Educ Stud Math (2012) 81:47–65
DOI 10.1007/s10649-011-9360-z

The effects of using drawings in developing young children's mathematical word problem solving: A design experiment with third-grade Hungarian students

Csaba Csíkos · Judit Szitányi · Rita Kelemen

Published online: 9 November 2011

Abstract The present study aims to investigate the effects of a design experiment developed for third-grade students in the field of mathematics word problems. The main focus of the program was developing students' knowledge about word problem solving strategies with an emphasis on the role of visual representations in mathematical modeling. The experiment involved five experimental and six control classes ($N=106$ and 138, respectively) of third-grade students. The experiment comprised 20 lessons with 73 word problems, providing a systematic overview of the basic word problem types. Teachers of the experimental classes received a booklet containing lesson plans and overhead transparencies with different types of visual representations attached to the word problems. Students themselves were invited to make drawings for each task, and group work and teacher-led discussion shaped their beliefs about the role of visual representations in word problem solving. The effect sizes of the experiment were calculated from the results of two tests: an arithmetic skill and a word problem test, and the unbiased estimates for Cohen's d proved to be 0.20 and 0.62. There were significant changes also in experimental group students' beliefs about mathematics. The experiment pointed to the possibility, feasibility, and importance of learning about visual representations in mathematical word problem solving as early as in grade 3 (around age 9–10).

Keywords Word problem · Visual representation · Design experiment · Elementary school mathematics

C. Csíkos (✉) · R. Kelemen
Institute of Education, University of Szeged, Szeged, Hungary
e-mail: csikoscs@edpsy.u-szeged.hu

J. Szitányi
Faculty of Elementary and Nursery School Teachers' Training, Eötvös Loránd University, Budapest, Hungary

1 Introduction

1.1 Intervention programs focusing on mathematical word problems

In the last decades, a number of educational intervention programs have been conducted in the field of mathematical word problems. These programs can be grouped according to several factors: students' age, the extent to which instructional methods and tasks differed from those of control groups', and whether the characteristics of word problems served as dependent or independent variables of the experiment. In the following, we present a review of the literature concerning research into the teaching of word problems in order to frame our study and warrant our chosen methods.

The importance of word problems in research on mathematics education originates in several factors. As Lave (1992) pointed out, the importance of word problems lies not in the mathematical structure they reflect but in activities of the school system, and—consequently—in learning theories concerning those situated actions. Another important feature of word problem research is the potential to comply with the norms of analytic scientific reasoning, i.e., the potential to change and alternate characteristics of word problems as independent variables of a laboratory-based or a classroom experiment (see Verschaffel, Greer & De Corte, 2000). Since general problem-solving skills are of central importance in school mathematics achievement (see, e.g., English & Halford, 1995), a third justification might be given by the fact that research on arithmetic word problems has shifted towards a general problem-solving perspective (Verschaffel, Greer & Torbeyns, 2006).

The role of word problem solving is highlighted in the Hungarian National Core Curriculum (Nemzeti alaptanterv, 2007) in that "the individual is able to apply basic mathematical principles and processes in acquiring knowledge and in solving problems in daily life, at home and at the workplace" (p. 9). The objectives for word problem solving (i.e., observing situations elaborated in word problems, separating relevant and irrelevant information etc.) are declared for school grades from 1 to 8, without detailing the steps or levels and without attaching them to grades or age groups. (It is the task of the local curricula to elaborate objectives for a given school grade.) In sum, the importance of word problems and problem solving in general is well recognized and emphasized in the National Core Curriculum, and what actually happens in classrooms depends on mainly the local curricula, on textbooks and on teachers' own view of the objectives (implemented curriculum).

As a consequence of these justifications, word problems play an important role in educational experiments and surveys. Solving word problems—at least when paper–pencil tests are used in an experiment—requires an adequate (or at least fledgling) level of reading skills, therefore the earliest possible school grade for intervention programs built around word problems should be grade 3. For those teachers and researchers, who may find it too early for third-grade students to encounter word problem-based interventions either because of the inappropriate level of reading skills or because of children struggling with arithmetic skills, we would like to underline that children around age 9, even when considered as low-achievers in mathematics "are capable of building sophisticated mathematical concepts" (English, 1996, p. 108.). Consequently, they should have the chance to meet problems that permit the use of various solution strategies. Indeed, there are experiments with third-grade students in the field of word problems.

Sáenz-Ludlow and Walgamuth's (1998) 1-year long experimental program addressed third-grade children's use and understanding of equality and the equality symbol. In their experiment, the idea of children's writing of their own word problems helped children overcome the difficulty of understanding contextual meaning of word problems. Making sense of contextual meaning may be considered as one source of difficulties children encounter when solving word problems.

In Selter's (1998) teaching experiment, the number line and the heterogeneity of solution strategies play a central role. It was a qualitative study that presented a case study sharing components of an experiment conducted with 28 children, illustrating the uniqueness and complexity of that study.

Several intervention programs with students from higher grades have characteristics that can be transferred to elementary school programs. Beyond similarities in research design, content variables of other studies can be transferred to possible experiments with third-grade students. One such rationale for research into word problems is eliminating students' "superficial coping strategies" (Verschaffel & De Corte, 1997a). By upper elementary school, students possess a variety of such strategies as searching for figures in the text, using keywords to decide which operation would be used or simply looking at the numbers that might indicate which operation is necessary to compute.

Over the last 15 years, a number of intervention programs have been successfully conducted with both upper elementary and secondary school students. Significantly, with respect to warranting the research reported below, experiments with fifth graders (English, 1997; Verschaffel & De Corte, 1997b; Verschaffel et al., 1999), interventions with seventh-grade students (Mevarech & Kramarski, 1997; Kramarski, Mevarech & Arami, 2002; Kramarski, Mevarech & Lieberman, 2001) all have emphasized realistic (or authentic) tasks, and higher level components of problem solving. In the particular case of Verschaffel et al. (1999) study, three major pillars of the learning environment were changed: (1) the use of carefully designed realistic word problems, (2) the use of powerful instructional techniques, and (3) creation of a stimulating classroom culture. A successful adaptation of the Flemish intervention program was realized in Hungary among fourth-grade students (Csíkos, 2005).

Beside changes in the learning environment, an explicit formulation of learning goals different from "traditional" aims was revealed in these experiments: the aim of developing students' metacognitive strategies and shaping their beliefs about mathematical problem solving. There is plenty of evidence to support the existence of metacognitive strategies during mathematics problem solving as early at the age of 10 (Verschaffel, De Corte & Lasure, 1994), making the development of intervention programs for third-grade students both feasible and desirable.

Further support of the idea that elementary school children may profit from learning by constructing drawings came from a study with fourth- and sixth-grade students. Van Meter, Aleksic, Schwartz and Garner's (2006) exploratory study concerning the role of student-generated drawings as a learning strategy can be considered important from the viewpoint of mathematics as well. Their study focused on expository texts (i.e., texts that belong to a given content area and are dense in information), nevertheless, the study was built around the so-called "generative theory of drawing construction". This theory focuses on the dual nature of verbal and nonverbal information, and a recursive process is described, i.e., drawing may require text reinspection supported by self-monitoring processes.

1.2 The role visualized models in mathematics problem solving

The term "visual imagery" was defined by Presmeg (1986) as a mental scheme that depicts visual or spatial information either existing with the presence of the object being visualized or without the presence of that object. According to Jonassen (2003, p. 269), "Successful problem solving requires the comprehension of relevant textual information, the capacity to visualize the data." Since mathematical concepts and relations are often based on visual mental representations attached to verbal information, the ability to generate, retain and manipulate abstract images is obviously important in mathematical problem solving. Goldin and Kaput (1996) analyzed the structure of internal mathematical representations, and found that the imagistic system (nonverbal configurations of objects, relations and transformations, including visual imagery and spatial representation) receives much less attention from educators than other systems of mathematical representations.

It is clear from Goldin and Kaput's (1996) argumentation that the ability to visualize data (and their relations) in a mathematical problem may contribute to mathematics problem solving. Indeed, Hegarty and Kozhevnikov (1999, p. 688) have found that "some visual-spatial representations promote problem-solving success" among sixth grade students (there were only boys participating in their experiment). Furthermore, it was revealed that it was possible to teach students how to produce appropriate visual representations. That is, representations focused on relevant data and relations, and not irrelevant iconic or pictorial representations. Following the verbalizer–visualizer dichotomy, Kozhevnikov, Hegarty and Mayer (2002, p. 47.) designate people who "rely primarily on imagery processes when attempting to perform cognitive tasks" as visualizers. Visualizers belong to two groups: spatial and iconic type visualizers. Kozhevnikov, Hegarty and Mayer bring evidence about the relevance of making distinctions between two types of images, and therefore, between two types of visualizers. One aspect is the neuropsychological and neuro-imaging differences, secondly the working memory literature is relevant, helping to distinguish between two types of images Presmeg (1986) also differentiated between what she called "pattern images" and "concrete images" and the former were more productive in solving mathematical problems. Furthermore, especially importantly from an educational point of view, Hegarty and Kozhevnikov (1999) found that it was possible to differentiate reliably between schematic and pictorial representations made by sixth grade children. It was also found that "the use of schematic spatial representations was associated with success in mathematics problem solving" (Kozhevnikov, Hegarty & Mayer, 2002, p. 51).

The idea that schematic representations are meaningful for word problem solving in terms of containing relevant data and relations calls forth the question whether using explanatory drawings generated by mathematics teachers and students in classroom situations is a powerful method of visualization that will improve students' problem solving ability. This broad hypothesis has antecedents in the literature of many frameworks including the field of worked examples and the cognitive load theory.

In Mwangi and Sweller's (1998) experiment, third-grade children were presented with a worked solution that contained a schematic visualization of the objects and quantities from the text of a word problem. The word problems were of consistent language, and students of both the experimental and control groups solved almost all tasks; it was the number of incorrect attempts that made a difference between the two groups, favoring the experimental group. The authors drew two important consequences. First, the results may be interpreted within the cognitive load framework, i.e., worked examples reduced working memory load. They point to the importance of coherence of the source materials children receive. Secondly, the possibility of an intervention using worked examples with schematic drawings as early as in the elementary school years has been demonstrated.

Elementary school children's capability to match types of schematic diagrams with a mathematics problem has been further evidenced by Diezmann (2005). In her study of third- and fifth-grade students' judgments on types of diagrams, she found that even 3rd graders were able to match diagram types with word problems in proportions unlikely to be due to chance. As she stated in her conclusion: "students need to be able to select the appropriate diagram for a particular problem and adequately justify their selection" (p. 286).

In order to justify the importance of schematic drawings, more relevant features of mathematics word problems must be taken into account. A widely recognized source of incorrect solutions is the use of direct translation strategy (see Hegarty, Mayer & Monk, 1995) often caused by inconsistent language, e.g., when “greater than” is matched with subtraction. Successful problem solvers often tend to stop and return during the phase of reading the text of a word problem (Mayer & Hegarty, 1996). Unsuccessful problem solvers, on the contrary, use a straightforward (direct) solution strategy like matching the hastily read key word (e.g., “less than” with an arithmetical operation (subtraction in this case)). Jonassen (2003) aimed at providing theoretical analysis of “story” word problems that prevent students from using the direct translation strategy. For this prevention, a classroom environment is needed that “requires that students first identify the class of problem … before they are allowed to transfer them [the components of the problem] to a quantitative representation” (p. 292). He emphasizes the role of worked examples in light of the cognitive load theory, similarly to the approach expressed in Mwangi and Sweller’s study.

With respect to how students’ visual representations can foster word problem solving competence, Van Meter and Garner (2005) reviewed the literature on student-generated drawings that were made to support learning goals. They express their hope that “the potential of drawing as a strategy to improve learning and problem solving will be realized” (p. 322). It has been revealed that adding diagrams to word problems per se does not foster understanding in a straightforward way: success of word problem solving may depend on a number of factors like the types of the problems and on different student characteristics like visuospatial abilities (Pantziara, Gagatsis & Elia, 2009).

What types of drawings can improve the understanding of mathematics word problems? Berends and van Lieshout (2009) used a system with four categories: (1) bare picture (e.g., symbols), (2) useless, (3) helpful, and (4) essential illustrations. The latter contain numerical information that is essential part of the word problem. The most important distinction between the useless and helpful drawings is that useless illustrations contain irrelevant visual information, whereas helpful ones represent numerical information from the problem text. In another taxonomy, Gagatsis and Elia (2004) coined the terms “decorative” and “informational” for naming the second and fourth types in Berends and van Lieshout’s system.

Since schematic drawings seem to play an important role in mathematics understanding, it is relevant to ask whether third-grade students possess the capabilities necessary to generate and evaluate schematic drawings for simple arithmetic problems. One archetypal branch of schematic drawings is the number line. If students at the age of 9 are already capable of appropriately representing the number line, any results suggesting that third-grade students have relatively accurate mental number lines will support the claim for early intervention programs using student-generated drawings for better understanding word problems.

Children’s developing visual representations of natural numbers and their relations (especially comparison) have been examined in investigations concerning the number line and its mental representations. In grade 3, the 1–100 part of the number line is of great importance both from curricular and metal representation reasons. According to Opfer and Siegler (2007), there is a shift in mental number line representation from kindergarten to second grade, from a logarithmic to a linear scale, and from second grade to sixth grade estimating the locations of numbers between 1 and 1,000 develops. Siegler and Opfer (2003) revealed that by second grade, children have both types of representations of the number line available. Further evidence was brought to the development of accuracy in mental visual representations of the 1–100 part of the number line by Schneider et al. (2008). In their experiment, children from first to third grade age added two-digit numbers where the sum was not bigger than 100. It has been shown that eye movements reflect children's developing knowledge about natural numbers and their spatial representations. In that experiment, an increasing accuracy of number line estimation was revealed from eye-tracking data, with larger differences between first and second grade, and slighter differences between second and third grade.

The usefulness of providing appropriately scaled number lines for a special type of word problems (one-step additive problems) in an educational setting has been documented by Elia, Gagatsis and Demetriou (2007). It has been revealed that achievement on this number line-added type of word problems relevantly increases at the age of 7–8.

1.3 Teachers' knowledge about and approach for visual representations in mathematics problem solving

Teachers' knowledge about word problems and the role of visual representation has been emphasized in various investigations. On the one hand, preservice elementary teachers tend to use word problem strategies that may lead to failures in more difficult problems (Van Dooren, Verschaffel & Onghena 2003). On the other hand, as revealed by Chapman (2006), it is the elementary school teachers (comparing to secondary school teachers) who tend to use the so-called narrative mode of conceptualization (in a Brunerian framework), and "create a classroom environment that was motivational for students to learn word problems" (p. 225). In Chapman's study, the narrative mode of conceptualization was opposed by the paradigmatic view. The latter refers to a cognitive functioning mode where the central question is how to know the truth. The narrative or humanistic mode centers on the meaning of experience. Whatever teaching style or cognitive functioning strategy is observable among teachers, it cannot be overemphasized that teachers and students may have strikingly different interpretations about a seemingly straightforward word problem.

Cai and Lester (2005) found remarkable differences in how Chinese and American teachers helped students to solve word problems. Chinese teachers preferred visual representations over verbal ones in their facilitation of students' understanding of word problems. These visual representations were not a required part of the solutions themselves. Cai and Lester give a typical example of how a Chinese teacher used a number line-like diagram to illustrate quantitative relations. On the contrary, US teachers usually used symbolic rather than visual representations in their classroom practice.

In another comparative study, Uesaka, Manalo and Ichikawa (2007) found that 13–15-year-old Japanese students—as compared to their New Zealand peers—had lower scores on the questionnaire item asking whether they are explicitly told or encouraged by their teachers to use diagrams. One possible reason can be that students tend to consider drawings as illustrations provided by the teacher in the process of problem solving, and "[Japanese] teachers need to spend more time in class teaching students how to actually use diagrams" (p. 333).

1.4 Characteristics of the intervention program

The characteristics of the current investigation can be summarized as follows:

– The most important characteristic of our intervention program is the use of visual representations in word problem solving. This use has several facets: (1) shaping students' belief about the importance of making drawings when solving word problems, (2) showing students drawings that illustrate possible ways of representing or modeling a problem, and (3) reassuring students in making drawings that befit their mental images and their understanding of the word problem text. In most of the lessons, both a schematic- and a pictorial-type drawing was shown by the teacher or generated by the students.

- The program has a maximum possible level of ecological validity, i.e., the system of tasks and visual transparencies will be usable in the future without requiring extra time and human resources, and without much change in curricular content.
- The program lasts for 20 units, which is about 5 weeks duration (at a rate of four lessons/week), and the program is similar in length to the usual duration of time dedicated to word problems in grade 3. The length is restricted by the aim of keeping the intervention intact and uniform in a relatively short time, avoiding the effects of possible changes in classroom and out-of-school characteristics that might interfere with experimental variables.
- Since, at least from a historical point of view (see Verschaffel, Greer & De Corte, 2000) mathematics word problems can be either the means for practicing arithmetic skills or the means for modeling reality, it seems to be necessary to make decisions about what the word problems of our program can be used for. We strove to select word problems with familiar and realistic content, trying to avoid even the appearance of mere drilling practice. However, we agree with Palm (2008, p. 55) in that "to practice real life task solving, then dismissing important conditions of real life … is not an efficient practice". Consequently, we selected realistic tasks as defined by Cooper and Harries (2002, p. 5.), i.e., "where the textual representation of the problem contains either persons or non-mathematical objects from 'everyday' settings such as shopping or sports", but we avoided selecting so-called 'more authentic' tasks as defined by Palm (2008).
- The intervention program focuses on several types of word problems in a systematically structured way: number of arithmetic operations to be computed (if operation should be computed at all), and whether the text has consistent or inconsistent language. This intention about covering a wide range of word problem types can be characterized as "building gradually" from simpler to more complex word problems. However, this feature of the program is not unique when compared with the textbooks from which the control group students learn, and their teachers plan their lessons

1.5 Aims and hypotheses

A number of hypotheses underpinned the research design. Firstly, that the experimental group would demonstrate (1) better results on word problems, (2) equal or better results on the arithmetic skill test, (3) changes in their beliefs concerning the importance of making drawings when solving mathematics word problems, gaining self-confidence about their mathematical knowledge, and possessing more positive attitude towards mathematics lessons. (4) It was also hypothesized that no significant sex differences would emerge on the word problem test. Although there may be differences between boys and girls in certain fields of mathematical knowledge (Geary, 1999) or in certain populations (Halpern & LaMay, 2000), the effect size of students' sex will be lower than participating in the intervention program. (5) Finally, it is assumed that experimental group students' drawing skills would have no effect on their achievement in the intervention program.

2 Methods

2.1 Subjects

The students involved were recruited from six schools located close to each other in Budapest, Hungary. Eleven classes participated, five experimental and six control classes. The experimental classes were randomly chosen from the pool of 11 classes. In each case, the whole class was designated as either a control or an experimental group. All participants were third-grade students whose mean age was 9 years in March, 2008. The experimental classes comprised 106 students (53 boys and 53 girls), and the control classes 138 students (63 boys and 75 girls).

According to recent legislation, there must be no relevant differences among schools within the same town concerning the rate of low social-economic status students in a school. Therefore, classes participating in the investigation either as experimental or control classes can be considered as representative groups of the capitol's inner city schools. From nationwide and international system-level surveys (see, e.g., Báthory, 2003), it can be inferred that this student population has above-average characteristics in academic achievement.

2.2 Pre- and post-tests of the experiment

A pre-test–post-test–control (PPC) design was adopted. We used two tests as both pre- and post-test, consequently the PPC design was realized. The first test was an *arithmetic skill test* comprised 32 items covering National Core Curriculum aims. The second test comprised six word problems, along with some questions concerning students' attitudes and learning aims.

Each student's score on the arithmetic skill test was calculated as the number of correctly solved items. The test contained 32 dichotomous items. The Cronbach-α reliability coefficient of the pre-test was 0.84 (N=228). Using this arithmetic skill test as post-test, the reliability coefficient proved to be 0.83 (N=237).

Scoring of the word problem test was carried out according to the scoring system of Verschaffel, De Corte and Lasure (1994), i.e., the score of the test was the number of realistic reactions. Realistic reactions may have different forms, namely, mathematically correct solutions taking realistic considerations into account, or explicit notions of the task being unsolvable or otherwise problematic. The reliability of this six item pre-test was 0.72 (N=230). Using the word problem test as post-test, the reliability proved to be 0.72 again (N=232)

In order to control whether experimental group students' drawing skill influenced how they could profit from the experiment, the first task of the Clark Drawing Test (CDT) was administered as a pre-test (Clark, 1989). According to the Hungarian standardization results of CDT (Kárpáti, 2001), we decided to use only the first task of CDT (drawing an interesting house as seen from the other side of the street). Since the test has a very strong inner consistency as measured by the Cronbach-α coefficient (0.97), a shortened version may save time without losing much diagnostic value about the overall level of drawing skill. This subtest can be considered as a neutral measure of drawing skill, i.e., independent of any mathematical content. Another modification was skipping the item of giving a title to the picture because there are only 0.2–0.3 correlations between the quality of giving title and other items of the task. Thus, the maximum score was 60 points, deriving from12 five-point items. The reliability coefficient proved to be 0.88 (N=100).

2.3 The intervention program

Brown's (1992) seminal study emphasized the need for conducting educational experiments where the complexity and systemic nature of learning can be studied in view of interwoven variables. It has been clear in the previous decades that it is almost impossible to conduct single-factor experiments where only one independent variable is tested and the other factors are kept constant. According to Bell (2004), there is a wide range of different types of design-based experiments. Taking examples from the two furthest points, there are multivariate, multifactorial designs and there are ethnographic observations, and both can be considered as design experiments. Our current research is much closer to the multifactorial experimental design approach, but since there are several possible interwoven independent variables, we regard it as design experiment.

The intervention program in our current investigation has several characteristics (see below), but none of these variables can be isolated from the others, so the intervention program is considered as a complex intervention package. This complexity enables (1) practical feasibility in dissemination (Brown, 1992), therefore assuring ecological validity (Gravemeijer & Cobb, 2006), and (2) possible contribution to theories of classroom learning. The current design experiment program created a learning environment that provided ample opportunity for children to encounter various types of word problems and various types of drawings that might help them to solve those problems.

2.3.1 Characteristics of the intervention program

The program had several interwoven characteristics. The basic idea of the intervention was built around the role of visual representations in word problem solving. For this aim, we first provided students with a systematic review of the types of word problems. The aspects of this review were: (1) whether the word problem can be solved by executing arithmetic computations. This aspect may be important to eliminate possible misconceptions or beliefs that lead to frequent searching for *the* appropriate operation. (2) The number of operations to be computed for solving a word problem is usually one in grade 3 but at the end of the intervention program, two-step problems are applied. (3) Since Mayer and Hegarty's (1996) work, the term "consistent language" has become widely accepted by the discourse community. This term indicates that the keywords in the text of a word problem correspond with the arithmetic operation to be computed. The structure of word problem types in the intervention is shown in Table 1.

The program contained 73 word problems altogether. Each task was presented to each student on a separate A4 sheet one at a time during lessons. Children had no booklet, because we wanted to avoid their prereviewing any of the tasks.

A booklet was prepared for teachers of the experimental classes. This booklet contained a short theoretical introduction about the nature of the experiment and its hypotheses and, for each lesson, the aim of that lesson, the type of tasks to be solved, the tasks themselves, instruction about the methods, and the use of supplementary materials.

Since one focus of the program was helping students become aware of the role and importance of making drawings related to mathematics word problems, teachers were invited to encourage students to make drawings even in the case of the simplest—or seemingly simple—tasks. Teachers were invited by us to initiate conversations during the task solving process allowing students to gradually become aware of the existence of

Table 1 Types of word problems in the intervention program

Lesson	Type of word problems			
	Can be solved by arithmetic computation	One definite solution	Can be solved by one arithmetic operation	Consistent language
1	No	Yes		
2	No	No		
3	No	Yes		
4	No	No		
5	Yes		Yes	Yes
6	Yes		Yes	Yes
7	Yes		Yes	No
8	Yes		Yes	No
9	Yes		Yes	Yes/no
10	Yes		Yes	Yes/no
11	Explicit analysis of drawings for arithmetic word problems			
12	No	No		
13	Yes		Yes	Yes
14	No	No		
15	Yes		Yes	No
16	Yes		Yes	Yes
17	Yes		Yes	Yes/no
18	Yes		Yes	Yes/no
19	Yes		No	Yes
20	Yes		No	No

different types of drawings, and albeit there might be individual difference in preference, usually it is the schematic drawings that are useful to sketch.

Another focus of the program was of instructional methodological nature. Changing the usual classroom practice had two main dimensions. First, teachers were asked to use the think-aloud technique in some of the tasks in order to demonstrate the possible branches and circles in the solution process. Second, in some of the lessons students worked in heterogeneous groups consisting of five to six students, learning to tolerate and discuss different solution plans.

We prepared colored transparencies for the majority of the tasks. (Overhead projector transparencies meant the common available tool for the experimental classes from an educational infrastructural aspect.) The aim of using transparencies was to standardize visual aid students may receive when copying, comparing or analyzing drawings created for a word problem. For example, word problem #16 in the sixth lesson had the following text: this morning, three uniform bunches of carnations were ordered. The gardener picked 21 carnations. How many carnations did each bunch contain? After individual seat work, students compared their drawings to three different types of drawings presented on the preprepared transparencies by the teacher (see Fig. 1). The drawings on the transparencies were planned by the first and third author of this study, and were prepared by a graphic artist. The lines and the use of colors imitated many characteristics of children's drawings except for being purposeful about

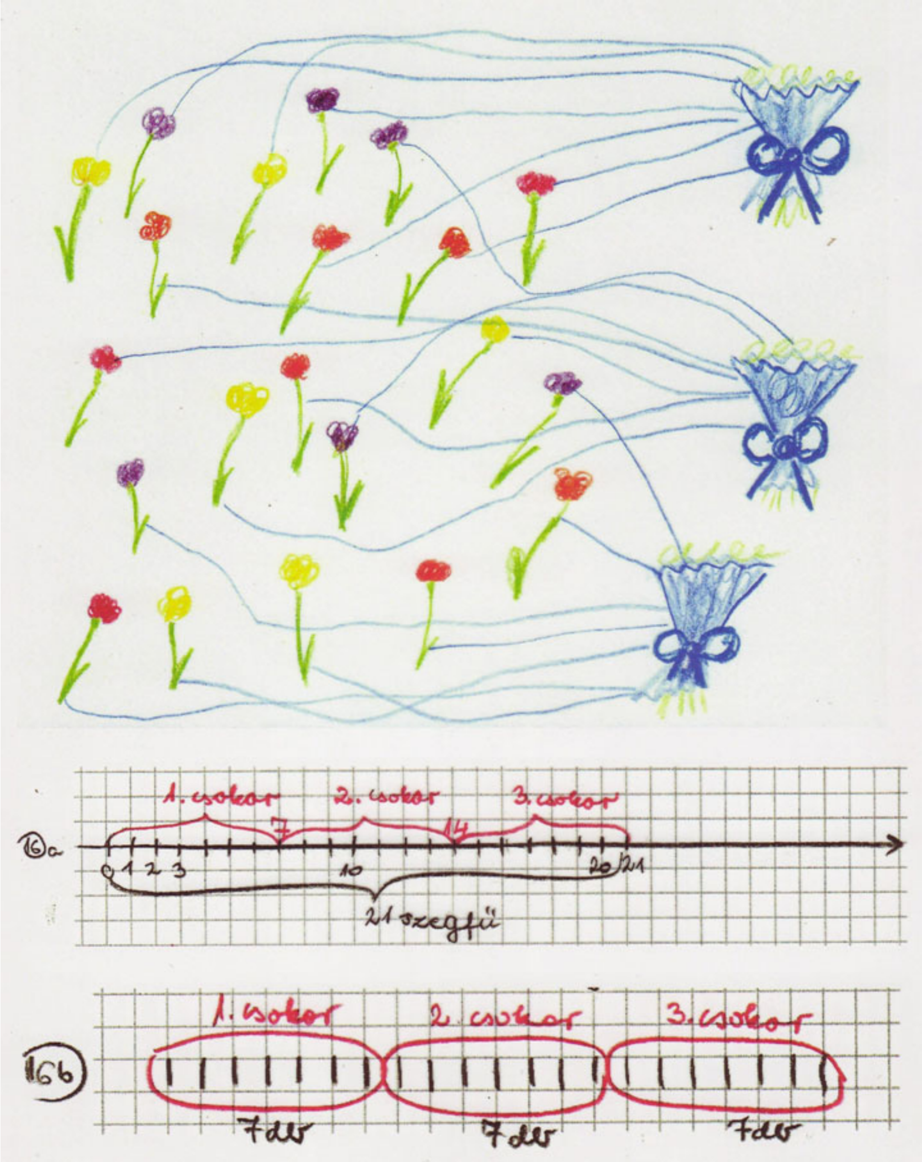

Fig. 1 Illustrations provided for word problem #16 of the sixth lesson unit

the mathematical–structural content. The lesson plans gave detailed instruction when and how to use the transparencies.

2.3.2 Teacher training and consultation, implementation of the intervention program

Teachers of the experimental classes were gathered for an oral discussion led by the first and second authors of this study. During this talk, we presented the aims of the experiment, our hypotheses, and the lesson plans were handed out. Later, during the weeks of the experiment, teachers' questions were received and answered via emails and telephone conversations.

The pre-test phase involved administering three tests in the experimental group in three consecutive lessons, and administering two tests in the control group. After the pre-tests, the experimental group received 20 lessons, at a rate of four per week. (Four mathematics lessons per week is a general characteristic in the third grade in Hungary.)

The lessons were taught by the regular classroom teachers who, as indicated above, were familiar with the aims of the experiment. The intervention lessons did not require the full 45 min of a mathematics lesson; the rest of the time could be freely used to make progress with the regular material and targets.

Teachers of the control groups received no information about the aims and hypotheses of the experiment, but they were informed about the importance of their contribution as control class teachers of an experiment. In the participating schools, during the time interval of the intervention program, the regular classroom material consisted of solving and practicing on word problems. These spring months, when the experiment took place are usually, country-wide, the season for allocating much time for drilling practice on word problems, i.e., solving dozens or hundreds of word problems; many of them can be labeled as routine and pseudo-realistic (a term borrowed from Boaler, 1994). The main difference between the experimental and control class practices—from the viewpoint of control group teachers—was allocating much more time for drilling practice on routine word problems. The post-tests were administered immediately after the intervention program.

3 Results

Only those students who completed both the arithmetic skill and word problem tests, as both pre- and post-tests, are taken into account. Since test administration took place in the students' classroom, a 5% rate of attrition, typically due to illness, seems reasonable and understandable. As we have seen, the reliability coefficients could be computed from 228 to 230 students' scores, and there were 216 students (97 from experimental and 119 from control classes) who completed all four tests.

3.1 Descriptive statistics and comparisons

Table 2 shows the basic statistics about the pre- and post-test results of the experimental and control groups. The data in Table 2 suggest that while there were differences between the

Table 2 Means and standard deviations of the experimental and control groups' test achievement and post-test—pre-test differences (PPD) on the pre- and post-tests

		Experimental group (N=97)			Control group (N=119)		
		pre-test	post-test	PPD	pre-test	post-test	PPD
Arithmetic skill test	Mean	25.31	27.00	1.69	26.74	27.56	1.56
	SD	4.61	4.35	3.04	4.25	4.00	1.43
Word problem test	Mean	2.57	4.12	0.82	3.51	4.06	0.55
	SD	1.68	1.68	3.13	1.56	1.67	1.25

experimental and control groups on the pre-test favoring the latter group, there were notably fewer differences on the post-tests results.

Table 3 shows the *t* test comparisons between experimental and control group pre-test achievement. The differences in pre-test achievement were significant in both tests. Table 4 shows similar data for the post-tests.

There were no significant differences between the experimental and control groups in either of the post-tests. The paired-samples *t* test provides information about the development of students' achievement during the 6-week period of the intervention program. On the arithmetic skill test, both the experimental and control groups significantly outperformed their pre-test achievement in the post-test. Statistical values are $t(118)=2.87$, $p<0.01$ for the control group and $t(96)=5.48$, $p<0.001$ for the experimental. Similarly, in the case of the word problem test, $t(118)=4.78$, $p<0.001$ for the control group and $t(96)=10.73$, $p<0.001$ for the experimental group indicate significant changes during the 6-week intervention period.

3.2 Effect sizes

The PPC design of this experiment calls forth the use of the modified Cohen's *d* effect size (Morris, 2005). Since it has been proven that both the experimental and control groups had significantly higher means at the end of the intervention period in both tests than they had at the beginning, it is this effect size that may indicate how much bigger development rate took place in the experimental group.

In case of the PPC design, Cohen's *d* index is the standardized mean change for the experimental and control groups, and is estimated by the following equation (see Morris, 2005):

$$\Delta = \frac{(M_{\text{post,exp}} - M_{\text{pre,exp}}) - (M_{\text{post,control}} - M_{\text{pre,control}})}{SD_{\text{pre,pooled}}}, \text{ where}$$

$$SD_{\text{pre,pooled}} = \sqrt{\frac{(n_{\text{exp}}-1)SD^2_{\text{pre,exp}} + (n_{\text{control}}-1)SD^2_{\text{pre,control}}}{n_{\text{exp}} + n_{\text{control}} - 2}}$$

The final *d* index will contain a c constant for unbiased estimation: $d=c\,\Delta$, where

$$c = 1 - \frac{3}{4(n_{\text{exp}} + n_{\text{control}} - 2) - 1}$$

The unbiased effect size *d* for the word problem test was 0.62, and the unbiased d for the arithmetic test is 0.20. According to Cohen (1969), $d=0.8$ can be considered as large effect size, $d=0.5$ is medium, and $d=0.2$ is small effect size. Following this, the intervention program had a small effect on the arithmetic skills, and a notably perceptible (between medium and large) effect size on the word problem test.

Table 3 *T* tests for comparisons between experimental and control group achievement on the pre-tests, and on each of the six tasks of the word problem test

	Levene's statistic		Two-sample *t* test	
	F	*p*	|*t*|	*p*
Arithmetic skill test	1.24	0.27	2.37	0.02
Word problem test	0.81	0.37	4.29	<0.001

Table 4 *T* tests for comparisons between experimental and control group achievement on the post-tests, and on the word problem test

	Levene's statistic		Two-sample t-test or Welch test	
	F	p	$\|t\|$	p
Arithmetic skill test	0.52	0.47	0.99	0.32
Word problem test	0.40	0.53	0.28	0.78

3.3 Connections with background variables

Beyond the two mathematics-related achievement tests, students from the experimental group completed the shortened Clark Drawing test, and both the experimental and control groups received additional questions about their attitudes towards and beliefs about mathematics. For each background variable, the first aspect of analysis will be whether that variable has a significant impact on the post-test results of the experiment.

The first task of the Clark Drawing test can be considered as a manifest variable of drawing skills. Since the effect of the intervention may partly lean on students' drawing skills, we computed correlation coefficients between CDT total score and the mathematics related post-tests of the experiment. Table 5 shows the Pearson correlation coefficients. None of the values are significant.

The CDT scores were used to divide the experimental group into two equal subsamples in number. Using a threshold score of 35 allows this; and eta-squared effect size can be computed to determine the proportion of the variance on the post test scores that can be attributed to drawing skills. For the arithmetic skill test, eta-squared proved to be 0.003 and for the word problem test, eta-squared was 0.018 or 1.8%. This value belongs to the "small" effect size category according to Cohen (1969).

Since there was no significant difference between the experimental and control groups but their pre-test achievement differed significantly; in the following analyses, the increment in achievement will serve as a measure of achievement gain. In this way, we can analyze the multiple analysis of variance (ANOVA) effects of different background variables.

To compare the effect of the intervention program and students' sex, and their possible interaction, a 2 (experiment or control)×2 (boy or girl) analysis of variance was conducted, first on the increment in word problem test scores. The eta-squared index of the experimental effect (explained variance) was 0.13 ($p<0.001$) for the experimental condition, and the gender effect was not significant: eta-squared=0.001 ($p=0.62$). Neither was the interaction of the two variables significant: eta-squared=0.01 ($p=0.09$). As for the increment in the arithmetic skill test scores, the same tendencies could be revealed. Eta squared and p values (in parentheses) for the experimental conditions, for gender effect, and for interaction were 0.02 (0.04), 0.001 (0.60), and 0.001 (0.66), respectively.

Table 5 Pearson correlation coefficients between the drawing skill test score and the post-tests of the experiment ($N=94$)

	Correlation	p
Arithmetic skill test	−0.04	0.73
Word problems test	−0.16	0.13

There were five questions about the beliefs and attitudes concerning mathematics learning. Out of the five questions, three were involved in both the pre- and post-tests. Table 6 shows the main data about these background variables.

Two-sample t tests show significant differences ($p<0.05$) in the first question (how are you getting on during mathematics lessons?) for both the pre- and post-test sessions. Paired-samples t tests shows that in both the experimental and control groups the judgment about the difficulty of mathematics changed significantly ($p<0.05$). In case of the experimental group, the mean change was 0.16 which is almost twice as much as the 0.09 change in the control group.

There were two questions that appeared only in the post-test session. In case of the question "To what extent has your mathematics knowledge developed during the last 2 months?" there was no significant difference between the experiment and control groups: $t(210)=0.11$, $p=0.91$. As for the question "Do you agree that having drawn a good drawing about a word problem, you will find the solution easier?" there was significant difference between the experimental and control groups: $t(209)=2.18$, $p=0.03$. These results on the last two background variables suggest that although awareness about the development of mathematics knowledge did not differ by the end of the intervention, an important declarative metacomponent of mathematical knowledge had developed by the end of the intervention program.

4 Discussion and conclusion

The first hypothesis of the experiment stated that the experimental group would have better results on the word problem post-test. Our results suggest that the intervention program proved to be successful in terms of significant development in the experimental group comparing to the control group. The control group had a significant advantage at the beginning of the intervention period in both the arithmetic skill and the word problem test. After the 6-week long intervention period, there were no significant differences on these tests between the two groups, therefore the success of the experiment lies in the difference of the experimental group gains. Consequently, the volume of gains in achievement was assessed by computing the experimental effect size.

The second hypothesis stated that there would be equal or better results achieved on the arithmetic skill test in the experimental group than in the control group. Similarly to the results on the word problem test, there was significant difference on the pre-test, and there was no significant difference on the post-test.

Table 6 Descriptive statistics for three background variables (1=highest or most positive option, 2=middle or neutral, 3=most negative)

		Pretest		Post-test	
		Exp.	Control	Exp.	Control
How are you getting on during Mathematics lessons?	Mean	1.65	1.43	1.61	1.41
	SD	0.70	0.59	0.69	0.63
How difficult do you find Mathematics?	Mean	1.85	1.87	1.69	1.78
	SD	0.51	0.45	0.51	0.46
How important is it for you to master Mathematics	Mean	1.07	1.02	1.09	1.05
	SD	0.30	0.13	0.36	0.26

To quantify the relative magnitude of achievement change, experimental effect sizes were estimated by means of the unbiased PPC design-modified version of Cohen's *d* statistics. The results suggest that the intervention program had a notable (above-medium level) effect on students' word problem solving, and a weak but positive effect on students' arithmetic skills.

The third hypothesis concerned students' math-related beliefs. The analysis revealed that (1) the experimental group was significantly more inclined to accept that making an appropriate drawing for a word problem will make it easier to solve that problem, and (2) there was no significant difference in their perceptions of their own mathematical development during the 2-month intervention period. These results indicate that the intervention program was successful in that students of the experimental group were more aware of the importance of making drawings for mathematics word problems.

The fourth hypothesis concerned gender differences. According to the {gender} × {experimental condition} two-way ANOVA results, students' sex was not a significant variable in the experiment, and the lack of interaction with the experimental condition shows that girls and boys could equally benefit from the intervention program. In similar vein (fifth hypothesis), the intervention program was equally successful for students with higher or lower level drawing skills as shown by the {drawing skill} × {experimental condition} two-way ANOVA results.

The results of our intervention program may allow the drawing of some inferences. The program aimed at substituting the tasks and lesson plans of 20 consecutive mathematics lessons with our tasks and lesson plans. The structure of the intervention program aimed at reviewing several types of word problems, while helping students to become aware of the importance of making drawings in mathematics problem solving. It was important for us to develop an ecologically valid intervention program, i.e., provide lesson plans and supplementary materials that can be easily embedded in current lesson plans and textbooks. Our program was successful in fostering the development of not only word problem solving but the development of arithmetic skills.

There is a debate among in-service teachers about the role of word problems in mathematics teaching and learning. First, word problems can be a "dressed up" version of simple mathematics operations, therefore the means for drilling practice in order to master arithmetic skills. The second viewpoint stresses the importance of modeling realistic mathematical situations by means of using word problems in the classroom. Our results may strengthen the second viewpoint, since by using a carefully selected pool of word problems with realistic content and relations; it is still possible to help students develop their arithmetic skills, without solving too many similar tasks. By saying farewell to the drilling practice in the world of word problems, we may offer new resources for efficient classroom methods that help students become aware of their mental processes and of the importance of using appropriate visualization methods in solving word problems.

Acknowledgments This research was supported by a grant from the Hungarian Scientific Research Fund (OTKA 63360) awarded to Csaba Csíkos and by the MTA-SZTE Research Group on the Development of Competencies. Thanks are due to Gabriella Pataky for her help in the use of the Clark Drawing Test. We would also like to thank Paul Andrews, Andrea Kárpáti, Julianna Szendrei, Malcolm Swan and Lieven Verschaffel for their constructive comments on an earlier draft of this paper. Parts of the present study have been presented at a conference (Csíkos, Szitányi & Kelemen, 2009). Also, part of the data have been published in Hungarian in Magyar Pedagógia (Csíkos, Szitányi & Kelemen, 2010).

References

Báthory, Z. (2003). Rendszerszintű pedagógiai felmérések. [System-level educational surveys]. *Iskolakultúra, 13*(8), 3–19.
Bell, P. (2004). On the theoretical breadth of design-based research in education. *Educational Psychologist, 39*, 243–253.
Berends, I. E., & van Lieshout, E. C. D. M. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction, 19*, 345–353.
Boaler, J. (1994). When do girls prefer football to fashion? A analysis of female underachievement in relation to 'realistic' mathematics context. *British Educational Research Journal, 20*, 551–564.
Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*, 141–178.
Cai, J., & Lester, F. K. (2005). Solution representations and pedagogical representations in Chinese and U.S. classrooms. *Journal of Mathematical Behavior, 24*, 221–237.
Chapman, O. (2006). Classroom practices for context of mathematics word problems. *Educational Studies in Mathematics, 62*, 211–230.
Clark, G. (1989). Screeining and identifying students talented in the visual arts. Clark's Drawing Abilities test. *Gifted Child Quaterly, 33*, 98–106.
Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic.
Cooper, B., & Harries, T. (2002). Children's responses to contrasting 'realistic' mathematics problems: Just how realistic are children ready to be? *Educational Studies in Mathematics, 49*, 1–23.
Csíkos, C. (2005). *A metacogniton-based training program in grade 4 in the fields of mathematics and reading*. Paper presented at the 11th European Conference for Research on Learning and Instruction, Nicosia, Cyprus, 23–27, August, 2005.
Csíkos, C., Szitányi, J., & Kelemen, R. (2009). *Promoting 3rd grade children's mathematical problem solving through learning about the role of visual representations*. Paper presented at the 13th European Conference for Research on Learning and Instruction held in Amsterdam, The Netherlands, August 25—August 29.
Csíkos, C., Szitányi, J., & Kelemen, R. (2010). Vizuális reprezentációk szerepe a matematikai problémamegoldásban. Egy 3. osztályos tanulók körében végzett fejlesztő kísérlet eredményei. [The role of visual representations in mathematical problem solving: Results of an intervention program with 3rd grade students]. *Magyar Pedagógia, 110*, 149–166.
Diezmann, C. M. (2005). Primary students' knowledge of the properties of spatially-oriented diagrams. In H. L. Chick & J. L. Vincent (Eds.), *Proceedings of the International Group for the Psychology of Mathematics Education* (pp. 281–288). Melbourne: PME.
Elia, I., Gagatsis, A., & Demetriou, A. (2007). The effects of different mode of representation on the solution of one-step additive problems. *Learning and Instruction, 17*, 658–672.
English, L. D. (1996). Children' construction of mathematical knowledge in solving novel isomorphic problems in concrete and written form. *Journal of Mathematical Behavior, 15*, 81–112.
English, L. D. (1997). The development of fifth-grade children's problem solving abilities. *Educational Studies in Mathematics, 34*, 183–217.
English, L. D., & Halford, G. S. (1995). *Mathematics education: Models and processes*. Mahwah, NJ: Erlbaum.
Gagatsis, A., & Elia, I. (2004). The effects of different modes of representations on mathematical problem solving. In M. Johnsen-Hoines & A. Berit-Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education: Vol. 2* (pp. 447–454). Bergen, Norway: Bergen University College.
Geary, D. C. (1999). Sex differences in mathematical abilities: Commentary on the math-fact retrieval hypothesis. *Contemporary Educational Psychology, 24*, 267–274.
Goldin, G. A., & Kaput, J. (1996). A joint perspective on the idea of representation in learning and doing mathematics. In L. Steffe, P. Nesher, P. Cobb, G. Goldin, & B. Greer (Eds.), *Theories of mathematical learning* (pp. 397–430). Hillsdale: Erlbaum.
Gravemeijer, K., & Cobb, P. (2006). Design research from a learning design perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 17–51). London: Routledge.
Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12*, 229–246.
Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*, 684–689.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*, 18–32.

Jonassen, D. H. (2003). Designing research-based instruction for story problems. *Educational Psychology Review, 15*, 267–296.

Kárpáti, A. (2001). *Firkák, formák, figurák. A vizuális nyelv fejlődése a kisgyermekkortól a serdülőkorig. [Scribbles shapes and figures. The visual language of children]*. Budapest Hungary: Dialóg Campus Kiadó.

Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer–verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction, 20*, 47–77.

Kramarski, B., Mevarech, Z. R., & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. *Educational Studies in Mathematics, 49*, 225–250.

Kramarski, B., Mevarech, Z. R., & Lieberman, A. (2001). Effects of multilevel versus unilevel metacognitive training on mathematical reasoning. *The Journal of Educational Research, 94*, 292–300.

Lave, J. (1992). Word problems: A microcosm of theories of learning. In P. Light & G. Butterworth (Eds.), *Context and cognition. Ways of learning and knowing* (pp. 74–92). Hillsdale, NJ: Erlbaum.

Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 29–53). Mahwah, NJ: Erlbaum.

Mevarech, Z. R., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 2*, 365–394.

Morris, S. B. (2005). *Effect size estimation from prestest–posttest–control designs with heterogeneous variances.* Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction, 16*, 173–199.

Nemzeti Alaptanterv (2007) [National Core Curriculum] Available from: http://www.okm.gov.hu/letolt/kozokt/nat_070926.pdf.

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55*, 169–195.

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 67*, 37–58.

Pantziara, M., Gagatsis, A., & Elia, I. (2009). Using diagrams as tools for the solution of non-routine mathematical problems. *Educational Studies in Mathematics, 72*, 39–60.

Presmeg, N. (1986). Visualisation in high school mathematics. *For the Learning of Mathematics, 6*, 42–46.

Sáenz-Ludlow, A., & Walgamuth, C. (1998). Third-graders interpretations of equality and the equal symbol. *Educational Studies in Mathematics, 35*, 153–187.

Schneider, M., Heine, A., Thaler, V., Tornbeyns, J., De Smedt, B., Verschaffel, L., et al. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development, 23*, 409–422.

Selter, C. (1998). Building on children's mathematics—a teaching experiment in grade three. *Educational Studies in Mathematics, 36*, 1–27.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237–243.

Uesaka, Y., Manalo, E., & Ichikawa, S. (2007). What kinds of perceptions and daily learning behaviors promote students' use of diagrams in mathematics problem solving? *Learning and Instruction, 17*, 322–335.

Van Dooren, W., Verschaffel, L., & Onghena, P. (2003). Pre-service teachers' preferred strategies for solving arithmetic and algebra word problems. *Journal of Mathematics Teacher Education, 6*, 27–52.

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Edcuational Pschology, 31*, 142–166.

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review, 17*, 285–325.

Verschaffel, L., & De Corte, E. (1997a). Word problems: A vehicle for promoting authentic mathematical understanding and problem solving in the primary school? In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 69–97). Hove, UK: Psychology Press.

Verschaffel, L., & De Corte, E. (1997b). Teaching mathematical modeling and problem-solving in the elementary school. A teaching experiment with fifth graders. *Journal for Research in Mathematics Education, 28*, 577–601.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modelling of school arithmetic word problems. *Learning and Instruction, 7*, 339–359.

Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Design and evaluation of a learning environment for mathematical modeling and problem solving in upper elementary school children. *Mathematical Thinking and Learning, 1*, 195–229.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, The Netherlands: Swets & Zeitlinger.

Verschaffel, L., Greer, B. & Torbeyns, J. (2006). Numerical thinking. In Boero, P. & Gutiérrez, A. (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 51–82). Rotterdam, the Netherlands: Sense Publishers.