Paper–and–pencil assessment that provides footholds for further instruction needs to break with a number of taboos in assessing mathematical knowledge

Marja van den Heuvel–Panhuizen
Freudenthal Institute, Utrecht University
The Netherlands

Introduction

One of the objections to traditional written mathematics tests, like multiple-choice and short-answers tests, is that they do not provide enough information about the students' learning process. Specifically, they do not inform about the various strategies that students employ when solving problems, and do not reveal, for instance, how an incorrect answer came about. As a consequence, it is no wonder that, for instance, the "Assessment Standards" (NCTM, 1995) caution that evidence acquired exclusively from short-answer and multiple-choice problems may lead to inappropriate inferences. Joffe (1990) even wonders:

> ... what kind of teaching would be guided by the results of tests which assess only the things accessible by timed multiple-choice tests. (Joffe, 1990, p. 158)

It has long been thought that instead of written tests, individual interviews are the best and only possible means for getting real insight into student's understanding and thinking strategies. Even the root of the word "assessment" refers to this. It means that the assessor has to "sit with" a learner in order to be sure that the student's answer really means what it seems to mean (Wiggins, 1989). So, it is not surprising that interviews immediately come to the fore if one likes to make diagnostic conclusions with more certainty. See, for instance, the following remark made by Clements (1980):

> It is obvious that any inferences about a child's thinking drawn from his written response alone represent little more than guesswork on the part of the researcher. Written responses can suggest to a researcher, or teacher, reasons why a child is making errors, but structured interviews must be conducted with the child before consistent patterns of errors can be determined with any degree of certainty. (Clements, 1980, p. 7)

In spite of widespread criticism of written assessment, this assessment method has not been ruled out in classroom practice. Nor has it been abandoned in educational research aimed at the development of assessment alternatives that are in tune with the changes in mathematics education. Joffe (1990), for example, begins her outline of how assessment methods might be improved with a section-heading that reads:

> Getting more from existing written tests. (Joffe, 1990, p. 147)

In her opinion, a more creative exploitation of the potential of such tests is necessary. The same ideas are also expressed by others. Ginsburg et al. (1992), for instance, still see opportunities for written assessment.

> The point of the story is that even a dumb test can be exploited as a useful assessment technique, provided children are encouraged to reflect on and reveal the solution processes employed. (Ginsburg et al., 1992, p. 286)

The question that remains is how written assessment methods could be improved in such a way that they can help teachers to make informed instructional decisions. The present article will address this question within the framework of "Realistic Mathematics Education" (RME) — the approach to mathematics education in the Netherlands — and will show some of the "didactical" alternatives for the traditional written assessment that have been developed within this approach. The focus will be on the alternatives developed for primary education.

RME assessment in short

About 30 years ago, Freudenthal and his colleagues from the former IOWO (which now is called Freudenthal Institute) took the first steps in the direction of this new approach to mathematics education called "Realistic Mathematics Education". An a posteriori description of the theoretical framework of this approach has been made by Treffers (1987). Characteristic of RME is the rejection of the mechanistic, procedure-focused way of teaching in which the learning content is atomized into 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 and insight, and its applications, RME is aimed at the development of all these, by guided-reinvention. This means that RME has a more complex and meaningful conceptualization of teaching and learning. The students are considered to be active participants in the teaching-learning process in which classroom interaction plays an important role.

Crucial for the learning process are situations and contexts by which the students can constitute their mathematical knowledge and tools. In this learning process the students pass through various levels of understanding; starting at an informal context-connected level and eventually reaching a more formal level. In RME, models serve as an important device for bridging the gap between different levels of understanding.

Today, RME is still in development. Refinements continue to be made and emphases are altered on the basis of developmental research. One of the topics for which this is the case is assessment.

The initial phase of the development of RME was characterized by considerable opposition to what was seen as unsound testing both in the Netherlands and abroad. Freudenthal and the other proponents of RME had severe objections against the then prevailing optimism regarding achievement tests which, for instance, was expressed by Bloom and Foskay (1967). Freudenthal and his colleagues did not agree with the existing goals of mathematics education and the way they were described, had difficulties with the taxonomies as a means for test construction, and objected to the one-sidedness of the psychometric foundation in which formal characteristics were valued and subject matter and educational content were ignored. Furthermore, there was also dissatisfaction with regard to the formalized design of the tests, the traps in test questions, and the way in which the students' answers were evaluated. A more detailed overview can be found in Van den Heuvel-Panhuizen (1996).

The often fierce campaign waged against the existing tests at that time may give the impression that the early years of RME were primarily ones of anti-assessment. Upon closer examination of the publications of that time, however, it becomes clear that, alongside opinions on what *not* to do, there were also some very clear ideas about what *should* be done. Within RME there was a high priority on observation. Furthermore, one stressed the integration of instruction and assessment, a holistic approach in assessment, open-ended test formats and the use of genuine application problems. Although there were many objections to the existing tests, assessment was considered a meaningful activity, by which the teacher should be able to check the influence of the teaching process, particularly in order to improve it. This means that assessment was not viewed in the narrow sense of determining what the students had learned, but that it also was regarded from the viewpoint of educational development. In other words, assessment was not only intended for looking back, but also for looking forward. It should guide further instruction. In other words, assessment *of* learning is extended with assessment *for* learning. This broadening of the range of assessment is still very current (see, for instance, Stiggins, 2002). Moreover, it is noteworthy that notwithstanding the preference for individual interviewing and observation, in the early stages of RME, written tests were not excluded from the assessment repertoire.

> The point is to test sensible, [...] and this means that the function, rather than the form of the assessment is of primary importance. (Freudenthal, 1976, p. 69)

In comparison with the early period of RME — when primary education in the Netherlands was threatening to become awash in a flood of tests and test-oriented education — at the end of the eighties, a shift has occurred in the attitude of RME developers with respect to assessment. This means that there is still considerable interest in assessment. In contrast with two decades ago, however, when most assessment was imported and was held in considerable disregard by the pioneers of RME, in recent years more interest has been shown from within.

A significant catalyst for this new interest in written assessment was provided by the endeavor to secure the new secondary school mathematics curriculum through appropriate exams. More or less obliged to do so by legislatively required exams, the development of assessment appropriate to RME was begun in the early nineteen-eighties, simultaneous with the reform of the secondary education mathematics curriculum. This secondary mathematics education reform was carried out in the framework of the HEWET project, which was established for this purpose. The dissertation of De Lange (1987) gives a detailed account of the results of the related assessment research.

Later, alternatives to the existing methods of assessment in primary education were also sought along the same lines. Just as the HEWET project heralded developmental research on assessment for secondary education, so did the MORE research project play an important role in similar research for primary education. This project was in fact a study of the implementation and effects of mathematics textbooks (see Gravemeijer et al., 1993). The development of tests, which were needed in order to compare the learning achievements of students who had used different textbooks, gradually began to expand and to focus more on the implications and possibilities of assessment within RME. In this respect, the MORE research also became a field of exploration into RME assessment. The major concern of this part of the MORE research project was to find new opportunities for paper–and–pencil tests, and especially short-tasks problems, for "didactical" assessment: assessment meant to guide educational decisions taken by the teacher.

In the next section, some of the findings of this developmental research on assessment are discussed. All examples of assessment problems shown here originate from the MORE research. A more detailed discussion can be found in Van den Heuvel–Panhuizen (1996). This research resulted in a large number of measures that can be taken to make written short-task assessment problems more informative.

Taboos to be broken in order to make assessment more informative

One of the main characteristics of the measures that can make written short-task assessment problems more informative is that they all break with some misconceptions about mathematical problems, namely that:

- mathematical problems are or should be solved in fixed manners;
- each mathematical problem has only one correct answer;
- the correct answer to a mathematical problem can always be determined;
- mathematical knowledge that has not yet been taught cannot be assessed;
- good assessment problems should be unambiguous.

These ideas about mathematical problems have long determined the design of assessment problems. As a matter of fact, for the standardized way of testing they are still the ruling assumptions. One might even say that in the assessment world there are taboos against thinking differently about these issues.

In concerning "Mathematical problems of which the correct answer cannot be determined" it is shown that breaking with these taboos opens the road for more informative assessment problems. Some examples will be worked out, other examples will only indicate what might be possible with an other type of assessment problem.

Notwithstanding the revealing capacity of these problems, teachers (and researchers) should always be alert and check their interpretations and conclusions. This is as true for these new problems as it was for the traditional ones. The difference, however, is that these new problems are not exclusively focused on certainty, but instead allow more uncertainty in order to make the problems more revealing and create space for an assessment practice on a human scale (see also Streefland and Van den Heuvel–Panhuizen, 1999).

Mathematical problems that can be solved by applying different strategies

Within the psychometric tradition of test design it is generally neglected that problems can be solved in different ways. Problems are seen as a black box the answer of which indicates that the student has achieved a certain ability. This connection is based on the assumption that solving a particular problem stands for having a particular understanding. Looking at strategies does not make sense in such a case. Information about how the problems are solved can only be of use for diagnosing why students arrived at a wrong answer and what is lacking in their understanding.

Completely different, however, is the situation if the taboo against assessment problems that can be solved in a variety of ways is broken. This means that one has acknowledged that mathematical problems can be solved by applying different strategies that can each refer to different levels of understanding and can be considered as expressions of different abilities. In this case the applied strategy is part of the "answer".

The foregoing has consequences for the design of assessment problems. The way in which the problems are presented to the students should prompt them to give information about the method applied to solve the problem. Various actions related to the wording of the problem can be taken to give the teacher (or researcher) access to the students' strategies:

I. Asking the students a direct question about the strategy, like for instance, "Explain how you came to your answer" or "Show your thinking".

II. Providing the students with strategy information and ask them to reflect on the adequacy of the given strategy or ask them to complete the given strategy; even the multiple-choice format is useful in this respect (see Section 4 of this article).

III. Asking the students about the strategy indirectly, for instance by asking the students to explain the solution of the problem to a friend or to a student who was ill and could not attend the mathematics class in which this problem was discussed.

IV. Putting a piece of scrap paper or a work area on the test sheet.

This last action, having some space on the worksheet to work on, is initially meant for the students' own use. They may use it for writing down intermediate results. As such this action might support the students' reasoning and might help them to calculate the answer. Aside from this, however, this "solidified student behavior" also gives important clues about how the students solved the problem.

Examples of strategies that can be revealed by putting a piece of scrap paper on the test sheet are shown in Figure 2. The student work belongs to the problem shown in Figure 1. This problem is about a word game (called Scrabble) in which one of the players made four words. The students are asked to determine the total number of points. The test problem is administered in November grade 3 (7–year–old students).

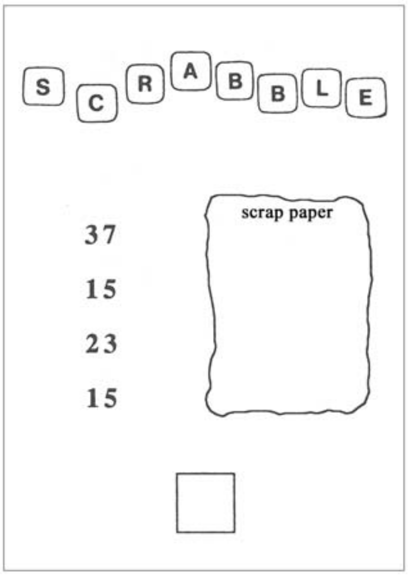

Figure 1. *Scrabble* problem

This simple short-task problem demonstrates that written assessment should not necessarily be restricted to providing information on whether the student is able to arrive at the correct answer. In addition to this, written assessment can also make visible the student's thinking and applied strategy. By doing this, the problem provides the teacher with an amount of valuable information for continuing instruction on smart mental calculation. The strategy data indicates that students (2), (3) and (6) still apply a kind of ciphering strategy (a digits-based algorithmic method of calculation). They apply this strategy either in a complete written mode (2 and 3) or in a more mental way (6) (seven and five is 12; and three is 15; and five is 20 …). The work of students (1) and (5), however, show how this ciphering strategy can evolve towards a smart mental calculation strategy. The information on the pieces of scrap paper provides the teacher with footholds for further instruction. The knowledge about the strategies and the possibilities for shifts to a higher level can immediately be used in tomorrow's teaching, for instance by putting the strategy of student (1) and student (2) on the blackboard and inviting the students to explain and compare both strategies.

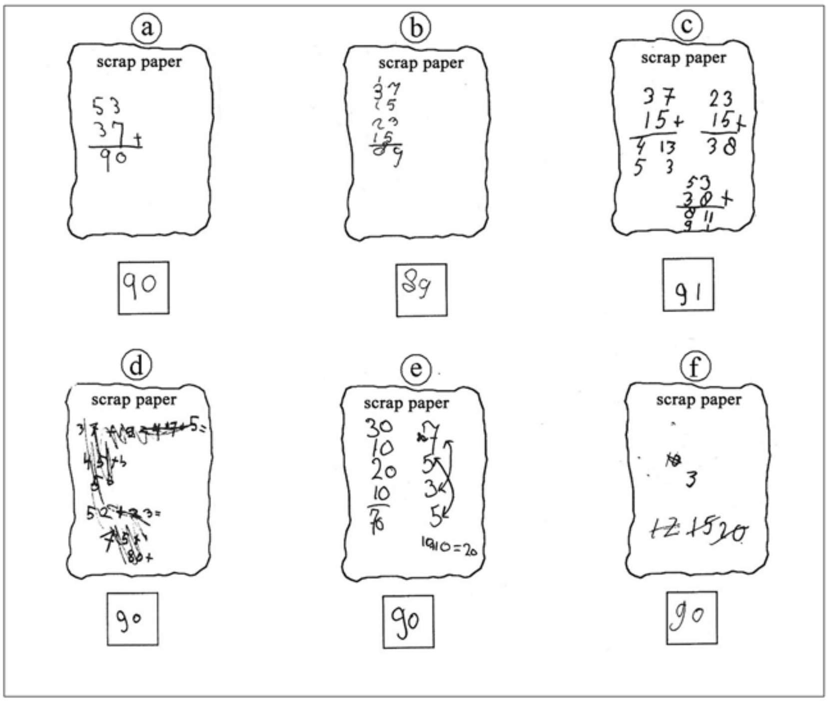

Figure 2. Sample of student work on the *Scrabble* problem

What this strategy data also makes clear is that answers on their own are not always sufficient to determine the level of understanding. The strategy itself is also a strong indicator of the students' achievement level (see also Van den Heuvel–Panhuizen and Fosnot, 2001).

Mathematical problems that may have more than one correct answer

Mathematical problems cannot only be solved in different ways, but may also have more than one correct answer. This is the second taboo that must be broken. Multi-correct problems give the students room to answer the problem on their own level and are at the same time very informative for the teachers. An example of such a problem is shown in Figure 3. It was also administered in November grade 3. The problem is about a mother, a father and two children who are going to the circus. They have to pay an admission fee of 50 guilders in total. The question is to specify the admission fee.

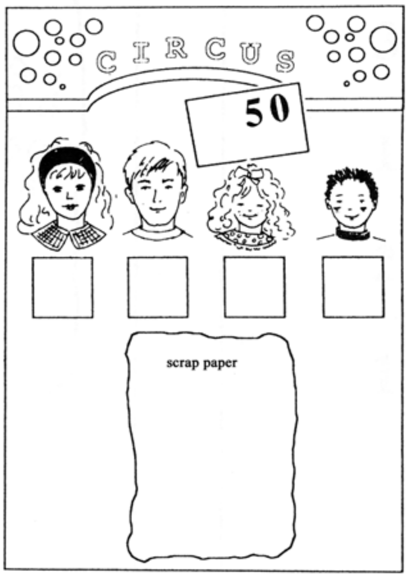

Figure 3. *Circus* problem

The tickets filled in by the students differed greatly. For instance, there were students who had three persons pay 10 guilders and the fourth person 20 guilders. Others students distinguished between adults and children and arrived at 13 and 12 guilders (see Figure 4a), or 15 and 10 guilders, respectively. Another group of students tried to divide the total fee equally over four persons. Not everybody succeeded in doing this. Nevertheless, "creating" some change (see Figure 4b) was not at all a bad solution. Even more surprising were the students who came up with the answers 121/2 or 12.50 (see Figure 4c). This was an unexpected answer because fractions and decimals had not yet been taught. The scrap paper of Figure 4c reveals that this student used a multiplying on (or more precisely an adding on) strategy and after several attempts he or she came to the answer 12.50. First, four times 15 was tried. After 30+15 the student stopped. Too much. Then, four times 11 was tried, with 44 as a result. Too little. After that, the move was made to four times 12; probably with two times 25 as a step in between. Finally, the remainder of two guilders was divided by four; that gave the final answer of 12.50.

Again the test results provide interesting issues to discuss in class and to work on further. Students who gave answers like student (a) could be asked whether other charges are also possible. Students who answered like student (b) could be told that there was no change, and students who came up with fractions could be asked to express their answer in a way in which fees are normally expressed (as a decimal instead of a fraction). Finally, student (c) could be asked to reflect on her or his strategy in order to find a shorter way to come to her or his answer. The teacher might suggest to this student (and the others) to start from two times 25.

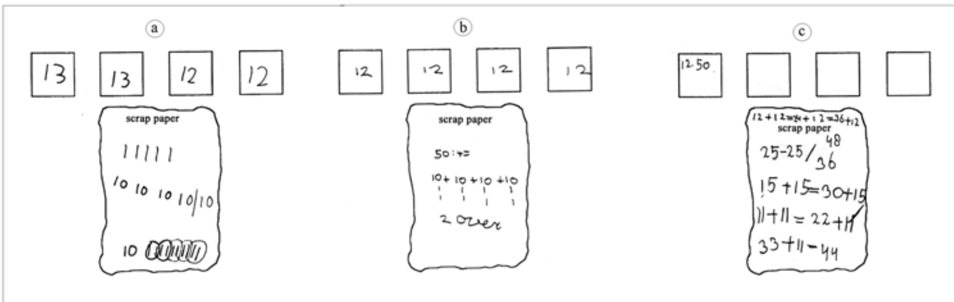

Figure 4. Sample of student work on the *Circus* problem

Making use of multi-correct problems in order to find footholds for further instruction is also very valuable in the lower grades where the students can have more difficulties in communicating about their strategy in writing and where it is not as feasible to work with scratch paper on the test pages. The *Candle* problem clearly demonstrates that footholds for further instruction can be found in the answers alone (see figure 5).

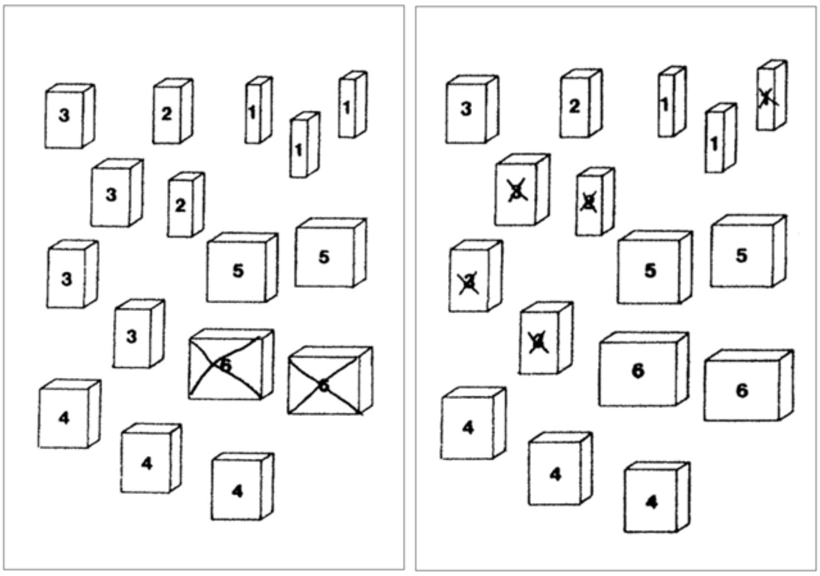

Figure 5. Examples of student work on the *Candle* problem

In this problem the students were asked to buy twelve candles. They could make their own choice of which boxes to buy. The solutions first graders gave in February show that this problem can expose relevant differences between students. The disparity that emerged may indicate a difference in degree of familiarity with the number structure.

Take, for instance, the two solutions shown in Figure 5. It would seem that the student whose work sheet is on the left already knows that twelve can be split into six and six, while the other student kept adding up until the total of twelve was reached.

Mathematical problems of which the correct answer cannot be determined

A very firm taboo in mathematical education is against problems in which not only several answers are possible, but in which it is not even clear what is a correct answer and what is not. Such problems are seldom presented to students, and in assessment they are absolutely rare. Yet these problems are both in tune with the new ideas about mathematics education and provide important information about the students, understanding and strategies. An example of such a problem is shown in Figure 6.

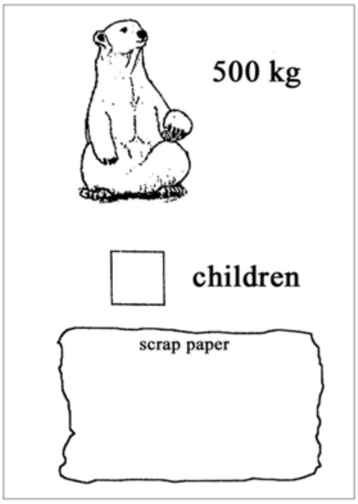

Figure 6. *Polar Bear* problem

The students are asked how many children it would take to equal the weight of one 500 kg polar bear. Due to the absence of information on the children's weight, this becomes a real problem. It gives no indication of what kind of procedure must be performed, and it implicitly requires the students to think up the weight of a typical child first. In both areas, the students are given room to make their own constructions, which may be considered the most important aspect of such problems. On the one hand, the students are provided the opportunity to show what they know, and on the other hand, the teachers thus acquire information on how their students tackle problems, on their knowledge of daily life measurements, on which "division" strategies they apply, and on which models and notation schemes they use to support their thought processes in a hitherto unfamiliar division problem. A selection of student work is shown in Figure 7.

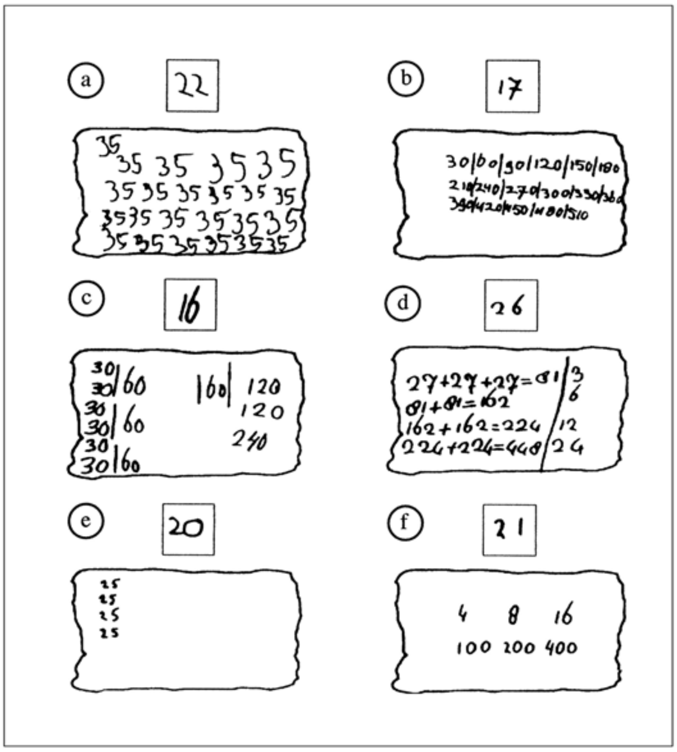

Figure 7. Sample of student work on the *Polar Bear* problem

Besides problems like the *Polar Bear* problem in which the students have to construct part of the problem by themselves (because not all the data is given and the solution procedure is not prescribed). If one needs information for informed instructional decisions, “own productions” in which students have to make up their own problems themselves are also very fruitful. Actually, one could say that students’ own productions are the jewels in the crown of assessment problems. As was shown in a developmental research project involving percentage problems, students’ own productions can provide the teacher both with a splendid cross-section of student understanding at a particular moment, and a longitudinal section of the learning path that students will generally follow with respect to percentage (Van den Heuvel–Panhuizen, Middleton, and Streefland, 1995). It is this last aspect that makes own productions an especially suitable source for deriving indications — and even educational material — for further instruction.

Assessing mathematical knowledge before it has been taught

To make written assessment more helpful in guiding the teaching-learning process, another taboo should be broken. This is the taboo against problems that assess mathematical knowledge that has not yet been taught. This is a misconception that has far-reaching consequences for educational practice, because it implies that assessment is necessarily restricted to looking back. However, if mathematics education is considered — as in RME — as building on the informal knowledge of children, looking back is not enough. On the contrary, in such an approach to mathematics education one has to foresee where and how one can anticipate from a distance that which is just coming into view (see Streefland, 1985). This means that assessment should also include looking forward. Crucial for this is that the students are presented with problems which are accessible and do not contain obstacles in the form of specific terms or procedures. If a student never learned how to do long division, one cannot expect that she or he is able to do a long division. On the other hand, the same student might be able to solve division problems by applying informal methods of division, such as repeated subtraction or repeated adding-up. See, for instance, the student work on the *Polar Bear* problem in Figure 7. Another example of “advance testing” can be found in the *Bead* problem and the corresponding problem with bare numbers shown in Figure 8.

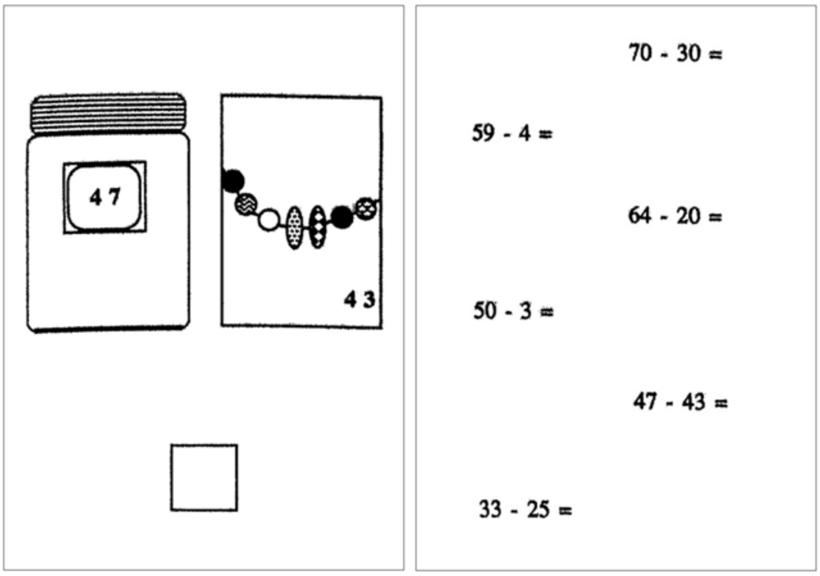

Figure 8. *Bead* problem and corresponding problem with bare information

The context problem is that a jar contains 47 beads, 43 of which are used to make a necklace; and the question is how many are left. At the time the test was administered (November grade 2) problems like 47–43 were not dealt with in class extensively. Only 38% of the students found the correct answer. Yet, 60% of the students solved the context problem correctly. The facilitating role of the context apparently led them to calculate differently here than in the case of the bare problem. Instead of the often laborious subtraction procedure that is usually taught for problems like 47–43, the students apparently used the informal strategy of "adding-up" that was elicited by the context. The results make clear that this strategy supported by a particular context situation should not be forgotten when proceeding with teaching the students how to do subtraction problems.

Using "elastic" assessment problems

By breaking the taboos mentioned above, an open attitude is created that is necessary if what students can do and how they go about doing it, is to become apparent. However, if students are to have an optimal opportunity for showing what they can do, then there should be another condition too, namely, that "all-or-nothing" testing be avoided as much as possible. This means that problems should be used that can be solved by the quick students and by the slower students as well. In other words, one should use assessment problems that have some latitude. Most of the previous examples do have such elasticity. By choosing their own strategy in the Scrabble problem, by selecting particular boxes in the *Candle* problem, by dividing the total fee in the *Circus* problem in a certain way, and by choosing a particular average weight of a child in the *Polar Bear* problem, the students can adjust the problems to their own level. Perhaps the greatest assessment taboo of all is against giving the students this freedom. Particularly within the world of psychometric assessment this is a great taboo. Breaking with this taboo implies namely that a break must be made with the prevailing psychometric idea that assessment problems should be unequivocal, the idea that students must interpret the problem in the same way and that it must elicit the same way of thinking or the same procedure, namely the one that is assessed. Although a portion of the assumed certainty is indeed lost with the introduction of this elasticity, such problems with latitude provide a wealth of information—particularly for daily classroom practice.

Open-mindedness is necessary in assessment reform

To improve assessment, not only the taboos of psychometricians, but also the taboos of didacticians involved in assessment reform, must to broken. As was indicated in the beginning of this article, the reform aimed at making assessment more in tune with the new goals and approach to mathematics education is often focused on the abandonment of multiple-choice tests. The next example, however, demonstrates that multiple-choice problems can be quite eye-opening. The example concerns a series of multiple-choice problems (see Figure 9) in which beginning Dutch fifth-grade students were asked how they would calculate three different arithmetic problems if they were given little time to do so. In each problem, they could choose from three different methods: calculating on paper, mental arithmetic, or using a calculator.

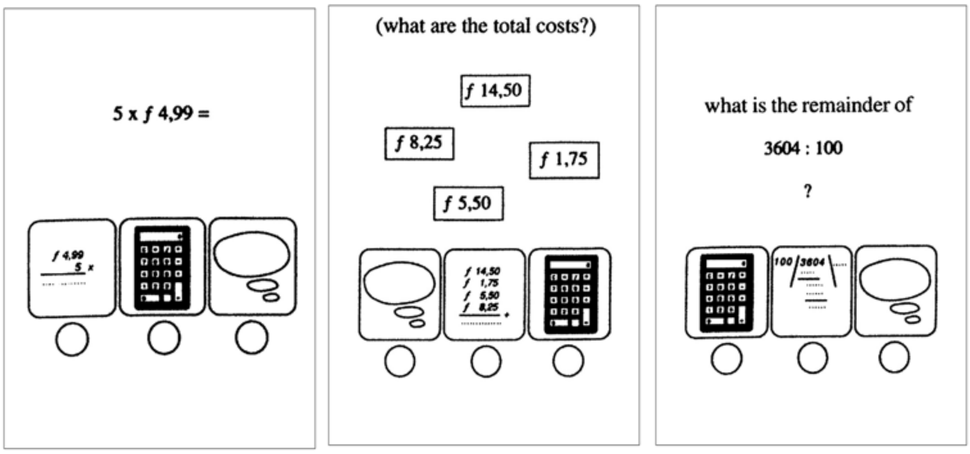

Figure 9. Three multiple-choice problems in which the strategy is requested

The strategies the students chose for each of the problems revealed a great deal about their insight into numbers and operations.

grade 5, Sep (n = 376)	5 x ƒ 4,99	ƒ 14,50 + ƒ 8,25 + ƒ 5,50 + ƒ 1,75	the remainder of 3604 + 100
strategy	**% selected**	**% selected**	**% selected**
mental calculation	25	14	24
column calculation	26	42	33
calculator	48	43	43
unknown	1	1	1

Table 1. Strategies selected by the fifth–grade students

As can be seen in Table 1, it is clear that many of the students did not recognize the special properties of the numbers and operations in question, and, consequently, would have had difficulty with clever mental calculation. Only one-fourth of the students chose to calculate 5×4.99 mentally, and more than 80% of them clearly were not aware that 14.50 and 5.50, and 8.25 and 1.75 fit nicely together. Moreover, the problems revealed quite a bit about the students' (lack of) familiarity with calculators. Consider, for instance, that 43% of the students chose to use the calculator for finding the remainder in the division problem.

This example demonstrates quite plainly that the format itself is not the cause of bad assessment. Even the currently renounced multiple-choice format can elicit very useful information. What is needed, however, is a good match; in other words, the assessment format, and the goal and content of the assessment should be well attuned to one another (see also Lamon and Lesh, 1992).

Another thing that is stressed by this example is that open-mindedness is crucial for the development of new methods of assessment. Only by this was it possible to revalue old-fashioned forms like paper-and-pencil tests and, instead of removing them from the assessment repertoire, give them a new chance.

Final remark

The above mentioned "didactical" alternatives for the traditional written test problems have a lot in common with what is called "diagnostic assessment". In the end, both are meant to provide the teacher with information needed for making instructional decisions on how to proceed. On the other hand, there are also differences between the two. Didactical assessment, as proposed here, has strong links with the didactics of mathematics education, and is as such more integrated into the teaching-learning process and connected with the content, whereas diagnostic assessment has emerged more from the psychological approach to education. The consequence of this is that diagnostic assessment is aimed more at how a student is proceeding along certain learning paths and whether the prerequisites for each new stage in development are fulfilled. This implies that diagnostic assessment is more looking back than looking forward. Finding footholds for further instruction, however, needs looking forward.

References

Bloom, B. S. and A. W. Foskay (1967). Formulation of hypotheses. In T. Husen (Ed.), *International Study of Achievement in Mathematics. A comparison in twelve countries* (I, pp. 64–76). Stockholm: Almqvist and Wiskell.

Clements, M. A. (1980). Analyzing children's errors on written mathematical tasks, *Educational Studies in Mathematics, 11*, 1–21.

De Lange, J. (1987). *Mathematics, insight and meaning.* Utrecht: OW&OC, Utrecht University.

Freudenthal, H. (1976). De wereld van de toetsen [The world of tests], *Rekenschap, 23,* 60–72.

Ginsburg, H.P., L. S. Lopez, S. Mukhopadhyay, T. Yamamoto, M. Willis and M. S. Kelly (1992). Assessing Understanding of Arithmetic. In R. Lesh and S. J. Lamon (Eds.), *Assessment of Authentic Performance in School Mathematics* (pp. 265–289). Washington: AAAS Press.

Gravemeijer, K., M. van den Heuvel-Panhuizen, G. van Donselaar, N. Ruesink, L. Streefland, W. Vermeulen, E. te Woerd, D. van der Ploeg (1993). *Methoden in het reken-wiskundeonderwijs, een rijke context voor vergelijkend onderzoek* [Textbook series in mathematics education, a rich context for comparative research]. Urecht: CD-B Press / Freudenthal Institute, Utrecht University.

Joffe, L. S. (1990). Evaluating assessment: examining alternatives. In S. Willis (Ed.), *Being Numerate: What Counts?* (pp. 138–161). Hawthom, Victoria: Australian Council for Educational Research.

Lamon, S. J. and R. Lesh (1992). Interpreting responses to problems with several levels and types of correct answers. In R. Lesh and S. J. Lamon (Eds.), *Assessment of Authentic Performance in School Mathematics* (pp. 319–342). Washington: AAAS Press.

NCTM (1995). *Assessment Standards for School Mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Streefland, L. (1985). Vorgreifendes Lernen zum Steuern Langfristiger Lernprozesse [Anticipatory learning to steer long-term learning processes]. In W. Dorfler and R. Fischer (Eds.), Empirische Untersuchungen zum Lehren und Lernen von Mathematik. Beitrage zum 4. Internationalen Symposium für Didaktik der Mathematik in Klagenfurt in 1984 (pp. 271–285). Wien: Hölder-Pichler-Tempsky.

Streefland, L. and M. van den Heuvel-Panhuizen (1999). Uncertainty. A metaphor for mathematics education. *The Journal of Mathematical Behavior, 17*(4), 393–397.

Stiggings, R. (2002). Assessment for learning. *Education Week*, 13, March, 2002.

Treffers, A. (1987). Three Dimensions. A Model of Goal and Theory: Description in Mathematics Instruction — the Wiskobas Project. Dordrecht: Reidel Publishing Company.

Van den Heuvel–Panhuizen, M., J. A. Middleton, and L. Streefland (1995). Student-Generated Problems: Easy and Difficult Problems on Percentage, *For the Learning of Mathematics, 15*, 3, 21–27.

Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education.* Utrecht: CD-B Press / Freudenthal Institute, Utrecht University.

Van den Heuvel-Panhuizen, M. and C. T. Fosnot (2001). Assessment of mathematics achievements: not only the answers count. In M. van den Heuvel-Panhuizen (ed.), *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education.* Volume 4 (pp. 335–342). Utrecht: Freudenthal Institute, Utrecht University.

Wiggins, G. (1989). A True Test: Towards More Authentic and Equitable Assessment, *Phi Delta Kappan, 70*, May, 703–713.

Resumo. Neste artigo, a avaliação é encarada numa perspectiva educacional. Geralmente, as questões de resposta curta ou os problemas de escolha múltipla são considerados pouco adequados para orientar as tomadas de decisões sobre o ensino. Mudar para entrevistas individuais pode ser uma solução, mas não é a única. Neste artigo são discutidas algumas medidas para quebrar certos preconceitos, tornando instrumentos de papel e lápis mais informativos. São apresentados diversos exemplos da avaliação sobre números, no ensino primário.

Palavras-chave: Avaliação, números, 1º ciclo.

Abstract. In this article, assessment is viewed from the perspective of education. Short-answer or multiple-choice problems are often considered to be not very suitable for informed instructional decision making. Moving to individual interviews is one possibility but is not the only solution. In this paper some taboo-breaking measures are discussed that can make paper-and-pencil assessment methods more informative. Several examples are shown from assessing number at primary school level.

Keywords: Assessment, number, primary school.

▪ ▪ ▪

MARJA VAN DEN HEUVEL–PANHUIZEN
Freudenthal Institute, Utrecht University
The Netherlands
`m.vandenheuvel@fi.uu.nl`