DOCUMENT RESUME

ED 316 850 CS 009 982

AUTHOR	De Corte, E.; And Others
TITLE	Solving Compare Problems: An Eye-Movement Test of Lewis and Mayer's Co'.sistency Hypothesis.
PUB DATE	Apr 90
NOTE	38p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Eye Movements; Foreign Countries; Grade 3; Higher Education; *Models; Primary Education; *Problem Solving; *Reading Comprehension; Reading Research
IDENTIFIERS	Belgium

ABSTRACT

This study tested a model (developed by A. Lewis and R. Mayer) that simulates the comprehension processes used when solving compare problems. The basis of the "consistency hypothesis" model is that young students and even adults are more likely to make comprehension errors when the order of the terms in the relational statement of the problem is not consistent with the preferred order. Subjects, 19 university students and 15 third-grade students, were administered a series of 1-step compare problems and had their eye movements recorded in 2 separate experiments. Results indicated that the model was supported by the data from the third graders but not supported by the data from the adults. (One note, seven tables of data, and two figures are included; 18 references are attached.) (RS)

Solving compare problems:

An eye-movement test of Lewis and Mayer's consistency hypothesis

E. De Corte, L. Verschaffel and A. Pauwels

Center for Instructional Psychology

University of Leuven, Belgium

April 1990

Paper presented in a session on "Mathematics learning and transfer" at the Annual Meeting of the American Educational Research Association, Boston, MA, April 16-20, 1990.

Authors Note

This research was made possible by grants from the Flemish Ministry of Education and from a Concerted Research Action contract on "Knowledge-based visual information processing". Lieven Verschaffel is a Senior Research Assistant of the National Fund for Scientific Research of Belgium.

The authors want to thank I. Gielen and D. Christiaens for their help in collecting and analyzing the eye-movement data discussed in this paper.

Correspondence concerning this article should be addressed to: Erik De Corte, Center for Instructional Psychology, Department of Education, University of Leuven, Vesaliusstraat 2, B-3000 Leuven, Belgium.

Solving compare problems:

An eye-movement test of Lewis and Mayer's consistency hypothesis

Abstract

In an attempt to explain pupils' and even adults' difficulties with particular types of compare problems, Lewis and Mayer (1987) have presented a model that simulates the comprehension processes when solving these problems. The basis of their model is the "consistency hypothesis", according to which students are more likely to make comprehension errors when the order of the terms in the relational statement of the problem is not consistent with the preferred order. To test this model, we carried out two eye-movement experiments in which adults (Experiment 1), and third graders (Experiment 2) were administered a series of one-step compare problems. While the data of the adult subjects revealed no evidence in favor of the consistency hypothesis, the results of the third graders provided good support of the Lewis and Mayer (1987) model.

Research has shown that problem solvers experience a lot of difficulties in representing and solving compare word problems, such as e.g. "Pete has 13 marbles. He has 7 more marbles than John. How many marbles does John have?". In an attempt to explain these difficulties, Lewis and Mayer (1987) constructed a model of word problem comprehension processes, that uses schemata concerning the preferred order of presentation of information as guides to understanding. In their conceptualization, students are more likely to make comprehension errors when the order of terms in the relational statement is not consistent with the preferred order of terms in their schemata. Although the performance data obtained by Lewis and Mayer (1987) are totally in agreement with this consistency hypothesis, those data can certainly not be conceived as strong empirical evidence in favor of the hypothetical comprehension and solution processes involved in their model.

In this paper two eye-movement experiments are reported that were carried out to provide a more thorough test of the model of Lewis and Mayer (1987) (LM-model).

In a first section we describe the LM-model, and the specific hypotheses that can be derived from it. Afterwards we present the design and the results of the first eye-movement experiment, in which university students participated. Then, we report the results of the second experiment with third graders. Finally we discuss the implications of these results for the LM-model.

Comprehension Processes of Compare Word Problems

Of all basic types of word problems, compare problems are undoubtedly the most difficult ones (De Corte, Verschaffel &

Verschueren, 1982; Pauwels, 1987; Riley, Greeno & Heller, 1983). This category involves a comparison between two sets. In their well-known classification schema of elementary addition and subtraction word problems, Riley et al. (1983) distinguish six different types of compare problems depending on the identity of the unknown (either the referent set, the compared set, or the difference between those two sets), and on the direction of the difference ("more" or "less").

The model of Lewis and Mayer (1987) concerns only the four compare problem types in which the quantity of one of the two sets has to be determined (i.e. compare 3 till 6 in the classification schema of Riley et al., 1983). Normally these problems start with an assignment sentence, which specifies a numerical value for a variable (e.g. "Joe's weight is 85 pounds."). This sentence is followed by a relational statement that defines one variable in terms of another (e.g. "He weighs 13 pounds less than Pete."). Finally, a question asks for the value of the unknown variable ("How much does Pete weigh?") (Mayer, 1982).

Within this kind of compare problems, Lewis and Mayer (1987) distinguish two forms, namely consistent and inconsistent language problems. In a consistent language (CL) problem (e.g. "Joe has 5 marbles. Tom has 3 more marbles than Joe. How many marbles does Tom have?") the unknown variable is the subject of the second sentence, and the relational term (in this case "more than") is consistent with the required arithmetic operation (namely addition). In an inconsistent language (IL) problem (e.g. "Joe has 8 marbles. He has 5 marbles less than Tom. How many marbles does Tom have?") the unknown variable is the object of the second sentence, and the relational term ("less than") is in conflict with the ihmetic operation (namely addition).

Most empirical studies on word problem solving have shown that IL problems are more difficult than CL problems (Briars & Larkin, 1984; Morales, Shute, & Pellegrino, 1985; Riley et al., 1983). To explain this finding, Lewis and Mayer (1987) have put forward the consistency hypothesis. Referring to previous work by Huttenlocher and Strauss (1968), they assume that problem solvers have a preference for a particular order in which problem information is presented. More specifically, they prefer the order involved in CL problems, in which the unknown variable is the subject of the second sentence. When given an IL problem, in which the unknown variable is the object of the relational sentence, pupils are assumed to mentally rearrange the relational sentence until it fits their preferred format. This rearrangement procedure consists of reversing the subject and the object of the relational sentence, as well as the arithmetic operation suggested by its relational term. Because the comprehension and solution process is more error prone when information must be rearranged, the probability of a reversal error (subtraction instead of addition, or the reverse) will be greater for IL than for CL problems.

In an attempt to investigate this hypothesis, Lewis and Mayer (1987) asked a group of 96 adult students to represent and solve a series of two-step addition, subtraction, multiplication and division compare problems (e.g. "At ARCO gas sells for 1.13 Dollar per gallon. Gas at Chevron is 5 cents more per gallon than gas at ARCO. How much do 5 gallons of gas cost at Chevron?"). The results were in line with the consistency hypothesis: Subjects were more likely to miscomprehend the relational statement, and thus to commit a reversal error when the unknown variable was the object of the relational sentence than when it was the subject. Moreover, the analysis yielded two additional interesting findings. First, the necessary operation had also a significant influence on the number of reversal errors: Problems that require an addition and a multiplication, produced more reversal errors than subtraction and division problems. Second, there was an interaction between language and operation: The difficulty of overcoming inconsistent language was enhanced in the case of addition and multiplication. In line with Clark's (1969) view, this interaction effect was explained by the fact that the relational term in an IL addition or multiplication problem is a marked[1] one; and since marked terms (i.e. "less") are more salient than unmarked terms (i.e. "more"), subjects would be more resistant to reverse them.

To state this general hypothetical explanation of the origin of students' reversal errors on compare problems in a more precise and formal way, the authors developed a process model of the schemata and the procedures needed to comprehend compare problems and to translate them into an output equation. This model is given in Figure 1. As an illustration, we will describe how the IL problem "Joe has 8 marbles. He has 5 marbles less than Tom. How many marbles does Tom have?" is internally processed.

Insert Figure 1 about here

Step 1-4 involve encoding the first sentence. In step 1 the subject scans the first sentence, and selects the assignment schema. In Step 2 the subject locates the name ("Joe"), and the number ("8") in the sentence. Finally, in Steps 3 and 4 the subject instantiates the assignment schema as Joe = 8.

Step 5-11 involve encoding the second sentence. In Step 5 the subject scans the second sentence, and selects the relational schema.

In Step 6 the subject locates the first name ("He" which refers to "Joe"), the number ("5"), the relational term ("less than"), and the second name ("Tom"). Next, in Step 7 the relational term is translated into an operational symbol ("less than" is translated into "-").

Before instantiating the relational schema, the subject conducts a test in Step 8 to determine whether the first name in the second sentence differs to the first name in the first sentence. If those names are different, the problem is in the preferred order, and the subject continues in Steps 9-11 to instantiate the relational schema. If the names match, as in our example, the subject must rearrange his or her encoding of the second sentence by jumping to the rearrangement subprocedure (Steps R1-R5).

In step R1 the subject reverses the ordering of the two names in the second sentence so that the unknown term ("Tom") becomes the subject, and the given term ("Joe") becomes the object. In the following step (R2), the subject reverses the operator (i.e., "-" is changed into "+"), but may fail to carry out this reversal with a certain probability. In Step R3 the subject tests whether the relational term is marked or unmarked. If, as in the example problem, the term is marked, the subject is more reluctant to reverse the operator and conversely the original operator will be retained in Step R4. In Step R5 the rearrangements are complete and the subject returns to Steps 9-11 to instantiate the relational schema. At Step 12, the subject creates either the correct output equation ("Tom = 8 + 5"), or the reversed equation ("Tom = 8 - 5").

This model implies that the subject only runs the risk of making a reversal error during the rearrangement procedure (in particular at step R2 in Figure 1), thus, only while solving an IL problem. Moreover, this risk is greater if that problem contains a marked term;

indeed, in this case the risk is situated at step R2 and at step R4.

The results of the Lewis and Mayer study (1987) are in accordance with the predictions of their model. However, these performance data can certainly not be conceived as sufficient empirical evidence in favor of the hypothetical schemata and internal processes involved in this model. In an attempt to test the hypotheses underlying the LM-model in a more straightforward way, we collected eye-movement data of students reading and solving compare problems. Eye-movement registration has already proved to be a valuable technique for collecting data on the processes that contribute to the comprehension of word problems, especially when specific hypotheses are formulated (De Corte & Verschaffel, 1987; De Corte, Verschaffel, & Pauwels, in press). The use of this technique in problem-solving research is based on the general assumption that there is a strong correspondence between what is fixated and what is cognitively processed (Just & Carpenter, 1987; Rayner & McConkie, 1976). Therefore the fixation time spent on a particular piece of visual information is assumed to be a good reflection of the time needed to process it.

Hypotheses

The hypotheses derived from the LM-model concern (a) the number of reversal errors, (b) response times, and (c) fixation times on the different sentences in a problem.

Number of reversal errors. The LM-model predicts that problem solvers will make more reversal errors on IL than on CL problems. Furthermore, an interaction effect between language consistency and the relational term is predicted: More reversal errors will occur on IL problems with a marked term than on those with an unmarked term;

this difference will not exist for CL problems.

In their study, Lewis and Mayer (1987) used two-step compare problems. For a reason that will be explained in the next section, we presented one-step problems. Because these one-step compare problems are rather easy tasks for adults, we did not expect any reversal errors at all in the first experiment. Consequently, no differences in the amount of reversal errors between IL and CL problems, nor between problems with an unmarked and a marked relational term were expected. But, the two hypotheses mentioned above were retained for the second experiment with the third graders.

Response times. In this respect the following hypothesis was stated. If students indeed have to reorganize the relational sentence in an IL problem to make it similar to their preferred format, this rearrangement will necessarily lead to an increase of response time for IL problems as compared to CL problems. The LM-model does not allow to formulate a prediction concerning the influence of the nature of the relational term (marked or unmarked) on the response time.

Fixation Times. As Figure 1 clearly shows, the difference in the solution process between CL and IL problems is situated in the processing of the second, relational sentence: In IL problems this sentence has to be reorganized, but this is not the case for CL problems. Therefore, we predict that the longer response times for the IL problems will be due to longer fixation times on the relational sentence. The fixation time on the first sentence, and on the question sentence will not differ systematically for both kind of problems.

According to Lewis and Mayer (1987), the rearrangement subprocedure is already initiated immediately after the first reading of the relational sentence, i.e. before the question sentence is read for the first time. This implies that the above-mentioned differences between

CL and IL problems in response time and in fixation time on the relational sentence, will already show up during the <u>initial reading of the first and the second sentence</u> (referred to further on as the first phase of the solution process). In particular, the duration of this first phase of the solution process will be longer for IL than for CL problems, and this will especially be due to the longer absolute fixation time on the second sentence in the problem.

EXPERIMENT 1

Method

Subjects

Nineteen university students volunteered to participate in the experiment.

Tasks

Each subject was administered 30 one-step addition and subtraction word problems: 16 target items complemented with 14 filler items to avoid stereotyped responses. The target items included eight CL and eight IL problems. Within each category half of the problems contained an unmarked relational term ("more"), the other half a marked term ("less"). Examples of the problems are given in Table 1.

Insert Table 1 about here

The different categories of problems were controlled for the number of sentences, the number of characters per sentence, the context, and the size of the numbers given in the problems.

As said before, we only presented one-step problems. The reason for using easier problems than Lewis and Mayer (1987) did, was that our eye-movement registration equipment allows only minor head movements of the subjects. Therefore, they were not given paper and pencil, but had to solve the problems in their head. In this respect two-step problems like those used in the LM-study could have caused an overload of working memory. Although the use of one-step instead of two-step problems certainly can bear an important influence on students' performances, we could see no reason why the predictions about the comprehension processes implied in the LM-model would not hold for one-step problems. Indeed, the gist of this model is the rearrangement of the second, relational sentence; whether this sentence is part of a one-step or a two-step problem does not really change this principle.

Procedure

All subjects were tested individually. They were asked to state orally the operation they would apply to solve the problem, and they were allowed to do this at their own pace. The word problems were presented on a television screen at about two meters distance from the subjects. While they were reading and solving the problems, their eye-movements were registered with DEBIC 80, a system that is based on the pupil center-corneal reflection method. Every 20 milliseconds the spatial coordinates of the subject's point of regard, and a time code are stored on an on-line PDP-11 computer. Furthermore, a videorecording of the subsequent points of regard on the displayed problems is also available: The subject's fixation point is represented as the intersection of a vertical and a horizontal axis superimposed on the problem (see Figure 2). For a detailed description of the eye-tracking system we refer to De Graef, Van Rensbergen and d'Ydewalle (1985).

Insert Figure 2 about here

The raw eye-movement data were transformed into consecutive fixations. A fixation was operationally defined as a time period of minimum 100 milliseconds during which the eye is close to immobile at a particular location (Rayner, 1978). Starting from those data, the absolute fixation times on the three different problem sentences were calculated. These variables, as well as the response times, were subjected to an analysis of variance with language consistency (consistent or inconsistent) and the nature of the relational term (marked or unmarked) as independent variables (2*2 randomized block factorial design). We point out that the response time is not the sum of the fixation times on the three different sentences of the problem. Indeed, response time includes also the intervals between two successsive fixations (the so-called "saccades"), as well as the fixations above, under and besides the problem sentences on the television screen.

To test the hypotheses concerning the first phase of the solution process, the eye-movement protocols were subdivided in two stages: a first stage including the initial readings of the first two sentences, and a second stage starting with the first fixation on the question and ending at the moment the answer is given. In these analyses we included only the data of those solution processes that showed several subsequent fixations on the first and on the second sentence before the reading of the question was started. Six out of the 304 solution processes did not meet this criterion.

Results

Success Rate

As expected, most of the problems were solved correctly. On a total of 304 answers, only 12 reversal errors occurred: five errors on CL problems and seven on IL problems. As these few errrors are probably due to lack of concentration, and not to comprehension difficulties, we will not discuss them any further.

Total Solution Process

Duration. We hypothesized that students would need more time to solve IL than CL problems. The analysis of variance with language and relational term as independent variables did not reveal a main effect of language consistency ($F(1,18) = 0.99$, $MSe = 280651.8$, n.s.). Compared to our prediction, means were even in the opposite direction: The students needed 9.39 s to solve CL problems, and only 8.78 s to solve the IL ones.

On the other hand, we found a main effect of the relational term ($F(1,18) = 12.58$, $MSe = 181085.0$, $p < 0.01$): Students needed significantly more time to solve the compare problems that contained a marked relational term (9.95 s), than for problems with an unmarked term (8.22 s).

There was no interaction effect between both variables ($F(1,18) = 3.20$, $MSe = 188718.8$, n.s.).

Fixation times on the different sentences. We predicted that the absolute fixation time on the second sentence would be longer for IL than for CL problems, while the fixation time on the two other sentences would not differ between both problem types. Mean absolute fixation times on the three problem sentences are given in Table 2.

Insert Table 2 about here

For none of the three sentences a significant effect of language consistency was found. For the second sentence, we found even a trend contrary to the prediction derived from the LM-model: The mean fixation time in IL problems (2433 ms) was shorter than in CL problems (2646 ms).

On the other hand, all three ANOVA's revealed a significant effect of the relational term [$F(1,18) = 9.67$, $MSe = 12138.0$, $p < 0.01$ for the first sentence; $F(1,18) = 14.91$, $MSe = 17300.3$, $p < 0.01$ for the second sentence; $F(1,18) = 5.45$, $MSe = 3345.6$, $p < 0.05$ for the question sentence]. Problems with a marked term elicited longer fixation times on all sentences than those with an unmarked term.

For the first sentence, the analysis also showed an unexpected interaction effect between the relational term and language consistency ($F(1,18) = 7.47$, $MSe = 14134.6$, $p = 0.01$): The difference between marked and unmarked terms was only found for the CL problems.

First Phase of the Solution Process

According to the LM-model (1987) as formalized in Figure 1, the rearrangement of the relational sentence occurs already during the initial reading of that sentence, i.e. before the subject starts reading the question sentence for the first time. Consequently, we hypothesized that it would last longer before students start reading the question in an IL problem than in a CL problem, and that this difference would be mainly caused by longer absolute fixation times on the second, relational sentence during this first stage of the solution process.

Duration of the first stage. The analysis on the duration of the first phase revealed a main effect of language consistency ($F(1,18) = 7.40$, $MSe = 8684.1$, $p = 0.01$). However, the means were not in the expected direction: Students needed significantly more time to read the first two sentences in a CL problem (5.91 s) than in an IL one (5.32 s).

The relational term had also a significant influence on this variable ($F(1,18) = 5.26$, $MSe = 8365.0$, $p < 0.05$): For problems with a marked term, the first stage of the solution process lasted longer than for problems with an unmarked term (5.85 compared to 5.32).

The interaction between both independent variables was not significant.

Fixation time on the different sentences during the first stage. We predicted that the influence of language consistency on the absolute fixation times would appear with respect to the second sentence, and not for the first one. The ANOVA's indeed revealed only a significant effect for the second sentence. But the means were again not in the predicted direction: CL problems (1648 ms) elicited higher absolute fixation on the second sentence than IL problems (1448 ms).

Insert Table 3 about here

Table 3 shows that the marked terms tended also to elicit longer fixation times on both sentences than the unmarked ones. But in this case, the differences were not significant.

Similar to the results of the total solution process, we also found an interaction effect of language consistency and relational term on the absolute fixation time on the first sentence ($F(1,18) = 10.67$, $MSe = 1269.9$, $p < 0.01$). The difference between marked and unmarked terms was only found for the CL problems.

In summary, the data obtained in the first experiment provide no empirical evidence for the consistency hypothesis put forward by Lewis and Mayer (1987). Mean response and fixation times were not in accordance with the hypotheses derived from the model: Response times were not longer for IL than for CL problems, and the students did not look longer to the relational sentence in a IL than in a CL problem. Moreover, for the first phase of the solution process we observed differences that are significant in the opposite direction.

Although the LM-model does not provide clear hypotheses concerning the effect of the nature of the relational term, we found that this variable had a significant influence on most of the dependent variables: Response and fixation times for the total process, as well as the duration of the first phase were longer for problems with a marked term than for problems with an unmarked term.

EXPERIMENT 2

Method

Subjects

The subjects were 15 pupils of a third-grade class of a local school.

Tasks and procedure

Each subject was administered 26 one-step addition and subtraction word problems: 16 target items, complemented with 10 filler items to avoid stereotyped responses. As in Experiment 1, the target items included eight CL and eight IL problems. Within each category half of the problems contained an unmarked relational term ("more"), the other half the marked term ("less").

Examples of the problems, that were controlled for the number of sentences, the number of characters per sentence, the context, and the the size of the numbers given in the problems, are presented in Table 4.

Insert Table 4 about here

The testing situation and the procedure for analyzing the data was the same as in Experiment 1, with the exception that for the third graders the number of reversal errors was also subjected to the analysis of variance.

Results

Reversal errors

The LM-model predicts more reversal errors on IL than on CL problems. Furthermore, it was expected that IL problems with a marked term would lead to more reversal errors than those with an unmarked term; for CL problems such a difference was not predicted.

An analysis of variance with language consistency and the relational term as independent variables, revealed indeed a significant main effect of language consistency on the occurrence of reversal errors ($F(1,14) = 19.64$, $MSe = 0.98$, $p < 0.01$). The percentage of reversal errors was 3.3% for CL, and 31.6% for IL problems.

Pupils made also more reversal errors on problems with a marked term than on problems with an unmarked term (22.5% and 12.5% respectively), but this difference was not significant.

While there was no difference between problems with an unmarked and a marked term within the CL problems - on both kind of problems pupils made 3.3% errors -, pupils made more errors on IL problems with a marked term (41.6%) than on those with an unmarked term (21.6%). However, the interaction effect between language consistency and relational term was also not significant ($F(1,14) = 3.50$, $MSe = 0.69$, $p < 0.09$).

Total Solution Process

Duration. We hypothesized that the pupils would need more time to solve the IL problems than the CL ones; and that the relational term would not affect this variable.

The results of the analysis of variance were in line with these predictions. We found a significant main effect of language consistency ($F(1,14) = 11.36$, $MSe = 947083.4$, $p < 0.01$): IL problems (16.6 s) elicited longer solution times than CL problems (12.4 s).

Contrary to the results of Experiment 1, the main effect of the relational term on this variable was not significant.

Again, no interaction effect was found.

Fixation time on the different sentences. According to the LM-model, the longer solution times for IL problems would be especially due to longer absolute fixation times on the second, relational sentence of the problem. Mean absolute fixation times on all three sentences are given in Table 5.

Insert Table 5 about here

The ANOVA on the absolute fixation time on the second sentence indeed revealed a significant effect of language consistency ($F(1,14) = 8.27$, $MSe = 85138.9$, $p < 0.02$): The third graders spent an average of 4841 ms on the relational sentence in an IL problem, and only 3601 ms on that sentence in a CL problem. However, the ANOVA's on the fixation time on both other problem sentences also revealed a main effect of language consistency: $F(1,14) = 7.71$, $MSe = 64285.4$, $p < 0.02$ for the first sentence, and $F(1,14) = 4.82$, $MSe = 10053.17$, $p < 0.05$ for the question.

The ANOVA's on these variables did not reveal a main effect of the relational term, nor an interaction effect between the independent variables.

First Phase of the Solution Process

Since it was possible to distinguish the first phase of the solution process - the initial systematic reading of the first two sentences - in all the eye-movement protocols, the entire data set was included in the analyses.

Duration. In accordance with the prediction, we found a main effect of language consistency ($F(1,14) = 12.91$, $MSe = 85841.6$, $p < 0.01$). As Table 6 shows, the initial phase lasted significantly longer for IL (9.6 s) than for CL problems (8.2 s).

Insert Table 6 about here

The main effect of the relational term did not reach significance ($F(1,14) = 4.46$, $MSe = 25211.3$, $p < 0.0$[illegible]). But the ANOVA revealed a significant interaction effect ($F(1,14) = 5.40$, $MSe = 10673.1$, $p < 0.05$). As Table 6 reveals, the difference between CL and IL problems is mainly due to the difference between both types of problems when they contain a marked relational term. This is not in agreement with the LM-model. According to Lewis and Mayer (1987) the relational sentence in an IL problem will always be reorganised, whatever the nature of the relational term. This implies that also for problems with an unmarked term a difference between CL and IL problems is expected.

Fixation time on the different sentences. In line with the LM-model, we expected that during the initial stage of the solution process, the effect of language consistency on the absolute fixation times would appear on the second, and not on the first sentence. Mean absolute fixation times on the first two sentences are given in Table 7.

Insert Table 7 about here

The fixation time on the second sentence was indeed significantly affected by language consistency ($F(1,14) = 25.51$, $MSe = 1922.6$, $p < 0.001$): The pupils looked for 3018 ms at the relational sentence in a IL problem, while only for 2446 ms in a CL problem. But, although less pronounced, the analysis also revealed a significant effect of language consistency on the fixation time on the first line of the problem ($F(1,14) = 7.26$, $MSe = 3869.1$, $p < 0.02$).

For both dependent variables, no main effect of the relational term, and no interaction effect was found.

Discussion

We reported two eye-movement experiments in which the consistency hypothesis of Lewis and Mayer (1987) was tested. According to this hypothesis, problem solvers have a preference for compare problems with consistent language. Therefore, when given a problem with inconsistent language, they will rearrange the relational sentence until it fits their preferred format.

The results of the first experiment with university students do not support this hypothesis. For several dependent variables we found means that are in the opposite direction; for the first phase of the solution process these differences are even significant. Although not predicted, we found a significant effect of the nature of the relational term on most of the variables.

On the contrary, the results of the experiment obtained for the third graders are in line with the hypotheses derived from the LM-model. First of all, IL problems elicited more reversal errors than CL problems. Furthermore, the duration of the total solution process, as well as that of the initial stage of that process, was longer for IL than for CL problems. The analysis of the eye-movement material revealed that these differences in solution times were not only due to different processing times for the second, relational sentence, but - albeit less pronounced - also for the other two sentences. The nature of the relational term had no significant influence on the dependent variables in this second experiment.

Looking at these findings one certainly wonders why we found empirical evidence in favor of the consistency hypothesis in Experiment 2 and not in Experiment 1. The most plausible explanation is that one-step compare problems are relatively easy problems for adults. This is proved by the lack of errors, and the rather low response times in the first experiment. The third graders made a lot more errors and needed more time to solve the problems. Probably, Lewis and Mayer's model only reflects the solution processes of subjects for whom the compare problems that are administered have a certain level of difficulty.

Another remarkable finding is that university students did not need less, but significantly more time to read the first two sentences in a CL problem than in an IL one. Research on discourse comprehension might provide an alternative interpretation of the process of understanding relational sentences, which seems more in line with the data of the present eye-movement study. According to Kintsch and Van Dijk (1978; see also Van Dijk & Kintsch, 1983), one of the fundamental properties of a discourse is its coherence. An effective way to connect what has been processed with new information, is to start with the agent that was introduced before. Moreover, "if an agent is mentioned in a proposition and if this agent is the initial proposition of an episode (...) it is plausible that the same agent will reappear in the following proposition. (...) In this case, the top-down and the bottom-up processes match, and the interpretation will be fast" (Van Dijk & Kintsch, 1983, p. 156). Applied to compare problems, this means that starting the second relational sentence with the same subject as the one mentioned in the preceding sentence (as in IL problems), will fasten its interpretation and its integration with the first sentence. It is reasonable to assume that third graders do not yet master the implicite rules concerning text coherence, so that this factor does not influence their performances.

The influence of the relational term on nearly all dependent variables in Experiment 1 was also not expected. Although Lewis and Mayer (1987) point to the fact that students have a preference for unmarked terms as compared to marked ones, and that this relational term has an effect on the number of reversal errors within the category of IL problems, the authors claim that the nature of the comprehension procedures (see Figure 1) is almost the same for marked and unmarked terms. The LM-model can certainly not account for the differences we observed between marked and unmarked terms <u>within the</u>

CL problems. It seems that Lewis and Mayer (1987) have underestimated the role of this variable in the comprehension process.

Finally, there might be other causes of children's reversal errors on IL compare problems than the one that derives from Lewis and Mayer's consistency hypothesis. The most well-known example is the so-called "key-word strategy" (Nesher & Teubal, 1975; Sowder, 1988). In this strategy the child's selection of an arithmetic operation is not based on a global semantic analysis of the problem situation, but guided by the occurrence of an isolated key word with which an arithmetic operation is associated. For example, the words "althogether" and "more" are associated with addition; the words "loose" and "less" with subtraction. In the context of the present study, applying this key-word strategy would lead to correct answers on CL problems, but produce reversal errors on IL ones. Some reversal errors made by the third graders in the present study might be due to the use of this strategy.

Note

[1] The difference between marked and unmarked terms refers to the fact that antonymous adjectives (like e.g. "good-bad" or "long-short") are often found to be asymmetric. The positive or unmarked member of such pairs can be neutralized in certain contexts (e.g. "How good is the food?"). Moreover, the unmarked member of the pair also serves as the name of the full scale (e.g. "goodness"). In sum, the unmarked term has two senses, but the marked only one. According to Clark (1969), unmarked terms are stored in memory in a less complex and more accessible form than their opposites.

References

Briars, D.J., & Larkin, J.H. (1984). An integrated model of skill in solving elementary word problems. Cognition and Instruction, 1, 245-296.

Clark, H.H. (1969). Linguistic processes in deductive reasoning. Psychological Review, 76, 387-404.

De Corte, E., & Verschaffel, L. (1987). First graders' eye-movements during elementary addition and subtraction word problem solving. In G. Lüer & U. Lass (Eds.), Fourth European Conference on Eye Movements, Volume 1: Proceedings (pp. 148-150). Toronto/Göttingen: Hogrefe.

De Corte, E., Verschaffel, L., & Pauwels, A. (in press). Influence of the semantic structure of word problems on second graders' eye-movements. Journal of Educational Psychology.

De Corte, E., Verschaffel, L., & Verschueren, J. (1982). First graders' solution processes in elementary word problems. In A. Vermandel (Ed.), Proceedings of the Sixth International Conference for the Psychology of Mathematical Education (pp. 91-96). Antwerp: University of Antwerp.

De Graef, P., Van Rensbergen, J., & d'Ydewalle, G. (1985). User's manual for the Leuven eye-movement registration system (Internal Report No. 52). Leuven: Laboratory for Experimental Psychology, University of Leuven.

Huttenlocher, J., & Strauss, S. (1968). Comprehension and a statement's relation to the situation it describes. Journal of Verbal Learning and Verbal Behavior, 7, 300-304.

Just, M.A., & Carpenter, P.A. (1987). The psychology of reading and language comprehension. Carnegie-Mellon University.

Kintsch, W., & Van Dijk, T.A. (1978). Toward a model of text comprehension and production. Psychological Review, 5, 363-394.

Lewis, A. , & Mayer, R. (1987). Students' miscomprehension of relational statements in arithmetic word problems. Journal of Educational Psychology, 79, 363-371.

Mayer, R.E. (1982). Memory for algebra story problems. Journal of Educational Psychology, 74, 199-216.

Morales, R.V., Shute, V.J., & Pellegrino, J.W. (1985). Developmental differences in understanding and solving simple word problems. Cognition and Instruction, 2, 41-57.

Nesher, P., & Teubal, E. (1975). Verbal cues as an interfering factor in verbal problem solving. Educational Studies in Mathematics, 6, 41-51.

Pauwels, A. (1987). Empirische toetsing van computermodellen over de ontwikkeling van de oplossingsvaardigheid bij eenvoudige redactie-opgaven over optellen en aftrekken. (Empirical validation of computer models of children's elementary addition and subtraction word problem solving) (Unpublished master's thesis.) Leuven: Center for Instructional Psychology, University of Leuven.

Rayner, K., & McConkie, G.W. (1976). What guides a reader's eye movements? Vision Research, 16, 618-660.

Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem-solving ability in arithmetic. In H.P. Ginsburgh (Ed.), The development of mathematical thinking (pp. 153-196). New York: Academic Press.

Sowder, L. (1988). Children's solutions of story problems. Journal of Mathematical Behavior, 7, 227-238.

Van Dijk, T.A., & Kintsch, W. (1983). Strategies of discourse comprehension. New York/London: Academic Press.

Schemata

Schema for Assignment Sentence: First Input Sentence
(SET A) = (value a)

Schema for Relational Sentence: Second Input Sentence
(SET B) = (value b) relation (SET A)

Schema for Output Equation
(SET B) = (value a) operator (value b)

Translation Procedures

Procedure for encoding First Sentence

1. Select assignment schema
2. Find (name 1) and (number 1)
3. Assign (name 1) to (SET A)
4. Assign (number 1) to (value a)

Procedure for Encoding Second Sentence

5. Select relational schema
6. Find (name 2), (number 2), (relation), and (name 3)
7. Assign value of (relation) to (operator)
8. If (name 2) is (Name 1) go to rearrangement subprocedure
9. Assign (name 2) to (SET B)
10. Assign (number 2) to (value b)
11. Assign (name 3) to (SET A)

Procedure for Creating Problem Representation

12. Create output equation using current values of SET B, value a, operator, and value b

Subprocedure for Rearranging Second Sentence

R1 Interchange (name 3) and (name 2)
R2 With probability = (1 - p), reverse (operator)
R3 If (relation) is marked
R4 With probability = p, assign value of (relation) to (operator)
R5 Go back to Procedure for Encoding Second Sentence

Figure 1
Model of schemata and translation procedures for representing compare problems (Lewis & Mayer, 1987, p. 368)

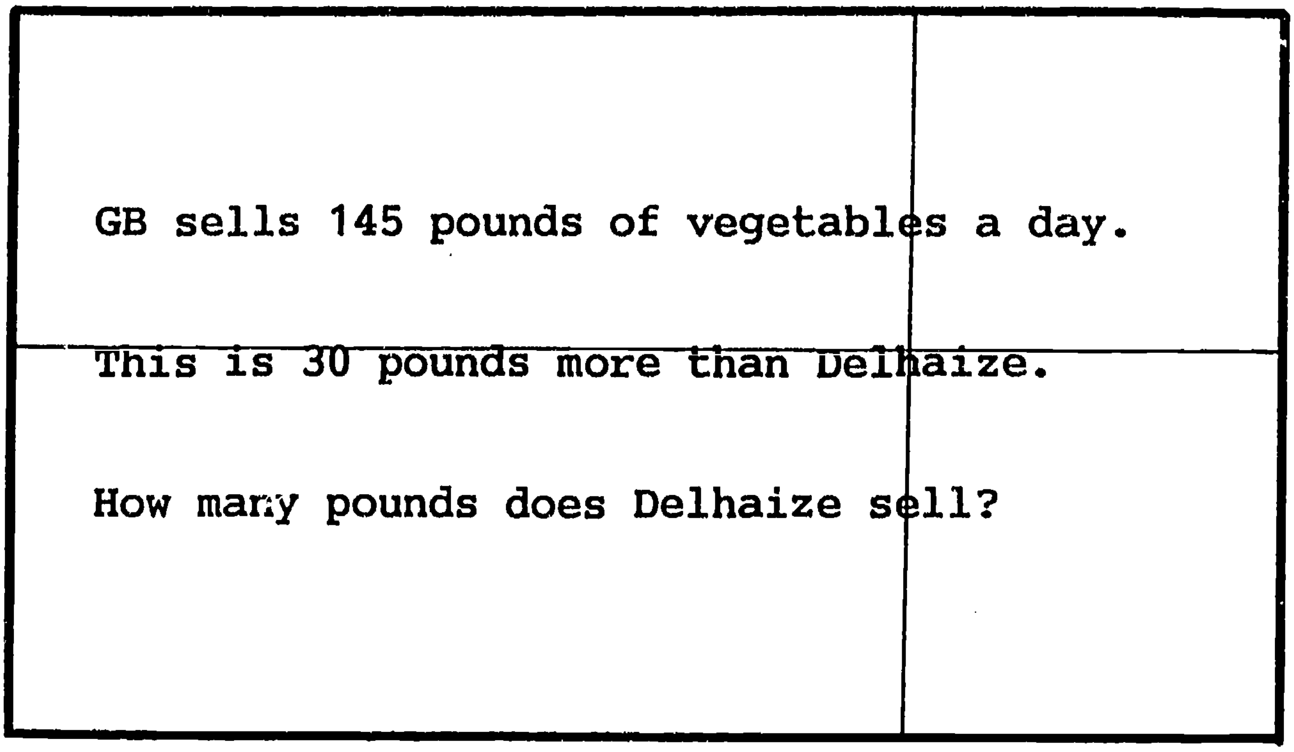

Figure 2
The problem presentation on the television screen.
The intersection of the axes indicates the subject's point of regard.

Table 1

Examples of the Presented Compare Problems in Experiment 1

Term	Language: Consistent	Language: Inconsistent
Unmarked	Brian's weight is 93 pounds. Peter weighs 18 pounds more than Brian. How much does Peter weigh?	Lisa's weight is 77 pounds. She weighs 12 pounds more than Kate. How much does Kate weigh?
	Colruyt sells 220 eggs a day. Aldi sells 75 eggs more than Colruyt. How many eggs does Aldi sell?	GB sells 145 pounds of vegetables a day. This is 30 pounds more than Delhaize. How many pounds does Delhaize sell?
Marked	Carol's weight is 77 pounds. Ann weighs 12 pounds less than Carol. How much does Ann weigh?	Robin's weight is 93 pounds. He weighs 18 pounds less than Alan. How much does Alan weigh?
	Delhaize sells 145 breads a day. GB sells 30 breads less than Delhaize. How many breads does GB sell?	Nopri sells 220 pounds of butter a day. This is 75 pounds less than Aldi. How many pounds does Aldi sell?

Note. Problems were presented in Dutch.

Table 2

<u>Mean Fixation Times (in msec) on the Sentences of the Problems in Experiment 1</u>

Sentence	Relational term	Language		
		CL	IL	Total
Sentence 1				
	Unmarked	2000	2128	2064
	Marked	2766	2148	2457
	Total	2383	2138	
Sentence 2				
	Unmarked	2280	2216	2248
	Marked	3012	2649	2831
	Total	2646	2433	
Sentence 3				
	Unmarked	632	731	681
	Marked	821	852	836
	Total	726	791	

Table 3

Mean Fixation Times (in msec) on Sentence 1 and 2 during the First Phase of the Solution Process in Experiment 1

Sentence	Relational term	Language		
		CL	IL	Total
Sentence 1				
	Unmarked	1698	1824	1761
	Marked	2110	1702	1906
	Total	1904	1763	
Sentence 2				
	Unmarked	1589	1371	1480
	Marked	1706	1525	1615
	Total	1648	1448	

Table 4

Examples of the Presented Compare Problems in Experiment 2

	Language	
Term	Consistent	Inconsistent
Unmarked	Linda's weight is 24 pounds. Peter weighs 33 pounds more than Linda. How much does Peter weigh?	Simon's weight is 33 pounds. He weighs 16 pounds more than Kate. How much does Kate weigh?
	Brian has 32 books. Ralph has 13 books more than Brian. How many books does Ralph have?	Pete has 28 pencils. He has 17 pencils more than Dick. How many pencils does Dick have?
Marked	John's weight is 37 pounds. Tim weighs 24 pounds less than John. How much does Tim weigh?	Robin's weight is 29 pounds. He weighs 14 pounds less than Alan. How much does Alan weigh?
	Carol has 35 dolls. Ann has 29 dolls less than Carol. How many dolls does Ann have?	Mary has 38 rings. She has 25 rings less than Joan. How many rings does Joan have?

Note. Problems were presented in Dutch.

Table 5

Mean Fixation Times (in msec) on the Sentences of the Problems in Experiment 2

Sentence	Relational term	Language		
			IL	Total
Sentence 1				
	Unmarked	2906	3319	3113
	Marked	2764	4169	3466
	Total	2835	3744	
Sentence 2				
	Unmarked	3759	4455	4107
	Marked	3443	5227	4335
	Total	3601	4841	
Sentence 3				
	Unmarked	870	1175	1023
	Marked	748	1011	880
	Total	809	1093	

Table 6

Mean Duration of the First Phase of the Solution Process (in s) in Experiment 2

Relational Term	Language		
	CL	IL	Total
Unmarked	8.51	8.89	8.70
Marked	7.96	10.30	9.13
Total	8.24	9.60	

Table 7

Mean Fixation Times (in msec) on Sentence 1 and 2 during the First Phase of the Solution Process in Experiment 2

Sentence	Relational term	Language		
		CL	IL	Total
Sentence 1				
	Unmarked	2190	2393	2292
	Marked	2011	2673	2342
	Total	2100	2533	
Sentence 2				
	Unmarked	2494	2836	2665
	Marked	2398	3200	2799
	Total	2446	3018	