25-25-25-25-15 - 75%: "100×.75=75%" - 1/4: "Divide it into groups of that size and see how big it fits in"
e.	No answer	4	- Twax changed into Twix - "Don't understand instructions"

CONCLUDING REMARKS

The 'Per Sense' problems turned out to be most suitable to document the achievements of the students. They can be of great help for making decisions about further instruction. As was illustrated by the last problem, they even yielded good teaching material. Yet, one should keep in mind

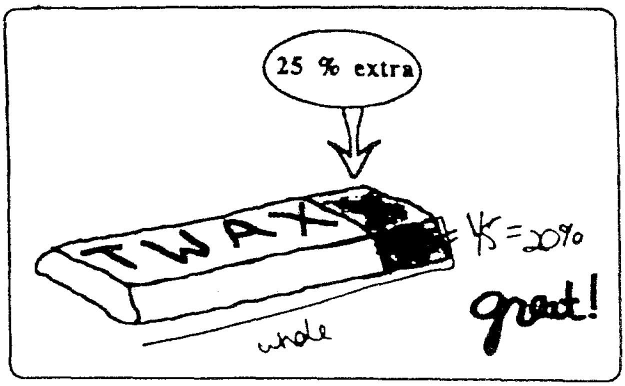

Fig. 7. Making use of the picture

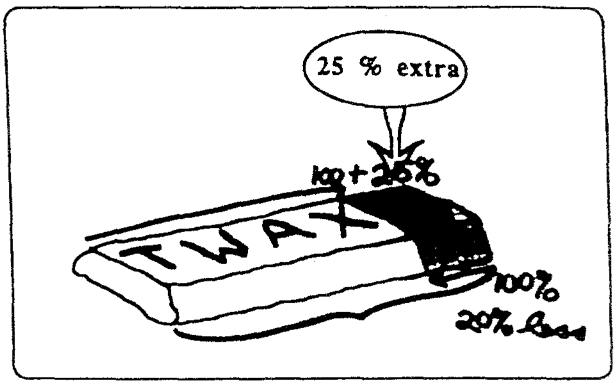

Fig. 8. Teacher's explanation

that all this was accomplished by means of a written test. Although it is clear that assessment must involve much more than mere testing and that multiple sources of information are needed, written tests as such are not all that bad. It depends on which problems are used. Different kinds of problems will evoke different results. One gets what one deserves. The richer and more open the problems, the more they will reveal the students' understanding and abilities. A consequence is, however, that the responses may be more difficult to interpret than closed, bare problems. Therefore it is better to confine oneself to a few good problems. In the end they will reveal more than a large number that are easy to grade. This study tried to develop some good problems. It was more a start than a final stage of the search for better assessment problems. Further developmental research that builds on this study is already being pursued.

NOTES

(1) This article is a revised version of a paper presented at PME XVII in Japan (Van den Heuvel-Panhuizen, 1993b).

(2) Later called OW&0C, and after Freudenthal's death re-named the Freudenthal Institute.

REFERENCES

Bell, A., H. Burkhardt and M. Swan: 1992, 'Balanced assessment of mathematical performance', in R. Lesh and S. Lamon (eds.), *Assessment of Authentic Performance in School Mathematics*, AAAS Press, Washington.

Burns, M., and C. McLaughlin: 1990, *A Collection of Math Lessons - From Grades 6 through 8*, The Math Solution Publications/Cuisenaire Company of America Inc, New Rochelle, NY.

Clarke, D.J., D. M. Clarke and Ch. J. Lovitt: 1990, 'Changes in mathematics teaching call for assessment alternatives', in T. J. Cooney and C. R. Hirsch (eds.), *Teaching and Learning Mathematics in the 1990's. 1990 Yearbook* NCTM, NCTM, Reston, VA.

Davis, R.: 1988, 'Is 'Percent' a Number?', *Journal of Mathematical Behavior* 7, 299–302.

De Lange, J.: 1987, *Mathematics, Insight and Meaning*, OW&CO, Utrecht.

De Lange, J.: in press, 'Assessment: No Change without Problems', in T. A. Romberg (ed.), *Reform in School Mathematics and Authentic Assessment*, SUNY Press, Albany, NY.

Freudenthal, H.: 1968, Why to Teach Mathematics so as to Be Useful?, *Educational Studies in Mathematics* 1, 3–8.

Freudenthal, H.: 1973, *Mathematics as an Educational Task*, Riedel Publishing Company, Dordrecht.

Freudenthal, H.: 1983, *Didactical Phenomenology of Mathematical Structures*, Riedel Publishing Company, Dordrecht.

Gravemeijer, K.: 1982, 'Het gebruik van contexten', *Willem Bartjens*, 1, 4, 192–197.

Gravemeijer, K.: 1990, 'Context Problems and Realistic Mathematics Instruction', in K. Gravemeijer, M. van den Heuvel, and L. Streefland, *Contexts, Free Productions, Tests, and Geometry in Realistic Mathematics Education*, OW&OC, Utrecht, 10–32.

Graue, M. E. and S. Z. Smith: 1992, *A Conceptual Framework for Instructional Assessment*, National Center for Research in Mathematical Sciences Education, Madison, WI.

Grouws D. A., and S. L. Meier: 1992, 'Teaching and assessment relationships in mathematics instruction', in G. Leder (ed.), *Assessment and Learning of Mathematics*, ACER, Victoria.

Kouba, V., C. Brown, T. Carpenter, M. Lindquist, E. Silver and J. Swafford: 1988, 'Results of the Fourth NAEP Assessment of Mathematics: Number, Operations, and Word Problems', *Arithmetic Teacher* **36**(8), 14–19.

NCTM: 1989, *Curriculum and Evaluation Standards for School Mathematics*, NCTM, Reston VA.

Romberg, T.A., E. A. Zarinnia and S. Williams: 1988, *The Influence of Mandated Testing on Mathematics Instruction: Grade 8 Teachers' Perceptions*, NCRMSE, Madison, WI.

Romberg, Th. A., E. A. Zarinnia and K. F. Collis: 1990, 'A new world view of assessment in mathematics', in G. Kulm (ed.), *Assessing Higher Order Thinking in Mathematics*, AAAS, Washington DC, 21–38.

Schoemaker, G.: 1993, 'Is Mathematics meant to be understood?', in P. Bero (ed.), *Proceedings of the 3rd Bratislava International Symposium on Mathematical Education*, Comenius University, Bratislava, 119–124.

Streefland, L.: 1991, *Fractions in Realistic Mathematics Education. A Paradigm of Developmental Research*, Kluwer Academic Publisher, Dordrecht.

Streefland, L.: 1992, 'Thinking strategies in mathematics instruction: How is testing possible?', in R. Lesh and S. Lamon (eds.), *Assessment of Authentic Performance in School Mathematics*, AAAS Press, Washington.

Streefland, L., and M. Van den Heuvel-Panhuizen: 1992, 'Evoking pupils' informal knowledge on percents', in International Group for the Psychology of Mathematics Education, *Proceedings of the Sixteenth PME Conference*. Volume III, University of New Hampshire, Durham, 51–57.

Stenmark, J. (ed.): 1992, *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions*, NCTM, Reston, VA.

Sullivan, P., and D. Clarke: 1987, 'Good teaching and good questions', in W. Caughey (ed.), *From Now to the Future*, The Mathematics Association of Victoria, Parkville.

Sullivan, P., and D. J. Clarke: 1991, 'Catering to all abilities through the use of 'Good' questions', *Arithmetic Teacher* **39**, 2, 14–21.

Treffers, A.: 1987, *Three Dimensions. A Model of Goal and Theory Description in Mathematics Instruction - the Wiskobas Project*, Reidel, Dordrecht.

Treffers, A.: 1993, personal communication.

Van den Brink, J.: 1984, 'Numbers in contextual frameworks', *Educational Studies in Mathematics* **15**, 239–257.

Van den Heuvel-Panhuizen, M.: 1990, 'Realistic arithmetic/mathematics instruction and tests', in K. Gravemeijer, M. van den Heuvel, and L. Streefland, *Context, Free Productions, Tests, and Geometry in Realistic Mathematics Education*, OW&OC, Utrecht, p. 53–78.

Van den Heuvel-Panhuizen, M., and K. Gravemeijer: 1991, 'Tests are not all bad', in L. Streefland (ed.), *Realistic Mathematics Education in Primary School*, CD-B Press/Freudenthal Institute, Utrecht (see also Van den Heuvel-Panhuizen, M., and Gravemeijer, K., 1993, 'Tests aren't all bad. An attempt to change the face of written tests in primary school mathematics', in N. L. Webb and A. F. Coxford, *Assessment in the Mathematics Classroom*, 1993 Yearbook, NCTM, Reston VA, 54–64).

Van den Heuvel-Panhuizen, M., L. Streefland, M. Meyer, J. A. Middleton, and J. Browne: in press, 'Per Sense', in T. A. Romberg (ed.), *Mathematics in Contexts: A Connected Curriculum for Grade 5–8*, Encyclopaedia Britannica Educational Corporation' Chicago.

Van den Heuvel-Panhuizen, M.: 1992, *Per Sense Test*, Freudenthal Institute, Utrecht (internal publication - draft version).

Van den Heuvel-Panhuizen, M.: 1993a, 'New chances for paper-and-pencil tests', paper presented at the CIEAEM 45, Sardinia, Italy.

Van den Heuvel-Panhuizen, M.: 1993b, 'New forms of assessment, but don't forget the problems', *Proceedings of the Seventeenth International Conference for the Psychology of Mathematics Education*, University of Tsukuba, Tsukuba, Ibaraki, Japan, Volume III, 186–193.

Freudenthal Institute
University of Utrecht,
Tiberdreef 4,
3561 GG Utrecht,
The Netherlands