Shortcomings of Mathematics Education Reform in The Netherlands: A Paradigm Case?

Koeno Gravemeijer, Geeke Bruin-Muurling, Jean-Marie Kraemer & Irene van Stiphout

To cite this article: Koeno Gravemeijer, Geeke Bruin-Muurling, Jean-Marie Kraemer & Irene van Stiphout (2016) Shortcomings of Mathematics Education Reform in The Netherlands: A Paradigm Case?, Mathematical Thinking and Learning, 18:1, 25-44, DOI: 10.1080/10986065.2016.1107821

To link to this article: http://dx.doi.org/10.1080/10986065.2016.1107821
Shortcomings of Mathematics Education Reform in The Netherlands: A Paradigm Case?

Koeno Gravemeijera, G geeke Bruin-Muurlingb, Jean-Marie Kraemerc, and Irene van Stiphoutc

aEindhoven School of Education; bHAN University of Applied Sciences; cCito (Dutch Institute for Test Development)

ABSTRACT
This article offers a reflection on the findings of three PhD studies, in the domains of, respectively, subtraction under 100, fractions, and algebra, which independently of each other showed that Dutch students’ proficiency fell short of what might be expected of reform in mathematics education aiming at conceptual understanding. In all three cases, the disappointing results appeared to be caused by a deviation from the original intentions of the reform, resulting from the textbooks' focus on individual tasks. It is suggested that this “task propensity”, together with a lack of attention for more advanced conceptual mathematical goals, constitutes a general barrier for mathematics education reform. This observation transcends the realm of textbooks, since more advanced conceptual mathematical understandings are underexposed as curriculum goals. It is argued that to foster successful reform, a conscious effort is needed to counteract task propensity and promote more advanced conceptual mathematical understandings as curriculum goals.

Introduction

In many Western countries controversies have arisen about mathematics education, between proponents of respectively reform mathematics and traditional mathematics education (see, for instance Baroody, 2003). Those discussions were induced by the results of national and international assessments that put the proficiency of the students in the spotlight. How these results are interpreted depends of course on one’s view on the goals of mathematics education and how it should be taught. In The Netherlands, national assessment, which showed a decline in the mastery of basic skills,1 has evoked a media hype on “failing” mathematics education. Leading are proponents of a traditional approach of teaching algorithms. They criticize mathematics education reform and the underlying theory of realistic mathematics education (RME). In response to this debate three self-contained PhD studies were carried out, aiming to establish whether such criticism was justified by gathering more detailed information on student proficiency and the actual curriculum. The studies, which focused on three topics—subtraction up to 100, fractions, and algebra—showed remarkable similarities in their findings. The goal of this article is to draw lessons from these three studies that transcend the individual cases. Thus even though we will present data from those studies, this is not a report on empirical research. Instead, the results of these studies form the starting point for an analysis of what might be going wrong with the implementation of reform mathematics.

We will start by laying out the context of the Dutch reform, which we see as representative for the reform efforts in many countries that build on the idea that students should be supported in actively constructing their own knowledge. Against this background, we will discuss the common pattern that emerges from the three studies. In each case student performance did not match expectations, and in each case textbooks analysis revealed a focus on teaching effective procedures for specific
tasks, instead of developing mathematical insights at a more advanced conceptual level as the reform intended. We will argue that this is at least in part due to what we call, “task propensity,” which we define as the tendency to think of instruction in terms of individual tasks that have to be mastered by students. This task propensity entices teachers and textbook authors to capitalize on procedures that can quickly generate correct answers, instead of investing in the underlying mathematics while accepting that fluency may come later. We will next discuss the relationship between task propensity and the lack of attention for more advanced conceptual goals. We assume a reflexive relationship, and we will argue that task propensity can flourish because more advanced conceptual mathematical understandings are not formulated as curriculum goals. Building on the literature, we delineate more advanced conceptual mathematical goals as end points of the reification processes that are lacking in the researched instructional sequences. We illuminate this framework with a rough sketch of these goals for the domains of the three original studies. In closing, we put our findings in perspective, and we will argue that in order to secure a truthful enactment of reform mathematics a conscious effort has to be made in counteracting task propensity, and explicating and advocating more advanced conceptual mathematical understandings as key curriculum goals.

Mathematics education reform in The Netherlands

The groundwork for the mathematics education reform in The Netherlands was carried out by the Freudenthal Institute and its predecessors. The work of the institute resulted in both a series of prototypical instructional sequences and an overarching (domain-specific) instruction theory, known as realistic mathematics education, or RME (Gravemeijer, 2008; Treffers, 1987). RME fits under the wider umbrella of what we—for lack of a better word—will call reform mathematics. Where we may broadly characterize reform mathematics as forms of mathematics education that try to do justice to the idea that, in order to develop a deep understanding of mathematics, students have to have an active role in constructing their own knowledge.

The Dutch reform is rooted in Freudenthal’s (1973) adagio that students should experience mathematics as a human activity. He describes that activity as solving problems, looking for problems, and organizing subject matter (Freudenthal, 1971). The latter activity, organizing, has later been denoted mathematizing, which refers to organizing subject matter to make it more mathematical. According to Freudenthal (1971), students would be able to reinvent conventional mathematics by mathematizing both subject matter from reality and mathematical matter—especially concerning their own mathematical activity—under guidance of the teacher. In his view, starting points should be experientially real for the students and from thereon they would have to be supported in a process of guided reinvention, with sophisticated, conventional mathematics as its potential end points. When Treffers (1987) later formulated RME theory, he linked guided reinvention to progressive mathematization, which he places in the broader framework of Van Hiele’s theory on levels of thought in mathematics education (van Hiele, 1973, 1986), and Freudenthal’s (1983) didactical phenomenology. According to the latter, one should look for phenomena that might compel the learner to constitute the mental object that corresponds to the concept under consideration. In this manner, those phenomena do not only function in applications but also as a source of concept formation or constitution of mental objects.

One of the tasks of the first predecessor of the Freudenthal Institute (the Institute for the Development of Mathematics Education) was to elaborate these ideas in exemplary instructional practices and teaching materials. In doing so it produced prototypical instructional sequences for various topics, which were adopted and adapted by primary- and secondary-school textbook authors. This was a dynamic process that started in the 1980s. The institute produced a series of curriculum documents, in which the results of various design experiments were worked out. Especially influential were the publications by Treffers, de Moor, and Streefland that offered specimens of curricula for various topics for primary school (Treffers & Moor, 1989, 1990; Treffers & Streefland, 1994; Treffers, Streefland, & Moor, 1996), and the publications that followed under the banner of “TAL,” which
stands for “Teaching-and-learning trajectories annex intermediate attainment targets” (TALteam, 2007; Treffers, Van den Heuvel-Panhuizen & Buys, 1999; Van den Heuvel-Panhuizen, 2001; Van Galen et al., 2008). Even though they did not have an official status, these curriculum publications were perceived as the mandated curriculum, especially since the statutory curriculum goals were aligned with those publications.²

From the 1980s onward, RME-inspired mathematics textbooks quickly began to dominate the market—although teacher support tailored to the innovation was scarce to virtually lacking. And, for several decades the RME approach was well-received, both in the educational community and in the Dutch society in general. In the beginning of this century, however, when national tests showed a decline in key basic skills, criticism started to be heard. This was amplified by the media. The teaching, the textbooks, and by implication RME theory and the Freudenthal Institute were blamed for what some considered to be highly unsatisfactory results. The basis of the critique was in the results of national assessments, such as the Periodical Assessment of the Level of Education, PPON (Janssen, Van der Schoot, & Hemker, 2005), and international assessments, such as PISA (Organisation for Economic Co-operation and Development, 2007). And even though test results are weighed differently by different groups, and the picture that emerges from the national tests is mixed, it is safe to say that the mathematics education reform in The Netherlands has not brought strong improvement when compared with the results that were achieved some decades ago.

### Studies on proficiency and solution methods

This raised the question: What was going on that might have caused the unsatisfactory results? This question was picked up in three studies, which independently of each other tried to get a better understanding of why and in what respect the results obtained by RME-inspired textbooks lagged behind expectations. Each study tried to answer this question for a different topic, respectively early arithmetic, fractions, and algebra. The reason for discussing those studies here is that the three studies showed similar results, revealing a consistent pattern across grade levels and mathematical content. The studies were comparable in that they paired detailed information on the proficiency and the solution procedures of students with a textbook analysis, although the methods varied.

In the following we will start with a description of the findings on the abilities of the students. Next, we discuss the textbook analyses. Then we will focus on the common thread that emerges from the three studies.

#### Subtraction up to 100

Subtraction up to 100 is one of the topics of the so-called PPON surveys that the National Institute for test development, Cito, periodically carries out on behalf of the Ministry of Education. In conjunction with the PPON-4 survey (Kraemer, Van der Schoot, & Hemker, 2005), Kraemer researched the solution procedures of 300 grade 3 students (Kraemer, 2011; see also 2009).³ Those students were sorted in groups of low, middle, and high level of proficiency. During individual interviews the students were presented with subtraction problems from the PPON items set, corresponding to their level of numeracy, and interviewed about their solution procedures. The solution procedures were classified according to a two-dimensional grouping of solution procedures, based on Beishuizen’s (1993) categorization, consisting of two dimensions, strategies, and computing methods:

- strategies:
  - subtraction
  - indirect addition
and

- computing methods:
  - jumping; or skip counting
    
    \[
    (\text{e.g., } 65 - 38 = \ldots, \text{via } 65 - 30 = 35, 35 - 5 = 30, 30 - 3 = 27)
    \]
  - splitting; splitting tens and ones
    
    \[
    (\text{e.g., } 68 - 45 = \ldots; 60 - 40 = 20, 8 - 5 = 3, \text{answer } 20 + 3 = 23)
    \]
  - reasoning; deriving number facts using arithmetic properties
    
    \[
    (\text{e.g., } 62 - 48 \text{ equals } 64 - 50 = 14; \text{or } 62 - 48 = \ldots \text{ via } 62 - 50 = 12, \text{thus } 62 - 48 = 12 + 2 = 14)
    \]
  - knowing; reproducing known facts

The analysis of the frequency of use and the correct (or incorrect) execution of the four computing methods showed that jumping was both the most used method (between 57% and 74%) and the most effective method (between 82% and 91% correct)—and was applied flexibly. Splitting and reasoning generated many incorrect answers (splitting between 65% and 42% correct, and reasoning between 50% and 31% correct). This might explain why the results of the fourth PPON did not show improvement in comparison to the third PPON.

An analysis of the relationship between strategies, methods, and task characteristics revealed that a large group of students flexibly varied between subtraction and indirect addition when modeling various types of contextual problems. However, when they had to solve subtraction tasks in the form of number sentences, between 90% and 99% of the students clung to subtraction. Based on the literature on the role of inversion in the development of addition and subtraction proficiency, we may conclude that these students had generalized indirect addition for story problems, but not yet for decontextualized subtraction tasks.

Two problems were identified as the main causes of the incorrect answers that occurred, when the low- and middle-level students did not use jumping methods. First, a substantial part of those students seemingly routinely applied the well-known buggy algorithm (e.g., \(62 - 48 = 20 + 6\)) (Fuson et al., 1997). Second, many students who operated on the middle level tended to use incorrect forms of splitting, and reasoning when solving context problems. Further, the Rasch analysis of the fourth PPON—in which the interview items were included—showed that the tasks that involved bridging ten were too difficult for most students (Kraemer et al., 2005). This, together with the sources of incorrect answers, points to a lack of proficiency in coordinating tens and ones.

In summary, the results showed that the students often used jumping methods, with which they were flexible and successful. However, the results also revealed structural problems, which appear to be rooted in a limited understanding of the inverse relationship and a lack of proficiency in coordinating tens and ones. In other words, the students acquired restricted set of computing methods and reached only a limited level of conceptual understanding.

**Fractions**

In 2006–2009 a comprehensive research project was carried out on the proficiency in the domain of fractions, concerning students in grade 6 through 9 (Bruin-Muurling, 2010). For this purpose a series of tests was constructed to measure both procedural and conceptual aspects of proficiency in the fraction domain. Based on literature, textbooks, and an analysis of the mathematics at hand, a number of subdomains, underlying big ideas and complexity factors, were defined, which together
formed a framework for item construction. Each of the grades worked on an appropriate subset of these tasks. In total, 1498 of these tests were administered. A Rasch analysis of the data resulted in one scale on which both item difficulty and student proficiency could be placed. Of interest here are the results of the ninth graders in the sample ($N = 347$) (higher general secondary education [HAVO] stream, and pre-university education [VWO] stream). The students did master addition and subtraction of fractions with common or related denominators, but only the best students were successful in adding or subtracting fractions with unrelated denominators. Multiplication proved to be difficult; only tasks such as 4/7 part of 35 euro, and $5 \times 4 \frac{1}{5}$, were mastered by half of the students. Division problems could be solved by half of the students when posed in a context. More formal division items were too difficult for about three quarters of the students. Problems involving ratio/rate in the context of density were mastered by only a small portion of the students.

A conceptual analysis was carried out by comparing the positions on the Rasch scale of items, which were rather similar but differed in some respects. This lead to the conclusion that the students did not grasp the underlying big ideas, such as unit, fraction as a number, and the relationship between fractions, multiplication, and division. We may illustrate this with the results of some test items.

The majority of the students were able to say which part of a rectangle or circle was shaded. However, if this shape was divided into unequal sized parts—with the shaded part depicted as one third of one half, for instance—a large portion of students was unable to name the corresponding fraction. This can be interpreted as a limited understanding of the unit.

As a second example we may mention that 50% of the students more or less mastered $10 \times \frac{2}{5}$, but only 10% could solve $31 \times \frac{17}{31}$ correctly. An analysis of written answers showed that the majority of the students started by multiplying 31 and 17, apparently not realizing that multiplying by $1/31$ is the inverse of multiplying by 31. This suggests that the majority of the students did not grasp the relationship between fractions and the operations of multiplication and division.

In summary we may conclude that—similar to the case of subtraction—the students acquired a number of solution procedures that worked well for specific tasks. However, they were not able to handle more complex variations on those tasks, and surely they did not reach the conceptual level of big ideas, and relationships between fractions, multiplication, and division.

**Algebra**

Van Stiphout (2011) did a study on algebra proficiency of 1020 students in grades 8 through 12 in 2008–2009. Students of each grade were assessed four times during one calendar year. The test items ranged from basic skills to symbol sense (Arcavi, 2005; Drijvers, 2003), covering customary algebraic skills taught in grades 7 thru 9 in The Netherlands. These skills concerned expanding brackets, simplifying expressions, and solving equations. A Rasch analysis applied on all data led to one scale on which both item difficulty and student proficiency could be placed. This allowed for a cross-sectional comparison of the student performances in grades 8 through 12, and an analysis of individual progress during the year in question. Both showed that the students made some progress, but did not reach the fluency and understanding for which voices in higher education ask (see also Van Stiphout, Drijvers, & Gravemeijer, 2013).

When we zoom in on the grade 12 students, the analysis showed the following. Most tasks related to negative numbers, addition of square roots, and simple algebra tasks, such as, “simplify $-2(4x - y) + 3 (-2y - 4) = \ldots$”, were mastered by about 75% of the students. The tasks, however, quickly became too difficult for most students. Almost all tasks that proved too difficult for most of the students asked for some deeper conceptual understanding of algebra. The students’ difficulties indicated a lack of algebraic understanding, which came to the fore in a lack of symbol sense and structure sense (Novotnà & Hoch, 2008), and the inability to shift between an unknown and a variable conception (e.g., Arcavi, 1994) or to assimilate Sfard’s (1991) process-object duality. We may illustrate the students’ difficulties with those aspects with the results on some test items, complemented with
information on how the students solved those items based on Van Stiphout’s analysis of the written solutions of the students.

One of those tasks was: “Simplify $-6 - (5 - 4)(-8) - 3$”. The difficulty of this task can be explained by the necessity to interpret the structure of the expression correctly as $\{-6\} - ((5-4)(-8)) - 3$. This task proved too difficult for 90% of the students. Many students interpreted the expression as a multiplication of two partial expressions; they first calculated the values of $-6 - (5 - 4)$ and $(-8) - 3$ yielding $-7$ and $-11$, and then multiplied both numbers. This can be seen as an illustration of a lack of symbol sense, as the students were not able to shuttle back and forth between considering the overall structure of the expression and doing the calculations.

A task that clearly asked for structure sense was: “Solve the equation $(x - 1)(x + 3)(x - 4) = 0$.” Here the students were expected to recognize the structure $A \cdot B \cdot C = 0$ as an extension of quadratic forms such as $(x - 1)(x + 3) = 0$. The students were familiar with solving quadratic equations by factorizing, but not with solving cubic equations. Thus the students had to generalize the structure $A \cdot B = 0 \rightarrow A = 0$ or $B = 0$ to $A \cdot B \cdot C = 0 \rightarrow A = 0$ or $B = 0$ or $C = 0$. This task was too difficult for more than 90% of the grade 12 students. The worked out solutions of the students showed that the students tended to expand the brackets, after which they were unable to find the factorization and did not arrive at the conclusion that $A = 0$ or $B = 0$ or $C = 0$. We may add that in order to make those generalizations, the students had to be able to conceptualize individual factors, such as $(x - 1)$, both as a process and as an object.

In an additional exploratory substudy, Van Stiphout (2011) tested the hypothesis that the difficulty of the tasks was determined by the two types of transitions Sfard and Linchevski (1994) discerned between purely operational algebra and structural algebra, and between an unknown and a variable conception. She designed a series of tasks with an ascending number of transitions, which were solved by 92 grade 11 students who were asked to explicate their reasoning. In line with the expectations, it showed that the success rate decreased when the number of transitions increased. An analysis of the written explanations and some individual interviews confirmed that the students had severe problems with the aforementioned transitions.

Recapitulating: The analysis showed that the students mastered simple tasks, but the tasks of the test became too complicated rather quickly. The majority of the students were not able to use the mathematical structure of expressions, and could not handle process-object duality. We may therefore conclude that—just as in the other studies—the students developed a restricted set of solution procedures, while the majority of them did not reach the level of more advanced conceptual mathematical understandings.

**General conclusion on student proficiency**

Looking at the findings of these three studies we may conclude that the students developed well-grounded ways of solving a specific set of problem types, but they could not manage tasks that transcended this set. We may note in passing that these data cannot be taken as an indication of a decline pertaining to some decades ago. These problems are not new, as the reform movement surely emerged as a response to shortcomings of the mathematics education of that time. Moreover, the Netherlands still belongs to the highest scoring countries on international tests such as PISA, and a comparison of consecutive PPON surveys (Janssen et al., 2005) showed a positive effect on a number of topics that RME innovators deem important (Van den Heuvel-Panhuizen, 2010). But this is not what these three studies tried to investigate; rather, they aimed at determining what the actual proficiency of the students was. What concerns us here is that results fell short of what might be expected of reform in mathematics education aiming at conceptual understanding. Apparently the students did master some basic solution methods but had not learned to reason on a more general, higher, level of mathematical understanding.

Concerning the critique on reform mathematics that motivated the studies, we may observe that the studies add valuable information to what could be gained from the (inter)national assessments.
There is proof of problems, but the frequently heard calls for more training on algorithms seem to be unjustified. The more pressing problem appears to be the weakness of the students' conceptual basis. In all three studies, this led to an investigation into the question of the cause of these results.

**Textbooks**

Since Dutch teachers are known to follow their textbooks faithfully (e.g., Olson, Martin, Mullis, Foy, Erberber, & Preuschoff, 2008), textbook analyses were carried out in all three cases in order to get a better understanding of how these disappointing results came about. We summarize the results of those textbook analysis in the following.

**Subtraction in textbooks**

Based on a textbook analysis by Menne (2004), Kraemer (2011) concluded that the solution procedures the students in his study used for subtracting up to 100, reflected the way these topics were introduced in the newer textbooks—which were used by 80% of the schools. The Dutch textbooks broadly followed the prototypical instructional sequences for addition and subtraction developed within the Freudenthal Institute by Treffers cum suis (Treffers & de Moor, 1990; see also Treffers & Buijs, 2001), which we describe next.

The prototypical instructional sequences capitalize on the two main informal solution procedures that come to the fore in the literature (Beishuizen, 1993): “jumping” (or “skip-counting”), and “splitting” (see also Verschaffel, Greer, & De Corte, 2007). Jumping is introduced first, while the introduction of splitting is postponed until the end of grade 2. To foster jumping methods student activities are suggested on counting, comparing, and adding or subtracting numbers of beads on a bead string of 100 beads organized in a pattern of ten dark beads, ten white beads, and so on (Whitney, 1988). While performing these tasks, students are expected to develop informal solution strategies in which they use multiples of ten as reference points. Subsequently, visualizations of the jumps by arcs on a number line are introduced—both as a means for communicating solution procedures, and as a means of scaffolding. Later on, these visualizations are replaced by a form of arrow language, which in turn is replaced by the formal arithmetical notation.

To foster splitting methods, contextual problems are proposed that involve working with money or packing objects in groups of ten. However, the prototypical instructional sequences do not encompass a detailed instructional sequence on splitting.

So-called conventional ways of subtracting (Treffers, 1991) are arrived at by stimulating students to gradually curtail their solution methods in a step-by-step manner. These conventional methods are characterized by standardization, which—for subtraction—involves first taking away the tens, then taking away the ones. In addition to those conventional methods, Treffers discerned “varied” solution methods (Treffers & Buijs, 2001; Treffers & Moor, 1990; Treffers, 1991). These concern indirect addition, and methods that Kraemer (2011) denoted “reasoning,” also known as “derived facts strategies” (Verschaffel et al., 2007). To facilitate indirect addition students are, for instance, presented with contextual problems that ask for finding the difference between two numbers. In addition they are offered subtraction tasks without context in which the difference between the subtrahend and minuend is small (such as 83–78). Reasoning methods involve compensating (62–48 is two more than 62–50), or changing both the subtrahend and minuend (62–48 equals 64–50).

A textbook analysis by Menne (2004) showed that textbook authors interpreted the recommendations of Treffers cum suis in different ways. She observed that with exemption of one textbook series, jumping and splitting were offered at about the same time. The way splitting and reasoning were treated in the textbooks varies. In general, the attention for splitting methods and reasoning methods was rather sparse. This may have to do with the fact that these methods were not well-elaborated by the designers/researchers. Another reason may have been the experience that jumping methods are easy to learn and highly effective.
The findings of Kraemer’s study suggest that over time, the students may have started to experiment with splitting and reasoning methods in combination with indirect addition to find more efficient ways to solve subtraction problems, while lacking a solid conceptual basis. Consequently, the efforts to use such methods generated many errors, especially with tasks that involved bridging ten. In summary we may conclude that the textbooks offer a sound grounding of particular informal strategies, but a continuation toward a higher level of understanding that integrates various ways of modeling and decontextualized calculation methods is missing. The latter may have been caused by the appeal of methods that routinely produce correct answers.

**Fractions in textbooks**

As part of the study of Bruin-Muurling (2010) four primary and two secondary textbook series were analyzed, which together covered almost the total market share in both primary and secondary education. First all fragments in the sixth- and seventh-grade textbooks, in which calculations with fractions were required or that handled fraction concepts that constitute these calculations, were analyzed. This analysis was directed at common structures in the textbooks, such as structure of the chapters and the use of examples and figures.

Next a more detailed analysis was made of the fragments involving fraction multiplication. This fine-grained analysis focused on the meaning of multiplication by selecting tasks that were representative for the complete set of fragments. The analysis showed that multiplying fractions was firmly grounded in contextual problems that could be considered experientially real for the students. Furthermore the commonly discerned conceptual aspects were addressed in the chosen contexts, such as the fraction as “part-whole,” “measure,” “quotient,” “ratio number,” and “multiplicative operator.” The various phenomenological manifestations of multiplication were used to explore the nature of multiplication in the fraction domain, such as repeated addition, part of a whole, and area. This introduction to fractions and fraction multiplication was phenomenologically rich, as might be expected from RME inspired textbooks. However, the informal strategies that emerged in this manner were turned into four distinct number-specific solution procedures.

The first procedure concerns the multiplication of a *whole number times a proper fraction*, which is solved by translating the multiplication in a repeated addition, for example, \(5 \times \frac{3}{4} = \frac{3}{4} + \frac{3}{4} + \frac{3}{4} + \frac{3}{4} + \frac{3}{4}\). In the textbooks, this procedure emerged as an informal strategy for contextual problems that could be modeled by repeated addition, for example, calculating the contents of nine bottles of three-fourths of a liter. The second procedure fits the multiplication of *proper fraction times a whole number*, for example, \(\frac{3}{4} \times 60 = \ldots \rightarrow \frac{3}{4} \times 60 = 60 \div 4 = 15; \frac{3}{4} \times 60 = 3 \times 15 = 45\). Here multiplying with a fraction is translated into “taking a part of.” In the textbooks this procedure was preceded by context problems concerning for instance, three-fourths of 28 tins, or two-fifths of 200 grams of flour. It has to be noted that the numbers in these problems were always chosen in such a way that the answer was a whole number. The third procedure is tailored to the *multiplication of two proper fractions*. This procedure is tied to a ready-made rectangular model that students had to interpret as the result of the following steps: (1) Identify the part of the rectangle corresponding to the second fraction (light grey). (2) Take the part of this part corresponding to the first fraction (dark grey). (3) Determine the size of this part of the rectangle, and equate this with the product of the two fractions (see Figure 1).

The fourth procedure designed for *multiplying mixed numbers*, varies over textbooks. Either splitting the mixed number in a whole number and a proper fraction and using the distributive property (e.g., \(4\frac{1}{2} \times \frac{3}{4} = 4 \times \frac{1}{2} + \frac{1}{2} \times \frac{3}{4}\)), or splitting the mixed number in a suitable whole number and a mixed remainder (e.g., \(1/3 \times 7\frac{1}{2} = 1/3 \times 6 + 1/3 \times 1\frac{1}{2}\)).

Mark that in all four cases the textbooks systematically used numbers that allowed for a smooth operation of the procedure the students were expected to use.

In summary, we may conclude that primary-school textbooks offered multiple inroads for the multiplication of fractions. A shortcoming, however, was that these approaches were not integrated at some higher conceptual level. Instead, the number-specific solution procedures were practiced as
separate, independent, procedures. In other words, fraction multiplication was split in four types of tasks based on the type of fractions involved, and students were taught specific solution procedures for each of these types of tasks. Subsequently, the standard procedure for multiplying fractions (multiplying nominators and multiplying denominators) was introduced in the secondary school textbooks without much ado. As a consequence, the students then had five different and unconnected procedures at their disposal, which they started to mix up.

We may observe that the pattern is similar to that of what was found for subtraction, the textbooks offer a sound grounding in informal strategies and contexts, but instead of working toward more advanced conceptual mathematical goals, the focus is on teaching procedures that offer short-term success for specific tasks.

**Algebra in textbooks**

Van Stiphout (2011) analyzed how linear relationships and linear equations were taught in the two secondary school, pre-university stream (VWO) textbooks, which together had an estimated market share of more than 95%. Based on the theory of emergent modeling (Gravemeijer, 1999) tasks and text segments were initially classified in three categories, which encompassed activities that (1) allowed for informal strategies within contexts, (2) fostered a shift in focus from informal to more general solutions and the mathematical relations involved, and (3) offered students support in constructing mathematical objects that derive their meaning from the emerging network of mathematical relationships. However, an additional fourth category was needed to classify activities that did not fit into the previous categories. Those concerned an unconnected, formal, introduction of the general formula for a straight line and the process of solving linear equations in both textbook series.

The analysis showed that the first two categories were well represented in both textbook series, which paid extensive attention to phenomenological exploration, with a considerable amount of activities that focus on the exploration of contextual problems. For example, in both textbook series, linear equations are introduced by means of informal strategies such as the cover-up method and the balance model. Figure 2 illustrates the use of the balance model in a grade 7 textbook.

These informal strategies may serve as background for solving linear equations and for interpreting the solution. In Figure 2, for instance, the balance model at the left-hand side supports the manipulations of the equation in the right-hand side, thus legitimating the different steps in the procedure of solving the equation.

However, both textbook series hardly support the step in which equations have to become reified to objects that are part of a network of mathematical relations that justify how to operate on these objects, for example, when adding and subtracting numbers or variables on both sides of an equation, or multiplying both sides of an equation by the same number. Rather, they abruptly switch to a more traditional approach in which new concepts are introduced as ready-made mathematics—without making the link to the aforementioned balance model. We may take Figure 3 as an example. In this fragment of a grade 9 textbook the rule “change sides then change signs” is introduced by turning a pattern observed in two examples into a rule.
However, this rule is not grounded in preceding situated knowledge, which might have been acquired with the balance model, for instance. Therefore, the introduction of the rule “change sides—change signs” does not fit in the approach of progressive mathematization. Instead, the rule “change sides—change signs” is presented as a given fact.
In summary we may conclude that the most crucial step in the formation of concepts, in which students build on their understanding of linear contexts to construct mathematical relationships and mathematical objects, was hardly supported. Instead, two distinct didactical tracks were found: The textbooks started with an RME approach, but shifted halfway to a more traditional approach. New rules were introduced, and practiced, as ready-made mathematics.

Again the phenomenological exploration and informal solution methods were not expanded into more sophisticated mathematics. Instead, there was a forced turn to traditional standard forms. We may assume that this shift was motivated by the wish to teach ready-made solution procedures that offer quick, immediate, results; similar to the other cases.

**Common trend in textbooks**

A comparison of these studies shows a common trend in the three instructional sequences. It shows that in all three cases, the connection with the informal knowledge and strategies was well cared for; the textbooks offered a well-considered phenomenological exploration, resulting in a rich variety of context-related solution procedures. However the students were not supported in building on their situated knowledge to construct more formal mathematics—which would have enhanced their ability to flexibly tackle a larger variety of tasks. In fact, the ambition to reach more advanced mathematical endpoints seemed to be lacking. The instructional sequences did not seem to aim at more advanced conceptual mathematical understandings. Instead, the emphasis seemed to be on using solution procedures that proved to be effective.

- In early arithmetic, these short-term effective solution procedures concerned so-called jumping procedures for both subtraction and indirect addition up to 100 (Kraemer, 2011).
- For multiplying fractions, the short-term effective solution procedures in primary education concerned number-specific procedures that could be traced back to informal context-bound solution procedures. The secondary education textbooks simply introduced the formal rule, even though the conceptual base was insufficient (Bruin-Muurling, 2010).
- In algebra, the situation was somewhat different in that the informal solution procedures were not turned into routines. Instead, the secondary-school textbooks ignored the insights the students developed when solving problems in meaningful contexts, and turned toward ready-made solution methods, which might offer short-term success. (Van Stiphout, 2011).

Returning to the controversies that prompted this research, we may conclude that claims about the inadequacy of RME theory are premature. For RME theory is not implemented in correspondence with its principles.

**Task propensity**

Apparently, more advanced conceptual mathematical goals were not taken into consideration by the textbook authors, who instead went for solution procedures that proved to be effective on the short-term. The progressive mathematization toward the higher Van Hiele levels, to which Treffers (1987) pointed, is lacking. And phenomena that were explored did not function as a source of concept formation. We may speculate on the causes of this deviation from the original goals. One of the causes of this deviation might be, what we may call, “task propensity”: the tendency to think of mathematics education in terms of individual tasks that have to be mastered by students. Resnick and Hall (1998) also observed this type of tendency, which they link to a popular view of learning that harks back to associative theories of learning of Thorndike’s time. According to such theories, knowledge consists of a collection of connections between pairs of mental entities or external stimuli and internal mental responses. Learning then is a matter of creating and strengthening the right bonds, which can be fostered by practice and feedback. Key here is that this form of instruction asks
for frequent testing of individual items in order to determine which bonds are mastered, or are not yet mastered. Resnick and Hall (1998) went on to say that the tenets of associationist instruction theory have been absorbed into the core pedagogy of American schools, and have become the basis for the pedagogical standard operating procedures of schools—and we would argue, not just in the United States but in many Western countries: “These are the familiar practices that teachers continue to use and the families and communities still recognize” (Resnick & Hall, 1998, p. 96).

This pedagogy translates into a sometimes explicit but often implicit task propensity. Central in this view is the tendency to link learning to the ability of students to successfully carry out well-defined tasks. “Getting right answers,” is seen as an immediate goal and not as a long-term spin-off of a deeper mathematical understanding. Consequently, tasks are likely to be considered in isolation of other tasks, while the underlying mathematical issues (Cobb, 1997) are disregarded. This resonates with Daro’s (2011) observation that teachers in the United States when considering a given task, ask themselves: “How can I teach my kids to get the answer to this problem?” instead of “What is the mathematics they are supposed to learn, working on this problem?” In line with the former, there is a preference for procedures with which students can get correct answers to given tasks in a direct manner. This may result in teaching specific solution procedures for each type of task. This fits with the pattern that we observed in the Dutch textbooks. We may add that this might also explain the phenomenon we mentioned earlier, that Dutch teachers hardly deviate from their textbooks (Olson et al., 2008). Most likely, many Dutch teachers also aim for quick results on individual tasks.

We cannot, however, blame the textbooks or the teachers for not aiming for more advanced mathematical understandings, for more advanced conceptual mathematical understandings were not formulated as educational goals in the mandated curriculum in The Netherlands. Also the curriculum guidelines of the Freudenthal Institute that informed the textbook authors showed little attention for more advanced conceptual mathematical goals, as the focus was on reinventing standard procedures. This we should add reflected the conventional goals of that time.

Moreover, most forms of mandated assessment measure student success in terms of right answers on (series of) individual tasks. And we may assume that the habit of focusing on individual test items influences our way of thinking about educational goals. There seems to be a reflexive relationship, in which task propensity influences our thinking about goals and the lack of attention for advanced conceptual mathematical goals allows for task propensity. This implies that to counter task propensity we have to put forward the importance of more advanced conceptual understandings. Note that what we understand by the latter differs from both higher-order thinking and conceptual understanding as such; even though those are important goals in and of themselves. Before elaborating on this difference in the following, we want to point to the tendency to describe conceptual or process goals in curriculum documents independently from the content goals. Mostly with the unfortunate result that only the content goals are tested.

**More advanced conceptual mathematical understandings**

Note that the issue is not that instructional designers, teachers, and textbook authors are not aware of the importance of conceptual understanding. The three examples show that a lot of time and energy is put into providing for a sound conceptual basis that is grounded in the students’ informal knowledge. The problem is that the instructional sequences end too early and are not carried through to reach the conceptual understandings that are needed for the next instructional phase. The typical pattern is one of laying a conceptual foundation, fostering informal solution procedures, and turning these into effective routines. In our view, however, mastering a set of isolated skills does not offer a solid basis for progressing from one instructional phase to the next. In this respect, we find ourselves at odds with the opponents of reform mathematics who claim that the mastery of “basic skills” is sufficient for smooth progress. This difference in opinion finds its origin in the thesis underlying mathematics education reform that learning mathematics is primarily a process of constructing. The literature on the latter shows that long-term learning processes in mathematics
involve more than a mere accumulation of skills. Or as Tall (2008, p. 13) put it: “Procedural learning (…) needs the additional compression into thinkable concepts to enable the long-term development of increasingly sophisticated mathematical thinking.” We will elucidate this point in the following.

According to many scholars the process of constructing mathematics is characterized by a series of transitions in which mathematical processes are transformed into objects, which in turn become part of new processes (e.g., Dubinsky, 1991; Freudenthal, 1991; Pirie & Kieren, 1994; Sfard, 1991; Tall & Thomas, 1991). What is missing from the instructional sequences is precisely this phase of reification. The instructional sequences lack a completing phase in which the students are supported in constructing mathematical objects. Consequently they will not be able to perform the new processes that are the starting points for the next phase in their learning process. Note, however, that the new mathematical objects should not be separated from their genesis. For, as Sfard (1991) expressed with her “dual nature of mathematics,” one often needs both a structural conception—involving mathematical objects—and an operational conception—which concerns processes, algorithms, and actions. Gray and Tall (1991, 1994) also pointed to this duality while introducing the term “procept” to express the interweaving of process and object.

We may assume that the instructional sequences lack the finishing phase described because abstract mathematical objects were not considered as potential end points by the textbook authors. The underlying problem, in our view, is that those goals are not explicitly addressed outside the scientific literature. We therefore advocate explicit attention for this type of advanced conceptual mathematical understandings in curriculum documents and related publications.

To further elucidate this idea of advanced conceptual mathematical understandings, we will briefly sketch the potential end points of the instructional sequences under consideration. For the sake of brevity, we will not address issues of applications and representations in the following. We want to stress, however, that both are essential constituting parts of the more advanced conceptual mathematical goals we have in mind.

**Goals for subtraction, multiplication of fractions, and algebra**

For the instructional sequence on (addition and) subtraction up to 100 the completing phase would have to aim at supporting students in constructing numbers as object-like entities. Here we may refer to the work of Gray and Tall (1991, 1994) and Van Hiele (1973), who spoke of “junctions in a network of number relations.” Instead of following jump or split procedures, they are to use the number relations that are ready to hand to them, and flexibly combine them on the basis of well-understood arithmetical properties. Eventually they will have to come to see subtraction and (indirect) addition as two sides of the same coin, and thus start to think of additions and subtractions that fit together as coherent sets of number relations (for instance, $34 + 28 = 62, 28 + 34 = 62, 62 - 28 = 34$ and $62 - 34 = 28$).

In short, numbers, sums, and difference have to become mental objects that can be composed and decomposed in different ways, which in turn forms the basis for flexible arithmetic and an expansion to bigger numbers, and later for integers and algebra.

Similar to natural numbers, students will also have to come to see fractions as mathematical objects that derive their meaning from a network of number relationships. This in turn has to be grounded in the subconstructs, “part-whole,” “measure,” “quotient,” “ratio number,” and “multiplicative operator” (Behr, Wachsmuth, Post, & Lesh, 1984). In order to come to grips with multiplying fractions, students have to reconstruct what it means to multiply. They are used to thinking of multiplying as repeated addition. This fits nicely with a multiplication such as $5 \times \frac{1}{4}$, but not with $\frac{1}{4} \times 5$, which cannot be construed as repeated addition. For tasks like $\frac{1}{4} \times 5$, students are taught to interpret the multiplication as taking a part of. With some experience, students may come to accept the equality of $5 \times \frac{1}{4}$ and $\frac{1}{4} \times 5$ on the basis of the equal results, an area model, or by assuming that the commutative property holds. However, to come to grips with multiplying fractions on a higher conceptual level, more is needed. Students have to see that this equality holds on the basis mathematical reasoning, and understand how
these various interpretations of multiplication (such as repeated addition, area, and part of) relate. This involves reasoning with relationships between (operations with) fractions, and the operations of multiplication and division. That is, grasping the dual conceptualization of a fraction as a number and as an operation has to evolve into the ability to flexibly switch between the two conceptions when multiplying fractions. This means, for example, that the students have to be able to think of “multiplying with a given fraction” as “multiplying with its nominator and dividing with its denominator” (e.g., “... x ¾”, equals “(... x 3) ÷ 4”), and vice versa. Eventually students have to come to see products of two rational numbers as objects that can be acted and reasoned with, but which also can be decomposed in a series of operations. This would imply that relationships between operations and fractions, such 15 × ¾ = 15 × (3 ÷ 4), 15 × ¾ = (15 ÷ 4) × 3 or 15 × ¾ = (15 × 3) ÷ 4, would be readily available for the students. Those relationships would also encompass inverse relations (e.g., 15 ÷ 4 = 15 × ¼, or 4 × ¼ = 1), including the phenomenological roots of those relationships.

In relationship to the algebra instructional sequence, we may observe that the need for conceptualizing functions as both processes of computing output values from input values, and as objects that can be acted on, is well documented (Dubinsky, 1991; Gray & Tall, 1991, 1994; Sfard, 1991). Typical here is that algebraic expressions can be composed of, or decomposed in, smaller entities, which can be perceived as individual functions or variables. In relation to this the notions of symbol sense (Arcavi, 2005) and structure sense (Novotná & Hoch, 2008) may be brought to bear, which refer to the ability and attitude to flexibly shuttle back and forth between more static and more dynamic interpretations of algebraic expressions and their constituting parts.

In “2x+7,” for example, they have to be able to see this expression in a variety of ways, and to flexibly switch between those meanings. First, “2x+7” has to signify a computational prescription that can be executed for a given value of x to generate a value. Second, “2x+7” has to signify an object that consists of a set of number pairs, which can be associated with a linear graph and that can be interpreted dynamically. Third, “2x+7” has to signify an object that can be isolated and manipulated in the composition of algebraic expressions such as (2x+7)³. This structural conception also has to encompass the ability to compare expressions as objects, noting, for instance, that “2x+17” is 10 more than “2x+7,” and that “4x+14” is twice “2x+7.”

Switching between the different conceptions comes to the fore when composing functions f (x) = x³ and g(x) = 2x+7 and computing the derivative of (f o g)(x) = f(g(x)) = (2x + 7)³ for instance. Using the chain rule, students first have to treat “2x+7” as an object in computing the derivative of g and then switch to the procedural mode to compute the derivative of 2x+7. And finally switch to a completely operational mode when simplifying 3(2x+7)² · 2 to 6(2x+7)² or to 24x² + 168x+294.

Conceptualizing functions and variables this way allows for looking at functions dynamically, investigating the role of parameters, and eventually constructing limits, integrals, and derivatives.

**Assessment**

We realize that advocating for the goals we just described will not have much impact, unless this plea is accompanied by suggestions for ways of assessing these goals in large-scale assessment. Building on the fractions and algebra studies and a repeat study on subtraction (Kraemer & Hop, 2012), however, we may argue that the method of analyzing patterns in the results on individual items on a Rasch scale has potential. It showed that patterns in item difficulty could be linked to mastery of more advanced conceptual mathematical goals (or lack thereof), which in the algebra study was backed up by a qualitative study on solution procedures of individual students (Van Stiphout et al., 2013). This suggests that perhaps test formats do not have to change immediately. What is important here is the shift from looking at individual items to analyzing patterns—and the first challenge will be to develop informative sets of test items that allow for analyses of patterns in item difficulty. Looking at the future, however, we join Daro, Mosher, and Corcoran (2011) in arguing that we will have to invest in developing new forms of assessments. A promising avenue might be in Lobato’s...
press) approach to developing “grounded conceptual learning goals,” which takes its starting point in psychological analyses of fine-grained empirical studies of student thinking.

**Conclusion and discussion**

The debate on the results of mathematics reform in The Netherlands created the need for more fine-grained information than national and international tests offered. In this article we discussed three independent PhD studies which each assembled such information by zooming in on a specific domain. An overview of the results of the three studies showed that there are indeed shortcomings in the students’ capabilities in the researched domains (respectively subtraction up to 100, fractions, and algebra), and that the RME approach was only partly worked out in the commonly used Dutch textbooks. The instructional sequences broke off before the intended long-term learning processes were completed. Instead, the attention was shifted toward procedures that generated answers for specific tasks. Since Dutch teachers follow their textbooks very contentiously (Olson et al., 2008) and typically do not transcend what textbooks offer (Gravemeijer et al., 1993), the result was that the students did not reach proficiency on the level of more advanced conceptual understandings. In this respect, the findings about the use of RME theory in Dutch textbooks may come as a warning that a successful implementation of reform mathematics depends on a wide endorsement of more advanced mathematical understandings as curriculum goals. We have to caution, however, that this is not just an issue of teachers and textbook authors adapting reform ideas. Mathematics education research also harbors similar tendencies, because of the tests that are being used, and as a consequence of the necessary limited timespan of research projects, which may cause a lack of attention to long term learning processes.

The events in The Netherlands might be exemplary for reform efforts that are hampered by elaborations in textbooks and teaching approaches, which are not consistent with the intentions of the innovators. Such curriculum innovation problems are not unique for The Netherlands. Silver (2009), for example, observed that the TIMSS 1999 video study revealed that U.S. teachers rarely maintained the cognitive demands of tasks during instruction. In his view, this may explain why students in the United States proved to be weak in reasoning and problem solving, even though the intended mathematics curriculum aimed at problem solving and reasoning. Hodgen, Küchemann, Brown, and Coe (2009) also pointed to teachers and textbooks to explain why rises in examination results in mathematics in England—due to large-scale national curriculum initiatives—appear more a result of “teaching to the test,” than of an increase in genuine mathematical understanding. In Denmark Niss and Højgaard (2011) identified a similar problem, which they denote with the term *syllabusitis* (adopted from Jensen, 1995). They argued that the tendency to identify the mastery of mathematics with proficiency in mathematical subject matter, may lead to a syllabus that mainly consists of a list of guiding tasks for the written exam. Such a syllabus fails to acknowledge important aspects, such as problem solving, reasoning, proving, and modelling. They started the KOM project (Niss & Højgaard, 2011) to try to offer an alternative way of describing curriculum goals in the form of a matrix that has subject matter topic on one axis and competencies—such as mathematical thinking, problem tackling, modelling, and so forth on the other axis. In this manner they try to integrate the “what” and the “how,” which in many curricula are described separately.

**Paradigm case**

We believe that our analysis of the evolution of the Dutch innovation can be interpreted as a paradigm case by those who want to analyze, plan, or enact mathematics education innovations. One might want to look for similar patterns and similar underlying mechanisms. In doing so one could use the concepts that were developed as tools for analyzing what transpired in The Netherlands, such as the notions of task propensity and more advanced conceptual mathematical goals.

We identified _task propensity_—the tendency to think of mathematics education in terms of individual tasks that have to be mastered by students—as a strong inhibitory factor. The notion of
task propensity touches on many observations in the literature that teachers and students are preoccupied with answers on individual tasks. Often, however, this tendency is linked to curriculum goals and tests (Hodgen et al., 2009; Niss and Højgaard, 2011). We do this too, but we go one level deeper by linking this phenomenon to views on learning, following Resnick and Hall (1998) who pointed to the problem of a rather popular associationist view of learning. This analysis offers new leverage points for changing the attitude of teachers—and by extension of students—by supporting teachers in coming to grips with the conception of learning that underlies the intended mathematics curriculum. And by linking this conception of how students learn mathematics with more advanced conceptual mathematical goals, and how these can be reached.\textsuperscript{13}

The approach that Niss cum suis (Niss & Højgaard, 2011) took to counter the one-sidedness of curriculum documents that mainly consists of content goals is typical for many similar efforts. Our take on this problem, however, is different. We observed that the instructional sequences broke off too early, which led us to come up with the notion of more advanced conceptual mathematical goals in an effort to describe what was missing in those instructional sequences. Moreover, we elaborated those goals by building on the theoretical framework that underpins the mathematics education reform under consideration. Those goals differ fundamentally from either content goals or process goals, because they are grounded in the idea that constructing or reconstructing mathematics entails an ongoing sequence of processes being turned into objects, and objects becoming subject to new processes. And we may argue that the activities of problem solving, reasoning, proving and modeling, which are the motor of those construction processes, largely coincide with the process goals that are stated in various curriculum documents. Therewith it is not said that it cannot be very useful to identify and explicate process goals. It only explains why we did not feel the need to elaborate on process goals in our analysis.

Returning to the Dutch example we want to point to the fact that the more advanced conceptual mathematical understandings of the mathematical topics that were researched in the three studies were not formulated as educational goals in official curriculum documents in The Netherlands. Also the curriculum documents by the Freudenthal Institute, which laid the basis for the textbooks, showed a lack of attention for more advanced conceptual mathematical goals and a tendency to aim for algorithms as potential end points.\textsuperscript{14}

In contrast, we may observe a growing concern for more advanced conceptual mathematical goals in the United States, where the National Council of Teachers of Mathematics published the Curriculum Focal Points (National Council of Teachers of Mathematics, 2006) to address this issue. In addition, the Common Core State Standards (2010) draw much attention as an attempt to capture both more advanced mathematical goals and higher order thinking skills. In addition, efforts are being made to tie the Standards to what is known about learning trajectories, or learning progressions. Following Niss and Højgaard (2011), however, a weakness may be that the “how” and the “what” are described separately; the “mathematical practices” are separated from the content standards. This creates the danger that only the content goals will be tested and the more advanced goals will get lost in translation. As Daro and colleagues (2011) pointed out, the Standards use conventional names and phrases for topics in a subject, and these are likely to be interpreted in terms of conventional classroom practices. And, as we would put it, under the influence of task propensity, they will be translated into conventional ways of teaching, grading, and testing.

We therefore finish by stressing the need for a conscious effort to counteract task propensity, and to identify more advanced conceptual mathematical understandings as key curriculum goals. The more advanced conceptual mathematical understandings should not just be seen as long-term goals, but as integral parts of instructional sequences. They are a prerequisite for mathematical growth, a reason to transcend a short-term focus on individual tasks, and a means for achieving consistency between educational goals and the conception of learning that underlies reform mathematics curricula.
Notes

1. The so-called PPON survey of 2004 showed a significant decline in the mastery of whole number addition, subtraction, multiplication, and division. However, the results did not decline across the board. The survey also showed a slight progress in the area of number sense, estimation, and mental calculation, while the level of computing with fractions, ratios, and percentages was stable compared to 1998—after an earlier rise in the 1990s (Janssen et al., 2005).

2. Note that the TAL publications came too late to influence the textbooks we are reporting on here. For those textbooks the so-called specimens of Treffers, de Moor, and Streefland formed the main source of inspiration.

3. As part of the fifth PPON survey the study was repeated with similar results, except for a bigger difference between tasks with and without bridging ten (Kraemer & Hop, 2012).

4. In the literature, such differences in solution strategies are associated with the long-term development of mathematical proficiency (Kilpatrick, Swafford, & Findel, 2001), and more specifically with the flexible/adaptive use of the inverse relationship between addition and subtraction (Verschaffel, Bryant, & Torbeyns, 2012). Recent theoretical insights and empirical findings focus on the complexity, ambiguity, and longitudinal character of the constitution and use of this relationship over time. Greer (2012) and Vergnaud (2012) observed that students first have to generalize the use of indirect addition over different contexts and classes of situations (combine; change; compare; equalize). Next, they have to extend this relationship to decontextualized operations (such as subtractions with bare numbers).

5. We equate mastery with a chance of 80% or more to answer an item correctly according to the Rasch analysis.

6. The term symbol sense is used in the literature to describe strategic work, global focus, and algebraic reasoning (Drijvers, 2010). In relation to this, Arcavi (1994) spoke of behaviors related to skills that exceed basic manipulations, such as seeing the communicability and the power of symbols, which encompass flexible manipulation skills, such as the ability to cleverly select and use a symbolic representation, and to manipulate or to read through symbolic expressions depending on the problem at hand.

7. Structure sense denotes a collection of abilities, such as recognizing a structure, seeing part of an expression as a unit; dividing an expression into meaningful subexpressions; recognizing which manipulation is possible and useful to perform; and choosing appropriate manipulations that make the best use of the structure (Novotná & Hoch, 2008).

8. Whether the performance was better than some decades ago is a topic of heated discussion as the results on the national PPON surveys varies for distinct topics.

9. Mark that the interpretation of the multiplier is not universal. In The Netherlands $5 \times 3$ is interpreted as $3 + 3 + 3 + 3 + 3$, $5$ thus being the multiplier. In various other countries this is the other way around and is interpreted as $5 + 5 + 5$ with $3$ being the multiplier.

10. An additional factor that will have been of influence is that the current curriculum goals for primary school did not include multiplication of decontextualized fractions. The shift from contextual problems to bare problems was to be made in secondary school. However, primary school textbooks did not prepare for this shift, and the secondary textbooks did ignore that students were not ready for a formal approach.

11. Mark that even though Van Hiele (cited in Tall, 2004) questioned the applicability of his levels to the domains of arithmetic and algebra, he did work out his theory for number in his 1973 publication, which was very influential in The Netherlands. It is this broader interpretation that we follow here.


13. We realize of course that teachers understanding the conception of learning that underlies the curriculum, and adopting the more advanced conceptual mathematical goals alone will not be sufficient to make the innovation into a success. Reform mathematics requires the reorganization of a complex practice, as the envisioned instruction differs significantly from current practices (Cobb, Jackson, & Dunlap, 2014).

14. We may observe, however, that there are two new developments, one that concerns arithmetic, which is explicitly construed as an independent topic—different from mathematics (sec), and one that concerns “thinking activities” in the most recent curriculum documents on secondary mathematics in The Netherlands. In the former, the task propensity is very visible in the official curriculum documents and national tests (Meijerink, 2008). In respect to mathematics, so-called thinking activities are proposed (Commissie Toekomst WiskundeOnderwijs, 2007), and the national exam committee, CvE, is piloting tests to assess those abilities. We may note, however, that those thinking activities primarily refer to higher-order thinking skills, not to more advanced conceptual mathematical understandings as such.

Acknowledgments

We want to thank the reviewers for their helpful comments.
References


