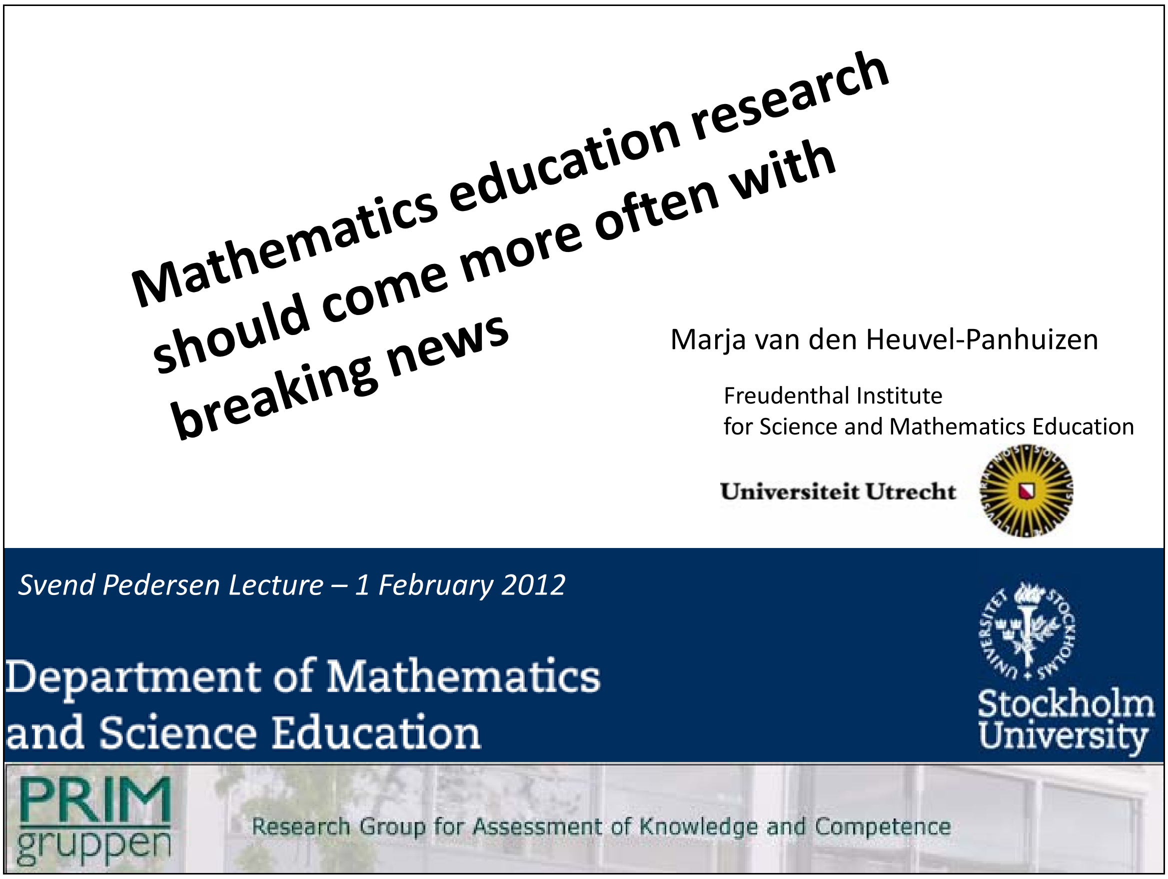


Lecture given on the occasion of receiving the *Svend Pedersen Lecture Award 2011*

MATHEMATICS EDUCATION RESEARCH SHOULD COME MORE OFTEN WITH BREAKING NEWS

Marja van den Heuvel-Panhuizen
Freudenthal Institute for Science and Mathematics Education, Utrecht University

I feel it as a very great honor to be selected for the Svend *Pedersen Lecture Award* and I would like to thank the Department of Mathematics and Science Education of Stockholm University for showing interest in my work and evaluating it in such a positive way that it is worthy of receiving this award named after Svend Pedersen who is so important for you in Sweden in the field of science education research – and I assume this also includes mathematics education research. In particular, I appreciate Svend Pedersen's effort to incorporate results from education research into the curricula of teacher education and give teachers and teacher educators access to results from education research.

1. SOME REFLECTIONS ON THE WORK OF SVEND PEDERSEN AS AN INTRODUCTION

When I recently read a number of articles coauthored by Svend Pedersen I actually experienced what this "given access to research findings" really means. I was truly moved by the clearness and cogency of these writings. These are the kinds of articles that the education field needs to make informed educational decisions. The findings reported in these articles have direct consequences for teaching. When you as a teacher are reading these articles, you immediately know what to do in your classroom the next day.

For example, in an article written by Molander, Pedersen, and Norell (2001) about deaf pupils' reasoning on scientific phenomena it was unmistakably revealed why deaf children may have difficulties in learning science in school. They do not have experiences like hearing children have with the popularized version of science through media, such as science programs on television. Through these media, although they may be short of scientific depth, hearing children have already encountered scientific terminology and explanations outside the school context; this, among other things, can generate children's interest in "how things work". This opportunity that hearing children have is lacking for deaf children, whose initial encounters with science will mostly take place in school, and from a stricter scientific perspective. Therefore, the authors of the article emphasize that instruction for deaf and hard-of-hearing students must bring informal science into the school context and must offer students opportunities to connect scientific reasoning to pre-instructional reasoning.

In another article, Pedersen and his coauthors Molander and Halldén (Molander, Halldén, & Pedersen, 2001) proved convincingly how important it is to take into account the setting in which students have to explain and interpret particular phenomena when investigating students' understanding of scientific concepts. Asking students to describe what happens when a glass is turned upside down into water, leads to different levels of reasoning depending on the setting in which this question is asked and the equipment that is used for presenting the question. Therefore, when teachers – or others – are designing tasks for science lessons they should be very conscious of whether or not to use laboratory equipment. As Pedersen and his colleagues have shown, specialized materials from the scientific domain which are common in scientific laboratory do not function as appropriate cues to elicit a student's reasoning as familiar materials from the life-world domain would.

My first thought after reading these articles was that the conclusions and recommendations of these studies sound very familiar to me. Reading these articles felt like coming home. In the Netherlands, in the early 1970s, Freudenthal and his colleagues from the Wiskobas group – of which Adri Treffers was a very prominent member – started to revise the then prevailing Dutch mathematics education which had a very mechanistic character and in which teaching directly began at a formal, symbolic level. Freudenthal and the Wiskobas group developed a new approach to mathematics education in which the idea of "mathematics in context" played a crucial role. This approach was later called "Realistic Mathematics Education" (RME). One of its characteristics is that children are offered rich contextual problems in which they can develop mathematical concepts and procedures for carrying out mathematical operations.

My second thought after reading the articles in which Svend Pedersen was involved, was: (1) Do Swedisch teachers know – and do their colleagues in other countries know – that teaching science to deaf children should start with giving them access to the popularized version of science, and keeping the textbooks with the scientific perspective closed for some time? (2) And do teachers know that laboratory equipment can hinder children's first scientific thinking, and that it is better to use equipment that is familiar to them; for example, that it is better to use a drinking glass than a test tube?

In fact, these questions brought me to the somewhat provocative title of my lecture. Findings of education research – at least when they are worthwhile enough – should be communicated more explicitly to teachers. And why not as breaking news? I think it would be in the spirit of Svend Pedersen. We have to ask ourselves: Are teachers really aware of the importance of a popular, informal science basis and the use of ordinary objects from daily life when we want students to bring to scientific thinking? And if not, who will tell them these things? Of course, when teachers attend in-service courses and read professional journals they can learn about them. But this is not enough. A broad societal basis is needed to support the implementation of these new findings in school. This means that parents and education policymakers should also know about these findings. However, I am afraid they have never heard of them. And this is not a surprise, because findings of education research only rarely appear in newspapers or in the news on radio and television. In other disciplines, however, publication of findings is very common. See, for example, what I found in Dutch newspapers and later could trace on the internet:

(1) A team of researchers at the University of Sussex found that, in colonies of Jatai bees, some are born soldiers. Until now such a caste system was unknown in bees.

(2) A team of researchers from Louisiana State University found the world's tiniest vertebrate in the New Guinea rainforest. It is a frog with an average adult size of 7.7 mm in length, less than half the diameter of a U.S. dime.

(3) The Scientific American's rubric 'Image of the Week' recently included a topographic map of the Italian Dolomites together with a stratigraphic column showing layers of a cave. According to the author of this Image of the Week message, this work, dating from 1739, is maybe the oldest figure of this kind. While looking at this drawing, we may think, why is there never a drawing by a student showing a clever solution as the image of the week?

(4) Finally, the exception of the rule. Sometimes, mathematics education is also breaking news, as was the case with the research of psychologist Susan Goldin-Meadow at the University of Chicago. She found that gesturing helps students develop new ways of understanding mathematics.

Although announcing results from research in the public domain is very common in all kinds of disciplines, it is not in ours: the didactics of mathematics – or taking it more broadly

– the didactics of science. And why this is not the case? Do we not have news that deserves the label breaking news? I think we have. For some teachers – and even more for parents, school administrators, members of parliament and so on – it will be a paradigm shift to introduce children in school to science by abandoning the science equipment and using ordinary objects from daily life to bring students to scientific thinking. Breaking news with new findings from education research presented in a concise and clear way with convincing examples and data that prove the findings can inform a professional audience as well as the general public. In the Netherlands, we have lately learned how crucial it is to keep the public thinking about education aligned with the professional thinking. However, that is another topic that occupies my mind, but that I will not address now (more about this topic can be found in Van den Heuvel-Panhuizen, 2010).

For this lecture I take the opportunity – in the spirit of Svend Pedersen – to share with you some research findings which I think everybody involved in education, and in particular in primary school mathematics education, should know. Both studies have been carried out at the Freudenthal Institute of Utrecht University.

The first finding is about reading picturebooks to kindergartners (4- to 6-year-olds) to support their learning of mathematics. The study is part of the picture book project, called *PICO* project (which stands for PIcture books and COncept development MAthematics), in which I collaborate with Alexander Robitzsch, a colleague whom I know from my work at IQB in Berlin and who is now in Austria at the Bundesinstitut für Bildungsforschung, Innovation & Entwicklung. The second person who is very important for this project is Iliada Elia, who is a lecturer at Cyprus University and who is working with me in several early childhood studies. Furthermore, two PhD-students have been involved in the PICO project.

The second finding that I will share with you is from the *IMPULSE* project (which stands for Inquiring Mathematical Power and Unexploited Learning of Special Education Students). In this project I work with my PhD-student Marjolijn Peltenburg and again with Alexander Robitzsch. The study I will present here is about whether or not students in special education who are weak in mathematics can apply alternative methods for solving subtraction problems.

2. USING PICTUREBOOKS TO SUPPORT KINDERGARTNERS' LEARNING OF MATHEMATICS

The picturebook study has shown that reading picturebooks to kindergarten classes:

- has educational potential for kindergartners' learning of mathematics
- is an effective whole-class activity
- is effective for children with different background characteristics
- is most gainful for the mathematical development of girls.

I think these are reasons enough for a news alert.

Before providing you with evidence for these findings I will give you some background information about using picturebooks for teaching mathematics in kindergarten classes.

2.1 Using picturebooks for teaching mathematics – some background information

For many children between the ages of four and six kindergarten is the first institutional educational setting in which they come across school subjects, including mathematics. The teaching of mathematics to children of such a young age already has a long history. In a way, we can say that it dates from 1631 when Comenius published his book *School of Infancy* in which he emphasized the importance for young children to observe and manipulate with objects, and in which he stimulated the creation of mathematics programs for young children. He even used a picturebook, called *Orbis Pictus* for assisting children to make impressions in the mind. Here a page of this book is shown which explains several forms of wind: (1) a cool breath of fresh air; (2) a stronger wind; (3) a storm that throws down trees; (4) a whirlwind that rotates; (5) an underground wind that causes an earthquake; (6) an earthquake that causes gaps in the earth and the falling down of houses.

Returning to our days, we can say that from the 1990s on linking mathematics instruction to children's literature has become increasingly popular for a variety of reasons. For example, it was recognized that children's literature can motivate students, connect mathematics to emotions, and provoke interest. The standards published by the National Council of Teachers of Mathematics (NCTM, 2000) gave a strong boost to the use of picturebooks in kindergarten by recommending the use of children's story books as an approach for introducing mathematical ideas.

These recommendations also received empirical support. Several studies investigated the effect of reading books to young children on their learning of mathematics. However, in most of these studies the book reading sessions in class were followed by other activities such as playing with story-related (mathematical) materials (Hong, 1996; Jennings, Jennings, Richey, & Dixon-Krauss, 1992; Young-Loveridge, 2004), singing mathematical rhymes (Young-Loveridge, 2004) or composing geometrical puzzles (Casey, Erkut, Ceder, & Mercer Young, 2008).

Most of these studies show positive effects of the combination of book reading and book-related activities on children's mathematical development. In particular, the studies found a positive effect on kindergartners' mathematics achievement (Jennings et al., 1992; Young-Loveridge, 2004), their attitude towards mathematics (Hong, 1996; Jennings et al., 1992), and their use of mathematical vocabulary (Jennings et al., 1992).

2.2 The setup of our picturebook study

Although the previously mentioned effect studies indicated that using children's literature might be a promising avenue in mathematics education, most of these studies just used the picturebooks as an introduction to a mathematics lesson or activity and did not inform us about the power of the picture book itself to support the kindergartners' understanding of mathematics. Thus, we decided to do a study on the effect of the book reading itself, i.e. without inclusion of additional (book-related) mathematical activities.

Our main research questions were:

(1) Can an intervention involving picturebook reading enhance kindergartners' mathematical understanding?

(2) What is the relationship between the intervention effect and characteristics of kindergartners, including kindergarten year, age, gender, mathematics and language ability, home language, and socio-economic status?

2.2.1 Research design

To answer the research questions a classroom experiment was carried out with a pretest-posttest-control-group design and a picturebook reading program as an intervention. In the control group the teachers followed the regular program for mathematics and picturebook reading. We came to the sample by first looking for nine pairs of schools in the Utrecht area which were similar with respect to urbanization level, school size and the socio-economic status (SES) of the school population. Then, of each school one kindergarten class was randomly assigned to the experimental or the control group. This resulted in nine experimental and nine control kindergarten classes. The classes were all mixed classes consisting of children who were in their first year of kindergarten (K1) (4- to 5-year-olds) and children in their second year of kindergarten (K2) (5- to 6-year-olds).

2.2.2 Participants

In total there were 384 children involved in the study, of which 199 children in the experimental group (106 boys and 93 girls). Of that experimental group 84 were in K1 and 115 in K2. The average age at the time of the pretest was 5 years and 3 months. Of the 199 children, 171 had Dutch as their home language, and 176 children had an average or higher SES.

The control group consisted of 185 children (95 boys and 90 girls), of whom 66 were in K1 and 119 in K2. The average age at the time of the pretest was 5 years and 4 months. Of the 185 children, 164 children had Dutch as their home language and 161 children had an average or higher SES.

2.2.3 Assessment instruments

The mathematics performance of the children was measured by a project test and a standardized test. The project test, called the PICO test, contains 42 multiple-choice items split up over two booklets. Every item covers one page and contains a large illustration depicting the situation that the question is about and four small illustrations that represent the possible answers; for example, an item shows a front view of a sitting mouse; the teacher asks: “How would Mouse look if you looked down on him like a bird?” After a test item is read aloud to the children, they can answer it by underlining the correct answer.

The standardized test used to measure children’s general mathematical ability was the so-called Cito test ‘Ordering’. The Cito test ‘Language for Kindergartners’ was used to measure the kindergartners’ language skills.

2.2.4 Picturebook program

The books used in the intervention are picturebooks, by which we mean books in which the illustrations are essential for telling the story, although they may contain text as well (Arizpe & Styles, 2003; Nicolajeva & Scott, 2000). All books are trade books of high literary quality which were not purposely written for teaching mathematics, but which even so have mathematics-related content and characteristics that may support the learning of mathematics. These learning-supportive characteristics were earlier identified by a literature review and an expert consultation. A paper about this sub-study is in press (see Van den Heuvel-Panhuizen & Elia, in press). To cover a rich variety of mathematical domains, we chose picturebooks dealing with number, measurement, and geometry. In total, the picturebook reading program consisted of 24 picturebooks which were read aloud in class by the teachers over twelve consecutive weeks (2 books every week).

To inform the teachers on how to read the picturebooks, we developed a reading key for every book, describing for each book page how it should be read. The reading guidelines are suggesting teacher behavior like (1) asking oneself a question out loud about the mathematics, (2) playing dumb, and (3) showing an inquiring expression. In general, the reading keys requested the teachers to maintain a reserved attitude and not to take each aspect of the story as a starting point for a class discussion, since lengthy or frequent intermissions could break the flow of being in the story and consequently diminish the story's own power to contribute to the mathematical development of the children. The leading principle in the development of the reading keys was to use the book's potential and "let the book do the work".

2.3 Some classroom impressions

Now I will show you some snapshots from the book-reading sessions which were taken from the pilot phase of the project.

2.3.1 Asking oneself a question

The picture book *22 Wezen* [22 Orphans] is about twenty-two parentless children who live in an orphanage. On pages 5 and 6 (see Figure 1) the stern lady principal takes the children to bed. In the dormitory, eight double-decker beds are visible. The lady principal's huge body blocks the view on the other beds.

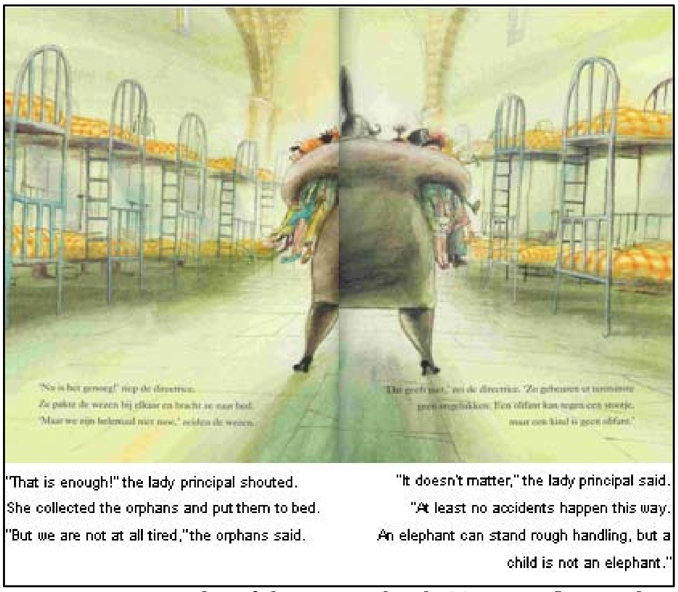


Figure 1. Pages 5 and 6 of the picture book *22 Wezen* [22 Orphans].

While looking at this picture the teacher asked herself out loud: "Is there room for everybody?" Some children thought there would be enough sleeping-places; others rejected this, which made children count the beds. In doing so, different strategies were used, including counting the sleeping-places separately or taking the double-deckers as a unit. Some children made clever use of the structure of the double-deckers; they counted in twos and pointed to both decks at the same time with two fingers: "two, four, six …" and so on. Moreover, the children included the invisible beds as well.

2.3.2 Playing dumb

In the picture book *Sssst!* [Ssh!], the children sneak – so to speak – through a giant's castle, trying not to wake him up. On pages 3 and 4 (see Figure 2), the front of the castle is shown from a "frog's eye view". After reading aloud the text in the book: "Ssh. This is the castle of a giant!" the teacher pretended not to understand that the size of the towers is the result of the point of view that was taken. After she said "Quite small towers for a giant", some children reacted that the towers seem smaller than they really are because of the low position from which the castle is seen. One child said: "Maybe it is because it is very tall."

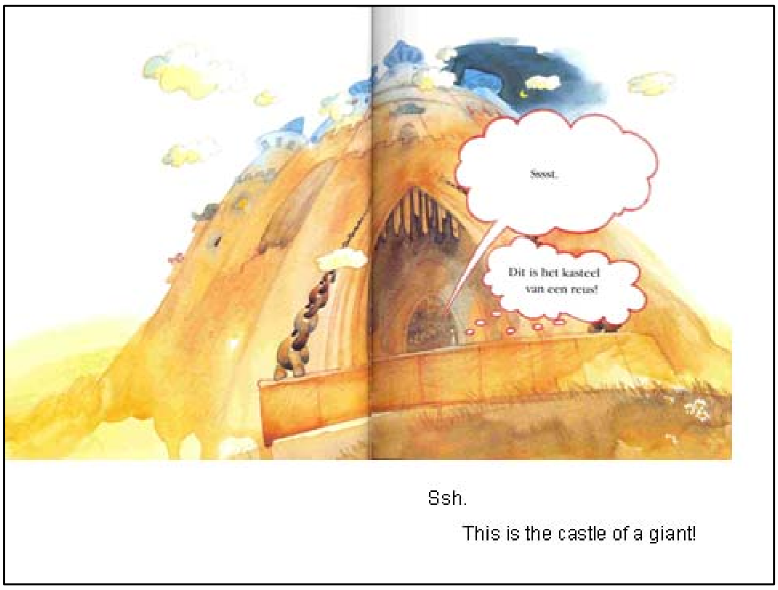


Figure 2. Pages 3 and 4 of the book *Sssst!* [Ssh!].

2.3.3 Showing an inquiring expression

The next picturebook is called *De prinses met de lange haren* [The princess with the long hair]. On the endpapers of the picture book the princess is depicted with a ponytail that is several times the length of the princess (see Figure 3).

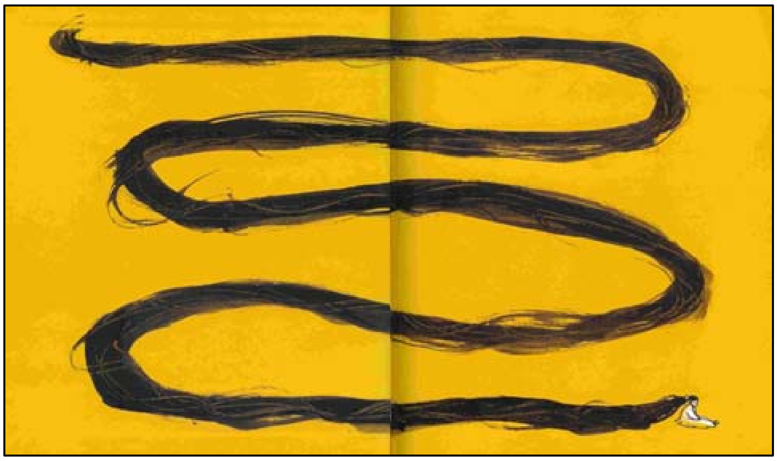


Figure 3. Endpapers at the beginning of the book *De prinses met de lange haren* [The princess with the long hair].

The ponytail is draped in a zigzag covering two pages. The teacher just looked amazed and waited for the children's reaction. The children were impressed by the enormous length of the hair and followed the zigzag with their index finger as if to experience how long the hair is. Later on, when the teacher asked how long her hair is in reality, the children ran around in the gym.

2.4. Our findings

Before we started to analyze the data for answering the research questions, we investigated the psychometric properties of the PICO test, the amount of missing data, and the comparability of the experimental and control group. I will leave out this technical part here –

recently, we have submitted a full article about this study to a journal (Van den Heuvel-Panhuizen, Robitzsch, & Elia, submitted) in which this technical part is described in detail – and will continue now with a short report about what we found regarding the effect of the book reading sessions.

Table 1
Description of the three regression models set up for analyzing the intervention effect

Investigated effect	Model 1	Model 2	Model 3
General intervention effect (effect in whole sample)	*Dependent variable*: PICO posttest *Covariate*: PICO pretest	Model 1 + *Covariates*: Cito Mathematics Cito Language	Model 2 + *Covariates*: Age Grade level Gender SES weight Home language Urbanization level
Specific intervention effect (effect in subgroups)	*Dependent variable*: PICO posttest *Covariate* ICO pretest	Model 1 + *Covariates*: Cito Mathematics Cito Language	Model 2 + *Covariates*: Age Grade level Gender SES weight Home language Urbanization level

In total, we did two analyses (see Table 1): one for the effect of the picturebook reading program in the whole sample and one in specific subgroups, including K1 and K2 children, girls and boys, children with as their home language Dutch and non-Dutch, children with a higher and a lower SES, older and younger children, children with a high and a low PICO pretest score, children with a high and a low general mathematical ability and children with a high and a low language ability. Within each analysis we investigated three different models: one with only the PICO pretest as a covariate (Model 1), one with the PICO pretest, Cito Mathematics and Cito language as covariates (Model 2), and one with all the nine covariates. In all three models, the PICO posttest score is a dependent variable.

The three regression models give similar results for the general intervention effect on the PICO posttest scores (see Table 2). Model 1 reveals a significant intervention effect, while Model 2, controlling for two extra covariates (Cito language and mathematics), shows a smaller but still nearly significant intervention effect. Here, I have to add that in the table the p-values for two-sided hypotheses are displayed. However, we assumed a positive effect of the intervention and formulated a one-sided hypothesis. This means that the criterion for significance can be set on .10. Finally, Model 3, which includes all nine covariates, indicates a significant intervention effect.

In order to investigate the size of the general intervention effect, we calculated the Cohen's d for each of the three models. For Model 1 we found an effect size $d = .15$, for Model 2 the effect size was $d = .11$ and for Model 3 it was $d = .12$.

At first sight, this effect size is perhaps not that large, but if we look at the mean increase from the PICO pretest to the PICO posttest in the control group, which has an effect size of $d = .59$, then the intervention effect is worth speaking of, because the influence of the intervention amounts to 25% (.15/.59 = .25) of the "regular" effect size in the control group. In Model 2 and Model 3, the increase in effect size was respectively 19% and 20%.

Table 2
Results of the regression analyses predicting PICO posttest scores for the whole sample.

	Model 1				Model 2				Model 3			
	B[a]	*SE*	*β*	*p*[b]	*B*	*SE*	*β*	*p*	*B*	*SE*	*β*	*p*
Intervention	.89	.36	.07	.01	.64	.39	.05	.11	.71	.39	.06	.07
PICO Pretest	.89	.03	.84	.00	.72	.05	.67	.00	.68	.05	.64	.00
Cito Mathematics					.09	.02	.20	.00	.08	.02	.18	.00
Cito Language					.01	.03	.02	.72	.01	.03	.01	.77
Grade level									.54	.66	.04	.42
Age									.04	.04	.05	.33
Gender									-.04	.35	.00	.90
SES Weight									.18	.78	.01	.82
Dutch Home Language									1.38	.61	.07	.02
Urbanization Level									.31	.24	.04	.20
R^2 (Explained Variance)	.70				.72				.73			
Cohen's *d*	.15				.11				.12			

[a] *B*: unstandardized regression coefficient of the intervention effect; *SE*: standard error of *B*; *β*: standardized regression coefficient.
[b] The *p*-values for two-sided hypotheses are displayed. This means that assuming a positive effect of the intervention as formulated in our one-sided Hypothesis 1, the significance level can be set on .10.

When looking at the effect of the intervention for the *specific groups of children* we found that the intervention effects for the various subgroups vary (see Table 3).

Table 3
Results of the regression analyses predicting PICO posttest scores for the whole sample.

Specific intervention effect	Model 1	Model 2	Model 3
Grade level (K1 / K2)	+ K2 ~ K1	n.s.	n.s.
Gender (boy / girls)	+ girls	+ girls	+ girls
PICO pretest (low-third / mid-third / high-third)	+ mid-third	+ mid-third	+ mid-third
Cito Language (low-third / mid-third / high-third)	+ low-third + high third	+ low-third + high third	+ low-third
Cito Mathematics (low-third / mid-third / high-third)	+ low-third	n.s.	n.s.
Home language (Dutch / non Dutch)	+ Dutch	+ non-Dutch	+ non-Dutch
Age (younger / average / older)	+ older	n.s.	n.s.
Socio-economic status (higher / lower)	+ higher	n.s.	n.s.

Regarding *grade level*, the results from Model 1 indicate a significant effect for K2 and a nearly-significant intervention effect for K1. However, the intervention effect becomes insignificant when controlling for extra covariates.

With respect to *gender*, all three models show a significant and relevant intervention effect for girls and not for boys.

For the children who had *a moderate PICO pretest score* the intervention effect was significant and rather large in all three models. However, we did not find this for the low- and high-achieving children in the PICO pretest.

With respect to the children's *language ability*, we found that the children who scored low in the Cito language test benefited considerably from the picturebook reading program in all three models. For the high-scoring children the intervention was significant in Model 1 and Model 2.

The split with respect to general mathematical ability showed that the children who scored in the low-third range of the Cito mathematics test showed a significant effect of the intervention, though controlling for the covariates in Model 2 and Model 3 led to non-significant effect sizes across the three subgroups.

With respect to the *Dutch home language* subgroup we found a significant intervention effect in Model 1, whereas when controlling for covariates we found a significant and rather large intervention effect in the subgroup of non-Dutch home language.

Regarding the subgroups of *age* we found in Model 1 a significant effect for the older children, but this disappeared in Model 2 and Model 3.

The same was the case for the children with the higher *socio-economic background*. In Model 1 there was a significant effect, which disappeared in Model 2 and Model 3.

2.5 Summary of the findings

Let me summarize our findings again:

2.5.1 Reading picturebooks has educational potential for kindergartners' learning of mathematics

Our study showed that a three-month picturebook reading program during which the teacher reads two picturebooks that contain mathematics-related content in class weekly, can have a positive effect on kindergartners' mathematical understanding. In fact, an increase in 25% more effect size between pretest and posttest in the experimental group than in the control group, as a result from a three-months program is quite a lot. And this is certainly the case when taking into account the spurt in cognitive growth children generally make at this age (Bowman, Donovan, & Burns, 2000), which can also be seen in the increase in performance of the children in the control group who received only the regular teaching.

2.5.2 Reading picturebooks is an effective whole-class activity

Although we found an intervention effect for the older children and K2 children, which might indicate that these children are more prepared to pick up the mathematical aspects in the picturebooks, this was only the case for Model 1. When we took into account other covariates, the intervention effect for age and grade level disappeared. In fact, our findings seem to suggest that picturebook reading sessions for supporting mathematical development can be effectively carried out as a whole-class activity in combined K1/K2 classes.

2.5.3 Reading picturebooks is effective for children with different background characteristics

We also found that reading picturebooks is effective regardless the children's general mathematical ability. Moreover, it turned out to be effective both for children with a high language score and a low language score. For home language the findings were ambivalent. Without controlling for covariates the children with Dutch as their home language benefitted more than those who do not speak Dutch at home, whereas with controlling for covariates, the non-Dutch home language group gained more from the program. With respect to socio-economic status an intervention effect was found only for the children with a higher socio-economic status score when there was no controlling for the covariates.

In sum, these findings indicate that reading picturebooks to kindergartners seems *not* to be effective for only particular children, such as children with high language ability or high mathematics ability, but for a broad range of children.

2.5.4 Reading picturebooks is most gainful for the mathematical development of girls

Only for one background characteristic of the children we found that the picturebook reading was not equally beneficial: the picturebook program was clearly advantageous to girls even when taking into account other characteristics of the children (see Model 2 and Model 3). Given other research findings that have revealed that girls in grade 1 sometimes already lag behind the mathematical development of boys (Carr & Davis, 2001; Penner & Paret, 2008), it is a good thing to know that reading picturebooks may contribute to giving girls a better start in mathematics.

So far what I liked to tell you about the picturebook study. To follow now is what we found in a study into the mathematical abilities of special education students.

3. SPECIAL EDUCATION STUDENTS' ABILITY TO APPLY ALTERNATIVE METHODS FOR SOLVING SUBTRACTION PROBLEMS

As I said earlier, this study I carried out together with my PhD student Marjolijn Peltenburg and with Alexander Robitzsch, who is giving us support for doing the statistical analysis of the data. The starting point of the *IMPULSE* project was that in the Netherlands – but also in other countries – there is a strong belief in circles of special education (SE) educators and psychologists that students who are weak in mathematics cannot handle different calculation methods. The idea is that it better to teach them only one fixed method for each number operation because otherwise they get confused.

3.1 Background of the study

3.1.1 RME not suitable for SE students?

Since the development of "Realistic Mathematics Education" SE educators and psychologists have been reserved towards implementing this approach to mathematics education in SE. In fact, the principles of RME are considered to be too difficult for students in SE. First of all, presenting *problems in context*, which, by the way, was often incorrectly operationalized as presenting wordy word problems, is not acknowledged as being of help. Using contexts is judged as too complex for these students, and therefore educators tell teachers in SE to focus on bare number problems. Another principle of RME is to build from *children's informal knowledge*. Again, this approach is not seen as very beneficial, because of the assumption that SE students are not able to, or at least have difficulties to, generate informal solutions on their own; see, for example, the warning in a publication by Gelderblom (2008, p. 36) that "letting students who are weak in mathematics to discover strategies themselves is fatal"). A third principle of RME that is connected to the two foregoing is that of *progressive schematization*, which SE educators and psychologists often see as a detour, because they think it is better to teach children directly at a formal level. A fourth principle of RME is to make students' aware of *choosing an appropriate solution method*. This approach is rejected because – as I said already – varying a solution method is considered to be confusing for SE students. Finally, RME stimulates *reflection on solution methods* and having class discussions about different ways of solving problems to deepen the children's understanding. Again, this fifth principle of RME is also not regarded as helpful for weak learners in mathematics.

3.1.2 Teaching only one fixed method?

The focus in the study presented here was on whether to teach SE students only one fixed method or not. To avoid difficulties for SE students the general advice of SE educators given to SE teachers is to keep it simple and teach weak students in mathematics only one fixed method. For example, in a summary of Timmermans' (2005) thesis (see http://www.nwo.nl/nwohome.nsf/pages/NWOP_6HKFLE), it is stated: "Weak performing students in arithmetic […] attain better results with the traditional approach in which they learn to solve number problems in one particular way" (translated from Dutch). Similar advice is given to SE teachers based on Milo's (2003) thesis (see http://www.nwo.nl/nwohome.nsf/pages/NWOP_5LEJNJ): "[P]upils at special schools for primary education can best learn arithmetic using one specific strategy." In addition, similar voices are also heard outside the Netherlands. For example, the U.S. National Mathematics Advisory Panel (2008) also suggests, that students who are weak in mathematics would benefit from being taught one prescribed way of solving calculations.

However, one may wonder whether this is good advice, because the idea of teaching only one method goes against the goal of developing numeracy in students. Being numerate implies that students should be able to choose a suitable method when solving number problems (Treffers, 1989; Van den Heuvel-Panhuizen, 2001; Warry, Galbraith, Carss, Grice, & Endean, 1992). A further objection against the one-method approach is that if students have to restrict themselves to only one way of solving problems, many problems would require an unnecessarily long solution path (see, e.g., Torbeyns, De Smedt, Stassens, Ghesquière, & Verschaffel, 2009). Moreover, if students have to use prescribed methods, this can lead to "didactical ballast" (Van den Heuvel-Panhuizen, 1986) for them. It can be a source of error for students because such prescribed methods are not grounded in the students' own thinking, i.e., the ownership is completely on the side of the teacher or textbook author.

Based on the foregoing arguments, one might decide to teach even students who are weak in mathematics the flexible use of solution methods. However, the critical point for making this decision is whether SE students are able to operate in such a flexible way. Therefore, we decided to investigate this. Our focus is on subtraction problems up to 100, because these problems are found to be very difficult for SE students (Kraemer, Van der Schoot, & Van Rijn, 2009).

3.1.3 Strategies and procedures for solving subtraction problems up to 100

Generally, three different types of strategies can be distinguished for solving subtraction problems with numbers up to 100: splitting, stringing, and varying (Van den Heuvel-Panhuizen, 2001). Although researchers do not always use the same wording – for example, other expressions can be found in Klein, Beishuizen, & Treffers (1998) – there is broad agreement about the general meaning of these strategies: in splitting, both numbers are decomposed in tens and ones; in stringing, one number is kept as a whole number; and in varying, one or both numbers are changed in order to get an easier problem.

Torbeyns, De Smedt, Stassens, Ghesquière, & Verschaffel (2009) describe subtraction in a different way. They distinguish (1) direct subtraction (DS), which means taking away the subtrahend from the minuend; (2) indirect addition (IA), which means adding on from the subtrahend until the minuend is reached; and finally a less common procedure called indirect subtraction (IS), which means taking away from the minuend until the subtrahend is reached. According to Torbeyns, De Smedt, Stassens, Ghesquière, & Verschaffel (2009), splitting, stringing, and varying belong to the class of DS procedures, whereas IA is considered as a separate class of procedures which do not fit to the three strategies.

However, we see this differently. Splitting, stringing, and varying can be considered strategies which all describe *how we deal with the numbers* involved: either splitting numbers, or keeping one number as a whole number, or changing one or both numbers. In contrast to these strategies, we can label DS, IA, IS as procedures which describe *how the operation is carried out*: (DS) taking away the subtrahend from the minuend, (IA) adding on from the subtrahend until the minuend is reached and (IS) decreasing the minuend until the subtrahend is reached, and finally (MO) using multiple operations (see Figure 4).

Although integrating the two existing ways to describe how to solve subtractions up to 100 was never the intention of this study, it emerged when thinking about how to code the students' solution methods.

		Strategies Number perspective		
		Splitting	Stringing	Varying
Procedures Operation perspective	DS Direct Subtraction	**63–31=** * 60–30=30 3– 1= 2 30+ 2=32	**63–47=** 63–40=23 23– 3=20 20– 4=16	
	IA Indirect Addition	**67–52=** ** 50+10=60 2+ 5= 7 10+ 5=15	**62–58=** 58+ 2=60 60+ 2=62 2+ 2= 4	
	IS Indirect Subtraction	**67–52=** ** 60–10=50 7– 5= 2 10+ 5=15	**62–58 =** 62– 2=60 60– 2=58 2+ 2= 4	
	MO Multiple operations			**77–29 =** 77–30=47 or 78–30=48 47+ 1=48

* The problem can be solved by the following calculation steps. The description of these steps does not necessarily reflect how the problems are or should be notated by students. The students' use of materials and models is left out as well.
** These calculation steps are not very common to solve this problem; they are only given to explain this particular combination of procedure and strategy.

Figure 4. Relation between procedures and strategies illustrated with problems

Suddenly, it was clear that the splitting, stringing, and varying strategies and the DS, IA, IS, and MO procedures complement each other. Together, they offer a complete framework for describing how students solve subtractions up to 100.

Thinking about a news alert, this finding might also deserve such a notification. Like with other remarkable findings, one may wonder why nobody else came up with this integration of methods before. It looks so obvious. Therefore it is no wonder that after the scheme was presented at a conference, it very quickly – although somewhat adapted – found its way in another publication; unfortunately without any reference.

3.1.4 Solving subtraction problems by indirect addition

Back to the focus of our study. Connected to the debate about whether or not teaching SE students one fixed method for solving number problems that was described earlier, there is also controversy on whether SE students should be taught IA to solve, for example, a problem like 62–58 by calculating 58+2=60, 60+2=62, so the answer is 4 or that they should be taught to calculate 62–50=12; 12–2=10 and finally 10–6=4. Existing research findings are not clear about it.

For example, a few recent intervention studies concluded that even students in regular primary education hardly ever use IA to solve subtraction problems (Torbeyns, De Smedt, Ghesquière, & Verschaffel, 2009; De Smedt, Torbeyns, Stassens, Ghesquière, & Verschaffel, 2010). However, these studies are challenged by other intervention studies that support the claim that already in the first grades of primary mathematics education, students with a wide range of mathematical abilities can learn to solve flexible subtraction problems by applying IA (Klein, Beishuizen, & Treffers, 1998; Menne, 2001).

3.1.5 Factors influencing students' procedure use

The next question is what factors may influence the use of IA.

Numbers involved

Several studies (e.g., Menne, 2001; Torbeyns, De Smedt, Stassens, Ghesquière, & Verschaffel, 2009) have indicated that subtraction problems that require crossing the ten and that have a small difference between the minuend and subtrahend, for example, 62–58 may evoke the use of IA. However, IA could also be an efficient procedure for solving large-difference subtraction problems with a relatively small difference around the tens and requiring crossing the ten (Torbeyns, De Smedt, Stassens, Ghesquière, & Verschaffel, 2009). For example, 82–29 may be easily solved by IA. Finally, research suggested that small-difference problems that do not require crossing the ten, for example, 47–43 may also evoke the use of IA (Gravemeijer et al., 1993).

Problem format

The problem format may also play a role. Several studies (e.g., De Smedt et al., 2010; Van den Heuvel-Panhuizen, 1996) revealed that bare number problems hardly evoke the use of IA. Context problems, on the contrary, have the possibility to open up both interpretations of subtraction, taking away and determining the difference (Van den Heuvel-Panhuizen, 2005; Van den Heuvel-Panhuizen &Treffers, 2009). The latter may be evoked by "adding on" contexts.

3.2 Set up of the study

3.2.1 Research questions

In our study we intended to answer the following research questions:

1. Can SE students make spontaneous use of IA for solving subtraction problems up to 100, and which conditions influence the use of IA?
2. Does the use of IA help SE students to solve subtraction problems up to 100 successfully, and under which conditions does IA use lead to successful problem solving?

3.2.2 Method

Participants

In total, 56 students from fourteen second-grade classes in three Dutch SE schools participated in the study. The participating students were 8-12 years old, with an average age of 10 years and 6 months (SD=10.4 months). This means that the students in our study were 1 to 4 years behind in mathematics compared to their peers in regular primary school.

The students' general mathematical ability level was established with the Cito Monitoring Test for Mathematics End Grade 2 (Janssen, Scheltens & Kraemer, 2005). Our students had score that ranged from 32 to 56 with an average of 47.8 (*SD*=6.8) while their second-grade peers in regular primary education have an average ability score of 56.4 (*SD*=14.6).

Instruments

To assess the students' ability in solving subtraction problems we developed an ICT-based *Subtraction test*, consisting of fifteen items. Furthermore, we developed an *online teacher questionnaire* with which we collected data about the students' prior instruction on subtraction problems.

In the ICT-based Subtraction test, the item characteristics were varied systematically over the fifteen items. These characteristics include number characteristics and format characteristics (see Table 4). The number characteristics refer to the size of the difference between the minuend and subtrahend, whether the tens have to be crossed and whether or not the minuend and the subtrahend are close to a ten. The format characteristics refer to whether or not the items are presented as a bare number problem (BN) or as a context problem. The latter can describe a taking-away situation (ConTA) or an adding-on situation (ConAO). The following item is an example of a ConAO item (Figure 5).

Table 4
Different types of subtraction items in ICT-assessment

	Number characteristics				Format characteristics			
	Difference	Crossing the ten	Both numbers close to ten	Example	Bare number	Context		Total
						Taking away	Adding on	
					Number of items			
A	Small (<7)	No	No	56-52	1	1	1	3
B	Small (<7)	Yes	Yes	31-29	1	1	1	3
C	Large (>11)	Yes	Yes	51-39	1	1	1	3
D	Large (>11)	No	No	47-15	1	1	1	3
E	Large (>11)	Yes	No	57-28	1	1	1	3
Total					5	5	5	15

Figure 5. *Album* item; the accompanying read aloud instruction is:
"The album has space for 51 cards. 49 are already included. How many more cards can be added?"

The fifteen items were displayed one per screen. The students could click to continue to the next item. The accompanying text was read out by the computer. By clicking on the ear button, the student could hear the spoken text again. After a short introduction, the students worked individually on a touch-screen notebook. Students were told that they were free to choose any solution method. After filling in an answer, they reported verbally how they found this answer. The students' on-screen work was recorded by Camtasia Studio software. All students and their parents gave their permission for collecting these records.

Data analysis

I will not go in detail now with respect to the analysis, but will continue with the results. More details can be found in our article in *Educational Studies in Mathematics* (Peltenburg, Van den Heuvel-Panhuizen, in press; DOI 10.1007/s10649-011-9351-0).

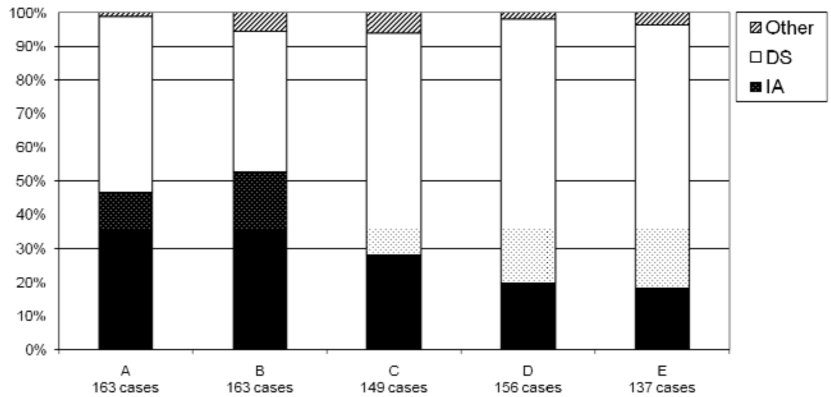
3.3 Our findings

3.3.1 SE students' spontaneous IA use

The data analysis was based on the cases in which the students gave an answer to a particular item. Of the 768 cases that could be analyzed, DS was used in 63% of the cases and IA was used in 34% of the cases. Of the fifteen subtraction problems, the total number of items in which the students applied IA ranged from 0 to 8 items (M=4.6 and SD=1.9).

Numbers involved and IA use

The following diagram (Figure 6) shows that IA was most frequently applied in small-difference problems without and with crossing the ten (categories A and B, respectively). A (single-level) logistic regression analysis showed a significant more frequent use of IA in A and B than in C, D, and E (B=1.24, SE=.16, p< .05).



(56–52) (31–29) (51–39) (47–15) (56–28)

Figure 6. Percentage of procedure use related to number characteristics of the items

Problem format and IA use

The next diagram (see Figure 7) shows that IA mainly appeared in the items with an adding-on context (ConAO) and that DS was most often used in items with a taking-away context (ConTA). A logistic regression analysis showed a significant difference in IA use between context and bare number problems (B=2.38, SE=.25, p< .05).

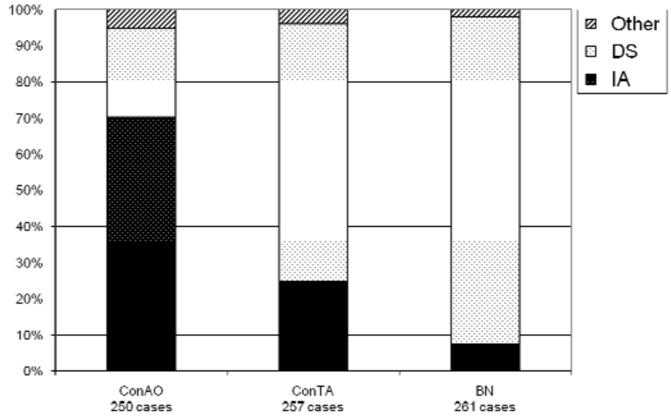


Figure 7. Percentage of procedure use related to the problem format

Prior instruction and IA use

The teachers' responses to the online questionnaire revealed that only three teachers taught both DS and IA. The sixteen students of these teachers applied IA in 29% of the total of 209 cases. The other 40 students who were not taught IA applied this procedure in 36% of the total of 559 cases.

3.3.2 Multilevel analysis with IA use as dependent variable

To examine the influence of the different conditions on IA use, we carried out a cross-classified multilevel analysis which revealed that the IA use is mainly an item characteristic (see Figure 8). The large *SD* (2.83) of the random item effect in Model 0 (that means the model without predictors) compared to the *SD* (.93) of the random student effect indicates that the application of IA is elicited by the nature of an item rather than by the specific preference of a student. Thus, students seemed to apply IA in a flexible, item-specific way. The *SD* (.41) of the random teacher effect is also small compared to the *SD* of the item effect. Nevertheless, it should be noted that there is a substantial variation between teachers whose instruction might have consequences for students' IA use.

In the cross-classified logistic multilevel analysis for Model 1, it was controlled for predictors at item level (numbers involved, problem format), student level (Cito ability score, gender), and teacher/class level (IA taught). In this Model 1, the *SD* (1.19) of the item effect was smaller than in Model 0, while the *SD* (.99) of the student effect and the *SD* (.44) of the teacher effect were the same as in Model 0 when there was no controlling for predictors.

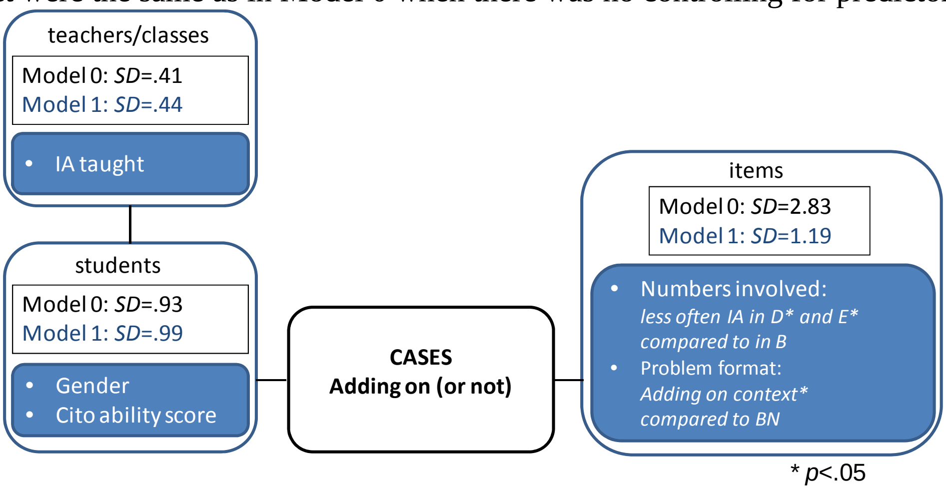


Figure 8. Results from cross-classified logistic multilevel regression analysis with IA use as dependent variable

At *item level*, for the predictor *numbers involved* we found that IA was significantly less applied for the items belonging to the categories D (47–15) (B=–3.35, SE=1.26, p<.05) and E (56–28) (B=–2.98, SE=1.17, p< .05) than for the items belonging to category B (31–29).

With respect to the *problem format*, we found that IA was significantly more often applied for items that involve a context problem that reflects adding on (ConAO) than for BN problems (B=4.74, SE=.93, p< .05). Such a significant difference was not found between items that involve context problems that reflect taking away (ConTA) and BN problems (B=1.42, SE= .99, p> .05). Furthermore, we found that there was a significant difference in IA use between the two types of context problems: ConAO and ConTA (B=3.33, SE= .86, p< .05).

When examining whether there is a difference in IA use between context problems (ConAO and ConTA) and BN problems, we found that IA was significantly more used in context problems (B=3.08, SE= .86, p< .05).

At *student level*, neither gender (B=.44, SE=.41, p> .05) nor the Cito ability score (B=–.01, SE= .03, p> .05) turned out to be a significant predictor for IA use.

At *teacher level*, we found that despite the variation between the teachers, the variable IA taught (B=–.44, SE= .55, p> .05) is not significant for IA use.

3.3.3 Findings with respect to success rate

Overall success rate

Of the fifteen subtraction problems, the students solved between one and fourteen items correctly (M=7.7, SD=3.5). In 68% of the 260 cases in which IA was applied and in 51% of the 480 cases in which DS was applied, the students' answers were correct. A logistic regression analysis showed that the higher success rate when using IA was significant (B= .82, SE= .17, p< .05).

Specific success rate for numbers involved

For items in category B, we found that the students' success rate was 87% when using IA, whereas it was 39% when using DS. However, for items in the category D (47–15) and E (56–28) we found respectively a difference of 11 and 4 percentage points between IA and DS in the advantage of DS. A logistic regression analysis showed that this difference in success rate between IA and DS for the different categories of numbers involved appeared to be significant (B=2.64, SE= .56, p< .05).

Specific success rate and prior instruction

The students who had received IA instruction correctly solved 77% of the 61 total cases for which they used IA. The students who did not receive IA instruction correctly solved 67% of the 199 total cases in which they applied IA. The difference in these percentages did not appear to be significant (B= .55, SE= .34, p> .05).

Multilevel analysis with success rate as dependent variable

To examine the influence of the different conditions on IA use, we carried out a cross-classified multilevel analysis which revealed that the success rate is mainly a student characteristic (see Figure 9). The SD (1.20) of the random student effect in Model 0 (without predictors) is larger than the SD (1.12) of the random item effect. This indicates that correctly solving an item is more student-related than item-related. In addition, the SD (.31) of the random teacher effect is quite small compared to the SD at the item level.

In Model 1, with the predictors included, the *SD* (1.07) of the student effect is smaller than in Model 0 and the same was found for the *SD* of the student and teacher effects.

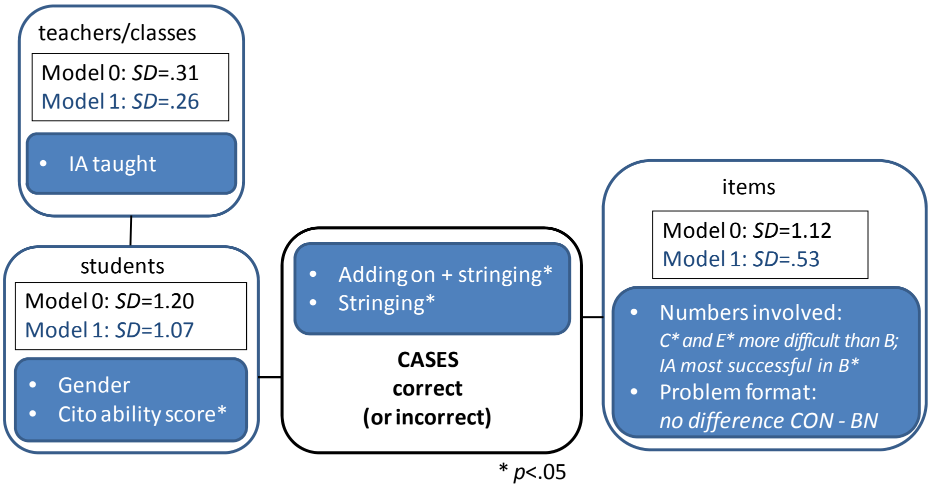


Figure 9. Results from cross-classified logistic multilevel regression analysis with success rate as dependent variable

At case level we found with respect to procedure use, that IA use on its own did not significantly predict success rate (*B*= –.40, *SE*= .52, *p*> .05). However, there was a significant and large effect of simultaneous IA use and stringing (*B*=1.17, *SE*= .55, *p*< .05). In fact, this combined use was the best predictor of reaching a correct answer. Furthermore, stringing on its own had also a significant effect on success rate (*B*= .72, *SE*= .28, *p*< .05).

At item level we found that the items belonging to the numbers involved categories C (51–39) (*B*=–1.39, *SE*= .52, *p*< .05) and E (56–28) (*B*=–1.30, *SE*= .55, *p*< .05) are significantly more difficult than items of category B (31–29). An additional analysis revealed that IA use is most successful when it is applied in small-difference problems with crossing the ten (category B) compared with all the other categories of numbers involved. Concerning correctness and problem format, we found that both types of context problems (ConAO and ConTA) did not significantly differ from the BN problems (*B*= .36, *SE*= .43, *p*> .05 and *B*= .05, *SE*= .45, *p*> .05 respectively).

At student level we found that students' success rate was positively related to the *general mathematics ability* (*B*=.12, *SE*= .03, *p*< .05), however, this was not the case for *gender* (*B*=–.20, *SE*= .39, *p*> .05).

Finally, *at teacher level* we found that IA taught was not a significant predictor of success rate (*B*= –.09, *SE*= .44, *p*> .05).

3.4 Summary of the findings

Our study showed that SE students:

- are able to use IA spontaneously to solve subtraction problems
- are rather flexible in applying IA to solve subtraction problems
- are quite successful when solving subtraction problems by IA

Again, I think these three main findings are reasons enough for a news alert. In addition, our study also revealed another essential lesson for teaching mathematics, namely the power teachers have by cleverly choosing particular problems. By intentionally including problems with specific characteristics, teachers can steer the learning process of their students. As our study has shown, the item characteristics – the numbers involved and the problem format –

were the main prompt for the students to use IA. Students mainly applied IA in small-difference problems with crossing the ten and most frequently employed IA in context problems that reflect adding on. This means that by offering students these kinds of problems, IA can be elicited "spontaneously" and then this student-generated method can be discussed in class to support a more general use of it in other problems that are also suited for this method.

This steering function of the problems also applies to assessment. Our findings made it clear that sensitive assessment tools, including test items designed with particular format and number characteristics enabled us to make visible, which some other studies could not make visible, namely that SE students are able to use IA.

A further finding that emerged from our study is the new mathematics-didactical structure of the constitutive content elements of subtraction up to 100 for analyzing student work and designing learning-teaching scenarios. Contrary to formerly used distinctions in methods for solving subtraction problems in which only one perspective is taken (mostly that of splitting, stringing, and varying) or in which splitting, stringing, and varying belong to the class of DS procedures, whereas IA is considered as a separate class of procedures which do not fit the three strategies, our new structure clearly distinguishes and integrates the two perspectives of the operations carried out (DS, IA, IS, and MO) and the way of dealing with the numbers (splitting, stringing, and varying). I think this new structure gives us a deeper understanding of the constitutive content elements of subtraction. This understanding is important for getting a better grip on the didactical ingredients that are necessary to support the learning processes of our students. Coding the students' responses according to these two perspectives allowed us to find that the best predictor of a correct answer is the combination of IA and stringing.

Of course, our study also has its limitations, so our results should be handled with care. Nevertheless, our findings showed that SE students are able to use IA. Providing evidence for this was the main goal of the study.

4. CONCLUDING REMARKS

A further warning to conclude. My plea for education research that brings us more breaking news is not an appeal for quick and dirty research. My plea is not to support the trend in psychological research towards a bit-size science that goes together with short and rapid publications resulting from the increased pressure on researchers to produce quantifiable output (Bertami & Munafò, 2012). Although shorter and quicker published articles may lead to a wider dissemination – even beyond the borders of a discipline – and will stimulate debates about the findings and a more active involvement of researchers, professionals and the general audience in the topic, I think that education research, like psychological research, is better off when there is what Ledgerwood and Sherman (2012) – in the same issue of *Perspectives on Psychological Science* call – a balance between the "hare" and the "tortoise" approach. In this latter approach, our understanding of a particular topic can steadily grow more slowly, which may avoid that we draw invalid conclusions, because there is more time to include more aspects of the topic under investigation, and to create a better theoretical integration.

The breaking news I am pleading for implies that our studies about mathematics and science education should end up with clear results. Researchers should always ask themselves what new understanding their study brought. Did anything emerge that every teacher should know? The key for revealing important knowledge is, of course, the quality of the study.

Badly conducted research can never result in something that is worthwhile to know. This means having good instruments for data collection and having good methods for analyzing the data. However, in my view the real key to get something valuable out is starting with good research questions. In fact, they form the midpoint of a sandglass in which previous empirical research findings and the underlying theoretical concepts come together in new queries to be solved and from which new data are collected and analyzed. After finding the results they can be disseminated, but that again is a real challenge.

It is generally acknowledged that it is not easy to reach the classroom with research findings. Although there are many reasons for this which I cannot elaborate on now, I read in Kaestle's (1993) article in *Educational Researcher* an interesting thought that gave me some thinking. In this article in which Kaestle discusses the "awful reputation of education research", he interviewed Raizen, a director of a center for improving science education, who contrasted scientists with school teachers. The first are, according to her, information-seeking while the second are not. They only have to teach their students the knowledge they have. To change this situation Raizen argued for transforming schools into learning communities of teachers. I like to add to this: Breaking news from education research might be helpful as well. It can help to make teachers curious and trigger a knowledge-seeking attitude in them.

Moreover, in that same article (Kaestle, 1993, p. 27), I also read that "[t]eachers need, 'as part of their pre-service education, to become sophisticated consumers of education research and development'." Is this not the same that Svend Pedersen was pleading for?

Thank you.

Literature

Arizpe, E., & M. Styles (2003). *Children reading pictures: Interpreting visual texts*. London: Routledge/Falmer.

Bertami, M., & Munafò, M. R., (2012). Bite-size science and its undesired side effects. *Perspectives on Psychological Science, 7*(1), 67-71.

Bowman, B., Donovan, M. S., & Burns, S. (Eds.). (2000). *Eager to learn. Educating our preschoolers*. Washington, DC: National Academy Press.

Carr, M., & Davis, H. (2001). Gender differences in arithmetic strategy use: A function of skill and preference. *Contemporary Educational Psychology, 26*, 330-347.

Casey, B., Erkut, S., Ceder, I., & Mercer Young, J. (2008). Use of a storytelling context to improve girls' and boys' geometry skills in kindergarten. *Journal of Applied Developmental Psychology, 29*, 29-48.

De Smedt, B., Torbeyns, J., Stassens, N., Ghesquière, P., & Verschaffel, L. (2010). Frequency, efficiency and flexibility of indirect addition in two learning environments. *Learning and Instruction, 20*, 205-215.

Gelderblom, G. (2008). *Effectief omgaan met zwakke rekenaars* [Effectively working with students who are weak in mathematics]. Amersfoort: CPS.

Gravemeijer, K., Van den Heuvel-Panhuizen, M., Van Donselaar, G., Ruesink, N., Streefland, L., Vermeulen, W., et al. (1993). *Methoden in het reken-wiskundeonderwijs, een rijke context voor vergelijkend onderzoek* [Textbook series in mathematics education, a rich context for comparative research]. Utrecht: CD-Beta Press.

Hong, H. (1996). Effects of mathematics learning through children's literature on math achievement and dispositional outcomes. *Early Childhood Research Quarterly, 11*, 477-494.

Jennings, C. M., Jennings, J.E., Richey, J., &, Dixon-Krauss, L. (1992). Increasing interest and achievement in mathematics through children's literature. *Early Childhood Research Quarterly, 7*, 263-276.

Janssen, J., Scheltens, F., & Kraemer, J. (2005). *Leerling-en onderwijsvolgsysteem. Rekenenwiskunde groep 4. Handleiding* [Student and education monitoring system. Mathematics grade 2. Teachers guide]. Arnhem: Cito.

Kaestle, C. F. (1993). The awful reputation of education research. *Educational Researcher, 22*(1), 23-31.

Klein, A. S., Beishuizen, M., & Treffers, A. (1998). The empty number line in Dutch second grades: Realistic versus gradual program design. *Journal for Research in Mathematics Education, 29*, 443-464

Kraemer, J., Van der Schoot, F., & Van Rijn, P. (2009). *Balans van het reken-en wiskundeonderwijs in het speciaal basisonderwijs 3* [Third assessment of mathematics education at special education schools]. Arnhem: Cito.

Ledgerwood, A., & Sherman, J. W. (2012). Short, Sweet, and Problematic? The Rise of the Short Report in Psychological Science. *Perspectives on Psychological Science, 7*(1), 60-66.

Menne, J. J. M. (2001). *Met sprongen vooruit. Een productief oefenprogramma voor zwakke rekenaars in het getallengebied tot 100 – Een onderwijsexperiment* [A productive training program for mathematically weak children in the number domain up to 100 – A design study]. Utrecht: CD-Beta Press.

Milo, B. (2003). *Mathematics instruction for special-needs students. Effects of instructional variants in addition and subtraction up to 100*. Heerenveen: Brouwer & Wielsma.

Molander, B.-O., Halldén, O., & Pedersen, S. (2001). Understanding a phenomenon in two domains as a result of contextualization. *Scandinavian Journal of Educational Research, 45*(2), 115-123.

Molander, B.-O., Pedersen, S., & Norell, K. (2001). Deaf pupils’ reasoning about scientific phenomena: School science as a framework for understanding or as fragments of factual knowledge. *Journal of Deaf Studies and Deaf Education, 6*(3), 200-211.

National Council of Teachers of Mathematics (NCTM) (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington: U.S. Department of Education.

Nikolajeva, M., & C. Scott (2000). The dynamics of picturebook communication. *Children’s Literature in Education, 31*(4), 225-239.

Peltenburg, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (in press). Special education students’ use of indirect addition in solving subtraction problems up to 100 – A proof of the didactical potential of an ignored procedur. *Educational Studies in Mathematics*. DOI 10.1007/s10649-011-9351-0

Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research, 37*, 239-253.

Timmermans, R. E. (2005). *Addition and subtraction strategies: Assessment and instruction*. Nijmegen: Radboud Universiteit Nijmegen.

Torbeyns, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009). Solving subtractions adaptively by means of indirect addition: Influence of task, subject, and instructional factors. *Mediterranean Journal for Research in Mathematics Education, 8*(2), 1-30.

Torbeyns, J., De Smedt, B., Stassens, N., Ghesquière, P., & Verschaffel, L. (2009). Solving subtraction problems by means of indirect addition. *Mathematical Thinking and Learning, 11*, 79-91.

Treffers, A. (1989). *Het voorkomen van ongecijferdheid op de basisschool* [Presence/Preventing innumeracy in primary school]. Utrecht: OW&OC, Utrecht University.

Van den Heuvel-Panhuizen, M. (1986). Reactie: Het rekenonderwijs op de lom-school opnieuw ter discussie [Reaction: Arithmetic education at schools for children with learning and behavioral difficulties again up for debate]. *Tijdschrift voor Orthopedagogiek, 25*(3), 137-145.
Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*. Utrecht: CD-Beta Press.
Van den Heuvel-Panhuizen, M. (Ed.). (2001). *Children learn mathematics*. Utrecht: Freudenthal Institute, Utrecht University.
Van den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics, 25*(2), 2-23.
Van den Heuvel-Panhuizen, M. (2010). Reform under attack – Forty Years of Working on Better Mathematics Education thrown on the Scrapheap? No Way! In 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* (pp. 1-25). Fremantle: MERGA.
Van den Heuvel-Panhuizen, M., & Elia, I. (in press). Developing a framework for the evaluation of picturebooks that support kindergartners' learning of mathematics. *Research in Mathematics Education*.
Van den Heuvel-Panhuizen, M., & Treffers, A. (2009). Mathe-didactical reflections on young children's understanding and application of subtraction-related principles. *Mathematical Thinking and Learning, 11*(1), 102-112.
Van den Heuvel-Panhuizen, M., Robitzsch, A., & Elia, I. (submitted). Effects of reading picturebooks on kindergartners' understanding of mathematics.
Warry, M. C., Galbraith, P. L., Carss, M. C., Grice, R. D., & Endean, L. (1992). Towards numeracy for the third millennium: A study of the future of mathematics and mathematics education. *Educational Studies in Mathematics*, *23*, 569-593.
Young-Loveridge, J. M. (2004). Effects on early numeracy of a program using number books and games. *Early Childhood Research Quarterly, 19*, 82-98.

Picture books cited

Veldkamp, T., & Hopman, Ph. (1999). *22 Wezen* [22 Orphans]. Rotterdam: Lemniscaat.
Grindley, S., & Utton, P. (2006). *Sssst!* [Ssh!]. Haarlem: Altamira Jeugd.
Van Haeringen, A. (1999). *De prinses met de lange haren* [The princess with the long hair] Amsterdam: Leopold.