Scandinavian Journal of Educational Research
Vol. 54, No. 4, August 2010, 335–355

Instructing Low-Achievers in Mathematical Word Problem Solving

Anu Kajamies, Marja Vauras, and Riitta Kinnunen
University of Turku

We describe the effects of an intervention designed to develop the mathematical word problem solving of low-achievers. The eight students participating in the intervention were selected from 429 10-year-olds on the basis of their difficulties in word problem solving. In the intervention, we combined intensive, systematic, and explicit teacher scaffolding in the cognitive, metacognitive, and motivational activities involved in skillful problem solving with carefully designed word problems embedded in a computer-supported adventure game. The results from the pre-test, post-test, and follow-up test indicate significant effects for the intervention students' word problem solving compared to the two control groups. A single-subject study describes the results also at the individual level.

Keywords: problem solving, low-achiever, intervention, computer-supported environment

Studies in mathematical word problem solving have convincingly documented that many students, especially low-achievers, do not even try to understand the problems, but jump immediately into calculations with the numbers mentioned in the problem and give the answer without confirming that it is meaningful (Bryant, Bryant, & Hammill, 2000; Mayer & Hegarty, 1996; Reusser & Stebler, 1997; Schoenfeld, 1988; Verschaffel, Greer, & De Corte, 2000). Intervention studies have indicated difficulties in developing the mathematical word problem solving skills of low-achievers (Fuchs & Fuchs, 2005; Jitendra & Xin, 1997; Kroesbergen & Van Luit, 2003; Verschaffel & De Corte, 1997a). Although average-level improvements in students' skills may follow from interventions, our continuing series of intervention studies (Lehtinen, Vauras, Salonen, Olkinuora, & Kinnunen, 1995) have shown that it is very demanding to develop the skills of students with deficient self-regulation and motivational vulnerability (Vauras, Rauhanummi, Kinnunen, & Lepola, 1999).

Components of Word Problem Solving

When trying to develop students' skills it is vital to remember that problem solving depends on the integrated application of cognitive, metacognitive, and motivational

Anu Kajamies, Centre for Learning Research and Department of Teacher Education, University of Turku; Marja Vauras, Centre for Learning Research and Department of Teacher Education, University of Turku; Riitta Kinnunen, Centre for Learning Research and Department of Teacher Education, University of Turku.

This research was supported by Grant 47369 and 114048 from the Council for Cultural and Social Science Research, the Academy of Finland. We wish to thank the participating headmasters, teachers, and students, who unfortunately must remain anonymous.

Correspondence concerning this article should be addressed to Anu Kajamies, Department of Teacher Education, University of Turku, FIN-20014 Turku, Finland. E-mail: Anu.Kajamies@utu.fi

ISSN 0031-3831 print/ISSN 1470-1170 online

DOI: 10.1080/00313831.2010.493341
http://www.informaworld.com

components. A well-organized and flexibly accessible knowledge base of mathematical facts, symbols, algorithms, concepts, and rules is the basis of word problem solving (Geary, 2004; Riley & Greeno, 1988; Verschaffel et al., 2000). When the use of these basic elements has become automatic, *more* resources can be devoted to word problem solving (Gersten & Chard, 1999). Word problem solving also requires reading comprehension (Bryant et al., 2000; De Corte, Verschaffel, & Pauwels, 1990; Nathan, Kintsch, & Young, 1992; Vauras, Kinnunen, & Rauhanummi, 1999), as well as real-world knowledge about the situations involved (Verschaffel et al., 2000). Students often know how to carry out the calculations, but have difficulties in understanding the situations (Mayer & Hegarty, 1996; Montague & Applegate, 1993).

When trying to solve complex word problems, many students do not spontaneously apply valuable heuristic strategies (Schoenfeld, 1988; Verschaffel et al., 1999), such as making a drawing of the problem situation (De Bock, Verschaffel, & Janssens, 1998). Metacognitive skillfulness, such as monitoring the solution process and evaluating its outcome, is absent in a majority of students' solution attempts (Carr & Biddlecomb, 1998; De Corte, Verschaffel, & Op't Eynde, 2000; Desoete, Roeyers, & Buysse, 2001). The failure of some students also results from inadequate mathematical beliefs, together with negative emotions (Kloosterman, 2002; Ma, 1999; McLeod, 1992; Op't Eynde, De Corte, & Verschaffel, 2001). When confronted with demanding tasks, the low-achievers typically lack task-orientation and show motivational vulnerability, for example, in the form of mental disengagement and helplessness (Salonen, Lehtinen, & Olkinuora, 1998).

Developing Word Problem Solving

Students' superficial word problem solving is deep-rooted, and radical interventions are needed to alter it (Reusser & Stebler, 1997). Verschaffel and De Corte (1997a) carried out an extensive study in which word problems were conceived as exercises in mathematical modeling. The study focused on the assumptions and appropriateness of the model underlying any proposed solution. It showed that more realistic mathematical modeling can be developed in elementary school students, but the improvement was more marked in the high-achievers (Verschaffel & De Corte, 1997a; see also Baxter, Woodward, Voorhies, & Wong, 2002). To find out how all students, in particular low-achievers, would benefit from interventions, more research is needed. The research so far has concluded that providing more intensive, systematic, and explicit coaching can also improve the performance of low-achievers (Fuchs & Fuchs, 2005; Kroesbergen & Van Luit, 2003; Swanson, 1999; Vauras, Rauhanummi, Kinnunen, & Lepola, 1999; Verschaffel & De Corte, 1997a; Xin & Jitendra, 1999). It is critical to find ways to support both task orientation and metacognition while teaching cognitive strategies (Boekaerts, 2002; Borkowski et al., 1992; Bottge, 2001; De Corte, Verschaffel, & Masui, 2004; Lehtinen et al., 1995; Mayer, 1998), because long-term metacognitive and motivational incompetence severely interferes with benefiting from instruction (Meichenbaum & Biemiller, 1998; Vauras, Rauhanummi, Kinnunen, & Lepola, 1999).

In this article our research question is whether low-achievers benefit from an intervention designed to develop students' word problem solving with the help of teacher scaffolding in the cognitive, metacognitive, and motivational activities involved in skillful problem solving, and carefully designed word problems embedded in a computer-supported adventure game.

Methods

Sample

Altogether there were 429 students participating in this study. The mean and median age of the students was 10 years and 4 months (SD 4 months) at the beginning of the study. All students spoke Finnish as their mother tongue and had parental permission to participate in the study. The students were from 12 schools and 21 classes. The schools were situated in cities, small towns, and rural communities in southern Finland. The students followed the mainstream curriculum of Finnish general education, including teaching in mathematics. The teachers reported in the interviews that they emphasized fluency in arithmetical skills rather than word problem solving. The analysis of teachers' answers in the interviews showed that in their teaching, guidance towards superficial word problem solving, rather than understanding, was salient.

Measures

Word problem solving.

Word problem solving was assessed with 15 one-step and multi-step problems (*group measurements*). Intervention students' were also tested by one one-step (approximately 43 words) and one three-step (approximately 105 words) problem (*single-subject measurements*). Problems in group measurements were different from those in single-subject measurements. All problems were constructed on the basis of problems used in earlier studies (Verschaffel et al., 2000) and reported in Kajamies, Vauras, Kinnunen, and Iiskala (2003). Problems demanded acute realistic consideration, not just straightforward application of arithmetical operations. The problems at different measurement points were structurally identical with different names and numbers to achieve an equal level of difficulty. There were no time limits in the tests. Examples of the problems are shown in Tables 1 and 2.

As an instruction, students were given an example as to how to do the tasks in the way required for maximum points. To make the solving process of the students as visible as possible, writing down the calculation steps was especially emphasized. In each problem, the student was given two points for each correct calculation step and for the correct answer. If the student made calculation errors in any step, only one point was given for that step. The total number of points from calculation steps and answers at one measurement point is here used as an indication of the student's word problem solving skills. The maximum score was 86 in group measurements and 12 in single-subject measurements. The alpha coefficient was .77 in group

Table 1
Examples of Problems in Group Measurements

Problems	Calculation steps	Answer
You are playing basketball with your friends. The team opposing you gets 24 points in the first period. It makes three points less than your own team. Both teams get the same amount of points in both periods. How many points were got in the whole game?	24 + 3 = 27, 27 + 24 = 51, 51 × 2 = 102	102 points
Twenty-two congressmen were taken to a presidential banquet by cab. One cab took four passengers. How many cabs were needed?	22/4 = 5, remainder 2	6 cabs

Table 2
Examples of Problems in Single-Subject Measurements

Problems	Calculation steps	Answer
The big brother of your friend takes a fitness test. He has once been able to run almost 200 meters during the test time. However, now he has not been keeping fit for a long time. He runs an average of seven metres and 10 centimeters per second. He has 20 seconds for the test run. How long a distance does he run during that time?	7.10 × 20 =142	142 metres
Your best friend's mother decides to throw a party for kids on Valentine's day, February the 14th. The party will surely be fun, since soap bubbles will be blown. All the boys and girls in the yard are invited. That makes eight children. The boys are very keen and blow 52 bubbles. The girls first admire the way the boys' bubbles glitter in the sun, and hence they find time to blow 16 bubbles less than the boys. The boys and the girls together break bubbles they blew by clapping their hands. Suppose the children make a deal: everyone can break the same amount of bubbles. How many bubbles will each child break? Everyone is so excited that all of the bubbles get broken. And before the children go and eat some tidbits, they blow 11 colorful balloons.	52 – 16 = 36, 52 + 36 = 88, 88/ 8 = 11	11 bubbles

measurements and .91 in single-subject measurements. To assess the rater agreement, the tests of the students in experimental groups (n = 24, 5.6%) in group measurements were scored, not only by the first author, but also by a trained researcher who was unaware of the purpose and the design of the study and the scoring made by the first author. Agreement percentages between 97 and 99 were obtained. Disagreements were solved by discussion.

Task orientation.

The amount of task orientation in students' typical classroom behavior was assessed with a Likert-type scale by the classroom teachers (Salonen, Kajamies, & Vauras, 2010). Examples of the items are "Works persistently," "Tries to solve problems independently," and "Is considering how things fit together." The minimum score was 1 (never) and the maximum 5 (very often). A mean score for eight items is used here as an indication of a student's task orientation. The alpha coefficient was .95.

Arithmetical skills.

Arithmetical skills were assessed with a time-limited (10 minutes) RMAT test (Räsänen, 2004). The RMAT starts with simple computations and ends with algebraic tasks. The RMAT is comparable to the WRAT-R (Jastak & Wilkinson, 1984) with similar instructions, but the RMAT follows the Finnish mathematics curriculum more closely (e.g., the role of fractions is small) (Räsänen, 1993). The RMAT contains more computational tasks and thus it can measure more basic arithmetical skills than the WRAT-R (correlations .547–.659, n = 2673) (Räsänen, 2004). The total number of correct solutions in the RMAT is here used as an indication of the student's arithmetical skills. The maximum score was 56. The alpha coefficient was .92–.95 (Räsänen, 2004).

Non-verbal intelligence.

Standard progressive matrices (Raven, Raven, & Court, 2000) were used to measure non-verbal intelligence, that is, a more general level of students' skills that is not dependent on linguistic skills. The total number of correct choices is here used as an indication of the student's non-verbal intelligence. The maximum score was 60. Raven et al. (2000) report that the test is reliable, whether in terms of internal consistency or retest reliability.

Reading comprehension.

Reading comprehension was assessed with the Finnish Standardized Reading Test (Lindeman, 1998) in which the student was given 48 multichoice questions about the four texts s/he had read. Reading comprehension was seen as an important measure of the linguistic skills of 4th graders. The total number of students' correct choices is used to classify the students into reading skill groups that are here used as indications of the student's reading comprehension skills. The following reading skill groups were formed: poor (1–3), average (4–6), and skilled (7–9). The Kuder-Richardson coefficient of internal consistency (CR20) is .87 (Lindeman, 1998).

Procedure

The research plan was accepted by the Ethical committee of the Academy of Finland. A pre-test, post-test, and follow-up test design with control groups was used. Before the intervention, all students participated in the pre-test on word problem solving. Parallel tests were given to all students after the intervention in the post-test and in the follow-up test. The test was always given in two parts. Before the intervention, students' task orientation, arithmetical skills, non-verbal intelligence, and reading comprehension were also measured. The measurements before the intervention were made in October–January when the students were in the 4th grade. Post-tests took place in May. Follow-up tests were carried out half a year later, in November–December. All tests were given in a classroom situation by trained researchers. If a student was absent in the classroom, tests were administered later individually.

Furthermore, we measured intervention students' word problem solving before the intervention (baseline phase, six sessions), during the intervention (treatment phase, 14 sessions), and after the intervention (post-treatment phase, six sessions). These parallel measurements were carried out twice a week. Both the baseline and the post-treatment phase lasted three weeks and included no instruction. The baseline phase started in the same week for all students and was followed by other phases without breaks. In the treatment phase, which lasted seven weeks, the measurements were followed by instruction in word problem solving. To allow more time for instruction, the one-step and the three-step problems were given in consecutive sessions.

Subgroups

To select the *intervention students*, those students who got low pre-test scores in word problem solving (below 38 points, cumulative percentage 32, $n = 138$) were selected from all the students. For practical reasons, the intervention students, four boys and four girls, were selected from these low-achievers from two classes (*n* = 16). Pairwise-matched, same-sex controls were selected from the remaining low-achievers (*n* = 130) on the basis of their scores in word problem solving, arithmetical skills, and non-verbal intelligence. Precautions were taken to eliminate large differences in task orientation within a pair. Two control students were selected for each intervention student. Half of the control students did not get any special attention during the intervention (*control*), while the other half took part in a reading comprehension strategy intervention that was carried out in 20 hours by special teachers in the Quest for Meaning project (*rc-control*), of which this study is a part. Among the low-achievers, 41 participated in the reading comprehension intervention. Eight of these students were selected as rc-control students to see whether special attention and scaffolding in reading comprehension are enough to develop word problem solving (cf. Borasi, Siegel, Fonzi, & Smith, 1998). Because the control students were from 11 different classes, an in-depth analysis of how their mathematics instruction was carried out was not possible. Intervention and rc-control students had parental permission to participate in the interventions. After the study, special teachers in the Quest for Meaning project were given the computer program used in the study and trained to implement these kinds of interventions to also give other students, e.g., control students, a possibility to participate.

Some characteristics of the intervention students are shown in Table 3. The intervention students seldom showed task orientation in their classroom behavior (mean 1.97). Peter and Jan never showed task orientation. Peter and Anna had the poorest arithmetical skills of the intervention students. Henry, Karl, and Tina's non-verbal intelligence was below average. Karl had poor reading comprehension skills.

We discuss the skills of four groups: intervention (*n* = 8), control (*n* = 8), rc-control (*n* = 8), and other students (*n* = 405), who served as a comparison group to establish the typical skill level. Differences between the groups were analyzed with one-way analysis of variance (ANOVA). Statistically significant differences were further analyzed with *post hoc* tests. In the light of group characteristics (see the bottom of Table 3), matching was

Table 3
Characteristics of the Intervention Students and the Groups

Intervention student	Task orientation	Arithmetical skills	Non-verbal intelligence	Reading comprehension
Peter	1.00	23	44	4
Henry	2.75	33	36	5
Karl	2.25	29	32	3
Jan	1.00	26	45	4
Tina	1.88	34	35	5
Nina	2.50	26	42	7
Monica	2.25	32	42	5
Anna	2.13	22	38	6
Group	Mean and (standard deviation)			
Intervention	1.97 (.65)	28.13 (4.58)	39.25 (4.68)	4.88 (1.25)
Control	2.69 (.64)	25.38 (3.89)	39.63 (5.78)	6.00 (1.20)
Rc-control	2.25 (.51)	28.00 (3.46)	38.38 (3.70)	4.00 (1.31)
Other	3.37 (.82)	32.63 (5.68)	41.53 (6.52)	6.32 (1.71)

successful, because the only difference between intervention, control, and rc-control students was found in reading comprehension $F(2, 21) = 5.137$, $MSE = 8.0$, $p < .05$. Since the main goal of the Quest for Meaning project was to develop the reading comprehension skills of the students with the severest difficulties, rc-control students performed more poorly than control students in reading comprehension ($p < .05$). Other students differed from intervention, control, and rc-control students in task orientation $F(3, 425) = 14.231$, $MSE = 9.3$, $p < .001$, arithmetical skills $F(3, 425) = 7.600$, $MSE = 238.7$, $p < .001$, and reading comprehension $F(3, 425) = 6.818$, $MSE = 19.4$, $p < .001$. Compared to other students, intervention and rc-control students were less task orientated; control students had poorer arithmetical skills; and rc-control students had poorer reading comprehension ($p < .01$).

Intervention

On the basis of the skills needed in word problem solving and earlier attempts at developing word problem solving, we designed an intervention to promote the word problem solving of low-achievers. The most important objective was to get the students to understand that skillful problem solving is a complex, cyclical process that includes understanding the situation described, constructing a mathematical model describing the relevant relationships involved, working through the model to identify the appropriate calculations, and evaluating and interpreting their outcomes in relation to the original situation (cf. Greer, Verschaffel, Van Dooren, & Mukhopadhyay, 2009; Verschaffel et al., 2000). In order to lay the basis for learning, the acquisition of a set of appropriate beliefs and positive attitudes with regard to word problems was also a vital aim (cf. Mason & Scrivani, 2004; Verschaffel et al., 2000). Students should become more active, strategic, and motivated solvers of word problems.

The intervention was carried out by the first author in February–April when the students were in the 4th grade. There were two students present in each session (Peter and Henry, Karl and Jan, Tina and Nina, or Monica and Anna), and they played the adventure game together. Intervention was implemented consistently across the students. Each session was arranged in a quiet room at the students' schools and lasted about 45 minutes. The intervention lasted seven weeks, and one of the two-weekly sessions was held outside regular school hours.

Word problems embedded in a game environment.

In the intervention, a pilot version of the computer-supported adventure game called the Quest of the Silver Owl (Vauras & Kinnunen, 2003) was used as a tool for developing students' word problem solving. The adventure goal in the game is to find the Silver Owl, which allows the two adventurers to save the Realm of Secret Numbers. The owl can be found after the adventurers have acquired enough points by solving word problems that reflect the spirit of the game. In the game, standard word problems were replaced by carefully designed problems to elicit negative feedback on superficial solving strategies, and to create a need for more realistic word problem solving (cf. Verschaffel & De Corte, 1997a). In contrast to many earlier studies, which only used simple one-step problems, we considered it important to develop complex word problem solving as well. The problems vary from simple (e.g., compare and combine, addition and subtraction) to complex tasks (e.g., multiplication, division and mixed tasks with equal measures, rate, area, conversion, and Cartesian product).

The adventurer in turn chooses the difficulty of the problem from four difficulty levels (Mistland, Rainland, Fireland, and Magicland). The more difficult the problem is, the more points the adventurer in turn can obtain from the correct answer. The other adventurer checks the answer, and s/he is given one point, if s/he has correctly assessed the other student's solution. If the program gives the feedback that the solution is wrong, the adventurer has the possibility to correct it and still get half of the points. The whole intervention session was spent playing the game. Two built-in game wizards gave game instructions, a picture of the problem, and verbal hints, if asked for. The adventurers often looked at the pictures. Verbal hints that tried to direct the adventurers' attention to the relevant information in the problem were seldom used, because if verbal hints were asked for, one point was lost. Asking for pictures did not influence the point situation.

On the basis of earlier research (Cognition and Technology Group at Vanderbilt, 1997; Vauras, Rauhanummi, Kinnunen, & Lepola, 1999), we assumed that the first successful step in developing the students' skills requires that students engage in innovative task environments in which they overcome their anxiety and become intellectually more involved. The game structure and goals, multiple feedback on progress, and an attractive graphic environment were designed as the main motivational incentives. A computer-supported environment was used because the reviews and meta-analyses have shown that students using computers learn more, and more quickly, than students in control groups, and they also show improved motivation (Lehtinen, 2003; McLeod, 1992). Admittedly, relatively few studies have produced conclusive findings on computer-supported mathematics instruction for low-achievers, and most of the existing findings concern the practising of computation skills, not word problem solving. Nevertheless, it can be concluded that computers cannot replace teachers in the instruction of low-achievers, but when combined with teacher support, as in this study, computers can be effective (Kroesbergen & Van Luit, 2003; Salomon, Perkins, & Globerson, 2000; Xin & Jitendra, 1999).

Instructional discussion.

A crucial element of the intervention was the instructional discussion between the students and the teacher (cf. Pressley et al. [1992] on transactional dialogue) while the students were playing the game. The role of the teacher was to scaffold the students to engage in and to reflect upon the cognitive, metacognitive, and motivational activities involved in the model of skillful problem solving (see Table 4).

Students often skipped steps in the model, and the teacher had to remind them about going carefully through each step. At the beginning of the intervention, students typically picked two numbers from the text and based their selection of the operation on a superficial understanding of the text. For example, they were guided by isolated keywords, such as "more" or "less" and associated them directly with operations instead of realizing their status in the problem (cf. Hegarty, Mayer, & Monk, 1995; Schoenfeld, 1988; Verschaffel & De Corte, 1997b). During the intervention, the students were encouraged to carefully read the whole problem aloud, and try to understand the problem situation in order to correctly select the mathematical operations. The teacher's support was also needed in carrying out the steps. For example, it was not enough that the teacher asked the students to draw a picture of the problem situation in their notebooks. The teacher had to help the students find

Table 4
The Model of Skillful Problem Solving the Students were Instructed to Follow

Step 1. Read the problem carefully from the beginning to the end.
Step 2. Construct a representation of the problem.
- Think step by step what is happening in the task.
- Think carefully what is asked.
- Distinguish the relevant from the irrelevant, e.g., by underlining.
- Use your previous knowledge and experiences.
- Draw a picture of the important relationships.

Step 3. Decide how to solve the problem.
- Think step by step what kind of calculation you need (what is decreasing, increasing, happening many times, or being divided).

Step 4. Execute the necessary calculations.
- Use aids, e.g., concrete materials and paper and pencil calculation, if needed.

Step 5. Interpret the outcome and formulate the answer.
- Think whether you have answered the question.

Step 6. Evaluate the solution.
- Read the task again.
- Think whether you have figured out the task correctly.
- Check all calculations.
- Check whether the answer is reasonable or not.

Source: Vauras, Kinnunen, Kajamies, and Iiskala [2003], modified from Montague, Warger, and Morgan [2000], Polya [1957], and Verschaffel et al. [1999].

the important relationships in the text and make a simplified picture or comic strip of them. Without help, the students easily found a way to avoid the task by drawing nice pictures that had little to do with the task. The teacher also helped the students to activate relevant previous knowledge and experiences. For example, she reminded the students of the things they had learned in earlier problems. Students needed help with calculations, particularly with multiplication and division with big numbers. In the calculations and in using the game, students sometimes got help from each other, but otherwise their collaboration was not very effective. One of the most difficult steps for the students was evaluating the solution. Typically, the students did not do this at all because they thought that it was the duty of the teacher. When the teacher asked them to evaluate, they only checked the calculations and found it very hard to understand why the teacher asked them to go through the whole solving process critically. The most productive discussions took place after the program had told the students that the solution was wrong and they had a possibility to correct.

Instruction consisted of scaffolding procedures such as questioning and modeling in students' zones of proximal development (Vygotsky, 1978). The teacher offered immediate and concrete feedback. As the student showed increasing mastery, the teacher faded out her role, and encouraged the student to take more responsibility. The teacher tried to get the students to understand that it is important to be able to describe their solution process instead of just giving an answer (cf. Gravemeijer, 1997). At the beginning of the intervention, students typically erased their calculations and started new ones when the teacher asked them to describe what they had done and why, but, little by little, the students learned that the teacher always asked them to describe how they had solved the task. When listening to students' descriptions, the teacher tried to understand why students made the errors they did in order to help them to solve the tasks more efficiently. The teacher repeatedly pointed out to the students that the most important thing is not to solve the task as quickly as possible. Instead, investing effort and time in the solving process is worthwhile.

Results

Development of Word Problem Solving in the Group Measurements

To test whether the intervention developed the intervention students' skills, we analyzed the word problem solving scores using a 3 (Time) × 4 (Group) univariate analysis of variance (ANOVA) with repeated measures from the pre-test, post-test and follow-up test. The Type III sums of squares were used because they are invariant with respect to the cell frequencies, and therefore useful for this kind of unbalanced model. The analysis revealed significant main effects for Time, $F(2, 850) = 26.56$, $MSE = 77.16$, $p < .001$, and for an interaction between Time and Group, $F(6, 850) = 3.11$, $MSE = 77.16$, $p < .01$. The powers of the test were 1.00 and 0.92, respectively. The means, standard deviations and confidence intervals for means are shown in Table 5.

Main and interaction effects were further analyzed with contrast analysis, where user-specified *a priori* contrasts were tested by *t*-statistics as a function of ANOVA. When the scores in the pre-test were contrasted, no differences were found between the intervention and the two control groups, but all these groups obtained lower scores than other students ($p < .001$). This result is based on the fact that word problem solving in the pre-test was used as the inclusion and matching criterion.

Table 5
Scores in Word Problems in Group Measurements

Student	Pre-test	Post-test	Follow-up
Peter	4	33	22
Henry	50	35	47
Karl	17	46	46
Jan	17	45	56
Tina	21	39	47
Nina	31	49	60
Monica	22	51	57
Anna	21	23	18
Group	**Mean and (standard deviation) 95% confidence interval for mean: lower-upper bound**		
Intervention	22.88 (13.28) 11.77–33.98	40.13 (9.46) 32.21–48.04	44.13 (15.82) 30.90–57.35
Control	22.00 (11.44) 12.44–31.56	30.13 (12.47) 19.70–40.55	36.00 (16.84) 21.93–50.07
Rc-control	23.00 (12.58) 12.48–33.52	31.63 (14.01) 19.91–43.34	35.63 (17.58) 20.93–50.32
Other	45.28 (16.06) 43.71–46.85	48.53 (15.92) 46.98–50.09	51.81 (15.64) 50.29–53.34

When the scores in the pre-test and post-test were contrasted, there was an increase in the scores of the intervention students and the other students ($p < .001$). The increase in rc-control students' scores was also significant ($p < .05$). In the control group the progress was not significant. When the scores in the post-test and delayed test were contrasted, an increase was found only in the scores of the other students ($p < .001$). When the gains were contrasted, the only difference was found between intervention students and other students. Intervention students' scores increased from pre-test to post-test significantly more than other students' scores ($p < .01$).

Interestingly, when the scores in the post-test and follow-up test were contrasted, the intervention students no longer differed from other students. Control and rc-control students still obtained lower scores than other students ($p < .01$). The differences between the intervention, control, and rc-control groups failed to reach significance.

Effect sizes were calculated by dividing the difference between control and intervention students' gain scores from pre-test to post-test by the pooled standard deviation of the pre-test (average of control and intervention students' pre-test standard deviations) (Cohen, 1988; Ives, 2003). Cohen's definition (1988) was used as the criterion for the magnitude of effect size: .20 in absolute value is a small, .50 a moderate, and .80 a large effect size. When comparing intervention students' scores to control and rc-control students' scores, moderate effect sizes were found (0.61 and 0.65, respectively). These results indicate that the intervention resulted in some positive and lasting effects, and that special attention and instruction in reading comprehension may develop the word problem solving more than the lack of special attention.

In Table 5, the scores of the intervention students in word problems are shown. All intervention students, except Henry, had higher scores in the post-test than in the pre-test. It is plausible to suggest that Henry's lower scores were due to his poor concentration in classroom test situations. Henry was selected for the intervention on the basis of his very low scores in the first part of the pre-test. Because of illness, he was not present in the classroom for the second part of the pre-test. After parental permission to take part in the intervention had been granted, we were able to administer the second part of the test to him. To our surprise, in this individual test situation, Henry was able to concentrate on the problems and solved nearly all of them correctly. This resulted in his above average scores in the pre-test. In the post-test and in the follow-up test, Henry was present in the classroom for measurements and, afterwards, proudly reported to the first author that he was the quickest student in the class. As we can see from his scores, he still had difficulties concentrating in classroom test situations. Since removing Henry and his controls from the variance analysis did not substantially influence the results, they were included in the analysis.

Peter, who got the lowest score in the pre-test, gained most; he obtained more than eight times more points in the post-test than in the pre-test. Karl, Jan, and Monica more than doubled their scores from pre-test to post-test. All intervention students, except Peter, maintained, in the follow-up test, the level they had gained in the post-test. Anna's scores were fairly stable from pre-test to follow-up test.

To see whether task orientation, arithmetical skills, non-verbal intelligence, or reading comprehension had an effect on the development of the word problem solving skills of the groups from pre-test to post-test, these control measures were used one at a time as a covariate in univariate analysis of variance (ANOVA). The analyses revealed that the groups differed in their development even though covariants were used ($p < .01$, after Bonferroni

Table 6
Scores from Correct Answers Only in Word Problems in Group Measurements

Group	Mean and (standard deviation) 95% confidence interval for mean: lower-upper bound		
Intervention	6.38 (5.37)	9.88 (4.36)	10.13 (6.08)
	1.89–10.86	6.23–13.52	5.04–15.21
Control	4.75 (3.24)	6.88 (4.52)	9.25 (6.23)
	2.04–7.46	3.10–10.65	4.04–14.46
Rc-control	4.38 (3.16)	6.75 (4.13)	9.25 (6.16)
	1.73–7.02	3.30–10.20	4.10–14.40
Other	12.74 (6.02)	13.64 (6.42)	14.81 (6.44)
	12.15–13.33	13.02–14.27	14.19–15.44

correction). When other students' development was contrasted with intervention, control, and rc-control students' development, contrasts showed that intervention students developed more than other students, unlike control and rc-control students. When intervention students' development was contrasted with control and rc-control students' development, contrasts showed that intervention students' development did not differ from control and rc-control students' development.

Non-parametric Kurskal Wallis tests showed that the development of the word problem solving skills of the groups from pre-test to post-test differed, $\chi^2 = 10.282$, $p < .05$. Parametric tests were used to find out the contrasts and whether covariants had an effect on the development.

To see whether the same results could be achieved by analyzing word problem solving scores from *correct answers only*, we used 3 (Time) × 4 (Group) univariate analysis of variance (ANOVA) with repeated measures from the pre-test, post-test and follow-up test. The analysis revealed no interaction between Time and Group, $F(6, 850) = 1.037$, $MSE = 12.63$, $p > .05$. The means, standard deviations and confidence intervals for means are shown in Table 6. This means that a more subtle way of analyzing the solutions, by also using points for calculation steps, was needed to see the differences in the developments of the groups. The aim of word problem solving is to find the correct answer, but for low-achievers this aim, particularly in complex problems, is so far that steps in the right direction must be valued.

Development of Word Problem Solving in the Single-Subject Measurements

The data from single-subject measurements are described by time-series figures with mean lines. Changes in level, trend, and variability were analyzed from the figures. Graphic presentation of the data is used because this has a long and strong history in single-subject research, and visual analysis can often provide snapshot views of the intervention effects (Franklin, Allison, & Gorman, 1996). The means, standard deviations, and effect sizes are given, since reliance on visual analysis alone does not establish whether interventions have produced changes that are greater than chance levels (Franklin, Gorman, Beasley, & Allison, 1996).

Although there is no consensus on procedures for calculating effect size for single-subject studies (Swansson, 1999, p. 34), the effect size was understood as the difference between means scores of the baseline and treatment phases divided by the standard deviation of the baseline phase (Kromrey & Foster-Johnson, 1996). Because autocorrelation can be a problem when calculating effect sizes for single-subject studies (Matyas & Greenwood, 1996; Robey, Schultz, Crawford, & Sinner, 1999), we corrected for the correlation between baseline and treatment phases by using the following denominator: $M_{SDBaselineSDTreatment}/[2\ (1 - r)]^{1/2}$ (Swansson, 1999, p. 34).

When looking at the single-subject measurements at different measurement points, a rather large variability in students' scores can be noted (see Figure 1 and Table 7). The large variability should be taken seriously because we want to draw some conclusions about the development of the students' skills. When we compare the student's scores during baseline to her/his scores in treatment and post-treatment phases, we can see that the word problem solving of Peter, Monica, and Anna, especially, has developed. Peter and Anna started from quite low scores. At one measurement point Anna managed to get eight points. During the treatment and post-treatment phases Monica got quite high scores, the maximum score three times. Effect sizes indicate large intervention effects for these students.

The differences between the scores from different measurement points were largest for Henry, Jan, and Karl. During the treatment phase, some stability appeared in Henry's behavior. Henry and Jan's means indicate some positive changes during the post-treatment phase. At the last measurement points, Jan's scores were near the maximum. Karl had already got quite high scores at baseline, and his means show that no positive changes occurred. The effect sizes for these boys indicate no positive intervention effects.

Tina's performance was quite stable during the different phases. On average, she got a little less than half of the scores at each measurement point. Effect size showed that she made moderate progress. Nina did not have enough patience to do the same kinds of tasks 19 times. The means and effect size indicate negative changes in her word problem solving.

Discussion

The aim of this study was to describe the effects of an intervention designed to promote the mathematical word problem solving of 10-year-old low-achievers. The results from group measurements indicate significant positive effects for intervention students, which is in line with earlier intervention studies (Kroesbergen & Van Luit, 2003; Swanson, 1999; Vauras, Rauhanummi, Kinnunen, & Lepola, 1999; Verschaffel et al., 1999; Xin & Jitendra, 1999). Although the lack of intensive, systematic, and explicit coaching has made it difficult to show positive results for low-achievers (Fuchs & Fuchs, 2005; Jitendra & Xin, 1997; Verschaffel & De Corte, 1997a), our work shows that even low-achievers progress when they are given the necessary coaching. Furthermore, our study, unlike many other studies (meta-analyses by Kroesbergen and Van Luit [2003] and Xin and Jitendra [1999]), includes a follow-up which shows that the effects of the intervention are maintained even once the intervention itself is over.

Intervention students' word problem solving scores increased significantly from pre-test to post-test, even significantly more than other students' scores. In the post-test and follow-up test, the intervention students no longer differed from other students in word problem solving, and this can be considered an important result (Swansson, 1999, pp. 245–246). Effect sizes indicate moderate intervention effects. The differences in the scores of the

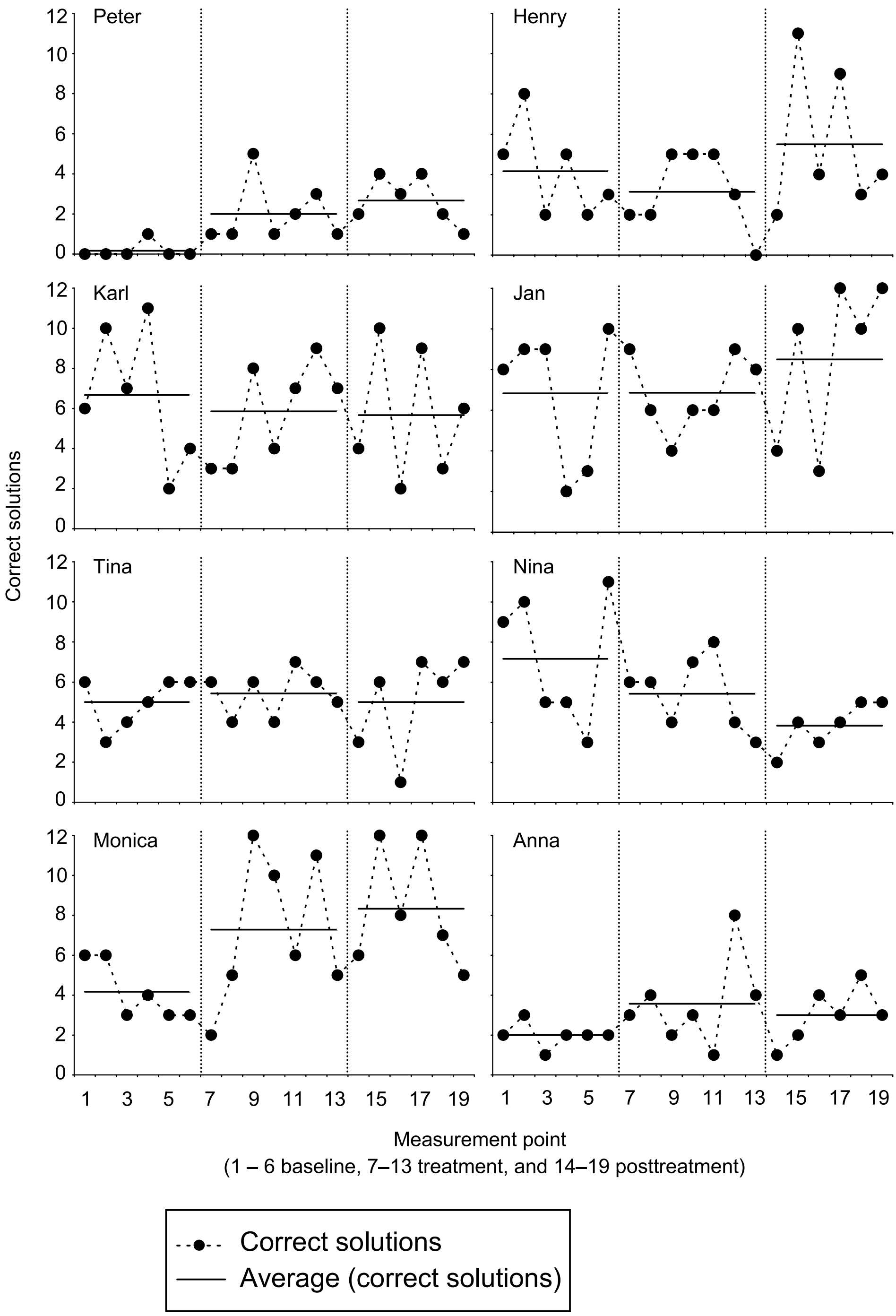

Figure 1. Scores in word problems in single-subject measurements.

Table 7
Scores in Word Problems in Single-Subject Measurements

Student	Baseline		Treatment		Post-treatment		ES[a]	ES-c[b]	r[c]
	M	SD	M	SD	M	SD			
Peter	.17	.41	2.00	1.53	2.67	1.21	4.46	2.41	.18
Henry	4.17	2.32	3.14	1.95	5.50	3.62	−.44	−.67	.04
Karl	6.67	3.44	5.86	2.48	5.67	3.27	−.24	−.38	.06
Jan	6.83	3.43	6.86	1.86	8.50	3.99	.01	.01	.18
Tina	5.00	1.26	5.43	1.13	5.00	2.45	.34	.53	−.09
Nina	7.17	3.25	5.43	1.81	3.83	1.17	−.54	−.98	−.02
Monica	4.17	1.47	7.29	3.73	8.33	3.01	2.12	1.32	.40
Anna	2.00	.63	3.57	2.23	3.00	1.41	2.49	1.60	−.06
Group	4.52	1.21	4.94	1.02	5.31	1.88			

[a] ES = Effect size = $M_{Treatment} - M_{Baseline}/\ SD_{Baseline}$. [b] ES-corrected = $M_{Treatment} - {}_{Baseline}/(M_{SDBaselineSDTreatment}/[2\ (1 - r)]^{1/2})$. [c] r = autocorrelation, lag 1, from baseline and treatment measurements.

intervention and the control groups in the post-test and in the follow-up test, failed to reach significance.

When evaluating the effectiveness of the intervention at the individual level, it was found that all intervention students, except Henry, progressed in their word problem solving. The improved level of word problem solving was maintained by all intervention students except Peter, for whom the intervention should perhaps have lasted much longer in order to bring about lasting learning (cf. Lehtinen et al., 1995; Xin & Jitendra, 1999). Peter's clear progress from pre-test to post-test indicates that continuing this kind of intervention could have been a good way to proceed further. Continuing the intervention would have been easy because all the students, especially Peter, enjoyed taking part in the intervention and would have liked to continue. In the future studies, longer interventions are also needed to achieve increases in correctly produced answers, not only in calculation steps.

Anna's skills in the measurements were fairly stable, and we argue that in order to get Anna actively involved in the instructional discussions, she would have needed more individual attention than was possible in a situation where two students were present. If we had been able to work with Anna only, it might have been easier to prevent her from shifting her responsibilities to her partner, Monica. In the future, it would also be important to analyze carefully the interactions and to find more effective ways to support the collaboration of low-achievers, because only gradually during the intervention did some of the students learn to benefit from their collaboration (cf. Jenkins & O'Connor, 2003). Our research with high-achieving students motivates this work by showing that collaboration between peers can produce high-level learning if the key conditions for effective collaboration are met (Iiskala, Vauras, Lehtinen, & Salonen, in press; Vauras, Iiskala, Kajamies, Kinnunen, & Lehtinen, 2003). Since both Peter and Anna also had problems in arithmetical skills, it might have been useful in the training to focus also on the automatization of arithmetical skills so that they could devote more attentional resources to word problem solving (cf. Gersten & Chard, 1999).

In single-subject measurements, improvement can be seen only in Peter, Monica, and Anna's performances. This is in contrast with the meta-analyses by Swanson (1999, p. 236) and Kroesbergen and Van Luit (2003), in which they concluded that effect sizes in single-subject studies were even higher than those in group studies. In this study, we have to take into account the large variability in the students' scores in the single-subject measurements, because we want to draw some conclusions about the development of the students' skills. This variability shows that when we make decisions that can have a major impact on the student's life, we should always measure the student's performance at different time points because low-achievers are typically very sensitive to different kinds of situational factors.

Because it is extremely difficult to construct parallel word problem solving tests (Verschaffel & De Corte, 1997b, p. 73), critical questions can be raised about the tests used. In future single subject studies, the difficulty level of the tests should be scrutinized with large samples. A greater number of measurement points would have made the results of single-subject measurements more reliable, but more testing would perhaps also have resulted in problems with the patience of other students, not just Nina. Furthermore, better ways of dealing with the serial dependence in the time series data are needed (Matyas & Greenwood, 1996). Because effect sizes are biased when the subject is used as his/her own scale, we would also need other estimates of intervention effects. In the future it would also be interesting to study more carefully the effect of repeated testing and training on the skills of the students. In this study we tried to minimize the effects of testing by giving the students no feedback on their solutions.

Because of the variety of skills needed in word problem solving and the complexity of developing the skills of the low-achievers, it is not meaningful to individually evaluate the importance of the different components of the intervention in the production of the effects. But it can be assumed that combining the strategy teaching with the supporting of task orientation and metacognition was responsible for the improvements, as has also been suggested in earlier studies (Borkowski et al., 1992; De Corte et al., 2004; Lehtinen et al., 1995; Xin & Jitendra, 1999). According to our observations, the instructional discussion between the students and the teacher was a crucial element in the intervention. This observation is in line with the conclusions about the effectiveness of computers alone in the instruction of low-achievers (Kroesbergen & Van Luit, 2003; Salomon, Perkins, & Globerson, 1991; Xin & Jitendra, 1999). It also lends support to our suggestion that, as an implication of this study, teachers could perhaps successfully instruct their students to follow the model of skillful problem solving presented here even without the computer game. In future intervention studies, the discussions between teacher and students should be analyzed carefully to find out the most effective ways to scaffold the low-achievers and ways to explain why some students progress and others do not.

It seems likely that the computer-supported adventure game played a role in making beliefs about word problems more positive (cf. McLeod, 1992). This kind of innovative task environment may have helped the intervention students to overcome their anxiety and become intellectually involved in the tasks (cf. Vauras, Rauhanummi, Kinnunen, & Lepola, 1999). Students who showed no or low task orientation in the classroom, and prior to the intervention used superficial strategies, could spend a whole intervention session skillfully solving just one problem. This kind of extraordinary engagement has also been noted in some earlier studies (Bottge, Heinrichs, Chan, & Serlin, 2001; Higgins, 1997). We can conclude that at least some progress was made in changing the didactical contract about how to solve mathematical word problems. In the future, it will be important to analyze the role of the computer in supporting learning in more detail.

Developing the word problem solving skills of low-achievers requires keeping in mind that students should learn to understand the problems in order to learn skills that are applicable in everyday situations. We should avoid the danger of strengthening students' superficial solving strategies, for example, by teaching them how to use keywords. If we practice only one or two operations in an intervention, as is commonly done (Xin & Jitendra, 1999), we can obtain impressive results in the post-test, without the students having really learned to understand the tasks. In this study, working through the skillful problem solving process was emphasized in every task. We highlighted the understanding by using complex multistep problems that could not be solved with superficial strategies, and required the use of all four operations.

When implementing these kinds of interventions in the future, it is important to include the classroom teacher in the intervention, because implementing an effective change requires a major change in teachers', as well as students', beliefs about word problems (Lehtinen et al., 1995; Verschaffel et al., 1999). Otherwise, more realistic word problem solving will not be encouraged by the teachers, as Verschaffel, De Corte, and Borghart (1997) report. Furthermore, engaging the classroom teacher in the intervention could give us effective ways to provide new theoretical ideas for the learning and teaching of word problem solving, one of the most difficult parts of mathematics (Bryant, Bryant, & Hammill, 2000). Finally, the aim should be to integrate realistic mathematical modeling into the entire mathematics curriculum in order to help each and every student to effectively apply mathematical skills to the problems of everyday life (Greer et al., 2009; National Council of Teachers of Mathematics, 2000; Verschaffel et al., 2000, xi). This aim is important because the analysis of students' word problem solving difficulties suggests that an explanation can be found in classroom cultures where the implicit assumption is that superficial working with the numbers mentioned in the problem is sufficient. Because almost all attention in classrooms is devoted to computational skills at the expense of modeling and interpreting skills, students start to believe that trying to understand the problems is not worthwhile (Reusser & Stebler, 1997; Schoenfeld, 1988; Verschaffel et al., 2000). Intervention studies dealing with the generalizability of our results are to be done in the future.

References

Baxter, J., Woodward, J., Voorhies, J., & Wong, J. (2002). We talk about it, but do they get it? *Learning Disabilities Research & Practice, 17,* 173–185.

Boekaerts, M. (2002). Bringing about change in the classroom: Strengths and weaknesses of the self-regulated learning approach—Earli presidential address, 2001. *Learning and Instruction, 12,* 589–604.

Borasi, R., Siegel, M., Fonzi, J., & Smith, C.F. (1998). Using transactional reading strategies to support sense-making and discussion in mathematics classrooms: An exploratory study. *Journal of Research in Mathematics Education, 29,* 275–305.

Borkowski, J.G., Day, J.D., Saenz, D., Dietmeyer, D., Estrada, T.M., & Groteluschen, A. (1992). Expanding the boundaries of cognitive interventions. In B. Wong (Ed.), *Intervention research with students with learning disabilities* (pp. 1–21). NY: Springer-Verlag.

Bottge, B.A. (2001). Reconceptualizing mathematics problem solving for low-achieving students. *Remedial and Special Education, 22,* 102–112.

Bottge, B.A., Heinrichs, M., Chan, S., & Serlin, R.C. (2001). Anchoring adolescents' understanding of math concepts in rich problem-solving environments. *Remedial and Special Education, 22,* 299–314.

Bryant, D.P., Bryant, B.R., & Hammill, D.D. (2000). Characteristic behaviors of students with LD who have teacher-identified math weakness. *Journal of Learning Disabilities, 33,* 168–177, 199.

Carr, M., & Biddlecomb, B. (1998). Metacognition in mathematics from a constructivist perspective. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 69–92). Hillsdale, NJ: Erlbaum.

Cognition and Technology Group at Vanderbilt. (1997). *The Jasper project: Lessons in curriculum, instruction, assessment, and professional development.* Mahwah, NJ: Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

De Bock, D., Verschaffel, L., & Janssens, D. (1998). The predominance of the linear model in secondary school students' solutions of word problems involving length and area of similar plane figures. *Educational Studies in Mathematics, 35,* 65–83.

De Corte, E., Verschaffel, L., & Masui, C. (2004). The CLIA-model: A framework for designing powerful learning environments for thinking and problem solving. *European Journal of Psychology of Education, 19,* 356–384.

De Corte, E., Verschaffel, L., & Op't Eynde, P. (2000). Self-regulation: A characteristic and a goal of mathematics education. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 687–726). San Diego, CA: Academic Press.

De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of Educational Psychology, 82,* 359–365.

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities, 34,* 435–449.

Franklin, R.D., Allison, D.B., & Gorman, B.S. (1996). Introduction. In R. Franklin, D. Allison, & B. Gorman (Eds.), *Design and analysis of single-case research* (pp. 1–11). Mahwah, NJ: Erlbaum.

Franklin, R.D., Gorman, B.S., Beasley, T.M., & Allison, D.B. (1996). Graphical display and visual analysis. In R. Franklin, D. Allison, & B. Gorman (Eds.), *Design and analysis of single-case research* (pp. 119–158). Mahwah, NJ: Erlbaum.

Fuchs, L.S., & Fuchs, D. (2005). Enhancing mathematical problem solving for students with disabilities. *Journal of Special Education, 39,* 45–57.

Geary, D.C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37,* 4–15.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33,* 18–28.

Gravemeijer, K. (1997). Mediating between concrete and abstract. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 315–345). Hove, UK: Psychology Press.

Greer, B., Verschaffel, L., Van Dooren, W., & Mukhopadhyay, S. (2009). Introduction. Making sense of word problems: Past, present, and future. In L. Verschaffel, B. Greer, W. Van Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds: Modelling verbal descriptions of situations* (pp. xi–xxvii). Rotterdam: Sense.

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

Higgins, K.M. (1997). The effect of year-long instruction in mathematical problem solving on middle-school students' attitudes, beliefs, and abilities. *Journal of Experimental Education, 66,* 5–28.

Iiskala, T., Vauras, M., Lehtinen, E., & Salonen, P. (in press). Socially shared metacognition of dyads of pupils in collaborative mathematical problem-solving processes. Learning and Instruction (2010), doi: 10.1016/j.learninstruc.2010.05.002.

Ives, B. (2003). Effect size use in studies of learning disabilities. *Journal of Learning Disabilities, 36,* 490–504.

Jastak, S., & Wilkinson, G. (1984). *Wide Range Achievement Test-Revised (WRAT-R).* Wilmington, DE: Jastak Associates.

Jenkins, J.R., & O'Connor, R.E. (2003). Cooperative learning for students with learning disabilities: Evidence from experiments, observations, and interviews. In H.L. Swansson, K.R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 417–430). NY: Guilford.

Jitendra, A., & Xin, Y.P. (1997). Mathematical word-problem-solving instruction for students with mild disabilities and students at risk for math failure: A research synthesis. *Journal of Special Education, 30,* 412–439.

Kajamies, A., Vauras, M., Kinnunen, R., & Iiskala, T. (2003). *MATTE—Matematiikan sanallisten tehtävien ratkaisutaidon ja laskutaidon arviointi* [Evaluation of the student's mathematical word problem solving and arithmetic skills; in Finnish]. Turku, Finland: University of Turku, Centre for Learning Research.

Kloosterman, P. (2002). Beliefs about mathematics and mathematics learning in the secondary school: Measurement and implications for motivation. In G.C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A hidden variable in mathematics education?* (pp. 247–269). Mathematics education library, 31. Netherlands: Kluwer.

Kroesbergen, E.H., & Van Luit, J.E.H. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education, 24,* 97–114.

Kromrey, J.D., & Foster-Johnson, L. (1996). Determining the efficacy of intervention: The use of effect sizes for data analysis in single-subject research. *Journal of Experimental Education, 65,* 73–93.

Lehtinen, E. (2003). Computer-supported collaborative learning: An approach to powerful learning environments. In E. De Corte, L. Verschaffel, N. Entwistle, & J. Van Merriëboer (Eds.), *Unravelling basic components and dimensions of powerful learning environments* (pp. 35–53). Amsterdam: Pergamon.

Lehtinen, E., Vauras, M., Salonen, P., Olkinuora, E., & Kinnunen, R. (1995). Long-term development of learning activity: Motivational, cognitive, and social interaction. *Educational Psychologist, 30,* 21–35.

Lindeman, J. (1998). *Ala-asteen lukutesti* [Comprehensive school reading test; in Finnish]. Turku, Finland: University of Turku, Centre for Learning Research.

Ma, X. (1999). A meta-analysis of the relationship between anxiety towards mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30,* 520–540.

Mason, L., & Scrivani, L. (2004). Enhancing students' mathematical beliefs: An intervention study. *Learning and Instruction, 14,* 153–176.

Matyas, T.A., & Greenwood, K.M. (1996). Serial dependency in single-case time series. In R. Franklin, D. Allison, & B. Gorman (Eds.), *Design and analysis of single-case research* (pp. 215–243). Mahwah, NJ: Erlbaum.

Mayer, R. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26,* 49–63.

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

McLeod, D. (1992). Research on affect in mathematics education. A reconceptualization. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575–596). NY: Macmillan.

Meichenbaum, D., & Biemiller, A. (1998). *Nurturing independent learners.* Cambridge, MA: Brookline Books.

Montague, M., & Applegate, B. (1993). Mathematical problem solving characteristics of middle school students with learning disabilities. *Journal of Special Education, 27,* 175–201.

Montague, M., Warger, C., & Morgan, T.M. (2000). Solve it! Strategy instruction to improve mathematical problem solving. *Learning Disabilities Research & Practice, 15,* 110–116.

Nathan, M.J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction, 9,* 329–389.

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

Op't Eynde, P., De Corte, E., & Verschaffel, L. (2001). "What to learn from what we feel?": The role of students' emotions in the mathematics classroom. In S. Volet & S. Järvelä (Eds.), *Motivation in learning contexts. Theoretical advances and methodological implications* (pp. 149–167). Oxford: Elsevier.

Polya, G. (1957). *How to solve it: A new aspect of mathematical method* (2nd ed.). Princeton, NJ: Princeton University Press.

Pressley, M., Beard El-Dinary, P., Gaskins, I., Schuder, T., Bergman, J.L., Almasi, J., et al. (1992). Beyond direct explanation: Transactional instruction in reading comprehension strategies. *Elementary School Journal, 92,* 511–554.

Raven, J., Raven, J.C., & Court, J.H. (2000). *Standard progressive matrices.* Oxford: Psychologist Press.

Reusser, K., & Stebler, R. (1997). Every word problem has a solution: The social rationality of mathematical modeling in schools. *Learning and Instruction, 7,* 309–327.

Riley, M., & Greeno, J.G. (1988). Developmental analysis of understanding language about quantities and solving problems. *Cognition and Instruction, 5,* 49–101.

Robey, R.R., Schultz, M.C., Crawford, A.B., & Sinner, C.A. (1999). Single-subject clinical outcome research: Design, data, effect sizes, and analyses. *Aphasiology, 13,* 445–473.

Räsänen, P. (2004). *RMAT: Laskutaidon testi 9–12 -vuotiaille* [RMAT: A test for arithmetical skills for ages 9 to 12; in Finnish]. Jyväskylä, Finland: Niilo Mäki Institute.

Räsänen, P. (1993, July). *Consistency between two mathematical screening tests and teacher evaluations.* Paper presented at the III European Congress of Psychology, Tampere, Finland.

Salomon, G., Perkins, D.N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher, 20*(3), 2–9.

Salonen, P., Lehtinen, E., & Olkinuora, E. (1998). Expectations and beyond: The development of motivation and learning in a classroom context. In J. Brophy (Ed.), *Advances in research on teaching. Vol 7: Expectations in the classroom* (pp. 111–150). Greenwich, CT: JAI Press.

Salonen, P., Kajamies, A., & Vauras, M. (2010). *MOSU: Motivationaalisen suuntautumisen arviointi* [Evaluation of the student's motivational orientation; in Finnish]. Scale in preparation. Turku, Finland: University of Turku, Centre for Learning Research.

Schoenfeld, A. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist, 23,* 145–166.

Swansson, H.L., with Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities. A meta-analysis of treatment outcomes.* New York: Guilford.

Vauras, M., Iiskala, T., Kajamies, A., Kinnunen, R., & Lehtinen, E. (2003). Shared regulation and motivation of collaborating peers: A case analysis. *Psychologia, an International Journal of Psychology in the Orient, 46,* 19–37.

Vauras, M., & Kinnunen, R. (2003). *HoPE: Hopeisen Pöllön Etsintä. Opetuspeli laskutaitojen ja matematiikan ongelmanratkaisun taitojen kehittämiseen* [Quest of the Silver Owl, Teaching game for improving arithmetic and mathematical problem solving skills; in Finnish]. Turku, Finland: University of Turku, Centre for Learning Research.

Vauras, M., Kinnunen, R., Kajamies, A., & Iiskala, T. (2003). *Opettajan opas oppimisen ohjaamiseen matematiikan opetuspelissä* [Guidebook for teachers. How to guide learning while playing the educational computer game; in Finnish]. Turku, Finland: University of Turku, Centre for Learning Research.

Vauras, M., Kinnunen, R., & Rauhanummi, T. (1999). The role of metacognition in the context of integrated strategy instruction. *European Journal of Psychology of Education, 14,* 555–569.

Vauras, M., Rauhanummi, T., Kinnunen, R., & Lepola, J. (1999). Motivational vulnerability as a challenge for educational interventions. *International Journal of Educational Research, 31,* 515–531.

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

Verschaffel, L., & De Corte, E. (1997b). 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., & Borghart, I. (1997). Pre-service teachers' conceptions and beliefs about the role of real word knowledge in mathematical modelling of school word problems. *Learning and Instruction, 4,* 339–359.

Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. *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.

Vygotsky, L. (1978). *Mind and society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Xin, Y.P., & Jitendra, A. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *Journal of Special Education, 32,* 207–225.