Reform under attack – Forty Years of Working on Better Mathematics Education thrown on the Scrapheap? No Way!

Marja van den Heuvel-Panhuizen
Freudenthal Institute, Utrecht University
<m.vandenheuvel@fi.uu.nl>

This paper addresses the reform of mathematics education in the Netherlands and the attacks that presently take place against this reform. The attacks concentrate on primary education and criticize in particular the program for teaching calculation skills with long division as a case in point. The paper gives an overview of what Realistic Mathematics Education (RME) stands for, and what mathematics education the reform-attackers have in mind. Furthermore, attention is paid to possible factors that could have triggered this attack, and what other countries may learn from it.

What comes after a reform of mathematics education? It is beginning to look as if the answer to that question is: War. In the late nineties of the previous century a so-called 'Math war' overran the United States, starting as a reaction to the reform-based curriculum and teaching approach in California (Becker & Jacob, 2000). Presently, we have a similar situation in the Netherlands. The attack concentrates on primary education. The focus of the assault is both on what students should learn and on didactical methods. Without any evidence from research, the main principles of 'Realistic Mathematics Education' (RME), which form the basis of the Dutch reform, are called 'didactical blunders'. At the same time an argument is made in favour of a form of education that is almost the polar opposite of what RME stands for. According to the reform-attackers, mathematics should not be taught in context, informal strategies should be avoided because they confuse children, progressive schematisation leads to a long, unnecessary detour and the focus should not be on understanding; because understanding comes automatically after training. Moreover, it is stated shamelessly that children do not need to think. In the mind of the reform-attackers the main content to be learned in primary school must consist of written algorithms.

The development of RME commenced in the late 1960s and is still developing to reach further maturity as an instructional theory and in its implementation into educational practice. Up to some forty years ago, this development process went on with trial and error, but in relative peace. Actually, it was a silent revolution; there was hardly a whisper in the media (Treffers, 1991a). There was very little opposition and no pressure from above. For example, in all those forty years there has been hardly any government involvement with the reform of mathematics education. The Ministry of Education was only involved in a facilitating sense. Government subsidy made it possible that an extensive infrastructure arose in the Netherlands allowing development, research and training to take place in mutual coherence and cooperation with the field of education. Where other educational researchers were blamed for the gap between their research and educational practice, we were held up as an example of how research should be done; see the report by the Education Council of the Netherlands (Onderwijsraad, 2003). Not only was there recognition in our own country, but we also inspired developments in many other countries. Our work was, and still is, in great demand all over the world, even if only perhaps that it gives these countries good hope of being able to attain such high test scores as the Dutch.

Looking back, the peace we had for forty years had everything to do with student performance. In 2004 and 2005, the first reports of disappointing results in both national tests

L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia.* Fremantle: MERGA.

and in TIMSS and PISA came in. From that time on, RME was coming under fire. Unfortunately, the discussion that was started was anything but academic, but was a veritable libel campaign that took place mainly in newspapers and on websites. It happened regularly that there were sneering articles in these media about the ruination of Dutch mathematics education by the Freudenthal Institute — or more precisely — about the ruination of the education in written arithmetic, because that was what the discussion focused on. Dutch students could no longer do written calculations. Long division was often used as an example.

What is in the Newspapers in the Netherlands?

The opponents of RME have as their leader, a professor in mathematics, who used to teach at a military academy. He and his supporters argue for mathematics education based on bare numbers, where the teacher demonstrates the problems and the students learn by imitation. For each operation there is one prescribed procedure, namely that of algorithmic addition, subtraction, multiplication and division. Using this algorithmic procedure starts already for problems up to 100. There must be a lot of practice, and that practice will automatically lead to insight (Van de Craats, 2007).

Along with this plea for returning to the mathematics education of forty years ago — or rather, to the one-sided view that the attackers of RME have of mathematics education in the past — a great number of misconceptions and inaccuracies about RME, is put forward in the media:

- If we are to believe the media, students do not get the opportunity to practice in RME. This, however, is a flagrant contradiction of RME's long tradition in including practice (see De Jong, Treffers, & Wijdeveld, 1975; De Moor, 1980; Treffers & De Moor, 1990; Van den Heuvel-Panhuizen & Treffers, 1998; Menne, 2001; Van Maanen, 2007). An aside here is that in RME this means practising with understanding and coherence, which is radically different from the isolated drill that the opponents of RME have in mind.
- RME is said to have abolished algorithmic calculations. Again this is simply not true. See what the main curriculum documents — the so-called 'Proeve' (Treffers & De Moor, 1990) and the TAL learning-teaching trajectory for whole number calculation (Van den Heuvel-Panhuizen, 2008b) that describes the learning-teaching trajectory for whole number calculation in primary school mathematics — say about algorithms. Moreover, traditional algorithms are being widely taught (Janssen, Van der Schoot, & Hemker, 2005). It should be said though, that the degree to which that is done differs for each textbook series. For example, the RME textbook series *Wereld in Getallen (WIG)* contains a total of around 3000 digit-based algorithmic problems, 1200 for addition and subtraction, 1000 for multiplication and 750 for division (Levering, 2009).

  A disturbing misapprehension that is being presented in this context in the media — and which clearly evidences a lack of didactical knowledge — is that the so-called 'traditional', digit-based algorithm and the 'new' method of whole-number-based written calculation (more on which later) are being showcased as two opposite end goals. The opponents of RME do not realise that the whole-number-based method is a transparent and insightful introduction to the digit-based algorithm and clearly show their didactical lack of understanding by presenting ridiculous examples of this approach in the media.
- RME supposedly only involves word problems. This is another unfounded assertion. One only needs to open an RME textbook to see that it is filled with a large amount of

bare number problems. Of course the amounts are different for each textbook, but there is not one RME textbook that does not have bare number problems.

However, there is one more reason that makes this insinuation far from the truth. Word problems have always been an object of suspicion within RME (Van den Heuvel-Panhuizen, 1996). Of course, linking problems to reality is important. This means that within RME students are presented problems which they can imagine and with which they have daily life experience, but this does not mean that word problems have a central role in RME. The crucial point is that the problems are presented in a meaningful and accessible context. Therefore they are often presented visually through pictures, models, and diagrams. Word problems with complicated ways of explaining a problem are avoided. They cannot be considered typical RME problems. However, our opponents do try to represent them as such.

- RME is said to teach students as many different calculation strategies as possible, which confuses students. Neither the first nor the second is true. RME starts teaching with following on from what students themselves come up with and do — which has natural variation — and from thereon gradually works towards a standard method, which is however not a straitjacket. The students must have an understanding of the numbers with which they calculate, and if possible use shortened calculation methods or smart strategies — which implies an intentional variation in strategy that reflects the high level of number understanding that RME wants students to reach.
- All RME textbooks are of low quality. Another inaccuracy. The outcomes of the large-scale studies (performed by Cito, the national institute for assessment in the Netherlands) into the effects of the textbooks belie this claim. The RME textbooks were among the best textbook series more often than the traditional ones (Janssen, Van der Schoot, Hemker, & Verhelst, 1999). Later, Cito (Janssen et al., 2005) concluded that the newer textbooks, despite the differences between them, have had a small, but positive summative effect on student performances. In other words, without these textbook series, performances would likely have been lower. More about these studies later.
- Supposedly, the average Dutch student at the end of primary school is incapable of calculating. Based on the latest Grade 4 data from TIMSS (Mullis, Martin, & Foy, 2008) it is clear that this statement is wholly unfounded. If it were in fact true, it would not only be the case for the Netherlands, but for all other Western countries that took part in TIMSS. I will return to this point later.

In addition to these misconceptions and inaccuracies about current Dutch mathematics education, the reform attackers regret deeply that a reform occurred and they worry themselves sick wondering why Dutch mathematics education was reformed. Utterances like these bear painful witness to a lack of any knowledge about the problems that existed in mathematics education forty years ago both in the Netherlands and internationally.

How RME Started and What It Stands For

Although the very beginning of RME can be placed at the end of the 60s, the name ‘Realistic’ was only used at the end of the 70s (Treffers, 1991a). The very beginning of the reform movement was the start, in 1968, of the Wiskobas project (meaning ‘mathematics in primary school’) initiated by Wijdeveld and Goffree, and joined not longer after by Treffers. It was these three who in fact built the foundation for RME. In 1971, when the IOWO Institute, with Freudenthal as its director, was established for the Wiskobas project and a similar project for secondary education, the movement received a new impulse to reform mathematics education.

In the 1960 the Netherlands wanted to abandon the then prevalent mechanistic approach to mathematics education. Characteristic of this approach is its focus on calculations with bare numbers, and the little attention that it pays to applications; which is certainly true for the beginning of the learning process. Mathematics is taught in an atomised way. Students learn procedures in a step-by-step way in which the teacher demonstrates how to solve a problem.

Conversely, mathematics education in England had an empiricist slant in those days. Typical of this type of education was that students were let free to discover much by themselves and were stimulated to carry out investigations. This method deviated greatly from the, at that time existing, structuralist approach derived from the ideas from Bourbaki group about mathematics as a discipline, and which in the US led to the so-called New Math movement. This is a method of teaching mathematics which focuses on abstract concepts such as set theory, functions and bases other than ten.

In its search for an alternative for the mechanistic approach, the Netherlands pursued neither the empiricist nor the structuralistic approach. In particular through Freudenthal's opposition to the structuralistic 'New Math' movement that washed over the Netherlands, there was an opportunity to go in another direction and end up at the RME approach.

To understand this way of teaching mathematics and recognise how it differs from other approaches to mathematics education which were manifest in the early days of RME, Treffers' (1978, 1987) distinction in horizontal and vertical mathematisation is very helpful. Horizontal mathematisation involves going from the world of real-life into the world of mathematics. This means that mathematical tools are used to organise and model, and solve problems situated in a real-life situations. Vertical mathematisation means moving within the world of mathematics. It refers to the process of reorganisation within the mathematical system resulting in shortcuts by making use of connections between concepts and strategies.

Treffers' (1987) scheme included in Table 1 shows how the four different approaches to mathematics education diverge.

Table 1
Types of Mathematisation in Mathematics Education (from Treffers, 1987, p. 251)

Approach to mathematics education	Mathematisation	
	Horizontal	Vertical
Realistic	+	+
Structuralistic	–	+
Empiricist	+	–
Mechanistic	–	–

Connected to this 'two-way mathematisation', RME can also be explained by a number of principles (Van den Heuvel-Panhuizen, 2001)[1]:

- The *activity principle* refers to the interpretation of mathematics as a human activity (Freudenthal, 1971, 1973). In RME, students are treated as active participants in the learning process. Transferring ready-made mathematics directly to students is an 'anti-didactic inversion' (Freudenthal, 1973) which does not work.

[1] This list of principles is an adapted version of the five tenets of the RME instruction theory distinguished by Treffers (1987): "phenomenological exploration by means of contexts", "bridging by vertical instruments", "pupils' own constructions and productions", "interactive instruction" and "intertwining of learning strands".

- The *reality principle* emphasises that RME is aimed at having students be capable of applying mathematics. However, this application of mathematical knowledge is not only considered as something that is situated at the end of a learning process, but also at the beginning. Rather than commencing with certain abstractions or definitions to be applied later, one must start with rich contexts that require mathematical organisation or, in other words, contexts that can be mathematised (Freudenthal, 1979, 1968).
- The *level principle* underlines that learning mathematics means that students pass various levels of understanding: from the ability to invent informal context-related solutions, to the creation of various levels of shortcuts and schematisations, to the acquisition of insight into how concepts and strategies are related. Models serve as an important device for bridging the gap between informal, context-related mathematics and the more formal mathematics. In order to fulfil this bridging function, models have to shift from a 'model of' a particular situation to a 'model for' all kinds of other, but equivalent, situations (Streefland, 1985, 1993, 1996; see also Van den Heuvel-Panhuizen, 2003). For teaching calculations the level principle is reflected in the didactical method of progressive schematisation (Treffers, 1982a, 1982b). I will return to this point later.
- The *intertwinement principle* means that mathematical domains such as number, geometry, measurement, and data handling are not considered as isolated curriculum chapters but as heavily integrated. Students are offered rich problems in which they can use various mathematical tools and knowledge. This principle also applies to topics within domains. For example, within the domain of number this means that number sense, mental arithmetic, estimation and algorithms are taught in close connection to each other.
- The *interactivity principle* of RME signifies that the learning of mathematics is not only a personal activity but also a social activity. Therefore, RME is in favour of 'whole-class teaching'. Education should offer students opportunities to share their strategies and inventions with other students. In this way they can get ideas for improving their strategies. Moreover, reflection is evoked, which enables them to reach a higher level of understanding.
- The *guidance principle* means that students are provided with a 'guided' opportunity to 're-invent' mathematics (Freudenthal, 1991). This implies that, in RME, teachers should have a pro-active role in students' learning and that educational programs should contain scenarios which have the potential to work as a lever to reach shifts in students' understanding. To realise this, the teaching and the programmes should be based on a coherent long-term teaching-learning trajectory.

Although it certainly is not the case that nowadays every class is taught according to the principles of RME or that every textbook which advertises itself as RME is designed according to the RME principles, it is clearly true that since the beginning of the development of RME, the nature of textbook series has changed dramatically. In the 1980s, the market share of mechanistic textbooks was 95% and that of RME ones 5%. In 1987 the market share of RME textbooks was around 15%. In 1992 this had increased to almost 40%, and 75% in 1997. In 2004 RME textbooks were in use in 100% of cases (sources: Treffers, 1991a; Janssen et al., 2005; Janssen et al., 1999).

In the following I will illustrate this development by zooming in on long division.

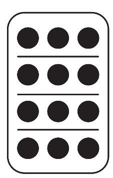
Reform Developments in the 1980s

Two Examples from the Classroom

In 1983 Fred Goffree had the idea to study how a mathematics lesson looks on an ordinary day in November at an ordinary primary school in the Netherlands. The day in question was Tuesday, November 15, 1983. Goffree asked me to join him in this study, which led to the publication 'Zo rekent Nederland' [This is how the Netherlands calculate] (Van den Heuvel-Panhuizen & Goffree, 1986).

Every teacher could be part of the study and was free to describe his or her own mathematics teaching that day. No one checked whether what had been written down did really take place. At first sight, from a sense of reliability, this does not appear to be a scientific approach. Of course, teachers could present their teaching in a better light than what really took place in their classroom, but what they told about that mathematics lesson, does in any case reflect their opinion about it. And this didactical baggage was in fact what we were looking for in this study, rather than what happened in practice. It is more than likely that the teachers' opinions about mathematics teaching and their didactical knowledge determined to a high degree the lesson they had decided on for this special day.

We received 161 lesson descriptions, containing a plethora of subjects, including long division. As it happens that day, one Grade 3 class was taught how to do long division (see Van den Heuvel-Panhuizen & Goffree, 1986, 66-67). The starting point was: dividing twelve into fours. As it was explained by the mechanistic textbook series *Naar Aanleg en Tempo* (Student book 6, Task 32, without year, published by Thieme-Zutphen) you can write this down as 'division' or as a 'division in the form' (see Figure 1).



	division	division in the form
	12 : 4 = 3	4 / 12 \ 3 12 0

Figure 1. Explanation of long division in mechanistic textbook

The teacher gave the following explanation:

> Imagine that these two slanting bars are tails. 4 out of 12 goes 3 times. You write down the 3. Then you take the final 3 x 4 is 12. Subtracted 0. Remember: subtract.

Then the teacher dictated a few problems to practice: 20÷5 and 30÷5. These were treated on the blackboard afterwards.

> Who had them right? Do you see how it works? A few more problems like that. The last 0 has to be under the right number. Do you remember how we call this division? Long division. In the book we call that 'in the form'.

And indeed, in this mechanistic approach to learning long division the lay out was all that mattered, learning outward details such as 'what is it called', 'how does it look', 'what you must do', and 'what you must say'. The teacher writes down a division problem and the students must turn it into long division. Characteristic of the mechanistic method is that it starts with small numbers and that the larger ones follow along. This structure is called ‘progressive complication’ (Treffers, 1982a, 1982b). Another characteristic of this method is that a start is made with calculations immediately.

Continuing with the classroom vignette from 1983, we see that nearly all students were able to do the long division problems they were given. The teacher should be pleased. Or maybe not?

The problem with using for instance 12÷4=3 to learn long division is that it is not very long, which is why the teacher was reduced to referring to the ‘tails’. The question is whether this will be any help to students when they are stuck: Long division? Oh, yeah, with the two tails. And what goes between them?

Happily there was also a lesson description (see Van den Heuvel-Panhuizen & Goffree, 1986, p. 68-69) that showed students learning long division with more understanding. In a Grade 5 class the students were given a row of long division problems that they had to solve in two ways: the whole-number based method that was used as an introduction in Grade 4 and the ‘regular’ digit-based method (see Figure 2).

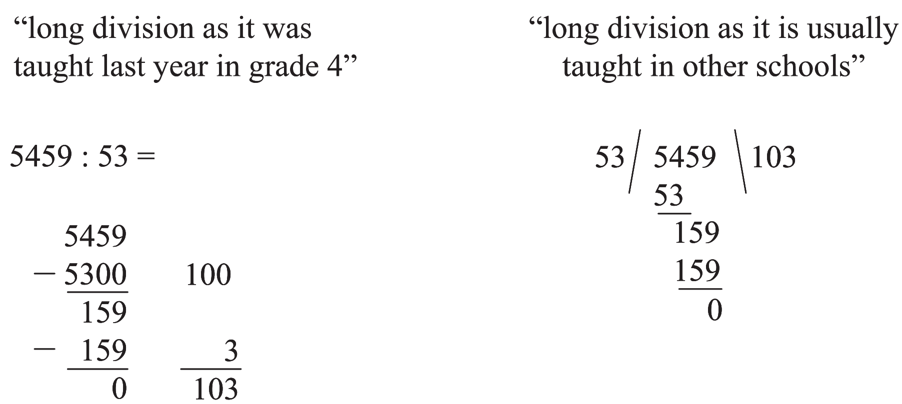


Figure 2. Combining whole-number-based division and digit-based division

In this school, which used a programme inspired by Wiskobas, the students had not been taught the shortest digit-based long division algorithm in Grade 4 immediately, but they started with a whole-number-based procedure of repeated subtraction. This clear and to the students natural approach of division, which starts with relatively large numbers immediately, gradually shortens the procedure until one arrives at the familiar standard algorithm; this is ‘progressive schematisation’ rather than ‘progressive complication’ (Treffers, 1982a; 1982b).

Although the whole-number-based long division can be called an icon of RME, it certainly was not invented within RME. Before this style of long division was included in the RME textbook series, it could already be found for instance in Dutch textbook series of the early 1960s that, like RME, had a broad approach to calculation, and did not limit themselves to algorithmic digit-based calculation. Outside the Netherlands, the history of this whole-number-based calculation and notation method goes back even further. At the beginning of the twentieth century this method of calculation with whole numbers rather than digits could already be found with the German mathematics didactician Kühnel (1925). In the NCTM Yearbook on developing computational skills, Hazekamp (1978) even refers to an example of this approach in a mathematics book from 1729.

Nowadays as well, whole-number-based written calculation, which is so under attack in the Netherlands, is not a procedure that is typical for RME. This stepping stone towards shortened long division is used in many places worldwide, such as in England (see http://nationalstrategies.standards.dcsf.gov.uk/node/19829; Thompson, 1999, 2008; Anghileri, 2001), the United States (Kilpatrick, Swafford, & Findell, 2001, p. 211-212) and Hong Kong (Leung, Wong, & Pang, 2006).

Unfortunately not everybody showed as much didactical insight in 1983 (see Van den Heuvel-Panhuizen & Goffree, 1986) as the teacher in Grade 5.

> A director of a lower vocational school mentioned that his daughter, who was at a teacher education college, had taught him the new method of long division. To the question of whether this method was used in his school as well, he answered: "No, because you cannot use tricks with it. Students are immediately unmasked with this method" (p. 69).

This seems upside-down: do not try to let students understand long division, but limit yourself to teaching the outward form. The problem with this kind of blind calculation procedure is that its success is limited. Despite the large amount of teaching time spent on calculation — often more than fifty hours of teaching time for long division (Treffers, De Moor, & Feijs, 1989; Wijnstra, 1988) — results were poor. One in three children had problems with long division, and over half stumbled on harder division problems with zeros in the result (Treffers & De Jong, 1984). American studies at the time also showed the huge problems students had with long division. For instance, Bright (1978) showed that in the National Longitudinal Study of Mathematical Abilities (NLSMA), that was performed by the School Mathematics Study Club at the end of the sixties, only 44% of ten-year old students gave a correct answer to a problem like 9792÷32 (result 306) and that the percentage for 482÷24 (result 20 rest 2) of correct answers was 61. In addition to difficulties with making bare number problems, students also had trouble with the aspect of applicability. This emerged from, for example, English studies, with Brown (1981) finding that students did not know which operations should be used in context problems.

Inception of a National Reform Plan

Although Wiskobas had already been active in the 1970s, publishing curriculum documents and background studies with examples of reformed mathematics education, at the end of the seventies and the start of the eighties mathematics education differed widely in both content and didactic approach. For that reason the Dutch Society for the Development of Mathematics Education (NVORWO) decided to commission the start of a national plan for mathematics education in primary school. Its goal was

> "to achieve a certain homogenisation on content in mathematics education, and to create favourable conditions for education, training, support, development and research, and the relationship between them" (Treffers & De Moor, 1984, p. 5).

A concept version of this plan was published in 1984 as '10 voor de basisvorming rekenen/wiskunde' (Treffers & De Moor, 1984) and presented to a large group of experts.

For algorithmic digit-based calculation it was proposed to:

- have it in a less central position in favour of mental calculation, estimation and number sense
- aim more at applicability, and
- not immediately teach students the most shortened forms of standard algorithms (working with digits), but to start with a notation using whole numbers; for long division this meant: starting with repeated subtraction.

Of the almost 300 respondents (among them around 70 teacher education teachers, 70 teacher counsellors and 70 primary school teachers) who were consulted about this concept plan for algorithmic calculation, 95% agreed with this proposal (Cadot & Vroegindeweij, 1986). Although it was also clear from the commentary that not all respondents were equally sanguine about the time gain that would result from this new approach, and concerns about implementation were expressed, in general there was an almost unanimous agreement with the reform of mathematics education as proposed in the concept plan. This was the case not just for this group of consulted experts. Another study (Ahlers, 1987) also showed that there was a desire to put algorithmic digit-based calculation on a new footing. Of teachers in grade 6 only 32% felt that algorithmic calculation was 'very important'. This percentage was slightly higher for parents, 43% chose this qualification for algorithmic calculation, while 56% of parents rated mental calculation as 'very important' and 62% indicated that they felt mathematics applied in daily life was 'very important'.

So, at the end of the 1980s the Netherlands was ready for a new direction for algorithmic digit-based calculation. It should be said immediately though that this was not the direction that had been decided upon in England after the publication of the Cockcroft Report in 1982 (DES, 1982). Although there was the intention to spend more time on solving problems with examples from daily life, unlike England the Netherlands was explicitly not going so far as to for example abolish long division (see Treffers & De Moor, 1990). All that NVORWO wanted was to get rid of the one-sided focus on algorithmic digit-based calculation, and at the same time it chose to have an insightful introduction to the shortest version of the algorithms. Alongside, a greater role was assigned to mental calculation, estimation and number sense. There were consequences to this new approach.

Consequences for Mathematics Achievements

The studies of the National Assessment of Educational Achievement (PPON) used to assess mathematics achievements of primary students in the Netherlands once every several years, clearly reflected the results of the new approach. As is shown in Figure 3, comparing the results in the years 1987, 1992, 1997 and 2004 of students in grade 6 (end primary school) (Janssen et al., 2005) revealed that the performance in the area of number sense and estimation have improved greatly. In comparison to the first assessment in 1987 these two topics show an increase of about 25 percentage points. In addition, mental addition and subtraction have also improved by about 10 percentage points. Aside from calculating with percentages, which has also improved by about 10 percentage points, achievements on 'relations, fractions and percentages' and 'measurement and geometry' (not included in Figure 3) have hardly changed between 1987 and 2004. However, Figure 3 also shows that achievements in written calculation have gone down substantially in the period 1987-2004. This is especially the case for multiplication and division. These scores have gone down by about 1.25 standard deviation. This means that the percentage of correct answers has gone down by about 30 percentage points. For composed written calculation the total reduction is 20 percentage points and for written calculation addition and subtraction about 15 percentage points.

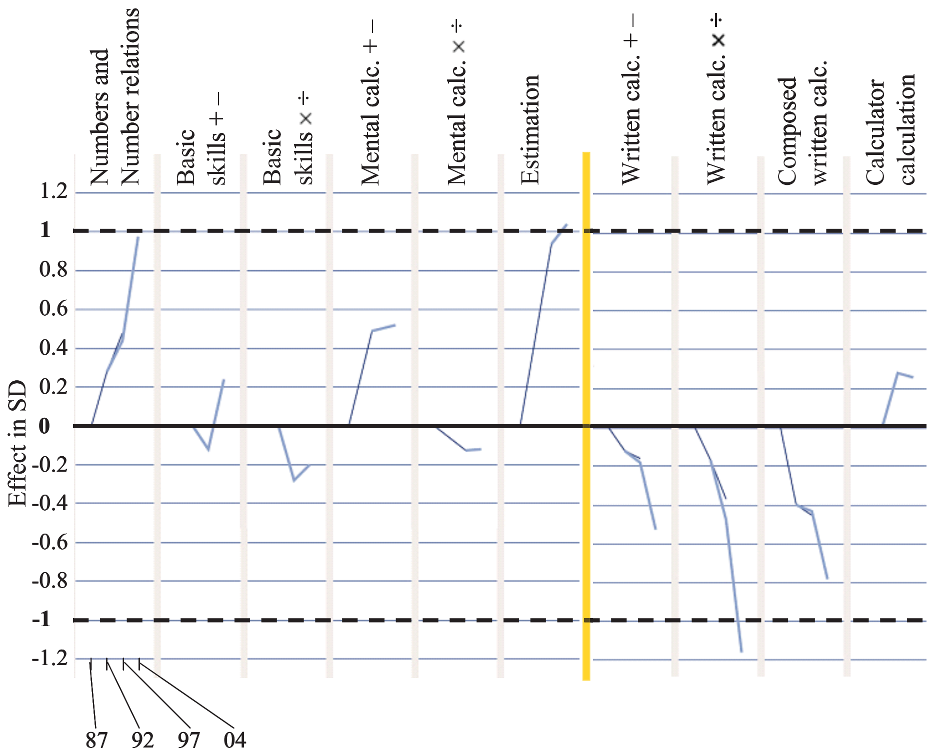


Figure 3. Effect sizes in changes in achievement in the domain of number at end primary school from 1987 to 2004 (based on Janssen et al., 2005)

So, according to PPON, written calculation has clearly become less over time. While this is the case, we might also say that what we see of the achievements in the domain of number in the Netherlands in 2004, does to some degree match the performance profile opted for twenty years ago. The reform that was proposed at the time has indeed taken place, and from within education without government intervention, something that is at the very least remarkable. Just as remarkable is however that the change in students' performance, especially in the public debate, has been taken as deterioration in mathematics achievements. Apparently, written calculation is identified more with mathematics than number sense, estimation, mental calculation and applications such as calculation with percentages.

A Critical Evaluation of the Assessment of Written Division

Although the lower achievement in written calculation can be taken partly as a result of the broadly-supported decision to spend less time on this topic, the difference in achievement is in fact higher than expected and intended. Before we place the blame for this difference on RME, a critical analysis of how these achievement scores have been established is called for.

Three points can be identified that give reason to question the assessment of written division (see Van den Heuvel-Panhuizen, Robitzsch, Köller & Treffers, 2009): (a) the problems used, (b) the time of measurement, and (c) the test instruction that was given.

The Problems Used

A total of nineteen items was used for assessing the topic 'Operations: multiplication and division' in 1997 and 2004. Of these items, only four were included in both assessments. These items were used to link the two measuring points. Unfortunately, three of these anchor items, two of which are shown in Figure 4, are more suited — especially with the improved number sense in 2004 — to mental than to written calculation.

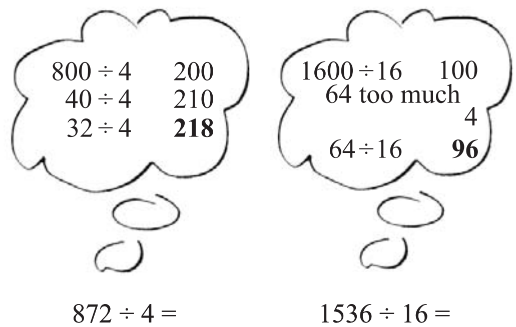


Figure 4. Two of the four anchor items for written division

Of the nineteen items, sixteen were context problem, with four focusing on the ability to interpret the remainder, which is not the same as being able to perform the division procedure.

Moreover, only one of the hardest items with whole numbers, those with a zero in the result (for example 64800÷16; see Figure 5), was included in the nineteen test items, although precisely for this type of problem the RME approach, applying a whole-number-based division, is less sensitive to errors than the traditional algorithm. In the latter approach you can take 16 out of 8 zero times, and then you must remember to put down a zero in the result. The same happens again at the end of the division procedure.

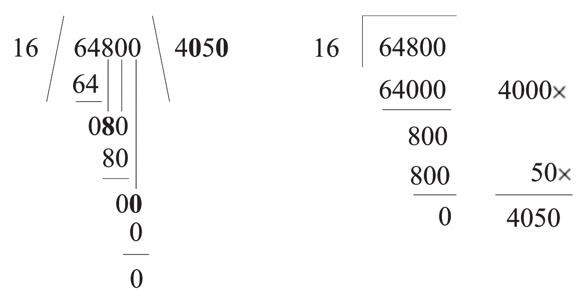


Figure 5. A division problem with a zero in the result

The Time at Which the Assessment Took Place

Because, in 2004, Cito did a reference study for the Cito Student Monitoring System, we do not only know how the students scored on the test items at the end of Grade 6 as collected in the PPON study, but we also know their scores halfway this grade (Janssen et al., 2005). It turns out that the decrease that has been found between the end of Grade 6 in 1997 and the end of grade 6 in 2004 is complicated. The downward movement in achievement for written calculation turns out to have occurred mostly in the second half of

Grade 6 (see Figure 6). There seems to be a turning point mid grade 6, with scores decreasing after the students have done the Cito End of Primary Test which is administered in the middle of grade 6.

Regrettably there are no data on the middle of grade 6 in 1997, but we do know the scores of the middle and the end of Grade 5 from the Cito Student Monitoring System reference study in 2004. These show an increase for the second half of Grade 5.

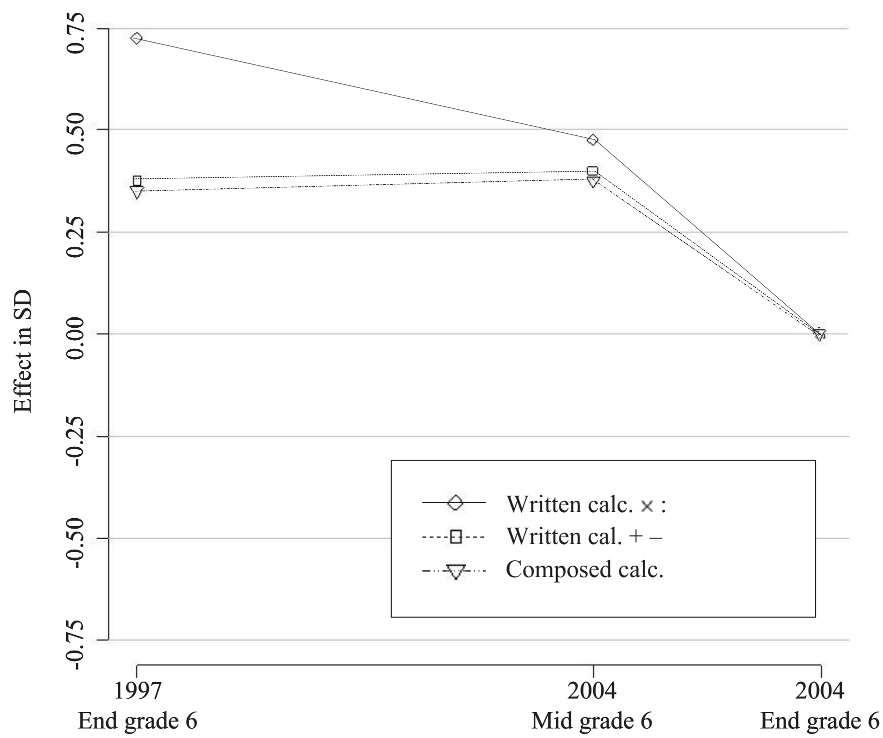


Figure 6. Achievements in written calculation end grade 6 compared to mid grade 6

Of course the same pattern with the notable decrease in Grade 6 may also have occurred in 1997. However, it is also possible that the trend to spend less time on mathematics instruction after administering the Cito End of Primary Test has become stronger over time. Add to that the fact that after primary school, students mostly switch to using a calculator for doing calculations and it is hardly surprising that written calculation skills in secondary school are not good.

The Test Instructions

The third point allowing questions to be raised about the assessment of written calculation and the conclusions based on it, concerns the given test instructions.

To study the use of strategy by students on the tested items, 140 students who took part in the written 2004 PPON test were also interviewed individually (see Van Putten, 2005; Hickendorff, Heiser, Van Putten, & Verhelst, 2009). Not only did this additional research allow study of the effect of the RME and traditional solution strategies , it also revealed, as in shown in Table 2, that there was a significant difference in correct scores between both testing formats. In 2004, the correct scores for the released items 736÷32 en 7849÷12 were about 30 percentage points higher for individual testing than in the class-administered written test. Compared to items from other topics that were tested in two ways, this is a large difference.

Another remarkable difference in 2004 is, that for the individual test format there was not a single student without written notations of the calculations. For the class-administered written test this was for the two items, respectively 30% and 35%.

Table 2
Percentage Strategy Use and Answers Correct in Two Test Formats

Item 9 736÷32 (in context)	1997	2004	
Strategy	Paper-and-pencil	Paper-and-pencil	Individual interview
Traditional Algorithm	42	19	29
Realistic[a]	24	33	71
No Written Working	22	30	0
Other	12	19	3
Answer correct	71	52	84
Item 10 7849÷12 (in context)	1997	2004	
Strategy	Paper-and-pencil	Paper-and-pencil	Individual interview
Traditional Algorithm	41	19	27
Realistic[a]	22	25	68
No Written Working	17	35	0
Other	20	21	5
Answer correct	44	29	60

[a] As defined by Hickendorff et al. (2009) including chunking and partitioning

Apparently, the instruction for the written test was not strong enough in comparison to the individual one to let the students do their calculations on paper in all cases. However, to establish whether Dutch students are capable of written calculation, and to make statements about changes over time in this ability, one should explicitly ask students to do their calculations on paper. Testing whether students can solve mathematical problems is something different than testing whether students can perform certain calculation procedures.

The three questionable points mentioned here (the items used, the time of assessment and the given test instructions) show that research into changes in achievement is not easy, certainly not if there is educational reform taking place at the same time. A problem like 736÷32 that would be tackled by using a division procedure on paper in 1997, is more likely to evoke mental calculation in 2004 — as a result of the greater emphasis on number sense: 736÷32, ... 20 already gives you 640, add 2 times 32, 64, that gets you to 704 and add one more 32, that gives you; together 23 times 32. However, if this mental calculation strategy turns out to be an overestimation of a student's skills, things will go wrong in the class-administered written test, while the individual test shows that 84% of students will get it right if they use written calculation.

What TIMSS Says About the Mathematics Achievements of Dutch students

The results of the Trends in International Mathematics and Science Study (TIMSS) 2007 (Mullis et al., 2008) provide the most recent facts on the mathematics achievements of Dutch primary school students. Because Dutch Grade 4 students participated in TIMSS in 1995, as well as in 2003 and 2007, the TIMSS scores show the development of mathematics achievements in recent years within an international context, like the PPON scores did on a national level. The most important conclusion of TIMSS is that there was a significant decrease in the whole period between 1995 and 2007, but that there was no significant change at the latter end, between 2003 and 2007. A problem with the first part of the conclusion is that in 1995 the Netherlands did not comply with the requirements for the sampling procedure. However, the TIMSS researchers did not see this as an obstacle for including the 1995 data in their trend analyses.

Despite the supposed decrease, Dutch fourth graders scored reasonably well for TIMSS in 2007. The Netherlands are in eighth place for the domain of number — which is what the current discussion is about — below four Asiatic participants (Singapore, Hong Kong, Taiwan, and Japan) and three participants from Eastern Europe (Kazakhstan, Russian Federation, and Latvia) (Mullis et al., 2008). The Netherlands have the highest score of the nine Western European countries that took part. Additionally, the Netherlands finished above the other Western countries that took part, including the United States, Australia and New Zealand. It would be justified to call the Netherlands 'Best of the West'. The media, however, see things differently. For example, a popular magazine for parents *j/m Voor Ouders* 'translated' the Dutch results as: "Internationally, our children have poor results in arithmetic."

The discrepancy between the achieved results and their perception becomes even larger if the opportunity to learn is taken into account. For example, in Singapore the test items used for the domain of number are covered for 91% by the taught curriculum. In the Netherlands, that is the case for only 64% of the items. Compared to the countries with a higher score, we have the lowest coverage percentage (although the coverage percentage for two countries is unknown) (Mullis et al., 2008). For instance, calculations with fractions and decimal numbers are not taught in Grade 4 in the Netherlands, while these topics were included in the TIMSS test.

Another point is the low spread in the mathematics scores of the Dutch students. The best Grade 4 students in Singapore may be better than our best Grade 4 students, but our weakest students are at around the same level as the weakest students in Singapore (Mullis et al., 2008). This low spread was not unique to TIMSS 2007, but emerged from earlier TIMSS and PISA studies as well. Even though this is a remarkable result that says something essential about our education system, it is given relatively little attention in the media and reports. Clearly, we do not want to establish ourselves as 'equal opportunities champion' (see Van Streun, 2009), while this is in a sense what we are, and while this also matches our thoughts of the mission we have with mathematics education.

Other data from TIMSS 2007 received equally little attention:

- diverging from most other countries, girls in the Netherlands do not do as well as boys
- the Netherlands are at the bottom of the league table for teacher participation in professional development
- the Netherlands has the highest percentage of time spent on working on problems individually without a teacher's guidance.

The type of classroom organisation that emerges from this last point, certainly does not match the RME model of interactive whole-class teaching. The newspapers and opponents of

RME do not mention this, but it is a point of concern to us. The lack of whole-class teaching may well match the limited ability of Dutch students to write down their calculations systematically, as this is something that is hard to teach a whole class through individual instruction. It is not clear why Dutch teachers make so little use of whole-class instruction. It may be an effect of the Inspection of Education 'teaching-to-size' policy that they pursued since the 1990s (see, e.g., Inspectie van het Onderwijs, 1998). What is remarkable here is that England which had the highest increase of all countries in 2007, had decided in favour of 'whole-class teaching' in the 1990s — within the framework of the National Numeracy Strategy (see Askew, 2002), which was partly inspired by the ideas of RME.

What is Better, Mechanistic or Realistic Mathematics Education?

RME Under Attack

Since 2007 a deluge of reports has flooded the Netherlands, all of them emphasising a downward trend in our mathematics achievements and observing that we do badly in comparison with other countries. There were only two sources for the findings in these reports in primary education: the PPON and TIMSS studies. While neither study is above criticism, as I have shown in the previous section, the reports that reference these studies do not show the necessary critical attitude. Even worse is the cumulative effect in reporting bad results. PPON and TIMSS show that achievements decrease. This result is then included in another report, with the effect that subsequent reports will then refer back to three, rather than two, sources showing that mathematics achievements of Dutch primary school students are falling behind. The next report mentions four sources, and so on.

A recurring element in the media is that the opponents of RME do know why achievements decreased so much. It is the fault of RME. Therefore they argue that RME should be dropped and we should return to the mechanistic teaching of before the reform; this would mean going back about forty years. In fact two new mechanistic textbook series are currently in production, while some RME textbooks series fear loss of market share and no longer call themselves RME or state explicitly that they have a balanced approach and combine RME characteristics and mechanistic characteristics.

An Arbitrator to Decide the Argument

To stop the debate, the Ministry of Education asked the highest academic body in the Netherlands, the Royal Netherlands Academy of Arts and Sciences (KNAW) to find out which approach to teaching mathematics is better: the RME approach or the traditional mechanistic manner of teaching. This latter approach includes those methods in which students are taught one standard algorithm per operation, teachers provide direct instruction and students learn by solving bare number problems.

The KNAW Commission was a mixture of both proponents and opponents of RME. To find an answer to the question of which teaching method is the best, the Commission did not carry out a study by itself, but instead looked at the empirical research conducted in the Netherlands in the past twenty years. In addition, a brief and general survey was done of studies conducted abroad. The conclusion of the Commission (KNAW, 2009) was that the

> empirical material is not unequivocal and does not permit any general, scientifically-grounded statements about the relationship between mathematics instructional approaches and mathematical proficiency. The research is limited and does not provide convincing empirical evidence for the claims made by either side of the debate about the effectiveness of traditional methods versus RME. (p. 14)

This 'not decided' conclusion by the KNAW Commission fits well into the Dutch policy that is known as the 'polder model'[2], which refers to the consensus policy in economics based on the tri-partite cooperation between employers' organisations, labour unions, and the government, aimed at defusing labour conflicts and avoiding strikes. It is a good result for the Ministry of Education. The worst is over. Both tabloids and serious newspapers are quiet again. But, really, this conclusion is not satisfactory. In fact our reform is back where we started. This is how a reform can end. Forty years of work for nothing. RME, based on the work of many researchers, developers, mathematics educators and teachers, is being compared with the opinion of a small group of opponents who have no development and research work to support them, and who only have some slogans going back to the past and at best a lean behaviouristic basis. It is difficult to name their approach a didactic of mathematics education. Asking which didactic is better, is to compare two unequal quantities with the result that the opponents of RME, despite of their low behaviour in the media, have been promoted to respected researchers of mathematics education.

Insufficient Evidence?

Measured by the current hype of *evidence-based educational* policy and decision making there is insufficient evidence for both the mechanistic approach and RME. On the one hand, there is no evidence available in the Netherlands at all for the first approach, simply because there has been no research, except for a couple, in some ways flawed studies into the effect on weak students of offering one single, fixed strategy. On the other hand, RME does have a long history of research — and in addition is supported by the huge body of knowledge about reformed approaches to mathematics education gathered by the international research community — but this research has not yet delivered the level of evidence that is required nowadays. The development and implementation of RME took place at a time when the emphasis was not yet on experiments with pretest-posttest designs with randomised control and experimental groups. RME was more interested in design experiments. First we had to find out what the reformed education should look like, how we could evoke certain learning processes in children and how we could raise the children to a higher level of understanding. The ideas for didactical approaches that emerged from this were often convincing enough in themselves. Everybody could test their didactical value every day in their own educational practice. These experiences were enough to implement RME in mathematics classrooms, teacher education and in-service courses, educational counselling activities and textbooks.

Examples of convincing didactical innovations of RME. Look for example at whole-number-based written division; a calculation procedure that, by the way, was not even a Dutch invention (see Hazekamp, 1978) and that was already in use in the Netherlands before the time of RME (see Van Gelder, 1959). The only research that was done to include this introduction to digit-based algorithmic calculation in RME was the study done by Rengerink (1983; see also Treffers & De Jong, 1984). This was a small-scale study with an experimental class of 21 students, an experimental programme of about 25 teaching hours, and a control class of 23 students following the regular programme.

Other examples of RME innovations that were introduced worldwide without randomised controlled trials are the empty number line and the corresponding stringing strategy (Treffers, 1991b; Van den Heuvel-Panhuizen, 2008a), the arithmetic rack with the two lines of beads in a 5-5 structure (Treffers, 1991b), the ratio table (Treffers, 1993;

---

[2] Although it must be said that some people wonder whether this is actually typically Dutch (De Bruijn, 2010).

Middleton & Van den Heuvel-Panhuizen, 1995), and the percentage bar (Van den Heuvel-Panhuizen, 2003).

Convincing large-scale and standardised research. If the success of these didactical innovations and the research that contributed to the development of them did not count in the eyes of the members of the KNAW Commission, and the worldwide support and appreciation that they received were also not important, then the KNAW Commission should at the least have taken a closer look at the few studies that were large-scale and standardised. These studies did in fact show that the RME textbook series that concretised RME — even if not all RME textbook series did so perfectly — did lead to higher achievement scores in mathematics in comparison with traditional, mechanistic textbook series.

For instance, the PPON analyses by Cito over the period 1992-2004 showed that, despite the strong decrease in written calculation, the newer mathematics textbooks as a whole contributed in a small, but positive way to the mathematics achievements of the students (Janssen et al., 2005). This argues in favour of the RME approach.

In addition, Cito did a comparison at textbook level using the achievement scores collected in 1987, 1992, and 1997 (Janssen et al., 1999). This was in fact the last time that it was possible to make a large-scale empirical comparison between RME and the mechanistic approach, since a part of the textbook market was still controlled by mechanistic textbooks. In 1997, RME textbooks had a market share of 75% with mechanistic textbooks at around 10%. The results of the comparison indicated that the RME textbooks were more often part of the best textbook series (with students in grade 6 obtaining the highest achievements in mathematics) and that the mechanistic methods were more often among the weakest textbook series (with students in Grade 6 obtaining the lowest achievements in mathematics). One illustrative point was that the textbook series *Wereld in Getallen (WIG)* — the oldest RME textbook still in use, and additionally seen as a good representative of RME — achieved the top place on 19 of the 24 mathematics topics that students were tested on, while the mechanistic series *Naar Zelfstandig Rekenen (NZR)* never achieved a top place in the category best textbook series. This method had the highest score in the category weakest textbook series on 13 topics.

Based on the data published by Cito (Janssen et al., 1999) another comparison can be made that gives an even better answer to the question of what way of teaching mathematics is better: the RME approach or the mechanistic manner of teaching. For this we can take the two RME textbooks that were included in the study then and which are still in use, virtually unchanged. These textbooks, which to a large degree determine the quality of current mathematics education with a combined market share of 70%, are the textbooks *Wereld in Getallen (WIG)* and *Pluspunt (PP)*. If we consider the achievement scores of grade 6 students with these two RME textbooks obtained for PPON in 1987, 1992, and 1997 against the scores of students who used the mechanistic textbook *Naar Zelfstandig Rekenen (NZR)*, then it is clear to see that the RME textbooks led to better results than the mechanistic textbook. Figure 7 shows that the RME textbooks *WIG* and *PP* outperform the mechanistic textbook *NZR* in nearly all topics within the domain of number.

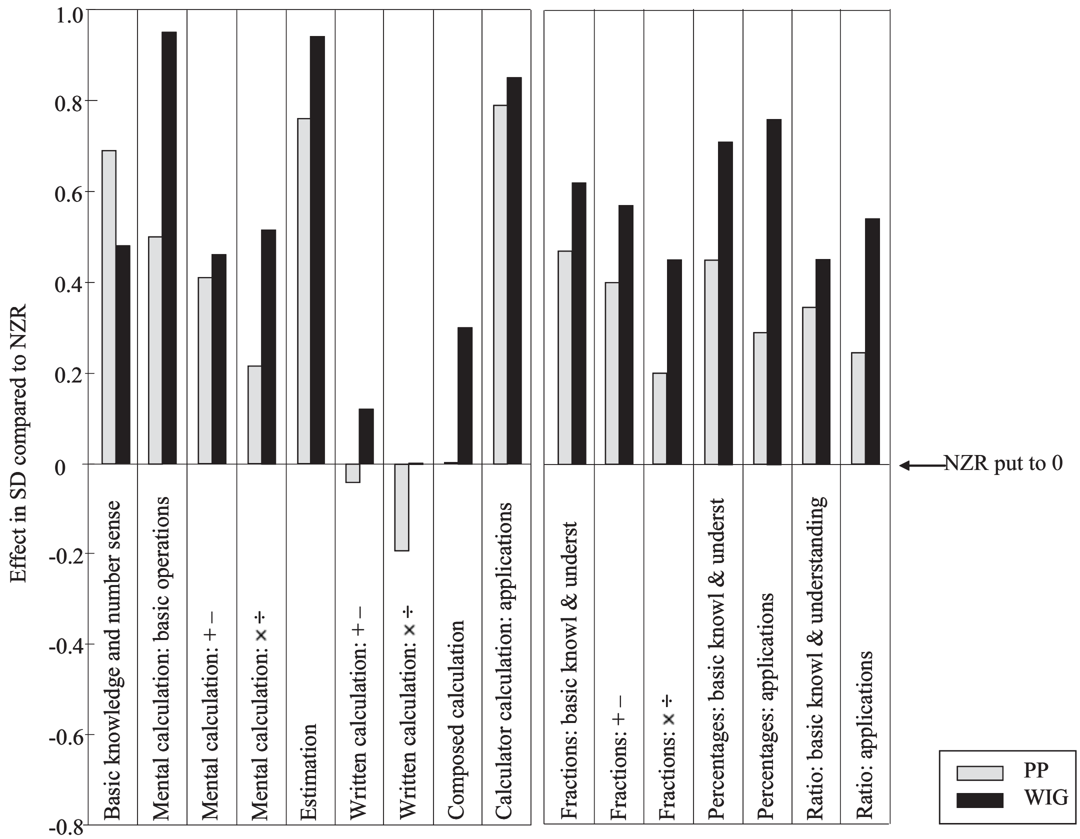


Figure 7. Effect sizes for the *PP* and *WIG* scores in comparison to the *NZR* scores

On basic knowledge and number sense, estimation and insightful use of the calculator *PP* scored 10 to 15 percentage points higher than *NZR*. On the other number topics the scores of PP are 5 to 10 percentage points higher. Only for written calculation is the score for *PP* slightly lower than for *NZR*. Two of the three topics for written calculation are at around the same level for *WIG* and *NZR*, though not the hardest topic, composed calculation. Here, *WIG* has a significantly positive textbook effect against *NZR* — and to consider that *NZR* strongly emphasises written calculation.

It is, however, not just written calculation where WIG does relatively well. The quality dominance of *WIG* against *NZR* shows especially in the fact that *WIG* also has significantly better scores than *NZR* on the other 16 number topics. The differences lie on the whole between 10 and 15 percentage points. *WIG* does especially well for basic knowledge and number sense, estimation, insightful use of the calculator and calculations with percentages.

Unfortunately, the KNAW Commission did not look at these PPON data in detail, but lumped together all RME textbooks and all traditional textbooks. This is especially detrimental for the traditional textbooks, of which there are two types: textbooks which only focus on digit-based algorithms and textbooks which have a broad interpretation of calculation. The first type does not include whole-number-based calculation (the insightful introduction to the algorithms), or (smart) mental calculation and estimation, while the textbooks of the second type do take these into account, and as a result are consistent with the RME textbooks with respect to the number domain. The students who worked with the traditional textbooks with a broad interpretation of calculation outperformed the students who were taught with the traditional textbooks that only focused on the algorithms (Treffers & Van den Heuvel-Panhuizen, 2010).

One can only guess at why the KNAW Commission did not take an in-depth look at the PPON data. At the same time it also remains unclear why this commission did not refer to the criticism of the assessment of written calculation that was mentioned earlier.

Fear of Monoculture

The most important question is still why the KNAW lent their name to comparing a serious reform movement which has a large national and international reputation with the opinion of a small group of people who think that we should return to the education of forty years ago. One may also question how the KNAW rates our discipline, the didactics of mathematics education. The distinguished university professor of mathematical physics who is the current chairman of the KNAW wonders in the foreword to the report (KNAW, 2009) "what [...] a scholarly gathering such as the Royal Netherlands Academy of Arts and Sciences has to do with the didactics of primary mathematics" (p. 5).

Something else that emerges from the foreword of the KNAW report (2009), is that in the eyes of the KNAW the battle between RME and mechanistic education could only end in a draw. Even the science of biology was quoted to justify the result, by warning that monocultures (read RME) are an impoverishment of an ecosystem.[3] It may be clear that the natural diversity within applications of RME in textbooks and classroom practice, the width of mathematising both horizontally and vertically and the interconnectedness of conceptual understanding and procedural knowledge that characterise RME, has not been recognised and acknowledged by the KNAW. It is therefore an illusion to think that we will ever be able to convince the opponents of RME.

Lessons to be Learned

Although the attack on educational reform described in this paper takes place in the Netherlands, it seems to me that there are lessons to be learned from it for every other country that engages in reform of mathematics education. Taking a step back to look at the situation in which we — in fact rather unexpectedly[4] — find ourselves from a more remote perspective, I reach the following conclusions.

This Math War is Based on Framing and Emotions Rather Than on Facts

The first conclusion that can be drawn is that, when a reform that has been going on for close to forty years, now suddenly calls up a counter movement that wants to recreate the traditional mechanistic method of teaching as it existed before the reform, the discussion is not about facts but about emotions, and framing techniques (Lakoff, 2004; Kuitenbrouwer, 2010) are used to convince people. One would not expect that anyone who is focused on knowing more about what mathematics children should learn and how this can best be

---

[3] This rejection of a monoculture is also mentioned by the Education Council of the Netherlands (Onderwijsraad, 2010), who argue in favour of two research and development institutes in the Netherlands: one for RME and one for a traditional approach to mathematics education.

[4] In 2003 the Education Council of the Netherlands (Onderwijsraad, 2003) still held up the Freudenthal Institute as an example of how development and dissemination of knowledge about education should take place: "This knowledge community has developed to a high level in three decades, and has many participants and contacts, both nationally and internationally" (p. 49).

taught, will seriously consider throwing all the knowledge that has been gathered in the last forty years on the scrapheap. Rationally, this makes no sense. Just imagine that we would do something similar in healthcare. The mathematics education debate is ruled by emotions, in this case the feeling that everything was better in the old days. It is not for nothing that the attack on current mathematics education is coming mainly from a few mathematics professors. Their educational career is the proof that everything was better in times past. All they forget is to wonder for how many children that was the case.

Reform vs. Restore Debates in Education are Universal

The discussion between traditional and reformed education is probably as old as education itself. Nussbaum (2002) provides the example of the comedy 'Clouds' from 424 BC, where the great ancient Greek comic playwright Aristophanes writes about the dangers of the new education:

> A young man, eager for the new learning, goes to a 'Think-Academy' near his home, run by that strange notorious figure Socrates. A debate is staged for him, contrasting the merits of traditional education with those of the new discipline of Socratic argument. The spokesman for the old education is a tough old soldier. He favors a highly disciplined patriotic regimen, with lots of memorization and not much room for questioning. He loves to recall a time that may never have existed – a time when young people obeyed their parents and wanted nothing more than to die for their country [...]) (p. 289).

The Style of Attack Reflects the Political Climate in the Netherlands

While the style of criticisms against RME in official reports stays within the social conventions that can be expected from professionals, the newspapers, including the so-called 'quality papers' as well as magazines and websites often contain crass and unfounded accusations, in which well-regarded scientists are pilloried. This is all 'normal' within the current political climate in the Netherlands. In recent years, for a variety of reasons — including the emergence of high-profile populist politicians — the tone of the public debate has been lowered; especially when the establishment is the target of the debate. The track record that the Freudenthal Institute has built since 1970 means that we are now part of that establishment, with all the consequences of that position.

Open Reform Movement Makes a Reform Vulnerable

In the same way that RME gives students room to work on their own solutions, there is also room for teachers, teacher educators, teacher counsellors, researchers and developers of mathematics education, and last but not least textbook authors, to include their own nuances in the core ideas of RME. There is no state didactics in the Netherlands, and there is no unified RME. Implementation accompanied with ownership for all involved in the reform is an essential value of RME. We succeeded in achieving that, but at the same time it does make us vulnerable. Any mathematics textbook can be qualified as RME by its authors, but at the same time the authors are free to include problems and instructional sequences which maybe go against RME. Such less successful interpretations of RME are grist to the mill of the opponents of RME. Similar situations can occur when enthusiastic teachers put their lessons online, and make the mistake of for instance explaining whole-number-based written calculation using numbers that are much too large, causing children to lose track. Opponents of RME use such examples to disqualify RME and argue in favour of the traditional, mechanistic approach.

These situations are hard to counter without changing the character of the reform. One way of regulating the implementation is performing textbook analyses. In such analyses, concretisations of RME can be examined and commentary and additions can be provided. This approach would allow constant adjustment. We stopped doing these textbook analyses at the end of the eighties. It might have been better if we had continued doing them.

A Reform Requires Professional Development of Teachers

Without a doubt, the driving forces behind the RME reform have been the textbooks. The textbook authors adopted the RME ideas and models of teaching particular topics and later on made use of the RME-based Proeve books (e.g., Treffers et al., 1989; Treffers, & De Moor, 1990) and TAL learning-teaching trajectories (e.g., Van den Heuvel-Panhuizen, 2008b). However, another driving force was missing in the Netherlands. Unlike many other countries, there is no obligatory in-service training for primary school teachers, nor is there a culture of in-service training (Mullis et al., 2008). This means that for the development of their knowledge of RME teachers had to depend on textbooks and on themselves. This left Dutch teachers with a weak foundation, with as a consequence that, although they have adopted RME, they can easily become uncertain about RME when it is criticised. They miss a profound and updated knowledge base. Research has shown that in-service programs for teachers are an essential factor in a change process (see, e.g., Clarke, 1997); especially teachers' beliefs about the teaching and learning of mathematics are considered as critical in determining the pace of curriculum reform in mathematics education (Handal & Herrington, 2003).

The Importance of the Inclusion of Parents

The parents of the children who are now in primary school were in primary school themselves in the 1980s, when RME was by no means ubiquitous. It is not unlikely that the mathematics education they received was given in a mechanistic way, with the consequence that there may be differences with what their children are learning, not just for content, but also for solution strategies, ways of notating and teaching methods. O'Toole and de Abreu (2005) who studied the gap between the past experiences of parents who went to school in England before the introduction of the National Curriculum and current mathematics education practice in their children's school, showed that this gap should be taken seriously. Many parents use their own past experiences as the main resource for understanding their children's current school learning. In addition, the parents' parents also play an important mediating role. This can be painfully observed in the Netherlands, in the plea to reintroduce granddad's mathematics. Unfortunately, this gap between the parents' and grandparents' mathematics education and that of their children has never been a focal point in RME. After seeing in 1987 that the parents agreed to the reform, there have been no further attempts to keep them involved. This is an omission.

Any Reform Will Have to Prove Itself at Some Point

The long history of RME reform made us almost complacent. As was said in the beginning of this paper, we never thought that the development and implementation of RME were finished or that we had found the ultimate answer to the best way for children to learn mathematics. We were never really satisfied and were always looking for further improvements. But we did so in relative peace. After the first twenty years, we no longer really considered that we still had to prove that RME was better than the mechanistic approach. It seemed so obvious for everybody. In fact, there only was a discussion with remedial educationalists on what mathematics education for special education students should look like. Meanwhile, teachers in special education were hard at work introducing the empty number line and other didactical innovations of RME. In view of these experiences, we considered the confrontation between RME and mechanistic teaching methods as a run race. This was a misconception that we have come to regret. When the wave of criticism started, we did not have our evidence readily at hand, at least not in the form that is currently demanded.

Take Care That the Assessment of Mathematics Achievement is in Order

The most vulnerable part of a reform movement is students' achievement scores. Everything seems to be fine, until scores decrease or are being perceived as less. Reform nearly always means opting for other content or other accents in content, resulting in different achievement profiles for students. The problem here is that no one will argue with the things that improve, but that is not the case for the things that decrease. Even if it has been agreed on that one topic is not so important, once the scores start going down, there is bound to be protest. Justifying the scores in retrospect does not seem to help. Therefore it must be recommended, when making choices, not to discuss only the input, but also, and especially the expected changes in the output. This makes them more 'real'.

The experience that changes in achievement profiles can become a breaking point for a reform, serves to emphasise the importance of how results are assessed. In the end, what matters is not the inspiring learning environment that can be achieved through RME, but the test scores. Therefore it is important for reformers to also look at tests. We did not really do this in the Netherlands, while we are now being judged on the basis of these test results.

RME Continues

I cannot end this paper without reporting a sign of recovery. Our opponents have taken the step of entering the market and found a publisher who dared to take the decision to publish a new traditional mechanistic textbook series, after the last one disappeared by the end of the last century. This publisher has the good practice of letting a panel of teachers judge prototypes of new products. The group had been established from proponents of restoring the mechanistic approach to mathematics education. However, their commentary was significant. No, this kind of textbook series was not what they needed: there should be illustrations and they missed the didactics telling them what to do when children do not understand; after all, you cannot go on infinitely with demonstrating.

The next step will certainly be that RME innovations such as the empty number line, the corresponding stringing strategy, the arithmetic rack with the two lines of beads in a 5-5 structure, the percentage bar, and probably even whole-number-based written division as an introduction to digit-based algorithmic calculation will show up in this mechanistic textbook series. Whether this also will happen with the broad interpretation of calculation, in which written calculation, mental calculation and estimation are integrated, remains to be seen.

What will in any case not appear in this mechanistic textbook series, and others that may follow, are problems that stimulate mathematical reasoning. The attackers of the RME reform state very explicitly that primary school students need to learn calculation skills and do not need to think. In contrast, within RME, getting the children to think mathematically is what it is all about. Unfortunately, this key goal is not reflected very strongly in the existing RME textbooks for primary school. Although these textbooks contain many RME characteristics, puzzle-like number problems in which children have to think are rare (Kolovou, Van den Heuvel-Panhuizen, & Bakker, 2009). In this respect, the RME textbooks truly do need improvement: there should be more mathematics in Dutch arithmetic education. This is a point we have to work on. The narrow focus on plain calculation can really threaten the mathematical competence of Dutch students, but this shortcoming in how RME is implemented in textbooks and maybe also in classrooms is beyond the view of our attackers. Their concerns are only about students' skills in one type of calculation, the algorithms.

References

Ahlers, J. (1987). Grote eensgezindheid over basisonderwijs. Onderzoek onder leraren en ouders. *School, 15*(4), 5-10.

Anghileri, J. (2001). Development of division strategies for Year 5 pupils in ten English schools. *British Educational Research Journal, 27*(1), 85-103.

Askew, M. (2002). The changing primary mathematics classroom – the challenge of the National Numeracy Strategy. In L. Haggarty (Ed.), *Aspects of teaching secondary mathematics: Perspectives on practice*. (pp. 3-17). London: RoutledgeFalmer.

Becker, J. P., & Jacob, B. (2000). The politics of California school mathematics: The anti-reform of 1997-99. *Phi Delta Kappan, 81*(7), 529-537.

Bright, G. W. (1978). Assessing the development of computation skills. In M. N. Suydam, & R. E. Reys (1978). *Developing computational skills* (pp. 148-241). Reston, VA: NCTM.

Brown, M. (1981). Number operations. In K. M. Hart (Ed.), *Children's understanding of mathematics: 11-16* (pp. 23-47). Londen: John Murray.

Cadot, J., & Vroegindeweij, D. (1986). *10 voor de basisvorming, Rekenen/wiskunde onderzocht. Op weg naar een nationaal plan voor het reken-wiskundeonderwijs op de basisschool en het gebruik van de computer daarbinnen*. Utrecht: OC&OW, Rijksuniversiteit Utrecht.

Clarke, D. M. (1997). The changing role of the mathematics teacher. *Journal for Research in Mathematics Education, 28*(3), 278-308.

De Bruijn, H. (2010). Polderen moet en de markt is er gewoon. *De Academische Boekengids, 80*, 3-5.

De Jong, R., Treffers, A., & Wijdeveld, E. (Eds.) (1975). *Overzicht van Wiskundeonderwijs op de Basisschool. Leerplanpublikatie 2*. Utrecht: IOWO, Rijksuniversiteit Utrecht.

De Moor, E. (1980). Gevarieerd rekenen. Leerplanpublikatie 11. *Wiskobas-Bulletin, 9*(1/2/3). Utrecht: IOWO, Rijksuniversiteit Utrecht.

Department of Education and Science (DES) (1982). *Mathematics Counts*, London: HMSO.

Freudenthal, H. (1968). Why to teach mathematics so as to be useful. *Educational Studies in Mathematics, 1*, 3-8.

Freudenthal, H. (1971). Geometry between the devil and the deep sea. *Educational Studies in Mathematics, 3*, 413-435.

Freudenthal, H. (1973). Mathematics as an educational task. Dordrecht: Reidel Publishing Company.

Freudenthal, H. (1979). Structuur der wiskunde en wiskundige structuren; een onderwijskundige analyse [Structure of mathematics and mathematical structures; an educational analysis]. *Pedagogische Studiën, 56*(2), 51-60.

Freudenthal, H. (1991). *Revisiting mathematics education. China lectures*. Dordrecht: Kluwer Academic Publishers.

Handal, B. & Herrington, A. (2003). Mathematics teachers' beliefs and curriculum reform. *Mathematics Education Research Journal, 15*(1), 59-69.

Hazekamp, D. (1978). Teaching multiplication and division algorithms. In M. N. Suydam, & R. E. Reys (1978). *Developing computational skills* (pp. 96-128). Reston, VA: NCTM.

Hickendorff, M., Heiser, W. J., Van Putten, C. M., & Verhelst, N. D. (2009). Solution strategies and achievement in Dutch complex arithmetic: Latent variable modeling of change. *Psychometrika, 74*(2), 331–350.

Inspectie van het Onderwijs (1998). *Verslag van de staat van het onderwijs in Nederland over het jaar 1997*. Utrecht: Inspectie van het Onderwijs.

Janssen, J., Van der Schoot, F., Hemker, B., & Verhelst, N. (1999*). Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 3*. Arnhem: Cito.

Janssen, J., Van der Schoot, F., & Hemker, B. (2005). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 4*. Arnhem: Cito.
Kilpatrick, J., Swafford, J., & Findell, B. (Eds.) (2001). *Adding it up: helping children learn mathematics*. Washington, DC: National Academy Press.
KNAW-Commissie Rekenonderwijs Basisschool (2009). *Rekenonderwijs op de basisschool*. Amsterdam: KNAW
Kolovou, A., Van den Heuvel-Panhuizen, M., & Bakker, A. (2009). Non-routine problem solving tasks in primary school mathematics textbooks – A needle in a haystack. *Mediterranean Journal for Research in Mathematics Education, 8*(2), 31-68.
Kühnel, J. (1925). *Neubau des Rechenunterricht II* (5. Druck). Leipzig: Klinkhardt.
Kuitenbrouwer, J. (2010). Niet de beste spreker wint de verkiezingen, maar hij die goed zijn woorden kiest. *NRC Handelsblad*, 15 May 2010,
Lakoff, G. (2004). *Don't think of an elephant!* White River Junction, VT: Chelsea Green.
Levering, B. (2009). Ons rekenonderwijs is Best of the West. *Pedagogiek in Praktijk Magazine, 47*, 14-20.
Leung, I., Wong, R., & Pang, W. (2006). *Departing from the Traditional Long Division Algorithm: An Experimental Study*. Paper presented at the 29th Annual Conference of Mathematics Education Research Group of Australasia. Canberra, Australia (http://www.merga.net.au/documents/RP382006.pdf)
Menne, J. J. M. (2001). *Met sprongen vooruit. Een productief oefenprogramma voor zwakke rekenaars in het getallengebied tot 100 – een onderwijsexperiment*. Utrecht: CD-beta Press, Universiteit Utrecht.
Middleton, J.A. & Van den Heuvel-Panhuizen, M. (1995). The ratio table. *Mathematics Teaching in the Middle School, 1*(4), 282-288.
Mullis, I.V.S., Martin, M.O., & Foy, P. (2008). TIMSS 2007 International Mathematics Report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
Nussbaum, M. C. (2002). Education for citizenship in an era of global connection. *Studies in Philosophy and Education, 21*, 289–303.
Onderwijsraad (2003). *Kennis van onderwijs: Ontwikkeling en benutting*. Den Haag: Onderwijsraad.
Onderwijsraad (2010). *Ontwikkeling en ondersteuning van onderwijs*. Den Haag: Onderwijsraad.
O'Toole, S. & de Abreu, G. (2005). Parents' past experiences as a mediational tool for understanding their child's current mathematical leaming. *European Journal of Psychology of Education, 20*(1), 75-89
Rengerink, J. (1983). *De staartdeling: Een geïntegreerde aanpak volgens het principe van progressieve schematisering*. Utrecht: OW&OC, Rijksuniversiteit Utrecht.
Streefland, L. (1985). Wiskunde als activiteit en de realiteit als bron. *Nieuwe Wiskrant, 5*(1), 60–67.
Streefland, L. (1993). The design of a mathematics course. A theoretical reflection. *Educational Studies in Mathematics, 25*(1–2), 109–135.
Streefland, L. (1996). Learning from history for teaching in the future. Regular lecture held at the ICME-8 in Sevilla, Spain; in 2003 posthumously published in *Educational Studies in Mathematics, 54*, 37–62.
Thompson, I. (1999). Written methods of calculation. In I. Thompson (Ed.), *Issues in teaching numeracy in primary schools* (pp. 169-183). Buckingham: Open University Press.
Thompson, I. (2008). Deconstructing calculation methods, Part 4: division. *Mathematics teaching, 208*, 6-8.
Treffers, A. (1978). *Wiskobas doelgericht*. Utrecht: IOWO, Rijksuniversiteit Utrecht.
Treffers, A. (1982a) Cijferen in het rekenonderwijs van toen en nu. *Pedagogische Studiën, 59*(3), 97-115.
Treffers, A. (1982b) Basisalgoritmen in het wiskunde-onderwijs op de basisschool. *Pedagogische Studiën, 59*(11), 471-483.
Treffers, A. (1987). *Three dimensions. A model of goal and theory description in mathematics instruction – the Wiskobas project*. Dordrecht: Reidel Publishing Company.
Treffers, A. (1991a). Realistic mathematics education in the Netherlands 1980-1990. In L. Streefland (Ed.), *Realistic Mathematics Education in primary school* (pp. 11-20). Utrecht: CD-β Press, Utrecht University.
Treffers, A. (1991b). Didactical background of a mathematics program for primary education. In L. Streefland (Ed.), *Realistic Mathematics Education in primary school* (pp. 21–56). Utrecht: CD-β Press, Utrecht University.
Treffers, A. (1993). Wiskobas and Freudenthal: Realistic Mathematics Education. *Educational Studies in Mathematics, 25*, 89-108.
Treffers, A., & De Jong, R. A. (1984). Nieuwe aanpak 'ouderwetse rekenkwast' voldoet uitstekend. *Didaktief, 14*(2), 2-4.
Treffers, A., & De Moor, E. (1984). *10 voor de basisvorming rekenen/wiskunde*. Werkboek. Utrecht: OW&OC, Rijksuniversiteit Utrecht.

Treffers, A., & De Moor, E. (1990). *Proeve van een nationaal programma voor het reken-wiskundeonderwijs op de basisschool. Deel 2. Basisvaardigheden en Cijferen*. Tilburg: Zwijsen.
Treffers, A., De Moor, E., & Feijs, E. (1989). *Proeve van een nationaal programma voor het reken-wiskundeonderwijs op de basisschool. Deel I. Overzicht einddoelen*. Tilburg: Zwijsen.
Treffers, A., & Van den Heuvel-Panhuizen, M. (2010). De rekenmethode telt. *Reken-wiskundeonderwijs: onderzoek, ontwikkeling, praktijk, 29*(1), 50-56.
Van de Craats, J. (2007). Waarom Daan en Sanne niet kunnen rekenen. *Nieuw Archief voorWiskunde, 5/8*(2), 132-136.
Van den Heuvel-Panhuizen, M. (1996). *Assessment and Realistic Mathematics Education*. Utrecht: CD-β Press, Utrecht University.
Van den Heuvel-Panhuizen, M. (2001). Realistic Mathematics Education in the Netherlands. In J. Anghileri (Ed.), *Principles and practices in arithmetic teaching: Innovative approaches for the primary classroom* (pp. 49-63). Buckingham/Philadelphia: Open University Press.
Van den Heuvel-Panhuizen, M. (2003). The didactical use of models in realistic mathematics education: an example from a longitudinal trajectory on percentage. *Educational Studies in Mathematics, 54*, 9–35.
Van den Heuvel-Panhuizen, M. (2008a) Learning from "didactikids": An impetus for revisiting the empty number line. *Mathematics Education Research Journal, 20*(3), 6-31.
Van den Heuvel-Panhuizen (Ed.), M. (2008b). *Children learn mathematics. A learning-teaching trajectory with intermediate attainment targets for calculation with whole numbers in primary school*. Rotterdam/Tapei: Sense Publishers.
Van den Heuvel-Panhuizen, M., & Goffree, F. (1986). *Zo rekent Nederland*. Enschede: SLO.
Van den Heuvel-Panhuizen, M., & Treffers, A. (1998). *NCRC-module Oefenen*. Utrecht: Freudenthal Instituut, Universiteit Utrecht.
Van den Heuvel-Panhuizen, M., Robitzsch, A., Treffers, A., & Köller, O. (2009). Large-scale assessment of change in student achievement: Dutch primary school students' results on written division in 1997 and 2004 as an example. *Psychometrika, 74*(2), 351–365.
Van Gelder, L. (1959). *Grondslagen van de rekendidactiek* (3e druk). Groningen: J.B. Wolters.
Van Maanen, J. (2007). *De koeiennon*. Utrecht: FIsme, Universiteit Utrecht.
Van Putten, C. M. (2005). Strategiegebruik bij het oplossen van deelsommen. In J. Janssen, F. van der Schoot, & B. Hemker, *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 4* (pp. 125-131). Arnhem: Cito.
Wijnstra, J. (Ed.)(1988). *Balans van het rekenonderwijs in de basisschool*. Arnhem: Cito.