Journal of the Korea Society of Mathematical Education Series D:
Research in Mathematical Education
Vol. 11, No. 1, March 2007, 1–31

韓國數學敎育學會誌 시리즈 D:
<數學敎育硏究>
제 11 권 제 1 호 2007 년 3 월, 1–31

The Relationship between Posing and Solving Arithmetic Word Problems among Chinese Elementary School Children

Chen, Limin
Center for Instructional Psychology and Technology, Katholieke Universiteit Leuven, Belgium and Elementary Education College, Tianjin Normal University, China

Van Dooren, Wim
Center for Instructional Psychology and Technology, Katholieke Universiteit Leuven, Belgium

Chen, Qi*
School of Psychology, Beijing Normal University, Beijing, 100875, China;
Email: chenqi@bnu.edu.cn, chenqi88@hotmail.com

Verschaffel, Lieven
Center for Instructional Psychology and Technology, Katholieke Universiteit Leuven, Belgium

(Received January 29, 2007. Accepted March 26, 2007)

Recent research has documented that there is a close relationship between problem posing and problem solving in arithmetic. However, most studies investigated the relationship between problem posing and problem solving only by means of standard problem situations. In order to overcome that shortcoming, a pilot study with Chinese fourth-graders was done to investigate this relationship using a non-standard, realistic problem situation. The results revealed a significant positive relationship between students' problem posing and solving abilities. Based on that pilot study, a more extensive and systematic ascertaining study was carried out to confirm the observed relationship between problem posing and problem solving among Chinese elementary school children. Results confirmed that there was indeed a close relationship between both skills.

Keywords: problem posing, problem solving, mathematics education, word problems, division

---

* Corresponding author

ZDM Classification: D50
MSC2000 Classification: 97-02

1. THEORETICAL AND EMPIRICAL BACKGROUND

Recent recommendations for the reform of school mathematics suggest an important role for problem posing. For example, the *Principles and Standards for School Mathematics* in the U. S. call for students to "formulate interesting problems based on a wide variety of situations, both within and outside mathematics" (NCTM 2000, p. 258) and recommend that students should make and investigate mathematical conjectures and learn how to generalize and extend problems by posing follow-up questions. Similarly, the Chinese *National Curriculum Standards on Mathematics* (NCSM) paid attention to the acquisition of problem posing ability (*cf.* Ministry of Education of Peoples' Republic of China 2001). For example, the standards for problem solving emphasize that students must be able to:

> (...) pose and understand problems mathematically, apply basic knowledge and skills to solve problems, develop application awareness; develop basic problem solving strategies, *reflect on the variety of strategies that exist for solving the same problem,* and develop practical ability and creativity; cooperate with others and articulate the process and the result of problem solving; develop evaluation and reflection awareness (NCSM 2001, p. 7) [a translation by the author of the Chinese version].

These documents show that the development of problem posing competency is generally recognized as an important goal of mathematics teaching and learning that lies at the heart of mathematical activity.

Since the late eighties, there is also growing interest among researchers for problem posing (see *e.g.*, Brown & Walter 1993; English 1998a; Kilpatrick 1987; Silver 1994). In those research articles, the most frequently cited motivation for curricular and instructional interest in problem posing is its perceived potential value in assisting students to become better problem solvers. To explore this potential value, several studies have been set up to investigate the relationship between word problem solving and word problem posing (*e.g.*, Cai & Hwang 2002; Ellerton 1986; Silver & Cai 1996). In these studies students were typically asked to generate one or more problems (sometimes of different levels of difficulty) starting from a given situational description, a picture, or a number sentence, and, afterwards, the quality of the mathematical problems generated by the students were compared with their problem solving capacities.

Silver & Cai (1996) analyzed the responses of 509 sixth- and seventh-grade middle school students to a task asking them to pose three questions (that could be answered using the given information) based on the following brief written story-problem description: "Jerome, Elliot, and Arturo took turns driving home from a trip. Arturo drove 80 miles more than Elliot. Elliot drove twice as many miles as Jerome. Jerome drove 50 miles." The questions posed by the students were examined according to their solvability (*i.e.*, is the goal posed compatible with the given information or is there sufficient information to solve the problems?), linguistic complexity (*i.e.*, occurrence of assignment, relational, or conditional phrases) and mathematical/semantic complexity (*e.g.*, the number of arithmetic operations, or the number of different semantic relations required to arrive at the answer), and these measures of problem posing were compared to students' performance on a problem-solving test involving eight mathematical problems coming from various mathematics content areas, such as fraction, geometry/measurement, number theory, patterns and relationships, ration/proportion, and statistics. It was found that students generated a large number of solvable problems, many of which were linguistically and semantically complex, and that nearly half of the students generated sets of closely related problems. As in most other studies (*e.g.*, Ellerton 1986), it was also found that good problem solvers generated more solvable problems and more mathematically/semantically complex problems than poor problem solvers did.

A more recent international comparative study by Cai & Hwang (2002) also revealed a close relationship between word problem solving and word problem posing. A sample of 155 Chinese sixth-grade students and a comparable sample of 98 U. S. sixth-grade students were administered three pairs of problem solving and problem posing tasks. Each pair of problem solving and problem posing tasks involved one mathematical situation (like a series of three figures representing, respectively, a 3×3, a 4×4, and a 5×5 square dot pattern). The problem posing task required students to generate one easy, one moderately difficult and one difficult problem from the three figures. The problem solving task required students to answer a number of mathematical problems about the same situation. As expected, they found a stronger link between problem posing and solving for Chinese students compared to U. S. students. For Chinese students, the variety and type of problems posed were clearly associated with their problem solving success; moreover, the number of extension problems a student posed (*i.e.*, a problem questioning the pattern beyond the given figures) was positively associated with the use of abstract problem solving strategies. For the U. S. sample, the relationship between the variety and type of problems posed and problem solving success was much weaker. Cai & Hwang (2002) suggested that the stronger link between the variety of posed problems and problem solving success for Chinese students might be attributable to the fact that the U. S. students used much less abstract problem solving strategies.

Besides ascertaining studies revealing the existence of a relationship between problem posing and solving, some design experiments wherein new instructional approaches that incorporate problem posing into the mathematics curriculum and use it as a vehicle to improve students' problem solving ability, have been designed, implemented and evaluated (English 1997a, 1997b; Rudnitsky, Etheredge, Freeman & Gilbert 1995; Verschaffel, De Corte, Lowyck, Dhert & Vandeput 2000; Yang & Wang 2004). In general, these latter studies revealed that having students engage in some activities related to problem posing (*e.g.*, making up word problems according to math stories) had indeed a positive influence on their word problem posing and problem solving abilities and their attitudes towards mathematical problem solving and mathematics in general.

Another topic that has attracted the attention of many researchers in mathematics education concerns the role of real-world knowledge and realistic considerations in pupils' solutions of arithmetic word problems. Since the nineties, several researchers have addressed this issue by carefully looking at pupils' approaches to and solutions of non-standard arithmetic word problems wherein — contrary to the traditional standard problems — the appropriate mathematical model or solution is neither obvious nor indisputable (at least if one seriously takes into account the realities of the context evoked by the problem statement) (Cai & Silver 1995; Silver; Shapiro & Deutsch 1993; Verschaffel; De Corte & Lasure 1994). By analyzing pupils' reactions to these non-routine problems, they provided strong empirical evidence that most elementary school children perceive school word problems as artificial, routine-based tasks that are unrelated to the real world, and, accordingly, approach these problems with superficial coping strategies that have little to do with authentic mathematical modeling and applied problem solving (see Verschaffel; Greer & De Corte 2000, for a review of this research).

One of the most frequently studied types of non-routine problems are the division with remainder (DWR) problems, *i.e.* problems wherein the computational work results in a non-integer outcome that needs to be interpreted and evaluated in function of the real-world constraints of the problem setting. This kind of problem was used for the first time with a stratified sample of 13-year-olds in the Third National Assessment of Educational Progress in the U.S. (Carpenter; Lindquist; Matthews & Silver 1983):

> "An army bus holds 36 soldiers. If 1128 soldiers are being bused to their training site, how many buses are needed?" Of 70% of the students who correctly carried out the division 1128 by 36 to get a quotient of 31 and remainder of 12, only 23% (of the total number of students) gave the answer as 32 buses, 19% gave the answer as 31 buses, and 29% gave the answer as "31, remainder 12".

A similar division problem appeared on the 1983 version of the California Assessment Program (CAP) Mathematics Test for Grade 6, and was answered correctly by about 35% of the sixth-graders in California (Silver et al. 1993). As in the NAEP assessment, most students erred by giving non-whole-number answers or rounding the outcome of the division to its nearest whole-number predecessor. Since then, numerous studies involving this type of problem have been done in different countries, basically leading to similar results and conclusions (for an overview see Verschaffel et al. 2000).

From the studies on problem posing, we can conclude that this activity has a potential value in assisting and improving children's problem solving capacities. However, an important feature of the problem posing research up to now is that it has investigated the relationship between problem posing and solving especially, even almost exclusively, for standard word problems that can be "unproblematically" modeled by one or more operations with the numbers given in the problem statement (Fuson 1992). Stated differently, little or no research investigated this relationship for non-standard, truly realistic problem situations, wherein students are invited or even forced to use their commonsense knowledge and experience when trying to understand and solve the problem (Verschaffel *et al.* 2000). This leads to the question whether these psychological models and instructional programs on problem posing and its relation to problem solving have succeeded in capturing the quintessence of genuine mathematical problem posing and problem solving.

2. A PILOT STUDY LEADING TO A NEW STUDY

Recently, we completed a pilot study about this topic with a group of fourth-grade pupils from a school in the City of Tianjin in China. In that study, we used problems requiring the proper realistic interpretation of division with remainder (DWR). The problem posing and problem solving tasks from that pilot study are described below.

THE PROBLEM-POSING (PP) TASK:

Invent three story problems that belong to the numerical problem 100 ÷ 8 = ______ such that the answer to the story problem is, respectively,

(a) 13
(b) 12
(c) 12.5.

THE PROBLEM-SOLVING (PS) TASK:

(a) 100 children are being transported by minibuses to a summer camp at the seaside. Each minibus can hold a maximum of 8 children. How many minibuses are needed? (100 ÷ 8 = 13 buses)
(b) They assemble in the sports center where the sporting materials are stored in big cases. These cases have to be carried to the sport fields. To carry one case, 8 children are needed. How many cases can be carried by the 100 children at one time, if they all cooperate? (100 ÷ 8 = 12 boxes)

(c) When the children have sported the whole day, they all are very hungry so they gather in the dining room. The cook has prepared 100 liters of soup in 8 equal big kettles, all filled to the rim. How much soup is contained in each kettle? (100 ÷ 8 = 12.5 liters)

It was found that, first, like their Western peers, these Chinese students held non-realistic perspectives in posing and solving these tasks. Second, there was a significant positive relationship between students' abilities to pose and solve realistic problems. And, finally, the order in which the problem posing and solving tasks were administered seemed to have a significant impact on students' task performance, in the sense that students who got the problem posing task after the problem solving task performed significantly better on the problem posing than those who got both tasks in the reversed order.

Although the findings from that pilot study suggested that the combined use of problem solving and problem posing tasks yields promises both to further our theoretical understanding of elementary school children's tendency to and beliefs about non-realistic mathematical modeling, and to apply and combine these two types of tasks in instructional attempts to improve children's disposition towards (more) thoughtful and (more) realistic mathematical modeling, this pilot study also suffered from some methodological weaknesses that jeopardized to some extent the internal and external validity of our results and conclusions. Firstly, taking into account the observed impact of the problem solving experience on the later problem posing task, which seemed to be mainly caused by students' tendency to copy more or less literally the corresponding *PS* item when doing the *PP* task, we decided to try to decrease this interference by considerably increasing the duration of the period between the two tasks. Secondly, a substantial number of students just wrote numerical answers without any further documentation or explanation of their solution steps, which made these responses very difficult to interpret and evaluate. To enrich the quality of our data, we decided to improve the test instructions and to ask the test administrator to check more carefully and systematically whether all students did more than just writing down their answers. Finally, in an attempt to probe students' mathematical thinking even more accurately, we decided to complement the paper-and-pencil test with individual interviews with a sub-sample of pupils.

In the new study, which was also conducted in the City of Tianjin in China, we further explored the relationship between posing and solving word problems with a more representative sample of subjects and with a methodology that overcame the above-mentioned methodological shortcomings of the pilot study. Based on the above literature review and the results of our pilot study, we raised the following predictions. Firstly, we expected a significant relationship between the success rate of problem posing and that of problem solving both in standard and non-standard situations. Secondly, we anticipated that there would be a significant relationship between pupils' success rate in problem solving and the level of linguistic and mathematical/semantic complexity of the problems being posed by these pupils. Finally, compared to the results of our pilot study, we anticipated that the impact of order (*i.e.*, problem posing first or problem solving first) on students' problem posing performance would no longer be significant, due to the considerable extension of the duration between the problem solving and the problem posing test, which would lead to a drastic decrease in the number of more or less literal copies of the *PS* items in the *PP* task.

3. METHOD

3.1 Subjects

Hundred-and-seventeen fourth-grade pupils (aged about 11.1 years old) participated in this study, which took place in the final term of the school year. These students belonged to four classes of two randomly selected elementary schools in the inner City of Tianjin and its countryside Tanggu. Tianjin is located in the northern part of China, very near to Beijing. So the participants are quite typical north urban Chinese students. According to the students' socio-economic status, their parents' educational background, and the school's technical and educational infrastructure, the two classes from the inner City of Tianjin can be considered as educationally rather high-level whereas the two classes from the countryside are of a lower educational level.

According to the National Curriculum Standards on Mathematics in China (Ministry of Education of Peoples' Republic of China 2001), at the time they were tested, these pupils had learnt the operations of addition, subtraction, multiplication and division, as well as simple and complex word problems based on these operations, in the fields of natural and rational numbers.

3.2 Instruments

Two pairs of problem posing (*PP*) and problem solving (*PS*) tasks were administered to the students: one pair involved a standard problem situation and the other was the above-mentioned non-standard, realistic problem situation from the pilot study.

PROBLEM POSING TASK OF PAIR 1 (*PP* 1)

On Children's Day, mother gives children candies. Xiaoli gets 30 candies from mother, Xiaoqiang gets 36 candies, and Xiaofang gets 60 candies. Xiaoli eats 3 candies per day, Xiaoqiang eats 2 candies per day and Xiaofang eats 4 candies per day.

Can you make up three questions based on the above situation, so that you get:

- An easy problem (PP 11)
- A moderately difficult problem (PP 12)
- A difficult problem. (PP 13)

PROBLEM SOLVING TASK OF PAIR 1 (*PS* 1)

On Children's Day, mother gives children candies. Xiaoli gets 30 candies from mother, she gets 6 candies less than Xiaoqiang, and Xiaofang gets 2 times candies as many as Xiaoli. Xiaoli eats 3 candies per day, Xiaoqiang eats 2 candies per day and Xiaofang eats 4 candies per day.

- How many candies does mother give Xiaoqiang? (PS 11)
- How many candies does mother give Xiaofang? (PS 12)
- How many candies should Xiaofang give Xiaoli and Xiaoqiang respectively to make them all three have the same number of candies after 5 days? (PS 13)

PROBLEM POSING TASK OF PAIR 2 (*PP* 2)

Invent three story problems that belong to the numerical problem 100 ÷ 8 = ______ such that the answer to the story problem is, respectively,

(a) 13 (*PP* 21)
(b) 12 (*PP* 22)
(c) 12.5 (*PP* 23)

PROBLEM SOLVING TASK OF PAIR 2 (*PS* 2)

(a) 100 children are being transported by minibuses to a summer camp at the seaside. Each minibus can hold a maximum of 8 children. How many minibuses are needed? (100 ÷ 8 = 13 buses) (*PS* 21)
(b) They assemble in the sports center where the sporting materials are stored in big cases. These cases have to be carried to the sport fields. To carry one case, 8 children are needed. How many cases can be carried by the 100 children at one time, if they all cooperate? (100 ÷ 8 = 12 boxes) (*PS* 22)
(c) When the children have sported the whole day, they all are very hungry so they gather in the dining room. The cook has prepared 100 liters of soup in 8 equal big kettles, all filled to the rim. How much soup is contained in each kettle? (100 ÷ 8 = 12.5 liters) (*PS* 23)

In the two problem posing tasks, students were asked to generate three questions or stories. In the two problem solving tasks, they were asked to solve three problems that were closely related to the questions or stories to be generated. Besides producing a numerical answer, they were also explicitly asked to write down the solution process or some explanation for each problem solving task.

3.3 Procedure and task administration

Particularly in the Chinese context, it is important to administer the test in a relaxed environment, since school examination results are considered as very important within the Chinese educational system (Zhang & Ren 1998). Therefore, it was explained at the outset that the test was not a regular school test and that the result would have no impact on students' regular school results. Using some worked-out examples, the test administrator showed the students what was expected from them in the problem posing tasks before starting the problem posing test, considering they may not be familiar to problem posing activities. Students were given, first, a verbal statement (*e.g.*, Xiaoli has 10 apples, Xiaoqiang has 6 apples) to make up three relevant mathematical questions and, second, a symbolic expression (*e.g.*, 76 + 28 = 104) and stimulated to search for different word problems that fitted with these expressions.

Examples of appropriate and inappropriate questions or problems were discussed with the whole class until all students understood the problem posing instruction and also what would be considered as a correct question or word problem for a given situation or numerical expression. Afterwards, the teacher asked the students to do the *PP* or the *PS* test themselves. Each test took about 40 minutes. During the test administration, the experimenter herself checked very carefully whether all students also articulated and explained their responses. The two tests were administered with an interval of twenty days in two schools (in stead of only seven days in the pilot study). And in each school, one class was administered the *PP* tasks first while the other class was administered the *PS* tasks first. Twenty days later the class that got the *PP* tasks first was administered the *PS* tasks while the class that got PS tasks first was administered the *PP* tasks.

The results of the collective paper-and-pencil part of the study provided empirical evidence for the students' problem posing and problem solving abilities. However, because this paper-and-pencil test procedure would not yield detailed information about pupils' actual thinking processes when posing and solving these problems and their perceptions of problem posing and solving, we included an individual interview with a subset of pupils as a complementary data-gathering technique. One week after the second collective test, 15 students randomly selected from each school were administered an interview, which took approximately 15 minutes. During the interview, the researcher tried to further explore students' actual thinking process when posing and solving problems and their accompanying beliefs. The standardized interview included several open-ended questions, some of which were selected from English (1998b) and Lowrie (2002). The actual questions asked to the students were:

1. Could you tell me more clearly how you posed these questions/that question?
2. Can you pose more questions besides the three already posed?
3. How do you think about the making of math stories?
4. What kind of help do you need from the teacher to make math stories?
5. Could you tell me more clearly how you solved this question?

Before starting the interview, the researcher explained to the student that she wanted to know what they were actually thinking about when posing/solving the problems. Afterwards, the researcher gave the student his/her own answer sheet so that he/she could review the problems and their answers he/she had generated during the paper-and-pencil test. Students' responses were recorded by a digital recorder, transcribed and analyzed afterwards. All students were asked the above five questions in the same order. Question 1 and question 2 focused on *PP* task 1, question 3 and question 4 focused on *PP* task 2, and question 5 on *PS* task 2. During the interview, the researcher tried to elicit students' ideas and feelings by using a kindly tone and providing some extra prompts through paraphrasing the above-mentioned questions.

3.4 Data coding

3.4.1 Task pair 1

The responses to the standard problem posing (*PP*) task 1 were coded with the analysis method used in Silver & Cai (1996) with some adaptations. The method is explained in detail below. It consists of four different (but somewhat related) ways of looking at a problem being posed.

Students' problem posing responses were first categorized as mathematical or non-mathematical questions. Non-mathematical problems are problems that can be solved without any mathematical operation, such as "How many candies does Xiaoli have?" The next step involved categorizing the mathematical problems as solvable or unsolvable. Problems were considered as unsolvable if they posed a question that was impossible to solve by means of the given information or was incompatible with it. For example, the question "How many more candies than Xiaoli does Xiaoqiang have?" was classified as unsolvable because actually according to the problem statement Xiaoqiang had 6 more candies than Xiaoli. A question like "Xiaofang gives Xiaoqiang some candies, how many candies does Xiaofang still have?" was also scored as a unsolvable one because the amount of candies that Xiaofang gives Xiaoqiang was neither given in the task nor supplied as part of the question by the student. Only solvable problems were subjected to a further analysis.

The second part of the analysis of the *PP* task 1 concerned the linguistic/syntactic complexity of the solvable questions. Distinction was made between questions containing assignment, relational and conditional statements. An assignment question is a statement such as "How many candies do they have altogether?" A relational proposition is a question such as "How much more candies does Xiaofang have than Xiaoli?" or "How many more candies do Xiaofang and Xiaoqiang have than Xiaoli?" A conditional question imposes a constraint, such as "If mother gives Xiaofang 5 more candies, then how many candies does Xiaofang have?"

The third part of the analysis focused on the mathematical complexity of the problem being generated. Problems what were solvable by means of a single arithmetic operation with the given numbers were scored as one-step problems, whereas problems for which the response requires at least two arithmetic operations were scored as multi-step problems. The question "How many candies does Xiaoqiang get more than Xiaoli?" was scored as a one-step problem, as it can be solved by doing 36–30 = 6, whereas a question like "How many more candies do Xiaoli and Xiaoqiang than Xiaofang?" was considered to be a multi-step problem, because its solution involves the following two calculations: 30+36=66, 66–60=6. One-step problems were classified into additive and multiplicative problems. A problem generated by means of the question "How many candies do Xiaoqiang and Xiaofang get altogether?" was scored as a one-step additive problem whereas a question like "How many days will Xiaoqiang eat up all his candies?" was scored as a one-step multiplicative problem. Multi-step problems were also further classified into addition/subtraction problems and multiplication/division problems. A problem created by asking "How many more candies do Xiaoqiang and Xiaoli have than Xiaofang?" was scored as a multi-step additive problem, while problems with at least one multiplicative component, like "After 3 days, how many candies will Xiaoqiang and Xiaofang eat altogether" were scored as a multi-step multiplication problem.

Finally, all solvable problems were classified with respect to semantic complexity, or the number of different semantic relations required for solution using a classification scheme of arithmetic word problems developed by English (1998a) and Verschaffel and De Corte (1996). Addition and subtraction problems (or problem parts) were coded as having a combine, change, compare or equalize structure, whereas multiplication and division problems (or problem parts) were categorized as equal group, multiplicative comparison, rectangular pattern or Cartesian product structures. Some examples are provided below:

Problems involving only one semantic relation:

- "How many candies do Xiaoqiang and Xiaoli get altogether?" (combine)
- "How many candies do Xiaoqiang, Xiaofang and Xiaoli get altogether?" (combine)

Problems involving multiple (= two or more than two) semantic relations:

- "How many more candies do Xiaoqiang and Xiaoli have than Xiaofang?" (combine/compare)
- "After 5 days, how many candies will Xiaoqiang and Xiaoli have eaten altogether?"
  (equal group/combine)
- "After 3 days, how many candies will Xiaoqiang and Xiaofang have altogether?" (change/combine/equal group)

Problems involving a greater number of calculation steps and problems including multiplicative components were considered to be more complex than those involving fewer steps and merely additive components. Similarly, problems involving different semantic relations were considered to be more complex than those involving only one type of semantic relations.

The responses to the three items of *PS* task 1 were simply coded as correct or wrong. For example, for *PS* 11 a solution like "$30 + 6 = 36$" was coded as a correct answer, while solution like "$30 - 6 = 24$" was coded as a wrong one. For *PS* 12 the solution "$30\times2 = 60$" was coded as correct whereas "$(30 + 6)\times2 = 72$" was coded as wrong. For *PS* 13, the solution "$30 - 3\times5 =15$, $30 + 6 - 2\times5 = 26$, $30\times2 - 4 \times 5 = 40$, $(15 + 26 + 40) \div3 = 27$, $27 - 26 = 1$, $27 - 15 = 12$" was coded as correct, while a solution like "$30 + (30 + 6) + (30\times2) = 126$, $126 \div 3 = 42$, $42 - 36 = 6$, $42 - 30 = 12$" was coded as a wrong one.

3.4.2 Task pair 2

The responses to the three items of non-standard *PP* task 2 were coded using a classification scheme that was adapted from De Corte & Verschaffel (1996). They investigated the number type effects on solving multiplicative problems by means of student-generated word problems for given number sentences with samples consisting of students from upper elementary schools, secondary schools, and institutes for elementary school teacher training. In that study, students were asked to generate word problems using multiplications and divisions (*e.g.*, 5.3×0.6 or $6 \div 4.8$). In their classification scheme, each response was coded as a realistic problem (RP), a contextually in-appropriate problem (CIP), a mathematically inappropriate problem (MIP), and no problem (NP).

Students' responses to the *PS* task 2 were coded using a classification scheme adapted from another study by Verschaffel & De Corte (1996; 1997). Each response was coded as a realistic answer (RA), a contextually inappropriate answer (CIA), a mathematically inappropriate answer (MIA) and no answer (NA).

A RP or a RA follows from the effective use of real-world context or arithmetic operation elicited by the problem statement. A CIP or CIA results from the anticipated straightforward and uncritical application of the real world context or arithmetic operation elicited by the problem statement. A MIP or MIA results from mathematical mistakes when generating or solving word problems (*e.g.*, wrong number, computation or operation error). A typical RP for the first *PP* task (100 ÷ 8 = 13) is "100 students take boats. One boat takes at most 8 students. How many boats are needed?" A typical CIP for the first *PP* task (100 ÷ 8 = 13) is "100 kilos of apples are shared by 8 children equally. How many kilos does each child have?" An example of a MIP for the first *PP* task (100 ÷ 8 = 13) is "Xiaoqiang and Xiaoli have 100 apples together. Xiaoqiang has 8 apples. Then how many apples does Xiaoli have?" Accordingly, a typical RA for the first item from the *PS* tasks (*i.e.*, the buses item) is "13 buses." Examples of CIAs for the corresponding *PS* task (bus item) are "12 buses" or "12.5 buses." A MIA is an error to a *PS* task resulting from a wrong or wrongly computed operation, such as, respectively, "100 + 8=108 buses" or "100 ÷ 8=25 buses" for the bus item.

3.4.3 The interview data

Analysis of the interview data followed interpretative techniques (Miles & Huberman, 1994). Several foreshadowed core categories (*e.g.*, students' thinking process, students' beliefs about problem posing and solving, textbook, and teaching) were discriminated before the actual data collection but were refined during the interpretative process.

4. RESULTS

The results are presented in three sections. The first and second section provide a summary of students' problem posing and problem solving responses, respectively, while the third section presents an analysis of the relationship between students' problem posing and problem solving performance.

4.1 Results for PP tasks

4.1.1 Results for PP task 1

Table 1 shows the percentage of pupils who produced, respectively, 3, 2, 1, and 0 appropriate responses on the standard *PP* task 1. As this table shows, students did rather well on this task.

Table 1. Distribution of the pupils (in percentage and absolute numbers) according to the numbers of appropriate responses to *PP* 1

3	2	1	0	Total
73.50% (86)	17.95% (21)	4.27% (5)	4.27% (5)	100% (117)

No significant differences were observed in the number of (mathematically) appropriate responses for *PP* 11, *PP* 12 and *PP* 13: A logistic regression predicting appropriate responses revealed no main 'item' effect, $x^2(2) = 4.47$, $p = 0.1073$. Percentages for *PP* 11, *PP* 12 and *PP* 13 were 90.60, 86.32 and 83.76, respectively.

Table 2 shows the distribution of percentages of the pupil-generated questions over the different categories of linguistic/syntactic complexity for *PP* 1. As shown in this table, the percentage of linguistically simplest questions, namely assignment questions, is decreasing from students' firstly generated problem (in response to the task "generate an easy problem") to their third problem (in reaction to the task "generate a very difficult problem"), whereas the opposite is true for the conditional questions, which are typically considered linguistically very complex (Mayer, Lewis & Hegarty 1992; Silver & Cai 1996).

Table 2. Distribution (in percentages and absolute numbers) of pupils' responses over the different categories of linguistic complexity in *PP* 1

	PP 11	*PP* 12	*PP* 13
Assignment	74.36% (87)	52.99% (62)	38.46% (45)
Relational	14.53% (17)	26.50% (31)	17.95% (21)
Conditional	1.71% (2)	6.84% (8)	27.35% (32)
Others (wrong questions)	9.40% (11)	13.68% (16)	16.24% (19)
Total	100% (117)	100% (117)	100% (117)

In Table 3 we present the distribution of the pupil-generated questions over the different categories of mathematical complexity for *PP* 1. As far as mathematical complexity is concerned, we did not observe the expected decrease in the percentage of one-step problems and the expected increase in the number of multi-step problems from *PP* 11 to *PP* 13.

This unexpected pattern was completely due to the students' over-whelming tendency to react to PP 11 with the following simple and straightforward two-step combine question: "How many candies do they have altogether?" Apart from this, within one-step and multi-step problems, the percentage of problems involving (only) additive computations generally went down from *PP* 11 to *PP* 13, whereas the opposite was true for problems involving (at least one) multiplicative step.

Table 3. Distribution (in percentages and absolute numbers) of pupils' responses over the different categories of mathematical complexity in *PP* 1

	PP 11	*PP* 12	*PP* 13
One-step addition/subtraction	26.50% (31)	30.77% (36)	6.83% (8)
One-step multiplication/division	10.26% (12)	16.24% (19)	19.65% (23)
Total of one-step	36.75% (43)	47.01% (55)	26.50% (31)
Multi-step addition/subtraction	52.99% (62)	23.93% (28)	17.95% (21)
Multi-step multiplication/division	0.85% (1)	15.33% (18)	39.32% (46)
Total of multi-step	53.85% (63)	39.32% (46)	57.26% (67)
Others (wrong questions)	9.40% (11)	13.68% (16)	16.24% (19)
Total	100% (117)	100% (117)	100% (117)

The distribution of the problems posed over the different semantic categories for *PP* 1 is given in Table 4. The percentage of problems involving only one semantic relation goes down from *PP* 11 ("create an easy problem") to *PP* 13 ("create a very difficult problem"), whereas the reverse is true for the multi-relation problems. Within one-relation problems, the most frequently occurring additive structures are combine and compare, while equal group and multiplicative comparison are the most popular ones for the multiplicative structures. Not surprisingly, we did not find any additive problem with a change structure even though this relation is known to be the easiest for children (because the static problem statement did not provide any reasonable starting point to generate a dynamic change problem); moreover, no multiplicative 'rectangular pattern' or 'Cartesian product' structures were observed (first, because these problem types are known to be less familiar to children and, second, because the initial problem statement did not support in any way the generation of a problem with such a semantic structure).

In line with what was found in the above mathematical analysis, the percentage of additive compare and combine questions generally goes down from *PP* 11 to *PP* 13, whereas the opposite is true for problems involving one of the equal group and multiplicative comparison. For the multi-relation problems, there is more variation in semantic structures in *PP* 13 than in *PP* 11. While there is only one multi-relation problem in *PP* 11, namely, "Who will eat up his/her own candies fastest?" (compare/equal group), in *PP* 13 we observed several more complex semantic problems like "After 2 days, how many more candies will Xiaoqiang and Xiaofang have than Xiaoli?" (change/combine/equal group/compare).

So, we can conclude that most students tended to actually pose more linguistically, mathematically, and semantically complex problems when instructed to generate a difficult question (*PP* 13) and more questions with a simple linguistic, semantic and mathematical structure when an easy question (*PP* 11) was required.

Table 4. Distribution (in percentages and absolute numbers) of pupils' responses over the different categories of semantic complexity in *PP* 1

	PP 11	*PP* 12	*PP* 13
One-relation additive combine	65.81% (77)	28.21% (33)	15.38% (18)
One-relation additive compare	13.68% (16)	22.22% (26)	4.27% (5)
One-relation equalize	0% (0)	0% (0)	0.85% (1)
One-relation multiplicative comparison	0% (0)	0% (0)	2.56% (3)
One-relation equal group	10.26% (12)	16.24% (19)	17.09% (20)
Total of one-relation	89.74% (105)	66.67% (78)	40.17% (47)
Multi-relation	0.85% (1)	19.66% (23)	43.59% (51)
Others (wrong questions)	9.40% (11)	13.68% (16)	16.24% (19)
Total	100% (117)	100% (117)	100% (117)

4.1.2 Results for PP task 2

Our results revealed that the three items of *PP* task 2, namely *PP* 21, *PP* 22 and *PP* 23, elicited, respectively, 22.22, 4.27 and 25.64% appropriate responses. A logistic regression on the number of appropriate responses to *PP* 2 items revealed a main effect of 'item', $x^2(2) = 46.93$, $p < 0.00015$. Pair wise comparisons showed that this main effect occurred because *PP* 22 got significantly appropriate responses than *PP* 21 and *PP* 23 ($p < 0.00015$), while there were no differences between *PP* 21 and *PP* 23. How can we explain that the percentage of RPs was significantly lower for *PP* 22 ($100 \div 8 = 12$) than for the two other *PP* items?

A possible reason why performance on *PP* 22 was worse than on *PP* 21 is that in recent textbooks and test materials division with remainder (DWR) problems like the item *PP* 21 (where one has to round up to the next integer rather than to round down to the previous one) have become already rather common. For instance, the Chinese National Curriculum Standards on Mathematics (NCSM) contains an example of a boat item that is very similar to the *PP* 21 item used in our study, but no examples of an item like our *PP* 22 item. A possible explanation for the fact that *PP* 23 was also significantly easier than *PP* 22 for students is that, after all, *PP* 23 is a rather standard one, in the sense that 12.5 is 'simply' the computational outcome of a division of 100 by 8.

Table 5 shows the distribution of the different other categories of math stories for the realistic *PP* task 2. This table reveals that NP ("no problem") was by far the largest error category and that the percentage of CIPs (contextually inappropriate problems) was larger than the percentage of MIPs (mathematically inappropriate problems). In general, these results were similar to, and even somewhat lower than, those from the pilot study.

A comparison of the overall results on the two *PP* tasks revealed that performance on the standard task *PP* 1 (86.98% appropriate responses) was much better than on the non-standard task *PP* 2 (only 17.38 % appropriate responses). A logistic regression revealed that this difference was significant, $x^2(1) = 97.61, p < 0.00015$.

This suggests that pupils' inability to generate appropriate math stories in *PP* 2 was primarily due to their confusion about the realistic modeling issues, and more particularly, to their inability to understand and accept that the mathematical expression 100 ÷ 8 can have (completely) different answers (namely 13, 12 and 12.5, respectively) and/or to their inability to generate (contextually proper) situations that fitted with these expressions, and not merely to their unfamiliarity with problem posing tasks as such.

Table 5. Distribution (in percentages and absolute numbers) of pupils' responses over the different response categories for the three items of *PP* task 2

Math story code	*PP* 21 (100 ÷ 8=13)	*PP* 22 (100 ÷ 8=12)	*PP* 23 (100 ÷ 8=12.5)
RP	22.22% (26)	4.27% (5)	25.64% (30)
CIP	22.22% (26)	26.50% (31)	9.40% (11)
MIP	12.82% (15)	19.66% (23)	16.24% (19)
NP	42.74% (50)	49.57% (58)	48.72% (57)
Total	100% (117)	100% (117)	100% (117)

Some examples of student-generated problems being scored in the first three categories (RP, CIP and MIP) are given below:

RP (realistic problems):

- "100 students take boats. One boat takes at most 8 students. How many boats are needed?" for *PP* task 100 ÷ 8 = 13.
- "Xiaoming will buy pens for his teachers on Teachers' Day. One pen is 8 yuan. How many teachers will receive a pen if he has 100 yuan?" for *PP* task 100/8 = 12.
- "100 kilos of apples are stored in 8 baskets equally, how many kilos of apples are there in each basket?" for *PP* task 100/8 = 12.5.

CIP (contextually inappropriate problems):

- "104 candies are shared by 8 children, how many candies does each child have?" for *PP* task 100/8 = 13.
- "8 students shared 100 kilos of apples, How many kilos did each student get?" for *PP* task 100/8 = 12.
- "There are 100 apples in 8 apples trees, how many apples are there in each tree?" for *PP* task 100/8 = 12.5.

MIP (mathematically inappropriate problems):

- "Students take 13 buses to travel, each bus has 8 seats, and four seats are left empty, how many students are there?" for *PP* task 100/8 = 13.
- 108 candies are shared by 8 children, how many candies does each student have?" for *PP* task 100/8=12.
- Xiaoming has 100 yuan and Xiaoli has 8 yuan, how much does Xiaoming have more than Xiaoli?" for *PP* task 100/8 = 12.5.

Table 6 shows the percentage of children succeeding in posing, respectively, three, two, one and zero realistic problems for *PP* task 2.

Table 6. Distribution of the pupils (in percentage and absolute numbers) according to the numbers of realistic math stories produced in *PP* 2

3	2	1	0	Total
2.56% (3)	7.69% (9)	29.06% (34)	60.68% (71)	100% (117)

As shown in this table, nearly 60 % of the students did not produce a single realistic word problem and only 10 % succeeded in posing at least two realistically appropriate problems (compared to the 90 % of the students who posed at least two questions successfully in *PP* task 1). So, confronted with a non-standard (realistic) problem posing task, only a small percentage of students could come up with appropriate problem situations for the mathematical expressions provided; most of them had difficulty in understanding the mathematical formulae 100 ÷ 8 = 13 and 100 ÷ 8 = 12 and mapping them to an appropriate problem situation. For example, they constructed inaccurate problems like "8 students shared 100 apples equally and each student got another half apple, then how many apples does each student have?" for 100 ÷ 8 = 13 or like "8 students shared 100 apples equally and each student ate half apple, then how many apples does each student have?" to for 100 ÷ 8=12.

4.2 Results for the PS tasks

4.2.1 Results for *PS* task 1

The overall percentage of correct answers on the three *PS* 1 items was 62.96 %. A comparison of this percentage with the percentage of appropriate responses on the three *PP* 1 items (86.98 %) revealed that the overall performance for the (less familiar) *PP* 1 was significantly better than for the (more familiar) *PS* 1, $x^2(1) = 50.40$, $p < 0.00015$. This difference in favour of *PP* 1 is, after all, not so surprising, since in *PP* 1 pupils were only invited to generate increasingly difficult problems (but did not fail if they did not do it, as long as their self-generated problems were mathematically appropriate), whereas in *PS* 1 they actually received problems of an increasing difficulty level.

Table 7 shows the success rate of the three items of *PS* task 1. From this table, we can see that students did well on the first two items, which were, respectively, of a low and intermediate difficulty level, but performed very weak on the third one, which was anticipated to be a very difficult one. A logistic regression predicting 'correct answers' for *PS* 1 revealed a significant main effect of 'item', $x^2(2) = 94.53, p < 0.00015$. Pair wise comparisons showed that this main effect occurred because performances on *PS* 13 were significantly lower than on *PS* 11 and *PS* 12 ($p < 0.00015$) with no difference between the latter two.

Table 7. Percentage and absolute number of correct answers on the three items of *PS* task 1

PS 11	*PS* 12	*PS* 13
88.03% (103)	88.89% (104)	11.97% (14)

Table 8 shows the distribution of students' performance on *PS* task 1. Compared to performance on *PP* 1, the percentage of students (11.97%) who solved three questions correctly is much less than that of students (73.50%) who posed three correct problems.

Table 8. Distribution of the pupils (in percentage and absolute numbers) according to the numbers of correct answers in *PS* 1

PS 11	*PS* 12	*PS* 13
88.03% (103)	88.89% (104)	11.97% (14)

So, overall performance on *PS* 1 was not so good, but this can be largely explained by the fact that *PS* 13, which is a multi-relation problem with an equalize component, – a problem type that is known to be very difficult for children of that age (Carey 1991; English 1998a).

4.2.2 Results for PS task 2

Compared with the realistic math stories posed (*PP* 2), the percentage of students who behaved realistically in *PS* 2 was somewhat better. Whereas the overall percentage of realistic reactions on the three *PP* 2 items was only 17.38, the *PS* 2 elicited an overall percentage of 35.61 realistic responses. A logistic regression analysis revealed that this difference was significant, $x^2(1) = 26.64, p < 0.00015$.

The percentage of realistic answers (RAs) on the three items of *PS* 2 differed considerably: respectively, 29.91, 16.24, and 60.68 (see Table 9). A logistic regression on the number of 'realistic answers' to *PS* 2 items revealed a main effect of 'item', $x^2(2) = 46.93$, $p < 0.00015$. Pair wise comparisons showed that *PS* 23 got significantly more correct answers that *PS* 21 ($p < 0.00015$), which in turn got more correct answers than *PP* 22 ($p < 0.0002$).

Table 9 shows that the percentage of realistic answers (RAs) in *PS* 23 is much more than that in *PS* 21 and *PS* 22, whereas the reverse is true for the CIA.

Table 9. Percentage (and absolute number) of different answers for the three items of *PS* task 2

Answer code	*PS* 21 (100 ÷ 8=13)	*PS* 22 (100 ÷ 8=12)	*PS* 23 (100 ÷ 8=12.5)
RA	29.91% (35)	16.24% (19)	60.68% (71)
CIA	47.86% (56)	42.74% (50)	7.69% (9)
MIA	9.40% (11)	20.51% (24)	11.97% (14)
NA	12.82% (15)	20.51% (24)	19.66% (23)
Total	100% (117)	100% (117)	100% (117)

This can be explained by the fact that *PS* 23 is actually a standard situation that can be solved "realistically" by simply responding with the outcome of the calculation work (namely $100 \div 8 = 12.5$), without any further interpretational activity.

Meanwhile, the performance on *PS* 21 was better than on *PS* 22, the reason being more or less the same as for the difference between *PP* 21 and *PP* 22. Namely, probably due to the frequent citation of the buses item in the scientific and professional literature, division with remainder (DWR) items like our item *PS* 21, where one has to round up to the next whole number, have become rather common in recent textbooks and test materials.

Finally, as shown in Table 10, the percentage of children who succeeded to solve three, two and one item from *PS* 2 was, respectively, 10.26, 13.68 and 48.72%. So, about 73% of all pupils produced at least one correct answer, which was much less than the percentage of students (about 96%) who successfully solved more than one word problem in *PS* 1.

Table 10. Percentage (and absolute number) of correct answers in *PS* 2

3	2	1	0	Total
10.26% (12)	13.68% (16)	48.72% (57)	27.35% (32)	100% (117)

It means that students performed much better on solving standard problems than on solving non-standard problems requiring realistic sense-making. On this latter type of tasks, pupils demonstrated a strong tendency to exclude real knowledge and contextual considerations from their problem solving activities due to their extensive experience with an impoverished diet of standard word problems, and to the lack of systematic attention to the mathematical modeling perspective in their mathematical lessons, as suggested in several other studies reviewed by Verschaffel *et al.* (2000).

4.3 Relationship between *PP* and *PS*

4.3.1 Relationship between PP 1 and PS 1

There appeared to be a significant relationship between *PP* 1 and *PS* 1 (Spearman Rho = 0.2308, $t(115) = 2.54$, $p = 0.0061$). The data of task pair 1 (*i.e.*, the standard problem situation) supported the hypothesis that there would be a significant relationship between the number of correct reactions in *PP* 1 and in *PS* 1. The distribution of the percentages is provided in Table 11.

A further analysis of the data of pair 1 also supported the hypotheses that students who posed a larger linguistic, mathematical, and semantic variety of problems also performed better on the problem solving tasks.

There was a significant relationship between the number of linguistically different question types posed in *PP* 1 and the number of correct answers in *PS* 1 (Spearman Rho = 0.2911, $t(115) = 3.26$, $p = .0007$), a significant relationship between the number of multiplication/division questions in *PP* 1 and the number of correct answers in *PS* 1 (Spearman Rho = 0.3112, $t(115) = 3.512$, $p = 0.0003$), and a significant relationship between the number of multi-relation problems in *PP* 1 and the number of correct answers in *PS* 1 (Spearman Rho = 0.1656, $t(115) = 1.801$, $p = 0.0372$).

Table 11. Relation between number of problems posed (*PP*) and number of problems solved (*PS*) correctly for task pair 1

Correct *PS* / Correct *PP*	0	1	2	3	Total
0	1.71% (2)	0.85% (1)	1.71% (2)	0% (0)	4.27% (5)
1	0% (0)	2.56% (3)	0.85% (1)	0.85% (1)	4.27% (5)
2	0.85% (1)	2.56% (3)	13.68% (16)	0.85% (1)	17.95% (21)
3	0.66% (1)	10.26% (12)	52.14% (61)	10.26% (12)	73.50% (86)
Total	3.42% (4)	16.24% (19)	68.38% (80)	11.97% (14)	100% (117)

4.3.2 Relationship between PP task 2 and PS task 2

The results of the overall comparison between students' performances on the two kinds of realistic tasks are given in Table 12.

Table 12. Relation between number of problems posed (*PP*) and number of problems solved (*PS*) for task pair 2

PS	*PP*				Total
	0 *RP*	1 *RP*	2 *RPs*	3 *RPs*	
0 RA	19.66% (23)	6.84% (8)	0.85% (1)	0% (0)	27.35% (32)
1 RA	28.21% (33)	16.24% (19)	3.42% (4)	0.85% (1)	48.72% (57)
2 RAs	10.26% (12)	1.71% (2)	0.85% (1)	0.85% (1)	13.68% (16)
3 RAs	2.56% (3)	4.27% (5)	2.56% (3)	0.85% (1)	10.26% (12)
Total	60.68% (71)	29.06% (34)	7.69% (9)	2.56% (3)	100% (117)

There was again, a significant relationship between the two variables, indicating that students with a higher global score on the three PP tasks performed better on the three *PS* tasks (Spearman Rho = 0.1990, t (115) = 2.182, p = 0.0157). The same result was observed in our pilot study.

4.4 The impact of order

As explained in the Methods section, because our pilot study revealed a highly significant impact of "order" (*i.e.*, problem posing first or problem solving first) on the success rate on the *PP* task 2, we drastically extended the period between the problem posing and the problem solving test up to twenty days.

To test whether the order still had an impact on pupils' performance on the problem posing task, we compared the overall *PP* and the *PS* performance data of the 62 pupils who received the *PP* tasks first and the 55 pupils who received the *PS* tasks first.

Table 13. Mean success rate (and standard deviations) for the *PP* tasks and the *PS* tasks in the *PP-PS* and the *PS-PP* group for task pair 1

	PP-PS group (n = 62)		*PS-PP* group (n = 55)		Total (n = 117)	
	Mean	SD	Mean	SD	Mean	SD
Mean score for *PP* 1 (max. score = 3)	2.65	0.75	2.56	0.79	2.61	0.77
Mean score for *PS* 1 (max. score = 3)	2.00	0.63	1.76	0.64	1.89	0.64

Table 14. Mean success rate (and standard deviations) for the *PP* tasks and the *PS* tasks in the *PP-PS* and the *PS-PP* group for task pair 2

	PP-PS group (n = 62)		*PS-PP* group (n = 55)		Total (n = 117)	
	Mean	SD	Mean	SD	Mean	SD
Mean score for *PP* 2 (max. score = 3)	0.53	0.74	0.51	0.77	0.52	0.75
Mean score for *PS* 2 (max. score = 3)	1.00	0.98	1.15	0.83	1.07	0.91

As we can see from Table 13 and 14, for both *PP* tasks the mean score of the *PP-PS* group and *PS-PP* group was similar. The logistic regression analysis revealed no significant 'order' effect, $x^2(1) = 0.34$, $p = 0.5573$ and $x^2(1) = 0.01$, $p < 0.9177$ for *PP* 1 and *PP* 2, respectively. Likewise, for the task *PS* 2 there was no significant difference between the performance in the two groups, $x^2(1) = 0.47$, $p = 0.4909$. Only for *PS* task 1, the effect of 'order' just reached significance, $x^2(1) = 4.20$, $p = 0.0404$, indicating that the group who got the *PS* task after they had already done the corresponding *PP* task obtained a slightly better problem solving score than those who had to start with the *PS* task. Apparently, increasing the interval between the different tests resulted in a drastic decrease of the order effect (observed in our pilot study).

A qualitative analysis of the realistic reactions to *PP* 2 revealed that, in contrast to our pilot study, only 2 problems were generated that were clearly copied from the *PS* task.

5. INTERVIEWS

Because it was impossible to interview all students and because we were only able to ask students a few general questions about their problem-solving and problem-posing experiences in general, we did not get very rich interview data and we could not directly link the obtained interview data to the paper-and-pencil data in a systematic way. Nevertheless, these interview data revealed several interesting general observations that helped us to understand and explain some major findings from the collective tests. Before reporting these general observations, we provide some illustrative excerpts from the interviews with the students relating to each of the five interview questions.

Question 1 (Why did you decide to pose these questions/that question):

Researcher: Why have you chosen to pose this question? For example, the question posed in *PP* task 1 "How many candies do they have altogether?"

Student 1: I posed this question because I knew I only need to add all the candies they have, then I will get the answer.

Student 2: I have often seen that kind of questions in textbooks.

Researcher: Can you solve the problems you posed?

Student: I am not sure.

Researcher: Have you ever thought of it? I mean, have you thought whether you can solve the problems posed?

Student: Frankly, I haven't thought whether I can solve them or not.

Question 2 (Can you pose more questions besides the three questions already posed?):

Researcher: In *PP* 1, you posed these three problems "How many candies do Xiaoqiang and Xiaoli get altogether?", "How many more candies does Xiaofang more than Xiaoli?" and "How many more candies do Xiaoqiang and Xiaoli have than Xiaofang?" Can you pose more questions besides these three?

Student 1: Surely, for example, "Five days later, how many candies will Xiaoqiang and Xiaoli eat together?"

Question 3 (How do you think about the making of math stories?):

Researcher: How do you think about the making math stories situation? Do you find it difficult or easy?

Student 1: It is very difficult. I don't understand how 100 ÷ 8 can sometimes equal 12, at another time 13, and at still another time 12.5, I have never seen that kind of task.

Student 2: The making stories like 100 ÷ 8=13 was difficult. In my opinion, the exact answer for 100 ÷ 8 is 12.5; so, one has to construct a mathematical expression 100 ÷ 8 + 0.5 if we want to get at 13. For example, "8 students shared 100 apples equally and each student got another half apple, then how many apples does each student have?"

Student 3: I found the second problem posing task very difficult. With respect to 100 ÷ 8 = 13, I thought 104 divided by 8 is 13, so I constructed a situation (100 + 4) ÷ 8. For example: 100 animals attend the running game, sometime later another 4 animals come. If all the animals play in 8 groups, how many animals are there in each group?

Question 4 (What kind of help do you need from the teacher to make math stories?)

Researcher: How can the teacher help you in making math stories that fit with the expressions 100 ÷ 8=13, 100 ÷ 8=12, or 100 ÷ 8=12.5?

Student: The teacher may give me some similar examples to explain when the answer is 12, when the answer is 13, and when the answer is 12.5.

Question 5 (Could you tell me more clearly how you solved this question?)

Researcher: Can you tell me more clearly how you solved the buses problem (= *PP* 21)?

Student: Firstly, I recognized I should use division to solve this problem. Then, I divided the total number of students (100) by the number of students in each bus (8), and then I got the number of buses (12.5).

Researcher: (using a picture comprising of a large number of buses) Now, suppose that you, as part of a group of 100 peers, are waiting for those buses. A first bus comes and 8 students enter, (writing number 8 on the picture of the first bus); then another bus comes and 8 more students enter (writing the number 8 on the picture of the second bus). Can you tell me how many buses will be needed to transport all students?

Student: (writing number 8 on the next buses until he has filled 12 buses), 12 buses are needed, and 4 students are left.

Researcher: How will you deal with these 4 remaining students?

Student: We need another one bus to transport these four remaining students. Now I understand, we need 13 buses to take 100 students, not 12.5 buses.

Researcher: Do you think that 12.5 can be the number of boxes in the moving boxes situation?

Student 1: No, I don't think 12.5 can be the correct number of boxes, but I don't know what the answer should be if 12.5 is not suitable.

Student 2: I think 12.5 can be the number of boxes, for maybe there are some big boxes and some small boxes, so a small one can be considered as a half one.

Student 3: I don't think 12.5 can be the number of boxes, but I didn't notice that. I just divided the total number of students by the number of students moving one box. So, I got at the answer 12.5 automatically.

Our analysis of these interviews led to the following conclusions. For the standard situations, most students reported that they felt quite comfortable when being confronted with *PS* 1, but also when being given *PP* 1. Moreover, most pupils reported (and in many cases also demonstrated) that they knew the answers to the questions they had generated themselves, although a small minority (two out of 30 students in the interview) acknowledged that they could not solve their own questions or that they did not know whether they would be able to solve them. These two pupils belonged to the group of most weakly performing pupils on *PP* 1. So, for these two pupils, *PP* 1 did not elicit the kind of mindful thinking expected by those advocating problem posing as a valuable instructional device in teaching word problem solving. Finally, many students could generate even more (appropriate) questions besides the three questions they had posed for the standard situation in *PP* 1. The non-standard (realistic) problem-posing situations, on the other hand, were considered by most students as very difficult. Typically, this difficulty was explained by referring to the unfamiliarity with the task (namely to generate problems that will lead to different numerical solutions for the same mathematical operation). According to the students, one or two illustrative examples of contextually appropriate problems given by the teacher would have been very helpful to accomplish this task properly. Concerning the non-standard (realistic) problem solving situations, most students said they had automatically given the answer 12.5 (in all three cases), because they had neglected to activate their real-world knowledge to interpret the outcome of their computational work in terms of the original problem situation, even though they demonstrated afterwards, during the interview, possession of the common-sense knowledge that there are no “half buses” or “half boxes”.

5. DISCUSSION

Recent research has documented that there is a close relationship between problem posing and problem solving in (elementary) arithmetic. However, most studies investigated the relationship between problem posing and problem solving only by means of standard problem situations. To overcome that shortcoming, a pilot study with Chinese fourth-graders was done to investigate this relationship using a non-standard, realistic problem situation (Chen *et al.* 2004, 2005). The results revealed a significant positive relationship between students’ problem posing and solving abilities. Based on that pilot study, a more systematic study was carried out, with a more representative sample of Chinese pupils, combining a standard and a realistic problem situation and using a combination of paper-and-pencil tests and interviews, to confirm and further unravel the relationship between problem posing and problem solving.

Results of the study revealed, firstly, that students performed rather well both on the task wherein they had to pose and to solve different standard word problems. Moreover, in the standard problem posing task, pupils in general were aware of the (relative) difficulty of their problems during problem posing, as they created more linguistically, mathematically and semantically complex problems when asked to generate a difficult question for a given situation and more problems with a simple linguistic, semantic or mathematical structure when asked to raise an easy one. While this latter finding can be interpreted as evidence that problem posing invites students to engage in metacognitive activity, the observation from the interviews that some pupils were unable to solve their self-generated questions or to say whether they would be able to solve them, suggests that not all students addressed the problem posing task mindfully.

The finding that students performed much worse on the realistic problem posing and problem solving task (notwithstanding its formal similarity with the standard version) confirmed the conclusion from much previous research (see Verschaffel *et al.* 2000), namely that students have great difficulty with problem situations that require the activation of realistic considerations and sense-making. The fact that these difficulties showed up not only in problem solving settings, but also, and in very similar ways, in settings of problem posing, yields further evidence for this conclusion. The interview data were helpful in further unraveling the solution processes and the accompanying beliefs that lie at the basis of pupils’ non-realistic reactions. As in some previous interview studies about non-realistic problem solving (for a review see Verschaffel *et al.* 2000), many students acknowledged during these interviews that they had automatically answered or generated the problems, without paying any attention to the contextual appropriateness of their reaction.

As far as the relationship between problem posing and problem solving is concerned, the integrated and comparative analysis of both kinds of data revealed that, both for the standard and the non-standard situation, students' problem solving performance was highly correlated with their problem posing performance. Moreover, the interviews suggested that most subjects tended to pose problems that they could solve, thereby providing empirical support to the theoretical argument offered by Kilpatrick (1987) that one's problem posing can be considered an (alternative) index of one's problem solving capacity. However, as said above, this relationship was certainly not perfect. For instance, there were also a few pupils who could not solve the questions posed or did not know if they would be able to solve them.

Several of the above-mentioned findings were already found in the pilot study. However, because of the very short interval between the administration of the problem solving and the problem posing task (and the manifest impact this short time interval had on the nature of the problems generated by the children), some of the findings from that pilot study were difficult to interpret. By working with a much larger interval and by demonstrating that test order (*i.e.*, problem posing first or problem solving first) had little or no impact on students' performance, this methodological problem was resolved in the present study.

The findings of this study provide some educational recommendations with a view to stimulate the development of students' disposition to realistic mathematical modeling, problem solving and problem posing. Firstly, the interview data suggest that many students had the ability to pose (standard) word problems even without much experience in posing problems before. Typically, they produced easy and familiar problems, comprising problem situations and questions they had encountered frequently in their previous word problem solving lessons. Truly original problems were very rare — a finding that was also reported in some other studies with Chinese students (Yang & Wang 2004; Zeng, Lv & Wang 2006). This finding suggests that teachers should not only give problem-posing tasks around standard word problems, but also that they should stimulate and help them to generate different questions, including (linguistically, mathematically, and semantically) less trivial ones and more complex ones. Secondly, our findings confirm that students experience great difficulty with dealing with linking realistic problem situations to (outcomes of) arithmetic operations.

While previous research has documented this difficulty for the link from problem situation to computation, our study shows that this difficulty exists, even to a stronger extent, when students are asked to find a problem situation that fits with a given number sentence (as in our *PP* task 2).

Several reactions from children during the individual interviews suggest that confronting students more with realistic problem solving and problem posing (De Lange 1993; English, 1997a) activities in the mathematics classroom will lead to a growing disposition towards realistic problem solving and problem posing.

REFERENCES

Brown, S. I. & Walter, M. I. (Eds.). (1993): *Problem posing: Reflections and applications.* Hillsdale, NJ: Lawrence Erlbaum Associates. MATHDI **1994e.**02587

Cai, J. & Hwang, S. (2002): Generalized and generative thinking in U.S. and Chinese students' mathematical problem solving and problem posing. *Journal of Mathematical Behavior* **21(4)**, 401–421. MATHDI **2004a.**00537

Cai, J. & Silver, E. A. (1995): Solution processes and interpretations of solutions in solving division-with-remainder story problems: Do Chinese and U.S. students have similar difficulties? *Journal for Research in Mathematics Education* **26(5)**, 491–497. MATHDI **1996e.**03617

Carey, D. A. (1991): Number sentences: Linking addition and subtraction word problems and symbols. *J. Res. Math. Educ.* **22(4)**, 266–280. MATHDI **1992j.**02931

Carpenter, T. P.; Lindquist, M. M.; Matthews, W. & Silver, E. A. (1983): Results of the third NAEP mathematics assessment: Secondary school. *Math. Teach.* **76(9)**, 652–659. MATHDI **1984x.**00303

Chen, L.; Van Dooren, W.; Chen, Q. & Verschaffel, L. (2005): The relationship between posing and solving division with remainder problems among Chinese elementary school children. *Mediterranean Journal for Research in Mathematics Education* **4(2)**, 85–109.

Chen, L.; Verschaffel, L. & Li, X. (2004): An investigation of the relation between posing and solving word problems (In Chinese). *Journal of Mathematics Education* **13(4)**, 67–71.

De Corte, E. & Verschaffel, L. (1996): An empirical test of the impact of operations intuitive models of operations on solving word problems with a multiplicative structure. *Learning and Instruction* **6(3)**, 219–242. MATHDI **1997e.**03055

De Lange, J. (1993): Between end and beginning. Educational Studies in Mathematics for 12-16 Year Olds: 1987–2002. *Educational Studies in Mathematics* **25(1–2)**, 137–160. MATHDI **1995b.**00895

Ellerton, N. F. (1986): Children's made-up mathematics problems — a new perspective on talented mathematicians. *Educ. Stud. Math.* **17(3)**, 261–271. MATHDI **1987a.**07695

English, L. D. (1997a): The development of fifth-grade children's problem-posing abilities.

Educational Studies in Mathematics **34(3)**, 183–217. MATHDI **1998c.**02348

_____ (1997b): Promoting a problem posing classroom. *Teaching Children Mathematics* **4(3)**, 172–178. MATHDI **1998e.**03387

_____ (1998a): Children's problem posing within formal and informal contexts. *Journal for Research in Mathematics Education* **29(1)**, 83–106. MATHDI **1998e.**04113

_____ (1998b): *Problem Posing in Middle School Classrooms.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Fuson, K. C. (1992): Research on whole number addition and subtraction. In: D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 243–275). New York: Macmillan. MATHDI **1993f.**01809

Kilpatrick, J. (1987): Problem formulating: Where do good problems come from? In: A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 123–147). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lowrie, T. (2002): Designing a framework for problem posing: Young children generating open-ended tasks. *Contemporary Issues in Early Childhood* **3(3)**, 354–364.

Mayer, R. E.; Lewis, A. B. & Hegarty, M. (1992): Mathematical misunderstandings: Qualitative reasoning about quantitative problems. In: J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 137–154). Amsterdam: Elsevier. MATHDI **1994b.**00678

Miles, M. B. & Huberman, A. M. (1994): *Qualitative Data Analysis.* Thousand Oaks, CA: Sage. Ministry of Education of Peoples' Republic of China, (NCSM) (2001). Chinese National Curriculum Standards on Mathematics. Beijing: Beijing Normal University Publishing House.

Ministry of Education of Peoples' Republic of China (2001): *National Curriculum Standards on Mathematics,* Experiment Draft (in Chinese). Beijing: Beijing Normal University Press.

National Council of Teachers of Mathematics. (2000): *Principles and standards for school mathematics.* Reston, VA: NCTM. MATHDI **1999f.**04754

Rudnitsky, A.; Etheredge, S.; Freeman, S. & Gilbert, T. (1995): Learning to solve addition and subtraction word problems through a structure-plus-writing approach. *Journal for Research in Mathematics Education* **26(5)**, 467–486. MATHDI **1996e.**03618

Silver, E. A. (1994): On mathematical problem posing. *For the Learning of Mathematics* **14(1)**, 19–28. MATHDI **1994f.**02882

Silver, E. A. & Cai, J. (1996): An analysis of arithmetic problem posing by middle school students. *Journal for Research in Mathematics Education* **27(5)**, 521–539. MATHDI **1997f.**04550

Silver, E. A.; Shapiro, L. J. & Deutsch, A. (1993): Sense making and the solution of division problems involving remainders: An examination of middle school students' solution processes and their interpretations of solutions. *Journal for Research in Mathematics Education* **24(2)**, 117–135. MATHDI **1993g.**00504

Verschaffel, L. & De Corte, E. (1996): Number and arithmetic. In: A. J. Bishop, K. Clements, C.

Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education*, Vol. 4, Part 1, (pp. 99–137). Dordrecht, Netherlands: Kluwer. MATHDI **1998f.**04678

_____ (1997): Teaching realistic mathematical modeling in the elementary school: a teaching experiment with fifth graders. *Journal for Research in Mathematics Education* **28(5)**, 577–601. MATHDI **1998c.**02357

Verschaffel, L.; De Corte, E. & Greer, B. (2000): *Making sense of word problems*. Lisse, Netherlands: Swets & Zeitlinger. 1998e.03645 MATHDI **2000c.**01654

Verschaffel, L.; De Corte, E. & Lasure, S. (1994): Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction* **4**, 273–294.

Verschaffel, L.; De Corte, E.; Lowyck, J.; Dhert, S. & Vandeput, L. (2000): *Supporting mathematical problem solving and posing in upper elementary school children by means of Knowledge Forum*. Deliverable of project No. 2017 CL-Net: Computer Supported Collaborative Learning Networks in Primary and Secondary Education.

Yang, X.; Lv, C. & Wang, B. (2003): Third exposition on mathematics learning by using mathematical situations and posing problems in high school and primary school (In Chinese). *Journal of Mathematics Education* **12(4)**, 76–79.

Yang, X. & Wang, B. (2004): Study about mathematics teaching by using the model of mathematical situations and posing problems in high school and primary school. *Journal of Mathematics Education* **13(4)**, 84–87.

Yoshida, H.; Verschaffel, L. & De Corte, E. (1997): Realistic Considerations in Solving Problematic Word Problems: Do Japanese and Belgian Children Have the Same Difficulties. *Learning and Instruction* **7(4)**, 329–338. MATHDI **1998e.**03645

Zeng, X.; Lv, C. & Wang, B. (2006): Actuality and countermeasure of "posing mathematical problem" on junior high school student. *Journal of Mathematics Education* **15(3)**, 51–53.

Zhang, X. & Ren, Z. (1998): Reactions To: Should High-Stakes Tests Drive Mathematics Curriculum and Instruction? High Stakes Testing from the Chinese Perspective. *Mathematics Education Dialogues* **1998(May-Jun)**, p. 4 and p. 6. ERIC EJ567989