ANNALS OF THE POLISH MATHEMATICAL SOCIETY
5TH SERIES: DIDACTICA MATHEMATICAE 33 (2010)

Elżbieta Mrożek
Institute of Mathematics, University of Gdańsk

Task variables in compare word problems*

Abstract: The paper is a report on the solutions of one-step additive and multiplicative compare problems, which were part of a larger test written in 2007 by 70 thousand 10-year-olds at the beginning of grade IV. A sample of 788 responses was analysed in detail and various types of students' difficulties were identified (the difficulty depended not only on the structure of the problem but also on many other factors).

Tasks in which additive problems alternated multiplicative ones caused serious troubles; in particular, many students (some 10%) correctly solved the first problem and then erroneously used the same operation to the next one, not reading the text attentively, or (also some 10%) used the result of the first task as if it were the given number in the next one ('chain effect'). Thus, the impact of the order of tasks was considerable.

Many solvers searched for key words and could not cope with problems which were formulated in inconsistent language or required inverting the relation (particularly in the case of multiplicative problems). Many students wrote an incorrect operation and a correct answer.

The paper begins with a comprehensive discussion of task variables (context variables, structure variables and format variables) related to possible types of one-step additive and multiplicative compare problems, relevant to children's difficulties. In particular, it is shown why a problem of the type *Joe has 8 marbles. Tom has 5 marbles more than Joe. How many marbles does Tom have?*, which is regarded as an arithmetical one-step problem, is actually a multi-step problem if the number of mental operations is considered.

---

*The paper is a part of the research project "Study of the mathematical knowledge of students after grade III of the elementary school" conceived, directed and supervised by Agnieszka Demby (Demby, 2009; Demby, 2010). This project was part of a larger one "Strategy for mathematics education in Poland" which was supported by grant R11 017 02 of the Polish Ministry of Science and Higher Education.
The author is indebted to Professor Zbigniew Semadeni for his help in preparing the English version of the paper.

1 Introduction

It is well known that students' difficulties with both additive and multiplicative one-step compare word problems are much more serious than those with dynamic word problems corresponding to the same arithmetical operations (this concerns not only children but also undergraduate students, see Lewis and Mayer, 1987; Bestgen, 2009).

Compare problems may concern cardinal numbers and magnitudes of various kind. A problem involving a phrase '3 more apples' or 'greater by 3' may be replaced by 'greater by 3 m', 'heavier by 3 kg', 'longer by 3 hours', etc. If the students grasp the common structure of such problems, their previous successful strategies can be applied to new problems. Otherwise there are too many types of problems to learn to deal with all the cases.

Freudenthal (1973, p. 283) has shown that, given the arithmetical operation $16 - 10$ and a specific context, many different real-life problems can be formulated. We will label them (F1)-(F6) for further reference.

(F1) John is 16 yers old. How long ago was he 10?
(F2) John is 10 years old. In how many years will he be 16?
(F3) John is 10 and Peter is 16. How much is Peter older than John?
(F4) John is 10 and Peter is 16. How much is John younger than Peter?
(F5) John is 10 and Peter is 16. How long ago was Peter as old as John is now?
(F6) John is 10 and Peter is 16. In how many years will John be as old as Peter is now?

Freudenthal also gave further examples which, in practice, are also solved by subtracting $16 - 10$:

(F7) In 1916 John was 10. When was he born?
(F8) John was born in 1910. How old was he in 1916?

Problems (F1)-(F8) illustrate the fact that it is impossible for students to learn all the procedures needed to solve all special cases of compare problems. Many semantic and linguistic nuances may occur in real-life compare problems, which – in school practice – are reduced to a few routine, oversimplified schemes.

A frequently used term is *the structure of a word problem*. However, it is known that the concept of a structure (in mathematics, biology, social sciences) may be intuitively clear and yet difficult to formalize (Van Hiele, 1986). When we consider word problems, doubts arise even in the case of one-step compare problems.

Let us give a sample of questions, without attempting to answer them. Is there a single common arithmetical structure of the problems (F1)-(F6)? If so, what is this structure? Or perhaps several structures should be identified in these problems? Or some of these problems differ only in context or in format (in the sense explained in 2.1 below)?

For example, can one definitely say what is the arithmetical structure of the problem (F2)? A tempting answer is that the structure of the problem is simply the corresponding equation. However, there may be several adequate equations. E.g., in order to solve (F2) one may choose between the equation $16 - 10 = x$ and the equation $10 + x = 16$.

Do problems (F3) and (F4) differ only in the way the problem is formulated? Suppose the question *How much is Peter older than John?* is replaced by *How much is John younger than Peter?*; does then the structure of the problem change to another?

As a geometric background to (F1)-(F8) one should consider the time axis where the age of a person (say, 10 years) is interpreted as the difference of points. In a problem analogical to (F3) starting with, e.g., *John has 10 books and Peter has 16 books...* there is no axis in the background. Should one consider the background time axis as an element of the structure of (F3)?

How to distinguish the difficulties of students with a compare problem which are related to its mathematical structure from difficulties related to its format?

Van Hiele (2002, p. 45) wrote, that: *The signposting along the road may be more or less strong, but the behavior of the other road users causes the whole situation to have an apparent feeble structure*. By analogy, we may say that the structure of a word problem may be strong or feeble, depending on details, particularly during preliminary mathematization (in the sense of Krygowska, 1980). Accordingly, we may say that the problems (F1)–(F6) have a common strong structure 'subtraction $16-10$', and several additional feeble structures.

Let us consider the following problems.

(A) *On Sunday Johnny and Anne found many conkers in a park. Later they used them to make figurines. Johny made 5 figurines and Anne made 3 more figurines than Johnny. How many figurines did Anne made?*

This may be regarded as a compare problem. However, compare problems are usually described as static whereas (A) is not static.

(M_1) *Johnny made 5 figurines. Anne made 3 times as many figurines as Johnny did. How many figurines did Anne make?*

(M_2) *Johnny made a set of 5 figurines. Anne made 3 such sets. How many figurines did Anne make?*

Is the mathematical structure of the compare problem (M_1) different from that of (M_2)? The latter lacks the characteristic words '3 times as many'. Or perhaps (M_1) is essentially the same problem as (M_1) but worded differently? Putting it differently, does replacing (M_1) by (M_2) results in a conceptual change or is it only a linguistic change? If a conceptual change is needed, suitable schemes should be constructed in the child's mind. If only a linguistic change is needed, students may be trained in replacing the compare wording by an usual one. In the sequel we will show that the evaluation of a student's response in a test may depend on which answer to such questions is accepted.

Although we do not answer the above questions, they form a background of our analysis of task variables – texts of one-step compare problems and their semantic and syntactic characteristics.

We will also deal with typical students' incorrect solutions of compare problems and with some sources of the difficulties. However, neither subject variables nor situation variables will be considered here.

2 Compare problems

2.1 Categories of variables in research on problem solving: subject variables, task variables, situation variables

Three basic categories of variables identified by Kilpatrick (1978) have been chosen as the framework of this paper. Specifically we use these categories in the form described by Kulm (1979, 1-6). These three categories are derived from the necessary components of problem-solving event, which are a problem solver (subject) solving a problem (task) under a set of conditions (situation). Any problem-solving event involves a complex interaction among the variables describing these three components. Main categories are the following:

a) **Subject Variables:**

- *Organismic Variables* – for example: age, sex, social-economic status, geographic residence,
- *Trait Variables* – for example: cognitive style, attitude, persistence, mathematical memory, ability to estimate offer promise of being closely associated with problem-solving performance,
- *Instructional History Variables* – for example: the schools attended, mathematical topics studied, problem-solving instruction received by the subject.

b) **Task Variables:**

— *Context Variables* include those which characterize the physical situation of the problem, as well as the language in which the problem is expressed. They are intended to describe the differences between problems having the same mathematical structure,

— *Structure Variables* are intended to describe the intrinsic mathematical structure of a problem. One way to do so is to employ a mathematical formula or relation. Two problems with the same formula could be said to have the same syntactic structure, evidently using this term to refer to the syntax of the formula or relation,

— *Format Variables* describe the different manners or settings in which a problem may be presented. For example, it may be presented along with other problems, with hints, or with the aid of some apparatus.

c) **Situation Variables:**

— *Physical Setting* – for example: type of space (classroom, laboratory, outdoors, etc.), the nature of space (comfortable, stimulating, familiar, etc.), available resources (calculators, measuring instruments, manipulative materials, or amount of time),

— *Psychological Setting* – for example: the purpose of the event (testing, instruction, practice, etc.), the type of procedure (evaluative, prescriptive, diagnostic, etc.), the nature of the learning environment (type of amount of feedback, quantity or quality of interaction).

In this paper we deal only with task variables: context variables (with special attention to the impact of wording on the solutions), structure variables (showing the inherent difficulties with the concept of the structure of a compare problem), and format variables (showing that if a compare problem is presented along with another compare problem of a different type then the difficulty increases).

2.2 Types of one-step compare problems

Comparison of two numbers or two magnitudes (such as length, weight etc.) means finding out whether they are equal or which of them is greater. The result may be expressed with one of the signs $=$, $<$, $>$.

Additional quantitative information is provided by the following two kinds of comparison. In the first, one gets the difference of the given numbers or magnitudes (by subtracting the smaller from the greater). In the second, one gets the quotient (ratio), provided that the magnitudes are of the same kind, i.e., their quotient is a number. One tacitly assumes that the quotient is positive; thus, a phrase of the type '−2 times as many' is not acceptable.

Textbook problems on comparison phrased as, e.g., 'greater by 3' or 'smaller by 3' concern the additive structure of numbers (or pertinent magnitudes) and are called *additive compare problems*; the involved operations are: addition or its inverse, subtraction. Problems phrased as, e.g., '3 times as many', 'threefold' or '3 times less' concern the multiplicative structure[1] and are called *multiplicative compare problems*; the involved operations are: multiplication or its inverse, division. Several authors (e.g., Christou and Philippou, 1998) describe compare problems as static problems concerning numerical relations between sets, comparing two disjoint sets rather than a set and its subsets as in typical problems on subtraction or division.

A theoretical wording of a general description of additive problems was often presented as follows: given that one set is compared to another set, one set is given a label of *the referent set* and the other set is called *the compared set*. The third component of a compare problem is called *the difference set*.

A conspicuous feature of this formulation is that sets are treated as if they were numbers. This attitude was popular in pedagogical and psychological papers in the first half of 20th century. Sets were called equal when they had the same number of elements. Traces of this language can be seen in the above description, but such identification of sets with numbers is incompatible with the language of set theory. E.g., in the situation: *Johnny has 4 apples, Kathy has 7 apples* the difference set is pointless because these sets of apples are disjoint. Probably it was intuitively meant that the difference set consists of 3 apples. But which of the 7 apples? For a mathematician such terminology is unacceptable.

In some descriptions of additive problems by the third component is meant the amount by which one set is greater than the other set. Zofia Cydzik (1978, 83-84) put it this way: *In an additive compare problem the first magnitude is concrete while the second magnitude (phrased, e.g., as 'greater by 2') is abstract and determines a numeric relation between the given magnitude and the unknown one.*

Nesher, Greno and Reley (1982) identified six types of one-step additive compare problems, labeled here 1a-6a.

[1]More precisely, the multiplicative structure of the set of numbers and the structure of multiplying magnitudes by numbers, respectively.

1a Joe has 8 marbles. Tom has 5 marbles more than Joe. How many marbles does Tom have?

2a Joe has 8 marbles. Tom has 5 marbles less than Joe. How many marbles does Tom have?

3a Joe has 8 marbles. Tom has 13 marbles. How many marbles does Tom have more than Joe?

4a Joe has 8 marbles. Tom has 13 marbles. How many marbles does Joe have less than Tom?

5a Joe has 8 marbles. He has 5 marbles more than Tom. How many marbles does Tom have?

6a Joe has 8 marbles. He has 5 marbles less than Tom. How many marbles does Tom have?

Analogously, one may identify[2] six types of one-step multiplicative compare problems, labeled here 1m-6m.

1m Joe has 12 marbles. Tom has 3 times as many marbles as Joe has. How many marbles does Tom have?

2m Joe has 12 marbles. Tom has 3 times less than Joe. How many marbles does Tom have?

3m Joe has 12 marbles. Tom has 3 marbles. How many times as many as Tom does Joe have?

4m Joe has 12 marbles. Tom has 3 marbles. How many times less than Joe does Tom have?

5m Joe has 12 marbles. Joe has 3 times as many marbles as Tom has. How many marbles does Tom have?

6m Joe has 12 marbles. Joe has 3 times less than Tom. How many marbles does Tom have?

Both additive and multiplicative compare problems were traditionally listed in the Polish national curricula for grades III and IV of primary school. In 1959 four types: 3a, 4a, 3m, 4m were listed explicitly. In 1983 these four types were augmented by another four: 1a, 2a, 1m, 2m. In 1999 unified national curricula were abandoned in Poland. The present list of expected competencies

[2] The letters *a* and *m* in the labels are abbreviations of the words 'additive' and 'multiplicative'.
Another list of types of multiplicative compare problems was given by Harel et al. (1988).

(worked out by the Ministry of Education) mentions: *solving additive compare problems after grade III and multiplicative compare problems after grade VI*, without specifying the types.

Problems of types 5a, 6a, 5m, 6m have not figured in Polish curricula.

2.3 Students' difficulties depending on task variables

For long teachers have learned from experience that compare problems are specifically difficult for children and are more difficult to solve than problems that do not contain compare phrases. This was confirmed by many researchers: Kinitsch & Greeno (1985), Valentin & Chap Sam (2004), Nesher, Greeno & Riley (1982). Yet in Poland, at the beginning of the 1970s, some educators argued that compare problems should not be more difficult than other one-step word problems involving the same arithmetical operations, because if students understand these operations they should solve compare problems as well. This view is definitely an oversimplification and the argument misses the point.

Multiplicative compare problems are generally more difficult that additive one for several reasons. Multiplication and division are more difficult (both conceptually and computationally) than addition and subtraction. Moreover, one should be aware that multiplicative compare problems are an introduction to learning percents, ratios, proportions. The phrase *3 times less* may be expressed as the ratio 1 : 3 or as $\frac{1}{3}$ of a magnitude while *3 times as many as* is equivalent to 300%. Children's difficulties with multiplicative compare problems are akin to those with ratios and percents which are also based on multiplication and division, but it would be naive to think that knowledge of these operations is enough for those topics.

Semadeni (1986, 335-340) divided students' difficulties with compare problems into conceptual and linguistic. Conceptual difficulties are of mathematical and psychological character. In compare problems no actions are performed, nothing happens. Students have to deal with verbal information concerning some static relation between numbers or magnitudes, e.g.,

(*) *Johnny has 4 apples. Kathy has 3 more apples than Johnny.*

From such a verbal information, at a language level, without manipulation of concrete objects, the child must infer what is the right operation to perform. Staub and Stern (1997) point out that, e.g., problem of the type 3a becomes much easier if the question *How many marbles does Tom have more than Joe?* is replaced by *How many marbles must Joe get in order to have the same amount of marbles as Tom?*

An additional difficulty with (*) is that two questions naturally arise:

(**) *How many apples does Kathy have?*

(***) *How many apples do they have together?*

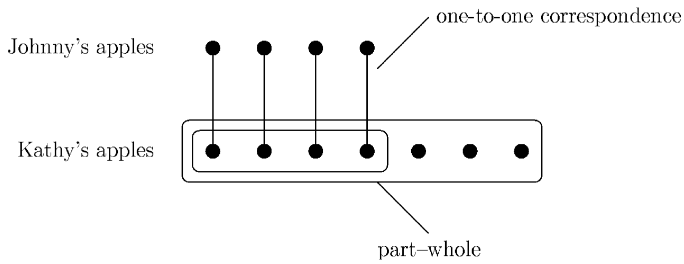
In part-whole situations and in dynamical ones usually only one question arises.

In didactical analyses the problem (*)-(**) is regarded as a one-step problem. In fact, its solution requires only *one arithmetical operation* $4+3$. What is often overlooked, however, is that it is a complex problem in the sense that a scheme with *two mental operations* is needed:

— one-to-one correspondence between the apples of Johnny and part of apples of Kathy,

— a part-whole operation.

Moreover, in such a case a standard schematic picture may be misleading.



This scheme – either explicitly drawn or being implicit in the student's mind – may suggest an incorrect answer to question (**). The teacher expects the answer: 7, whereas the student sees 11 elements as if (***) were the question.

Problems of types 5a, 6a, 5m, 6m require one more operation: inverting the given relation. E.g., in the problem 5a listed above the two sentences *Joe has 8 marbles. He has 5 marbles more than Tom* should be changed to *Joe has 8 marbles. Tom has 5 marbles less than Joe.* In this way the type 5a becomes the easier type 1a. Thus, to solve a problem of type 5a the student has to perform three successive operations at a verbal level.

Analyzing pertinent examples and theoretical constructs in Piaget's publications (e.g., Inhelder and Piaget, 1970, chapter XVII) and translating them (with unavoidable simplification) to a language understood by mathematicians, we may explain the point as follows. Suppose that a concrete-operational child is given two mental operations, say α and β. Even if she/he is able to perform the inversion β^{-1} and the composition $\alpha \cdot \beta$ separately, combining mentally them into $\alpha \cdot \beta^{-1}$ is beyond their capabilities; students can do this much later when they reach the level of formal operations[3]. It is clear that such

[3]P. Nowicki (1981) mentioned the results of Shayer et al. (1976). They carried out tests

composite operation $\alpha \cdot \beta^{-1}$ is needed for types 5a and 6a (unless students are sufficiently trained to learn the whole sequence of successive operations).

The conceptual difficulties are combined with linguistic ones. Many children mix up phrases which are specific to the additive compare problems with similarly sounding phrases specific to the multiplicative problems.

Invariably, in descriptions of the behavior of children solving compare problems, the problem of key words appears, particularly in the context of US schools. Typically, the student searches for characteristic words. If *more than* is found, almost automatically she/he chooses addition; similarly *less than* prompts subtraction while *times* prompts multiplication. Such solvers fixate on key words and numbers, neglecting the background information which is helpful in constructing a situation model of the problem.

Verschaffel (1994) warns against the automatic use of key words. Although in problems of types 1a, 2a, 1m, 2m the results of using key words are satisfactory, this method does not foster understanding. Generally, it has adverse effect and causes comprehension errors. Unsuccessful problem solvers do not build a mental model representing the situation described in the text, but devote their attention to key words and numbers. Stern (1993) points out that lack of access to flexible language makes compare problems so difficult.

Even if students are told that in many problems the key words should be properly inverted, learning key words for each kind of one-step problem is not feasible. Verschaffel points out that understanding the equivalence of conditions such as *a is less than b by c* and *b is greater than a by c* is indispensable for anybody solving compare problems.

Verschaffel (1994) emphasizes that many teachers are not aware that the compare problems which require inversion (i.e., problems of types 3a-6a and 3m-6m) are essentially more difficult than problems of types 1a-2a and 1m-2m.

2.4 The effect of mixing additive compare problems with multiplicative ones

Additive problems are taught first; serious troubles begin when children start to learn the multiplicative problems as well. This is particularly noticeable when both types appear during the same lesson. Additive problems, easier and apparently well learnt before, start to cause difficulties anew.

The adverse effect of mixing additive and multiplicative problems was confirmed by nation-wide survey of students' competences carried out in Poland

---

modeled on those of Piaget for 10 000 students of English schools. One of the results was that only 20% of students $14\frac{1}{2}$ years old had reached the level of (early) formal operations.

in 1981-1988 (Nowik, 1988). In one of the tasks in a test written by 5000 students of grade IV there were two questions concerning the same given data: the first question was of type 3a, while the second was of type 4m.

Specifically, one of the versions was: *Jimmy has 15 postage stamps and Michael has 5 stamps. How many stamps does Jimmy have more than Michael? How many times as many as Michael does Jimmy have?* It turned out that about 90% of the students solved (correctly) the additive part, whereas only some 50-60% solved the multiplicative one. Moreover, some 15% of the students attempted to solve the multiplicative part as if it were additive. Surprisingly, 10% of those who had version A of that test and 30% of those who had version B answered the second question combining both comparisons together.

In 2004-2008 competencies of students of grade III were studied by M. Dąbrowski. His conclusions (Dąbrowski, 2007, 2009) were different from those quoted above. He claimed that – in contrast to a dominated opinion – the results concerning multiplicative compare problems turned out better than those concerning additive ones. Specifically, he collated pairwise the results of the following additive and multiplicative compare problems:

a) One-step problems: *Karol and Ela picked conkers in a park. Karol picked 30 conkers and Ela picked 6 conkers more than Karol. How many conkers did Ela pick?* (79% of students solved this correctly)

   Bartek and Jurek picked conkers in a park. Bartek picked 15 conkers and Jurek picked 3 times as many conkers as Joe did. How many conkers did Jurek pick? (95% of students solved this correctly).

b) Chain-complex[4] problems: *There are two screening rooms in a cinema. There are 122 seats in the first room. In the second room there are 35 seats more than in the first. How many seats are there altogether in the cinema?* (48% of students solved this correctly)

   Ania has two shelves with books. There are 12 books on the upper shelf and 3 times as many books on the lower shelf. How many books are there altogether? (53% of students solved this correctly).

c) Non-chain-multi-step problems: *Ania and Bożena picked mushrooms in a forest. Ania picked 8 mushrooms more than Bożena. Altogether they picked 50 mushrooms. How many mushrooms did Ania pick?* (10% of students solved this correctly)

[4] A word problem is called *chain-complex* if it is a multi-step problem which can be solved by so-called *arithmetic method*, i.e., the method of successive arithmetical one-step inferences. In contrast, a *non-chain-multi-step* problem or an *algebraic word problem* cannot be decomposed so that the method of successive arithmetical inferences could be used.

There are automobile models on two shelves in a shop. There are 3 times as many models on the upper shelf than on the lower shelf. Altogether there are 129 models on the two shelves. How many models are there on the upper shelf? (12% of students solved this correctly).

Nevertheless some points should be made. The one-step problems and the parts of mutli-step problems used in that study were of types 1a and 1m only. Types 3a, 4a, 3m, 4m, which are the core of compare problems, were missed. None of the above problems required inverting mental operations.

To solve a problem of type 1m one has only to change the wording, e.g., a phrase of the form *find a number 3 times greater than* should be replaced by *multiply by 3*; an analogous compare phrase should be replaced by *add 3*. Consequently, learning to solve such compare problems does not require new mental schemes; learning new phrases will do.

In contrast, in types 3m, 4m one asks about the ratio of two numbers or magnitudes in the case where none of them is part of the other. When children learn division, they get partitive and quotitive problems, and then they only have to compare the number of elements of sets with those of their subsets; quotients of numbers of elements of disjoint sets never appear in such division problems. Thus, research on multiplicative compare problems must, first of all, concern types 3m and 4m.

Also another factor should be considered. Attention resources are limited and simultaneously required for computation and comprehension. Consequently, if one increases the complexity of the computation and/or the size of numbers in a word problem, this may capture resources needed for understanding the relations and make the comprehension of the problem more difficult (Bestgen, 2009). One may wonder whether the difficulty of operations $122+35+122$ in c) was equivalent to that of $3 \times 12 + 12$.

3 An empirical study on solutions of compare problems by Polish students after grade III

3.1 Organization of the study

The study consisted of two stages. The first was combined with part of a survey carried out by GFR[5]. Annually, at the beginning of the school year, GFR organizes a survey[6] for students of grades IV, V, VI in Poland. Teachers

[5]Gdańska Fundacja Rozwoju im. Adama Mysiora

[6]"Sesje z Plusem"

who wish to participate in it enrol themselves, receive the tasks for their students with hints how to evaluate the responses, and send the results of the evaluation – the number of points scored by each of their students (Demby, 2009; Demby, 2010) for each task – to an electronic database of GFR.

In 2007 about 70 000 students starting grade IV took part in the test (the total number of students of grade IV in Poland at that time was about 400 thousand).

The test consisted of eight problems concerning mental computations within the range 100, 1000 and 10 000, using inverse operations, measuring time, choosing operations to given one-step word problems, additive and multiplicative compare problems.

During the second stage of the study three of 16 Polish voivodeships were selected, namely those in which the 2006 results of the test were the closest to the nation-wide average. Teachers from these provinces were asked to send the paper sheets filled out by each of their students.

In this paper we analyse the responses to tasks No 2 and No 4 described below, which concerned compare problems[7].

3.2 Characteristics of the samples

Two sets of students are considered here. By *the whole group*, WhG, we mean the group of those 70 000 students whose scores have been sent to the database.

By *the group of students whose responses have been examined in detail*, GD, we mean the subset of WhG consisting of those students whose sheets with solutions have been sent by the teachers. There were 788 students in GD. Their papers were analysed both quantitatively and qualitatively in two ways:

1. with the same details that were available from the database (in order to compare the samples WhG and GD),

2. also in more detail, identifying competencies used by the respondents and types of correct, partially correct and incorrect solutions.

The crucial difference between WhG and GD was that in the former case only scores (assigned by the teachers) for each task were known whereas in the case of GD the responses written by each student could be analysed.

It turned out that the difference between the quantitative results of GD and those of WhG is not significant. Therefore the results of GD are presented with those of WhG in the background.

[7] I have published some of these results, addressed to teachers, in (Drewczynska, 2009) and (Mrożek, 2010).

Although the number of students in GD is large, we cannot claim that GD was a representative sample of students of grade IV in 2007. The way data were collected suggests a somehow biased sample because the teachers who volunteered to participate in the survey were likely to be more active and devoted. Yet, those teachers were just beginning to work with a new team of students. They wanted to have reliable knowledge of how these students had been prepared by teachers of grades I–III. Certainly they were not interested in overstating the students' competences.

3.3 Contents of tasks

There were two versions A and B of the tasks. Those concerning compare problems read as follows.

Task 2A (Task No 2, version A): In place of dots write down the operation and compute the result:

Kathy thinks of number 18.

a) *The number smaller by 2 than Kathy's number:* ....

b) *The number 3 times greater than Kathy's number:* ....

c) *The number greater by 9 than Kathy's number:* ....

d) *The number 6 times smaller than Kathy's number:* ....

Task 2B (Task No 2, version B): In place of dots write down the operation and compute the result:

Stan thinks of number 24.

a) *The number 4 times smaller than Stan's number:* ....

b) *The number greater by 8 than Stan's number:* ....

c) *The number 3 times greater than Stan's number:* ....

d) *The number smaller by 6 than Stan's number:* ....

Thus, Task 2 consisted of four subtasks of the following types:
Type 1a — subtask c) of task 2A and subtask b) of Task 2B,
Type 2a — subtask a) of task 2A and subtask d) of Task 2B,
Type 1m — subtask b) of task 2A and subtask c) of Task 2B,
Type 2m — subtask d) of task 2A and subtask a) of Task 2B.

Task 4A (Task No 4, version A):

The TV commercial of cakes 'Delicious' lasted 8 seconds and that of cakes 'Splendid' lasted 32 seconds[8].

[8]In this task (and in an analogous Task 4B) the context is different than that in the above

a) How many times longer was the commercial of cakes 'Splendid'?

Computation: ....

Answer: ....

b) By how many seconds shorter was the commercial of cakes 'Delicious'?

Computation: ....

Answer: ....

Task 4B (Task No 2, version B):

The TV commercial of washing powder 'Super' lasted 9 seconds and that of washing powder 'Extra' lasted 36 seconds.

a) By how many seconds shorter was the commercial of washing powder 'Super'?

Computation: ....

Answer: ....

b) How many times longer was the commercial of washing powder 'Extra'?

Computation: ....

Answer: ....

Thus, Task 4 consisted of two subtasks of the following types:
Type 4a – subtask b) of task 4A and subtask a) of Task 4B,
Type 3m – subtask a) of task 4A and subtask b) of Task 4B.

There were no tasks of the types 3a and 4m in the test.

The tasks were edited so that each subtask of version A had an exact equivalent in version B, with marginally different numbers. However, the order of the subtasks was deliberately changed. The rationale for the change was to prevent student copying. Those who had devised the tasks believed that the relative difficulty of both versions was identical, bearing in mind that the test was not written immediately after first lessons on the topic, but after several months of training. Yet, the results of this decision turned out surprising.

3.4 Quantitative results

Diagram 1 shows the percentages of correct responses to Tasks 2 and 4 in the test for WhG. Conspicuously, the results for some subtasks of versions A and B differ significantly.

problems: comparing time rather than comparing cardinal numbers. However, an analysis of the students' responses shows no hint that this fact influenced their solutions; they seemed to regard the words 'longer' and 'shorter' as 'more' and 'less'.

Percentage for types 1a, 2a, 1m, 2m are within the range of 60-75%. In A the first subtask concerned additive problems and the results for those types were a little better than those for multiplicative ones. In B – conversely – the first subtask concerned multiplicative problems and the results for those types were also a little better.

However, percentage for type 4a is 63% in A and 79% in B; for type 3m it is 40% in A and 32% in B. The first subtask in B concerned additive problems and the results for those types were much better; the first subtask in A concerned multiplicative problems and the results for those types were also much better.

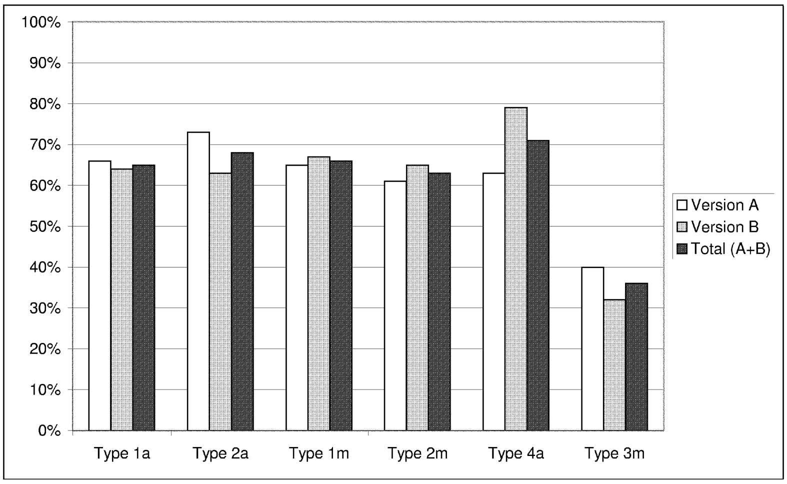


Diagram 1.

It is clear that those differences are the effect of the change of order of the subtasks. This question will be discussed in 3.6 below.

3.5 Types of solutions

We now systematically present (jointly for A and B) various kinds of the results for each subtask of types: 1a, 2a, 4a, 1m, 2m, 3m (those kinds which were found in less than 3-4% of solutions are not listed here).

Subtasks of Type 1a, i.e., c) in 2A and b) in 2B.

— most frequent kind (about 55% of responses): *Both correct operation and correct result.*

— about 20% of responses: *No operation, correct result.*

— about 15% of responses: *Incorrect operation* (most frequently it was multiplication instead of addition).

— about 5% of responses: *No operation, incorrect result.*

Subtask of Type 2a

— most frequent kind (50-60% of responses, more frequent in A): *Both correct operation and correct result.*

— about 15% of responses: *No operation, correct result.*

— about 10–15% of responses: *Incorrect operation* (more frequent in B).

— about 10% of responses: *No operation, incorrect result.*

— about 5%: *No response.*

Subtask of Type 1m

— most frequent kind (about 55% of responses): *Both correct operation and correct result.*

— about 20% of responses: *No operation, correct result.*

— about 10% of responses: *Incorrect operation* (most frequently it was multiplication instead of addition).

— about 10% of responses: *Correct operation, incorrect result.*

— about 5% of responses: *No operation, incorrect result.*

Subtask of Type 2m

— most frequent kind (50-55% of responses, more frequent in B): *Both correct operation and correct result.*

— about 15-20% of responses: *Incorrect operation* (most frequently it was subtraction, sometimes multiplication).

— about 15% of responses: *No operation, correct result.*

— about 10% of responses: *Correct operation, incorrect result.*

— about 5% of responses: *No operation, incorrect result.*

Only some 35% of students had complete correct responses to all the subtasks of Task 2.

Types 1a, 2a, 1m, 2m could be solved by choosing the suitable operation given a characteristic phrase: smaller by 2, 3 times grater, etc. The results for addition compare problems do not differ much from those for multiplicative ones. Thus, basing on these subtasks only, one cannot definitely say whether multiplicative problems are more or less difficult than the additive ones.

Subtask of type 4a

- most frequent kind (about 60% of responses, more frequent in B): *Correct operation, correct result, correct answer.*
- about 20% of responses: *Incorrect operation* (most frequently it was division).
- about 10% of responses: *No response* (particularly frequent in version A where this type was in the second subtask).
- about 5% of responses (more frequent in B): *Correct operation, incorrect result, correct answer.*

Subtask of type 3m

- most frequent kind (50% of responses): *Incorrect operation* (most frequently it was multiplication or addition).
- about 25% of responses (more frequent in A): *Correct operation, correct result, correct answer.*
- about 10% of responses: *Correct operation, correct result, incorrect answer.* Students had troubles with writing down the answer in the language of multiplicative compare, used words taken from the additive one or some irrelevant words.
- about 10%: *No response* (particularly frequent in version B, where this type was in the second subtask).

Problems of types 4a and 3m turned out difficult. Additive compare problems were correctly solved by 65-80% of students (depending on the version) whereas only 35-40% solved the multiplicative problems of type 3m, which turned out much more difficult than the additive type 4a.

Only 28% of students gave complete correct responses in all the subtasks of Task 4. Partial solutions of additive problems were found in 6-13% of papers, of multiplicative ones in 11%.

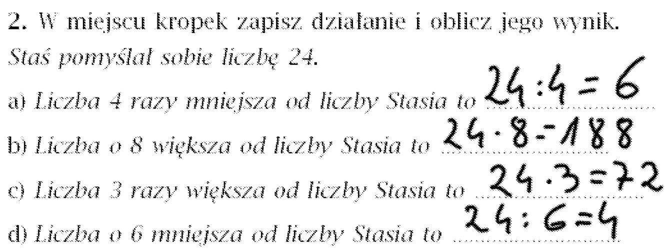
3.6 The impact of the order of tasks on the students' performance

In 3.3 we mentioned that the order of the subtasks was deliberately changed. We now present results showing the impact of preceding subtasks on the responses to the following ones.

The impact of the first subtask. A conspicuous phenomenon noted in the study was the impact of the first subtask on the way students solved the successive problems. As a typical example we reproduce a piece of work of a student with Task 2B (Example 1. "4 razy" means "4 times", "o 8" means "by 8".) In the first subtask a) one had to divide the given numbers; in the following subtasks b), c), d) several students either divided the numbers (when they read 'smaller') or multiplied them (when they read 'greater') even if the wording clearly suggested subtraction or addition.

In version A, conversely, in the first subtask a) the expected operation was subtraction. In the remaining subtasks several students performed only subtraction and addition even if the word 'times' clearly suggested a multiplicative problem.

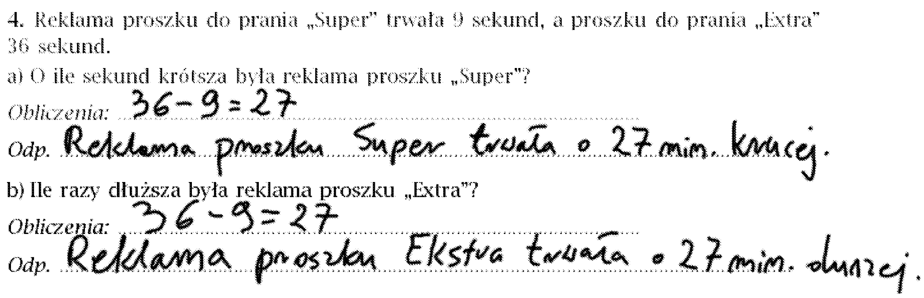
No fewer than 15% of students behaved this way, that is, responded to all four subtasks of Task 2 according to whether first of them was additive or multiplicative.



2. W miejscu kropek zapisz działanie i oblicz jego wynik.
Staś pomyślał sobie liczbę 24.
a) *Liczba 4 razy mniejsza od liczby Stasia to* 24:4 = 6
b) *Liczba o 8 większa od liczby Stasia to* 24·8=188
c) *Liczba 3 razy większa od liczby Stasia to* 24·3=72
d) *Liczba o 6 mniejsza od liczby Stasia to* 24:6=4

Example 1.

A similar phenomenon was noted in Task 4. In version B the first subtask was additive and 80% of students responded correctly. However, half of them (i.e., 40% of those who had version B) treated the second subtask as if it also were additive. In version A, conversely, the first subtask was multiplicative and 45% wrote the correct operation; yet, almost one forth of them (or 10% of those who had version A) wrote multiplication or division also in subtask (Example 2. „O ile sekund krótsza" means "By how many seconds shorter"; „Ile razy dłuższa".)



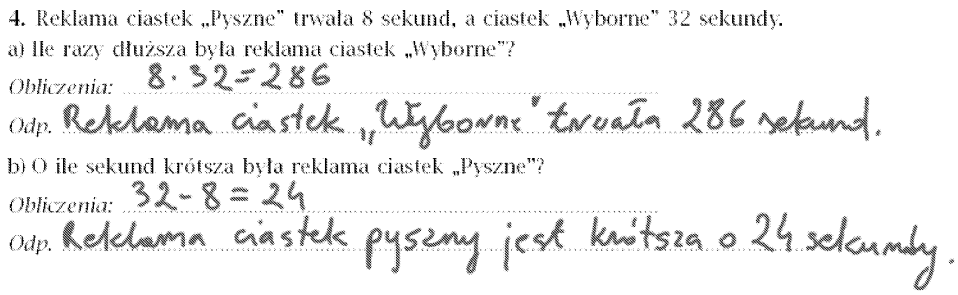
4. Reklama proszku do prania „Super" trwała 9 sekund, a proszku do prania „Extra" 36 sekund.
a) O ile sekund krótsza była reklama proszku „Super"?
Obliczenia: 36-9=27
Odp. Reklama proszku Super trwała o 27 min. krócej.
b) Ile razy dłuższa była reklama proszku „Extra"?
Obliczenia: 36-9=27
Odp. Reklama proszku Ekstra trwała o 27 min. dłużej.

Example 2.

Key words. Second phenomenon noted in the study was the technique of key words. In the interviews[9] students were arguing: *Since in the task is the word 'more', I should use addition or multiplication* and *Since in the task is the word 'less', I should use subtraction or division.* This strategy was correct in Task 2. This may be a reason that 40% of responses correctly solved whole Task 2.

The same strategy in Task 4 looks as follows:

1) In the subtask which contains the word 'shorter' students use subtraction (60% of the responses) or division (10% of responses) (see Example 3, subtask b).
2) In the subtask which contains 'longer' students use addition (10% of responses) or multiplication (10% of responses) (see Example 3, subtask a).


4. Reklama ciastek „Pyszne" trwała 8 sekund, a ciastek „Wyborne" 32 sekundy.
a) Ile razy dłuższa była reklama ciastek „Wyborne"?
Obliczenia: 8 · 32 = 286
Odp. Reklama ciastek „Wyborne" trwała 286 sekund.
b) O ile sekund krótsza była reklama ciastek „Pyszne"?
Obliczenia: 32 - 8 = 24
Odp. Reklama ciastek pyszny jest krótsza o 24 sekundy.

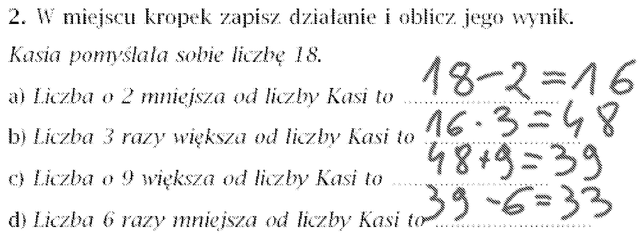
Example 3.

This strategy was successful only in one subtask of Task 4. Only connecting the word 'shorter' and subtraction was accurate. It is likely that less students would have solved this correctly if it had contained the question *By how many seconds longer was the commercial of cakes 'Delicious'?*

Chain effect. The third phenomenon was the procedure of using the number resulting from the first subtask as the given number (e.g., as the number that Kathy thought of) in the second one. Then the number resulting

[9]In September 2008 I additionally interviewed 65 forth-grade students, different from the ones I present in the article. They attended two schools in the Gdańsk-Gdynia agglomeration and two schools in the countryside in Kujawsko-Pomorskie Voivodship. Their role was to solve Task 2 and Task 4 and to explain them. During the interviews I observed all the phenomena described in this paper. In particular, students were often influenced by the formulation of the first subtask and used this (in various ways) in the following subtasks. They also used key words very often. For part of the students the multiplicative compare problems were totally incomprehensible; they responded they did not know how to solve the task.

from the second subtask was used as the given number in the third one, and so on. This way was used in 5% of responses in Task 2 and in 12% of responses in Task 4.


2. W miejscu kropek zapisz działanie i oblicz jego wynik.

Kasia pomyślała sobie liczbę 18.

a) *Liczba o 2 mniejsza od liczby Kasi to* 18-2=16

b) *Liczba 3 razy większa od liczby Kasi to* 16·3=48

c) *Liczba o 9 większa od liczby Kasi to* 48+9=39

d) *Liczba 6 razy mniejsza od liczby Kasi to* 39-6=33

Example 4.

Mathematical operations written in the responses. The following grading was used in Tasks 2 and 4: one point for each correct operation and one point for each correct result. Significant differences have been noted between the results concerning the two parts of the responses: the result was correct much more often than the operation. Such a phenomenon may suggest lack of the need for writing down the operation when the result is clearly visible rather than lack of competence.

Writing the answers. Among the responders there was a group students (about 5%) who wrote down the correct operation, but had troubles with writing down a correct sentence as the answer. The mistakes may be related to the linguistic complexity of phrases such as "4 times longer" or "longer by 4" as well as to inaccurate reading of the task[10]. Sometimes students added new words and objects which did not appear in the task at all (prices, weights, time).

The feeling of failure and lack of motivation to undertake the second subtask. Part of the students left the last subtask unsolved. This could be related to their feeling of failure (similarly, when the first subtask concerned multiplicative compare problems, some of the students did not attempt to solve the second subtask concerning an additive compare problem).

Some of the students solved the subtasks concerning additive compare problems and skipped the multiplicative subtasks.

[10]In her MSc thesis Beata Sieradzińska (1995) described a case of a child who argued, that *by three times more* means *four times more*. Normally at school such phrases are not used. However, it can be understood that the phrase "it became more expensive by 300%" is the same as "it became 400% more expensive"

4 Interpretation, implications, and comments

Researchers generally agree that compare problems are specifically difficult.

In the present study percentages of correct solutions for types 1a, 2a were close to those for types 1m, 2m (the difference did not exceed 5 percentage points).

For types 4a, 3m the differences were significant (10-15 percentage points, depending on which version A, B is considered). This vividly shows the significance of the way compare problems are formulated. Several authors (Lewis & Mayer, 1987; Verschaffel, 1994; Bestgen, 2009) have stressed the need for distinguishing *consistent* and *inconsistent* wordings of a problem. The latter means that the key words prompt an inappropriate operation (e.g., the word 'less' suggests subtraction while the right operation is addition), leading to a reversal error[11] . For any inconsistent compare problem it is possible to change the wording to get an inverse of the relational statement and to write down a consistent one that requires the same computation steps to solve it. Anyway, *consistent-inconsistent is a significant task variable.*

This variable might be essential for differences between results for type 3m (inconsistent language) and types 1a, 2a, 1m, 2m, and 4a (consistent language). The number of correct responses to tasks of type 3m was twofold lower than those of types 1m or 2m.

In fact, the difficulty of tasks of type 4a turned out much the same as that of tasks of types 1a, 2a (or even 4a was somehow easier). The point is that these tasks were formulated in a consistent language, whereas those of type 3m – in inconsistent. In problems of type 4a the word 'shorter' suggests 'less' and hence subtraction, whereas in 3m the word 'longer' suggests 'more' and may be misleading, suggesting multiplication.

It seems that there is no need to pay special attention to additive problems of types 1a, 2a in grades IV-VI. The main difficulties lie in the inverse operations and in inconsistent language.

The results of Nesher et al. (1982) differed from those presented above: types 3a, 4a were easier than 1a, 2a, but their study concerned preschool children who undoubtedly dealt with their tasks in ways different from those of Polish students in grade III. In fact, if the compared sets consist of concrete objects (manipulated or drawn) and the child forms pairs in a one-to-one correspondence, then the surplus elements are clearly visible and can be counted; no subtraction is needed.

[11] It is known (see, e.g., Fischbein et al., 1985) that children may choose a non-correct operation when they solve a word problem with numbers conflicting the rules of their primitive mental models.

The findings of the present study show a considerable impact of the order of tasks on the solutions. This confirms the sparse available data, in particular those worked out by Nowik (1988). Although he did not mention the impact explicitly, an analysis of his report shows an effect similar to that described above.

When students were asked to solve a problem in which they first read an additional compare problem and then a multiplicative one, the results of the first part did influence their responses to the other. Specifically, in the second part of the task they used the same arithmetic operation as in the first part (in spite of the text). Moreover, part of the students described in Nowik's report used the result of the first part in the irrelevant computations performed in the second part (the chain effect described above). Thanks to the fact that both additive and multiplicative compare questions appeared here in a single problem it was possible to identify these kinds of students' difficulties.

The findings confirm an long-standing opinion of many educators that both types of compare problems should not be given during a single lesson before the students can solve each type separately.

An controversial question if whether it is possible to work out a series of paradigmatic examples of compare problems which would effectively help students to solve such problems. Structural analysis of compare problems presented in this study shows that such a program is not feasible. The wide spectrum of pertinent problems, situations and ways of dealing with them cannot be reduced to a few schemes.

The scope of the study did not allow to examine the influence of the kind of the concrete situation of the problem on the students' performance. The only traces of this were found in responses to tasks of types 3m and 4a. Many solvers, writing down their answers, changed the context situation, e.g., in the tasks on the time of a commercial they mentioned irrelevant prices or weights. Perhaps they did it unthinkingly, fixed on numbers and operations, not used to paying attention to non-mathematical context of the problem.

We now gather the findings about students' errors and possible ways of preventing them.

1. Although a majority of researchers point out that the multiplicative compare problems are generally more difficult than the additive ones, such categorical statement is an oversimplification. The difficulty level of completing a compare problem depends on many task variables: on its mathematical structure, in particular on the computation steps needed for its solution and the size of numbers, on the real-life situation presented in the story, in particular whether it is static or dynamic, on

the wording of the problem, on the manner the problem is presented and whether it is a separate problem or is given along with other problems.

Results presented above give hints concerning the relative difficulty of some types of compare problems. Multiplicative problems which require inverting an operation are particularly difficult. Part of the students have not even started to solve such problems or used key words and chose an incorrect operation. Teachers in grades IV-VI should spend more time on various cases of inverting relations, particularly in multiplicative compare problems.

2. Tasks which included both additive and multiplicative compare problem turned out particularly difficult. Students, who apparently did not master any of these types, were confused by them. A significant number of students correctly chose the operation in the first subtask of the problem (according to which type, additive or multiplicative, was identified by them), and then – in spite of unambiguous wording – incorrectly used an operation of the same type in the following subtasks. For instance, in the first subtask of problem 2A the solvers had to write the number smaller by 2; some of them in the three consecutive subtasks wrote addition or subtraction (instead of multiplication or division) also in multiplicative subtasks.

   This may be a combined result of insufficient practice with such problems and perhaps of a hidden curriculum, of students' belief that each part of a school task must concern the same strictly specified topic.

   Therefore teachers should give all types of compare problems and series of mixed problems during a single lesson, but not before the students know well enough how to solve each task separately.

3. The study confirms the known fact that many students after grade III do not read problems carefully, paying attention to numbers and choosing the operations by intuition. Students should be accustomed to read texts of problems trying to understand what it is about. They should be discouraged to search for key words mechanically. An effective method is explaining the story or situation in the problem with the student's own words, with gestures or sketches, and perhaps even without specifying the numbers.

4. Several students had troubles with writing the arithmetical operation that properly corresponds to the solution of the problem. Teachers should pay attention to writing down operations even for easy problems, because later, when greater numbers, fractions or algebraic expressions appear

in the problems, lack of knowledge will be much more difficult to cope with.

Tasks, in which the students are presented with several operations and have to choose an operation leading to a solution of the given problem may be very effective (the numbers should be selected so that all the proposed operations should be typical and known to students).

5. Writing an answer was difficult for many students. Sometimes the answer was irrelevant, related neither to the performed operation nor to the content of the problem. This may be a result of not reading the problem carefully and not analysing it. Teachers should practice reading the text of the problem after having computed the solution, pinpointing various parts of the problem, with special attention to the problem question and additional assignment.
6. Multiplying one-digit number by a two-digit number was difficult for several students. This should be practiced, although it is time-consuming.

Skemp (see, e.g., Skemp 2002) distinguished between instrumental and relational understanding. Inspired by this, we may approach the question of compare problems in two ways:

a) *I have read the problem, I understand its sense, and I know how to solve it* or

b) *I have read the problem, I have found the key words, and I choose the operation indicated by the key words.*

Successful solvers tend to construct a meaningful and rich representations of the problem situations, their sense, and related intuitions. Helping students to construct suitable mental models is difficult and time-consuming. Teaching key words, taking didactical shortcuts is easier, saves time and requires less effort, but the resulting knowledge is practically worthless and causes failing when the relations should be reversed or otherwise treated in a flexible way. Moreover, such knowledge is not suitable for more advanced topics: percentages, ratios, map scale, etc.

References

B e s t g e n, Y.: 2009, Computational requirement and the misunderstanding of language inconsistent word problems, in: N. Taatgen and H. van Rijn (eds.), *Proceedings of CogSci 2009*, Cognitive Science Society Inc., 1500-1505.

C h r i s t o u, C., P f i l i p p o u, G.: 1998, The Developmental Nature of Ability to Solve One-Step Word Problems *Journal for Research in Mathematics Education* **29(4)**, 436-442.

C y d z i k, Z.: 1978, *Nauczanie matematyki w klasie pierwszej i drugiej szkoły podstawowej*, Wydawnictwa Szkolne i Pedagogiczne, Warszawa.

D ą b r o w s k i, M.: 2007, *Pozwólmy dzieciom myśleć*, Centralna Komisja Egzaminacyjna, Warszawa.

D ą b r o w s k i, M. (red.): 2009, *Badanie umiejętności podstawowych uczniów trzecich klas szkoły podstawowej. Trzecioklasista i jego nauczyciel. Raport z badań ilościowych 2008*, Centralna Komisja Egzaminacyjna, Warszawa.

D e m b y, A.: 2009, *Umiejętności uczniów na początku klasy IV (cz. 1)*, *Matematyka w Szkole. Czasopismo dla nauczycieli szkół podstawowych i gimnazjów*, GWO, **48**, 3-8.

D e m b y, A.: 2010, *Badanie wiedzy matematycznej uczniów po klasie III szkoły podstawowej*, in: Proceedings of the conference on mathematics education in Poland, 28-29 April 2010, Instytut Problemów Współczesnej Cywilizacji, Warszawa.

D r e w c z y n s k a, E.: 2009, Ile razy więcej?, *Matematyka w Szkole. Czasopismo dla nauczycieli szkół podstawowych i gimnazjów*, GWO, **48**, 9-11.

F i s c h b e i n, E., D e r i, M., N e l l o, M., M a r i n o, M.: 1985, The rule of implicit models in solving verbal problems in multiplication and division, *Journal of Research in Mathematics Education*, **16**, 3-17.

F r e u d e n t h a l, H.: 1973, *Mathematics as an educational task*, D. Reidel Publishing Company, Dordrecht-Holland.

H a r e l, G., P o s t, T., B e h r, M.: 1988, On the textual and semantic structures of mapping rule and multiplicative compare problems, in: A. Borbas (Ed.) *Proceedings of the XII International Congress, Psychology of Mathematics Education (PME)*, Volume II (pp. 372-379). Budapest: PME.

I n h e l d e r, B., P i a g e t, J.: 1970, *Od logiki dziecka do logiki młodzieży. Rozprawa o kształtowaniu się formalnych struktur operacyjnych* PWN, Warszawa.

K i l p a t r i c k, J.: 1978, Variables and methodologies in research on problem solving, A paper presented at the Research Workshop on Problem Solving, University of Georgia, 1975. Reprinted in L. L. Hatfied (Ed.), *Mathematical Problem solving: papers from a research workshop.*

K i n i t s c h, W., G r e e n o, J. G.: 1985, Understanding and solving word arithmetic problems, *Psychological Review*, **92(1)**, 109-129.

K r y g o w s k a, Z.: 1979, *Zarys dydaktyki matematyki*, Wydawnictwa Szkolne i Pedagogiczne, Warszawa.

K u l m, G.: 1979, The classification of problem-solving research variables, in: Goldin G. A., McClintock C. E. *Task Variables in Mathematical Problem Solving*, ERIC, Clearinghouse for Science, Mathematics and Environmental Education, College of Education The Ohio State University; reproduced 1984 by Franklin University Press.

L e w i s, A., M a y e r, R.: 1987, Students' miscomprehension of relational statements in arithmetic word problems, *Journal of Educational Psychology*, **79**, 363-371.

M r o ż e k, E.: 2010, O porównywaniu, *Matematyka w Szkole. Czasopismo dla nauczycieli szkół podstawowych i gimnazjów*, GWO, **53**, 32-34

N e s h e r, P., G r e e n o, J. G., R i l e y, M. S.: 1982, The development of semantic categories for addition and subtraction, *Educational Study in Mathematics*, **13**, 373-394.

N o w i c k i, P.: 1981, Uwagi o kształceniu matematycznym w szkole masowej *Wiadomości Matematyczne* **23**, 249-255.

N o w i k, J. (red.): 1988, *Wyniki ogólnopolskich badań osiągnięć uczniów, nauczycieli i szkół 1981-1988. Matematyka. Tom V*, Instytut Kształcenia Nauczycieli im. Władysława Spasowskiego w Warszawie, Warszawa.

S e m a d e n i, Z.: 1986, Porównywanie różnicowe i ilorazowe, w: Semadeni Z. (red.): *Nauczanie początkowe matematyki. Tom 3,* Wydawnictwa Szkolne i Pedagogiczne, Warszawa, 335-340.

S h a y e r, M., K ü c h e m a n n, E., W y l a m, H.: 1976, The distribution of Piagetian stages of thinking in British middle and secondary school children, *British Journal of Educational Psychology*, **46**, 164-173.

S i e r a d z i ń s k a, M.: 1995, *Porównywanie różnicowe i ilorazowe w nauczaniu arytmetyki.* Praca magisterska wykonana w Instytucie Matematyki UW pod kierunkiem prof. Z. Semadeniego.

S k e m p, R. R.: 2002, Relational Understanding and Instrumental Understanding, in: Tall D., Thomas M. *Intelligence, Learning and Understanding in Mathematics. A Tribute to Richard Skemp*, Post Pressed, Flaxton, Australia.

S t a u b, F. C., S t e r n, E.: 1997, Abstract reasoning with mathematical constructs, *Journal of Educational Research*, **27(1)**, 63-75.

S t e r n, E.: 1993, What makes certain arithmetic word problems involving the comparison of sets so difficult for children?, *Journal of Educational Psychology*, **85**, 7-23.

V e r s c h a f f e l, L.: 1994, Using retelling data to study elementary school children's representations and solutions of compare problems, *Journal for Research in Mathematics Education*, **25**, 141-165.

V a l e n t i n, J. D., C h a p S a m, L.: 2004, Roles of semantic structure of arithmetic word problems on pupils ability to identify the correct operation, *International Journal for Mathematics Teaching and Learning*, 1-14.

V a n H i e l e, P. M.: 1986, *Structure and Insight. A Theory of Mathematics Education*, Academic Press, London.
V a n H i e l e, P. M.: 2002, Similarities and Differences Between the Theory of Learning and Teaching of Skemp and the Van Hiele Levels of Thinking,in: D. Tall and M. Thomas (eds.), (2002) *Intelligence, Learning and Understanding in Mathematics. A tribute to Richard Skemp*, Post Presses, Flaxton, Australia, 27-47; Polish translation: Podobieństwa i różnice między teoria uczenia się i nauczania Skempa a poziomami myślenia van Hielego, *RToczniki Polskiego Towarzystwa Matematycznego, Seria V, Dydaktyka Matematyki*, **25**, 183-202.

Typologia zadań na porównywanie różnicowe i ilorazowe

S t r e s z c z e n i e

Praca ta jest raportem z badań dotyczących opanowania porównywania różnicowego i ilorazowego przez polskich uczniów po klasie III. Powszechnie wiadomo, że zadania dotyczące porównywania różnicowego oraz porównywania ilorazowego sprawiają uczniom ogromne trudności, znacznie większe niż dynamiczne zadania tekstowe o tej samej strukturze arytmetycznej. Ponadto, jeśli zastąpimy w zadaniu na porównywanie różnicowe zwrot np. *o 5 więcej* sformułowaniem typu *o 5 cm wyższy, o 5 zł droższy*, otrzymamy wiele dalszych typów zadań z nowymi trudnościami.

Zadania na porównywanie różnicowe dotyczą dodawania i odejmowania, a więc struktury addytywnej (zbioru liczb lub wielkości), natomiast porównywanie ilorazowe (np. *3 razy więcej* lub *trzykrotnie*) dotyczy struktury multyplikatywnej i stosunków. Zwrot *3 razy mniej* można wyrazić w postaci stosunku 1:3 lub ułamka $\frac{1}{3}$, natomiast zwrot *3 razy więcej* to 300% w języku procentów. Podczas rozwiązywania zadań na porównywanie ilorazowe ujawniają się trudności podobne do tych, które potem pojawiają się przy rozwiązywaniu zadań dotyczących stosunków czy procentów.

W pracy tej stosowane są oznaczenia 1a-4a podstawowych typów zadań jednodziałaniowych addytywnych: 1a – *o tyle więcej*, 2a – *o tyle mniej*, 3a – *O ile więcej?*, 4a – *O ile mniej?* (określenia słowne są zgodne z terminologią stosowaną dawniej w polskich programach nauczania). Analogiczne typy dla zadań jednodziałaniowych multyplikatywnych to 1m, 2m, 3m, 4m.

Wielu autorów podkreślało, że zadania na porównywanie różnicowe to zadania statyczne, w których jeden ze zbiorów porównywany jest z innym. Cydzik (1978, 83-84) pisze, że w zadaniu na porównywanie różnicowe pierwsza wielkość ma charakter konkretny, druga wielkość (np. „o 2 więcej") ma charakter

abstrakcyjny, oznacza związek ilościowy między wielkością daną w zadaniu i wielkością poszukiwaną. W zadaniach na porównywanie różnicowe i ilorazowe mowa jest nie o czynnościach, lecz o statycznych relacjach, podanych werbalnie; uczniowie mają dokonać obliczeń do zadania, w którym nic się nie dzieje.

Zadania typów 1a-4a i 1m-4m są jednodziałaniowe w tym sensie, że do ich rozwiązania wystarcza wykonanie jednego tylko działania arytmetycznego. Dokładniejsza analiza pokazuje jednak, że zadania te są złożone w tym sensie, że do ich rozwiązania niezbędne jest wykonanie więcej niż jednej operacji myślowej. Na przykład przy zadaniu najprostszego typu 1a: *Jaś ma 4 jabłka. Kasia ma o 3 jabłka więcej niż Jaś. Ile jabłek ma Kasia?* wprawdzie wystarczy wykonać tylko dodawanie 4+3, trzeba jednak wykonać dwie operacje myślowe, przedstawione poglądowo na rysunku w części 2.3 powyżej:

(a) wzajemnie jednoznaczna odpowiedniość między jabłkami Jasia i częścią jabłek Kasi,

(b) związek między zbiorem siedmiu jabłek Kasi a pewnym jego 4-elementowym podzbiorem.

W zwykłych zadaniach na dodawanie lub odejmowanie jest tylko zbiór i jego podzbiór.

Wiadomo, że najwięcej błędów przy zadaniach na porównywanie różnicowe i ilorazowe uczniowie popełniają wówczas, gdy poznają to drugie. Trudności widoczne są zwłaszcza wtedy, gdy oba typy porównywań pojawiają się na jednej lekcji. Wówczas nawet rozwiązywanie zadań na porównywanie różnicowe (łatwiejsze i wcześniej już opanowane), zaczyna sprawiać uczniom nowe kłopoty.

Specyficzne trudności to zadania typów 3a, 4a, 3m, 4m, w których chodzi o związki odwrotne do tych, które są w 1a, 2a, 1m, 2m, a więc o jeszcze jedną operację myślową. Konieczna jest świadomość ucznia, że zwroty *a jest o x mniejsze od b* oraz *b jest o x większe od a* są równoważne.

Na to wszystko nakładają się trudności językowe. Znaczna część błędów dzieci w klasach I-III polega na myleniu obu porównywań. Wiele osób nie jest świadomych, że zadania z odwracaniem (typy 3a, 4a, 3m, 4m) tak bardzo różnią się co do stopnia trudności od zadań typów 1a, 2a, 1m, 2m.

Uczniowie często stosują metodę słów kluczowych (*key words*), polegającą na wyszukiwaniu w tekście zadania pewnych charakterystycznych zwrotów i dobieraniu do nich działań matematycznych, np. widząc *więcej niż* wielu uczniów automatycznie wybiera dodawanie, przy *mniej niż* – odejmowanie, przy *razy* – mnożenie. Metoda ta prowadzi do błędów, gdy jest stosowana do typów 3a i 3m.

Przebieg badań

Etap pierwszy, organizowany przez Gdańską Fundację Rozwoju im. Adama Mysiora (GFR), polegał na zebraniu danych ilościowych od dużej grupy uczniów. Badania takie, „Sesje z Plusem", odbywają się co roku we wszystkich województwach w klasach IV; nauczyciele, którzy zgłoszą chęć uczestniczenia w tym badaniu, otrzymują od GFR zadania oraz kryteria oceniania prac uczniów. Po sprawdzeniu ich przesyłają wyniki do elektronicznej bazy GFR.

Etap drugi zbierania opisanych tu danych polegał na dotarciu do prac pisemnych uczniów, które zazwyczaj podczas badań GFR pozostają u nauczycieli. We wrześniu 2007 roku poproszono część nauczycieli z trzech województw o odesłanie prac uczniów do GFR (Demby, 2009; Demby, 2010).

Niniejsza praca dotyczy wyłącznie dwóch zadań z owego testu dotyczących porównywania różnicowego i ilorazowego. Treść analizowanych zadań można znaleźć w przykładach rozwiązań uczniowskich (przykłady 1-4) zamieszczonych w części 3.6 powyżej.

Badaniu ilościowemu poddana została tu „pełna grupa badanych uczniów" (PG) – tych, których oceny są w bazie GFR. Grupa ta liczy ok. 70 tysięcy uczniów.

Zarazem znacznie dokładniejszemu badaniu poddana została "grupa szczegółowo badanych uczniów" (GS) – podzbiór zbioru PG składający się z 788 uczniów, których prace nauczyciele przesłali do GFR. Prace te analizowano ilościowo (pod względem liczby poprawnych rozwiązań) na dwa sposoby: z tym samym stopniem szczegółowości, jaki był w bazie GFR (aby mieć możliwość porównania GS z PG) oraz bardziej szczegółowo, wyodrębniając w ocenie poprawności rozwiązań poszczególne kompetencje i identyfikując pewne typy poprawnych, częściowo poprawnych oraz niepoprawnych rozwiązań tych zadań.

Główne wyniki badawcze są następujące:

Wyniki procentowe (przedstawione na diagramie zamieszczonym w części 3.4 powyżej) dotyczące typów 1a, 2a, 1m oraz 2m zawierają się w granicach 60-75%. Porównując wyniki dla wersji A i wersji B testu (w pełni równoważnej, ale ze zmienioną kolejnością zadań), zaobserwowano, że w wersji A testu uczniowie nieco lepiej rozwiązali podpunkty na porównywanie różnicowe (w tej wersji testu pierwszy podpunkt dotyczył porównywania różnicowego) niż podpunkty na porównywanie ilorazowe. Natomiast w wersji B testu nieco lepiej rozwiązali podpunkty na porównywanie ilorazowe (tutaj pierwszy podpunkt dotyczył porównywania ilorazowego) niż podpunkty na porównywanie różnicowe. W pełni poprawne rozwiązanie całego zadania w obu wersjach testu podało jedynie około 35% uczniów.

Wyniki procentowe dla typu 4a wynoszą 63% (w wersji A) oraz 79% (dla wersji B), natomiast dla typu 3m wynoszą 40% (dla wersji A) oraz 32% (dla wersji B). Całkowicie poprawne rozwiązanie całego zadania podało jedynie 28% badanych uczniów. Uczniowie znacznie częściej rozwiązywali poprawnie podpunkt na porównywanie różnicowe w wersji B testu (gdzie porównywanie różnicowe pojawiło się w pierwszym podpunkcie) niż w wersji A testu (gdzie porównywanie różnicowe pojawiło się w drugim podpunkcie). Odwrotnie było w przypadku porównywania ilorazowego, co wskazuje, że te różnice nie są przypadkowe i są wynikiem zmiany kolejności zadań.

Jakkolwiek wielu autorów uznaje porównywanie ilorazowe jako trudniejsze dla uczniów niż porównywanie różnicowe, takie kategoryczne stwierdzenie byłoby nadmiernym uproszczeniem. Stopień trudności zależy bowiem od bardzo wielu czynników, nie tylko od struktury matematycznej zadania, ale też od użytego konkretu, sposobu jego zredagowania i okoliczności, w jakich uczeń dostaje zadania.

Wyniki przedstawione w pracy pozwoliły na ustalenie stopnia trudności typowych zadań na oba porównywania. Potwierdziło się, że zadania dotyczące odwracania porównywania ilorazowego sprawiły największe trudności. Ponadto, dodatkowe trudności sprawiło uczniom pomieszanie porównywania różnicowego i ilorazowego w jednej serii zadań. W konsekwencji tego zabiegu (poczynionego ze względów organizacyjnych, aby testy uczniów siedzących w jednej ławce różniły się) niektórzy uczniowie zadanie na porównywanie ilorazowe rozwiązywali tak, jak gdyby było to zadanie na porównywanie różnicowe, tzn. wykonując odejmowanie bądź dodawanie danych liczb. Sporo uczniów wyraźnie stosowało metodę słów kluczowych.

Część uczniów stosowała w kolejnych podpunktach zadania ten rodzaj działania, którego użyli w podpunkcie pierwszym. Przykładowo w wersji A testu w pierwszym podpunkcie należało podać liczbę o 2 mniejszą, a niektórzy uczniowie w kolejnych podpunktach (również dla porównywania ilorazowego) używali tylko odejmowania (w podpunktach, w których należało podać liczbę 4 razy mniejszą) i dodawania (w podpunktach, w których należało podać liczbę 3 razy większą oraz o 9 większą) zamiast mnożenia lub dzielenia. Potwierdził się tu też znany fakt, że znaczna część uczniów III klasy wydaje się nie czytać tekstu zadania, a skupiać się na podanych liczbach i do tych liczb dobierając na wyczucie działanie matematyczne.

Analizy badawcze prezentowane w pracy oparte są na kategoryzacji najważniejszych składników wszelkiego naukowego badania procesu rozwiązywania zadań tekstowych (Kilpatrick, 1978; Kulm, 1979), a mianowicie rozróżnienia między trzema rodzajami zmiennych, a mianowicie tych dotyczących:

1. osoby rozwiązującej (*subject variables*),
2. samych zadań (*task variables*),
3. warunków i okoliczności ich prezentacji (*situation variables*).

Z kolei zmienne dotyczące zadań dzielą się na:

a) zmienne dotyczące matematycznej struktury zadania (*structure variables*),
b) zmienne dotyczące konkretnej sytuacji użytej w zadaniu i języka użytego do jej opisania (*context variables*),
c) zmienne dotyczące sposobu, w jaki zadanie przedstawione jest rozwiązującemu, w tym: innych towarzyszących temu zadań, udzielonych wskazówek, możliwości użycia kalkulatora itp. (*format variables*).

Przedstawiona tu analiza możliwych typów zadań na porównywanie różnicowe i ilorazowe i analiza rozwiązań uczniowskich osadzona jest w powyższych kategoriach a), b), c).