Exploring the Episodic Structure of Algebra Story Problem Solving *

Rogers Hall
Dennis Kibler
Etienne Wenger
Chris Truxaw

Irvine Computational Intelligence Project
Department of Information and Computer Science
University of California
Irvine, CA 92717

Technical Report 86–24

November 24, 1986

*Supported by the Personnel Training and Research Division of the Office of Naval Research, contract N00014-85-K-0373.

Exploring the Episodic Structure of Algebra Story Problem Solving *

Rogers Hall
Dennis Kibler
Etienne Wenger
Chris Truxaw

Irvine Computational Intelligence Project
Department of Information and Computer Science
University of California
Irvine, CA 92717

November 24, 1986

1 Introduction and summary

This report describes an investigation of problem solving in the domain of algebra story problems. The work is part of an ongoing study of interactions between problem solving, learning, and teaching in task environments where the reasoner's background knowledge will have a significant influence on performance. Our eventual goal is to develop computational models of problem solving and learning which can be incorporated within a teaching environment. In this section, we briefly sketch the "problem" of algebra story problem solving, give an overview of the contents of the report, and

*Supported by the Personnel Training and Research Division of the Office of Naval Research, contract N00014-85-K-0373.

supported.

1.2 Itinerary

In this report we briefly examine some basic materials out of which algebra story problems can be constructed, and use these materials as background for an exploratory study of problem solving behavior in a sizable group of competent problem solvers. A major hypothesis in this work is that in order to generate a solution-enabling representation of a problem, reasoners must assemble quantitative constraints under the guidance of their understanding of the situational context (or story). This context serves not only as a vehicle for the quantitative problem, but also as a framework for justifying the existence of quantitative constraints and their interrelationships. Accordingly, we examine the quantitative and situational structure of a limited set of algebra story problems, and then use instances of these problems in the behavioral study.

In the behavioral study, we analyze the written protocols of 85 upper division computer science undergraduates instructed to show their work in solving four, representative algebra story problems. A scoring taxonomy is developed in which a written solution attempt is divided into a series of coherent problem solving episodes. Each of these episodes is coded along a set of categories reflecting tactical content, conceptual material, formal manipulative errors, and relationship to preceding episodes. Preliminary analyses of the scored protocols are used to give evidence for the frequency with which various problem solving behaviors occur within subjects' solution attempts, the content and outcome of the "final episodes" with which subjects conclude their efforts, and the role which "model-based reasoning" plays in solution attempts.

1.3 A pointer for the hurried reader

The domain analysis and empirical sections of the paper are developed in some detail. The hurried reader might at this point prefer to skip the concluding section entitled "Summary of findings" for an itemized overview of major points.

propose this framework as both an instructional medium and a hypothetical account of subjects' representations of algebra problems. Our interest in the framework is twofold. First, we will use the framework as a means for describing constraints essential for problem solution, although several limitations of the framework as a representational hypothesis will be explored. Second, we will employ some aspects of the framework in describing how an arbitrary pair of problems might be considered analogous for problem solving purposes.

Before pressing this framework into service, we give a brief description of its components with a set of four typical algebra story problems:

Motion: Opposite direction (*mod*).

> Two trains leave the same station at the same time. They travel in opposite directions. One train travels 60 km/h and the other 100 km/h. In how many hours will they be 880 km apart?

Motion: Round trip (*mrt*).

> George rode out of town on the bus at an average speed of 24 miles per hour and walked back at an average speed of 3 miles per hour. How far did he go if he was gone for six hours?

Work: Together absolute (*wta*).

> Mary can do a job in 5 hours and Jane can do the job in 4 hours. If they work together, how long will it take to do the job?

Work: Competitive (*wc*).

> Randy can fill a box with stamped envelopes in 5 minutes. His boss, Jo, can check a box of stamped envelopes in 2 minutes. Randy works filling boxes. When he is done, Jo starts checking his work. How many boxes were filled and checked if the entire project took 56 minutes?

These problems are used in an exploratory study of problem solving behavior, described in the second part of this paper.

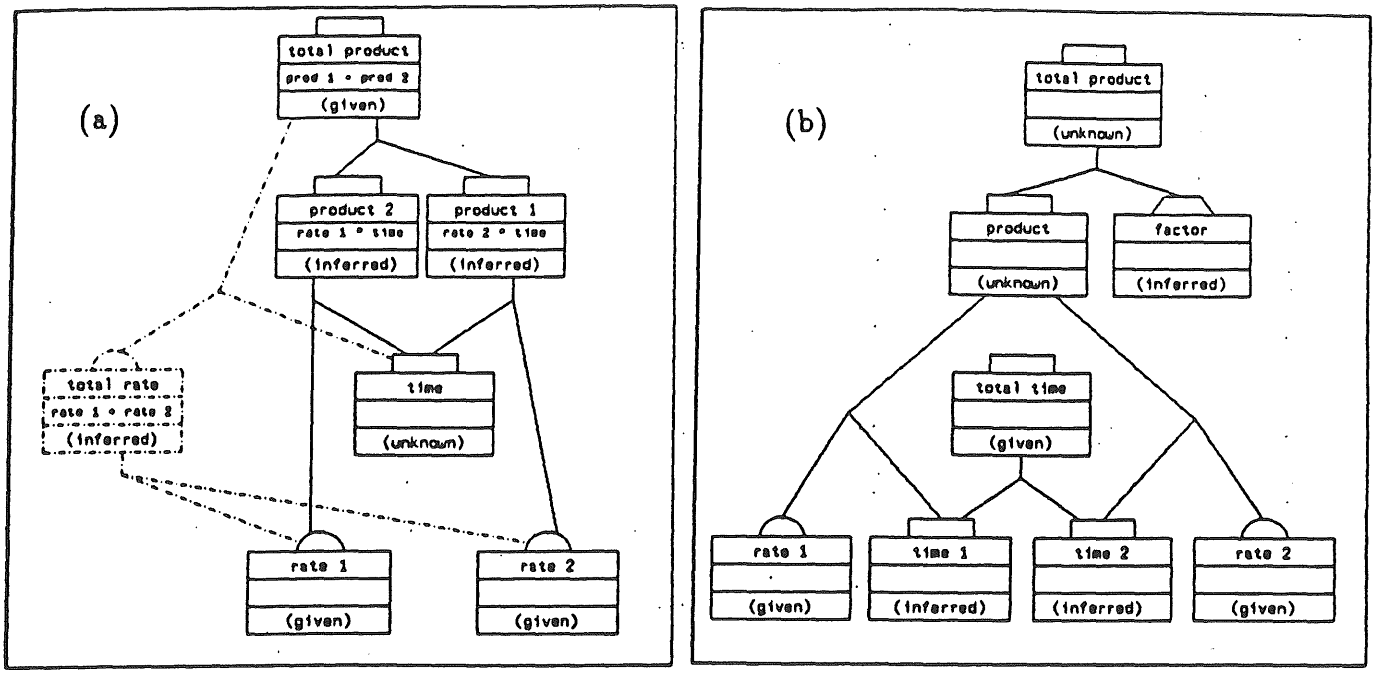

Figure 2: The quantitative structure of two problem classes: (a) contains problems *mod, wta* while (b) contains *mrt, wc.*

for quantitative elements may be annotated with phrases from the original problem text, algebraic expressions for unknown values, or the actual value for givens and intermediate calculations.

At a third level of structure, relational triads can be combined by sharing various quantities to yield "problem structures." These are generalized quantitative networks describing typed quantities and constraints among them. As shown with solid lines[4] in Figure 2(a), a single quantitative network can be used to graphically represent the problem of trains travelling in opposite directions. Sharing a common time, two rates combine through multiplicative triads to yield parts of the total distance. These parts are combined in an additive triad to give a single extensive quantity representing the total distance. Shown in Figure 2(b) is a quantitative network

[4]Portions of the network in dashed lines will be discussed shortly.

if one considers the space of possible problem structures which might be generated by an unconstrained ("weak") mechanism. Thus, one missing component at this level of analysis is that there is no **generative mechanism** for constructing quantitative networks given a problem text or a propositional representation of that text[7].

During empirical studies with this and similar problems, we have found considerable variety in the solution approaches taken by different subjects as well as by individual subjects within a single problem solving effort. As an example, consider elaborating the quantitative network shown in Figure 2(a) to include network components shown with dashed lines. Here we might imagine an elaborative reasoning process has inferred that the given rates can be added. The resulting combined rate (160 kph), when multiplied by the unknown time gives the total distance directly, without adding constituent distances. Hence, a second missing component in this analysis is that there is little **explanatory capacity** at the level of individual reasoning strategies.

We will claim that an explanatory theory which accurately predicts the generation of quantitative constraints during problem solving cannot (except in special, aberrant cases) be based solely within the level of mathematical entities. These networks give a particular quantitative formalism, but their information content is largely the result of processes which draw on other knowledge sources. These processes may include:

a. recognizing quantitative entities directly contained in or implied by the problem text,

b. composing these entities into local relational triads,

c. composing relational substructures into larger problem structures,

d. recognizing familiar substructural arrangements, and

e. detecting when constraints are sufficient for solution.

---

[7]Kintsch and Greeno (1985) present a model of solving arithmetic word problems in which *higher order schemata*, assumed to be available in a subject's memory, provide this generative mechanism. We are exploring related possibilities for algebra story problems, including how these schemata might be acquired.

lar substructures. A substructure, in this context is a subgraph within a larger quantitative network consisting of stated quantities, inferred quantities, and relationships among these quantities. For example, what Mayer (1981) describes as "current" problems are similar at a quantitative level because they share a facilitative–interference relationship[9] between the rate of the vehicle (steamer, canoe, etc.) and the rate of the medium in which it travels (current, tide, etc.). While other aspects of the quantitative structure for a pair of problems can be dissimilar, such a shared substructure may contribute to subjects' estimates of problem similarity demonstrated in empirical studies (e.g., Hinsley, Hayes and Simon, 1979).

In terms of mathematical expertise, similarity judgements at the level of "river" problems may seem paradoxical: problem solvers appear to acquire content-specific categorizations when the true pedagogical goal is to facilitate their learning of mathematical forms. A possible explanation is that quantitative substructures are learned through instruction and problem solving experience and thus form part of the underlying competence in this domain. Since particular substructures are correlated with problem types, the resulting categorizations may appear overly content–specific. However, there may be a functional or pragmatic basis for learning these problem classes. Such learning might occur in spite of the fact that these problems are dissimilar in overall mathematical structure.

2.2 Situational analysis

When exploring the quantitative structure of algebra story problems, we argued that the space of possible problem structures was too vast to allow a weak generative mechanism as a plausible explanation of what subjects actually consider during problem solving. In this section, we examine another level of domain abstraction – the situational setting of a story problem – as a source of constraint on a solution-enabling representation of an algebra story problem. As will become clear, our view of the situational content of an algebra story problem is not synonymous with what other researchers have called "surface content." Although surface materials (e.g., trains, buses or letters) are important problem constituents (this may be particu-

[9] Dellarosa (1985) uses this term.

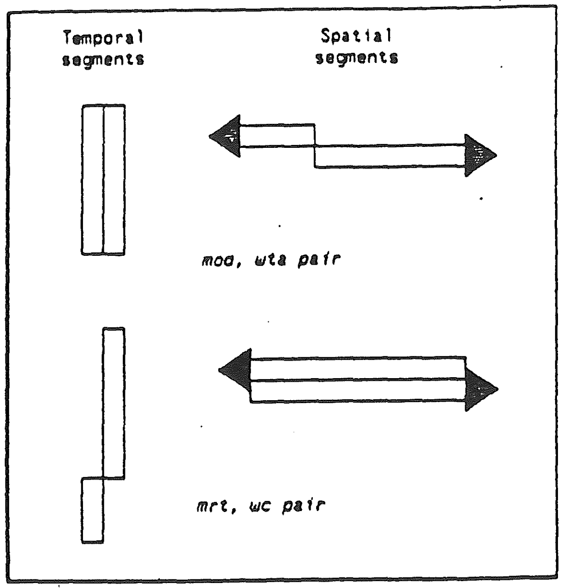

Figure 3: Situational contexts for pairs of isomorphic problems.

in the previous section could be constructed.

2.2.2 A space of situational contexts

Even if we restrict our analysis of compound motion problems to situations in which movement must be colinear and directed, a surprising variety of situational categories are possible. Taking two colinear, spatial segments we can define a set of spatial relationships (e.g., congruent, contained-in, adjacent, etc.) and combine these with directional orientation (same, opposite) to yield various spatial situations. Further considering relations over the temporal intervals during which these spatial segments occur, we find a relatively large space[11] of possible situational contexts. Although some are more plausible than others, these types of situational contexts provide the dimensional basis for algebra "stories" and, we will argue, an important component of world knowledge which can be used to guide the generation of quantitative structures while problem solving.

As with motion problems, we can examine the dimensional character of compound work problems. Here, work products can also be treated as

[11]The precise nature of this space is immaterial for present purposes.

project and will be the subject of later reports. Having described some aspects of these domain materials, we now turn to an exploratory study of problem solving competence.

3 A behavioral description of problem solving

In this section, we describe an exploratory study of problem solving behavior in a population which could be considered representative of competent problem solvers. This study was undertaken with two primary goals: first, to gain a more finely-detailed image of subjects' problem solving behavior, and second, to get an estimate of the frequency with which spontaneous transfer occurs between pairs of problems of varying similarity. This study involves minimal experimental intervention, beyond choice and ordering of problem materials. Results should be interpreted more as an exercise in hypothesis *generation* than in hypothesis confirmation.

3.1 Subjects

Subjects in this study were 85 undergraduate computer science students in their junior and senior years in the major. They were enrolled in an introductory course in Artificial Intelligence, and participated in the study as part of their classroom activities. These subjects could be viewed as "experts" in algebra story problem solving since they must have successfully completed courses in algebra during secondary schooling. In addition, as prerequisites to the course they must have completed three university level courses in calculus and must have finished or been enrolled concurrently in courses covering discrete mathematics. Thus the level of mathematical sophistication in this sample of problem solvers should be high. However, one might argue that these subjects may have at one time been expert algebra story problem solvers, but that their skills have in some sense been "retired" with the passage of time. As will be clear shortly, an image of smooth execution of practiced expertise does not fit the behavior we have observed in many members of this sample.

3.3 Procedure

Subjects were assigned to either of two groups[13], each of which worked through four algebra story problems. They were allowed a maximum of eight minutes to solve each problem, and all subjects worked through the problems at the same time. Subjects finishing early on an individual problem waited until the eight minute time limit expired before proceeding to the next problem. Before solving any problems, subjects were asked to "Show all of your work" in a written form, to "Work from top to bottom, writing new material below previous material," and not to erase after making a mistake. Instead, they were asked to mark through any mistake with a single line. Finally, subjects were instructed to "...draw a box around your answer." After solving all four problems, subjects were given 20 minutes to explain their solutions in writing on facing pages of the text booklet. Instructions for these written explanations are shown in Appendix xx.

The first group of subjects (group M, $n = 46$) saw problems in the following order: *mod, wta, wc, mrt.* The second group (W, $n = 39$) saw the following order: *wta, mod, mrt, wc.* Thus, each group solved pairs of problems which were isomorphic at the quantitative/situational level (*mod, wta* or *wc, mrt*) and also solved pairs of problems which were superficially similar but unrelated at the level of quantitative/situational structure (*wta, wc* or *mod, mrt*).

3.3.1 Data collection

The "behaviors" reported here, and all interpretations of them, are based entirely on subjects' written protocols. Relying solely on written protocols has several obvious disadvantages.

a. *There is no timing information.* We can neither determine how long a subject works on any single problem, nor how long any particular written episode (e.g., performing algebraic manipulation) lasts.

[13]Problem materials were distributed so that subjects with adjacent seating during data collection would be in different groups. Thus group membership was not randomly determined, but should reflect no systematic bias.

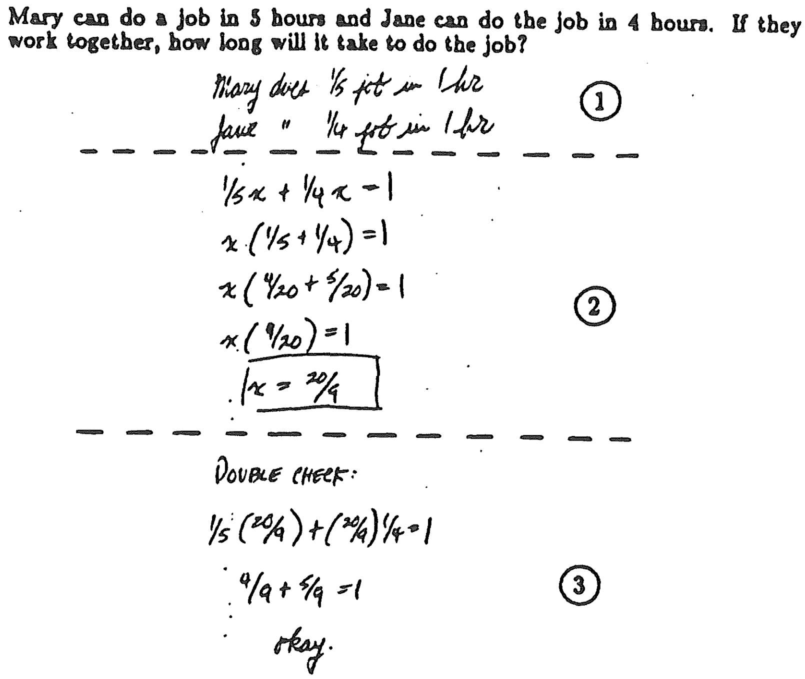

Figure 4: Protocol of subject m20 on the wta problem.

Although episodes divide problem solving into coherent chunks, the context created by earlier episodes is assumed to be inherited by later ones unless there is evidence that a reconceptualization has occurred. In the protocols given as illustrations of various categories in this section, episodes are separated by dashed lines and the sequence is shown with circled numbers. Subject m20 in Figure 4, for instance, goes through three episodes.

Our definition of an episode will become clearer in the following paragraphs as we specify in detail the scoring categories used to describe strategical, tactical, and conceptual aspects of the problem-solving process. The following description is organized under five headings covering different characteristics of any given episode: its strategical purpose, its tactical content, its conceptual content, the quality of formal manipulations and finally the relation of the episode to the overall sequence. This latter heading covers the relative correctness and the reason for transition to a new episode.

Strategical purpose

The strategical purpose of an episode is its relation to the ultimate goal of finding a solution. In this regard, our scoring distinguishes between three

George rode out of town on the bus at an average speed of 24 miles per hour and walked back at an average speed of 3 miles per hour. How far did he go if he was gone for six hours?

bus distance = (24 miles/hr)(x hours)

walking distance = (3 miles/hr)(6 − x hours)

bus distance = walking distance

(24 miles/hr)(x hours) = (3 miles/hr)(6 − x hours)

$24x = 18 - 3x$

(1) $27x = 18$

$x = \frac{18}{27} = \frac{2}{3}$ hours

bus distance = (24 miles/hr)($\frac{2}{3}$ hours)

bus distance = 16 miles = walking distance

One way = 16 miles
Round Trip = 32 miles

Figure 5: Protocol of subject m39 on the mrt problem.

- retrieval of formulas: the subject is remembering and writing down memorized formulas which seem relevant, (e.g., $v = \frac{d}{t}$, Figure 6, episode 4);
- diagram: the subject draws a pictorial representation of the problem situation (Figure 6, episode 1).

- *Algebra:* an episode is algebraic if it makes use of one or more equations placing constraints on the value of one or more variables. However, simple assignments were not treated as equations. Thus neither $100 + 60 = 160$ nor $d = 880$ were considered as equations while $d = 100 \times t$ was. As shown unusually clearly in the protocol of Figure 5, the tactical approach of the typical algebraist is to express constraints as a system of one or more equations (or proportions) and to solve for the appropriate unknown. However, we have found some cases of subjects trying equations in a generate-and-test fashion until, as one subject explained, an equation "looks good".

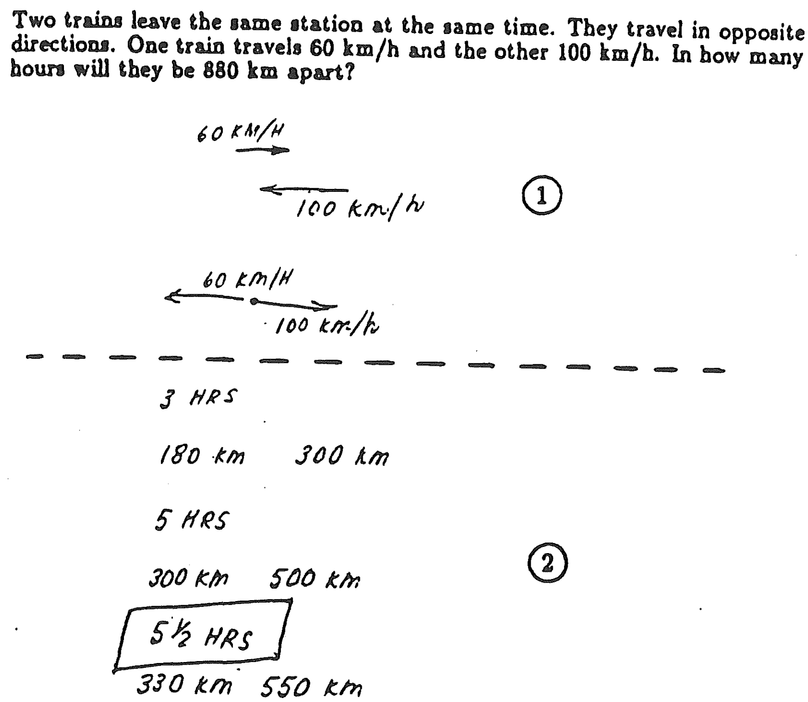

Figure 7: Protocol of subject m03 on the mod problem.

whose magnitude is determined at each iteration by estimations of closeness to the solution. The progression of this generate-and-test approach can be monotonic, as in episode 2 of Figure 7, or follow some form of interpolation search. After each generation of a value, the state of the problem situation being modeled is reconstructed and evaluated.

- *Ratio:* the "ratio" scoring covers a number of tactics by which relations of proportionality between quantities are used, sometimes providing clever "shortcuts" to a solution. This includes:
  - whole/part: the subject views a part as fitting a number of times into a whole quantity, as in episode 6 of Figure 10;
  - part/whole and part/part: these two types of ratios compare portions of entities. Use of the part/whole ratio is illustrated in episodes 2–4 of Figure 8, where the subject considers parts

George rode out of town on the bus at an average speed of 24 miles per hour and walked back at an average speed of 3 miles per hour. How far did he go if he was gone for six hours?

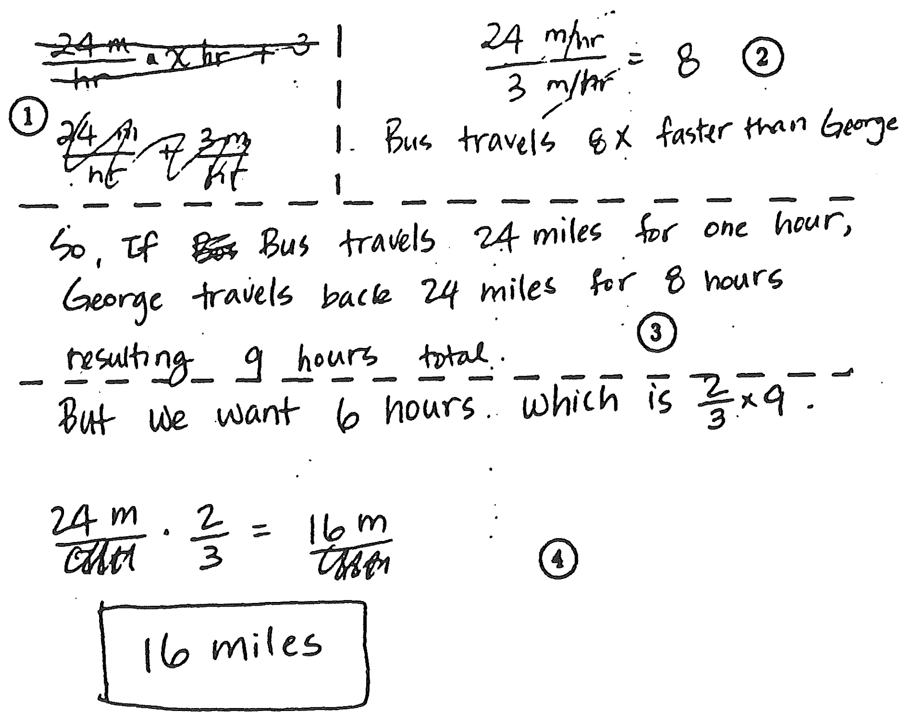

Figure 9: Protocol of subject w17 on the mrt problem.

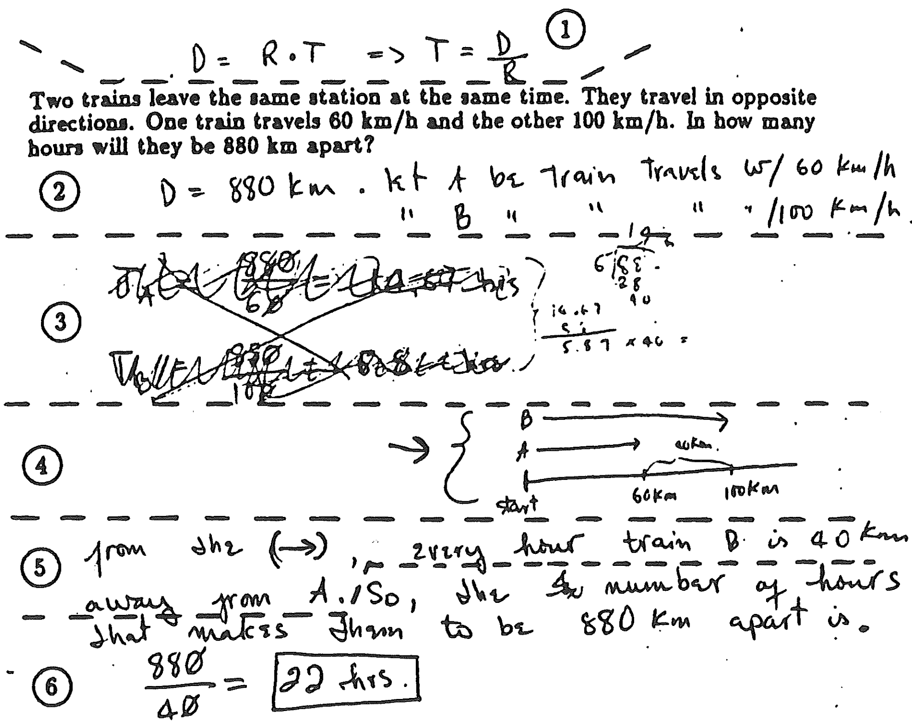

Figure 10: Protocol of subject m19 on the mod problem.

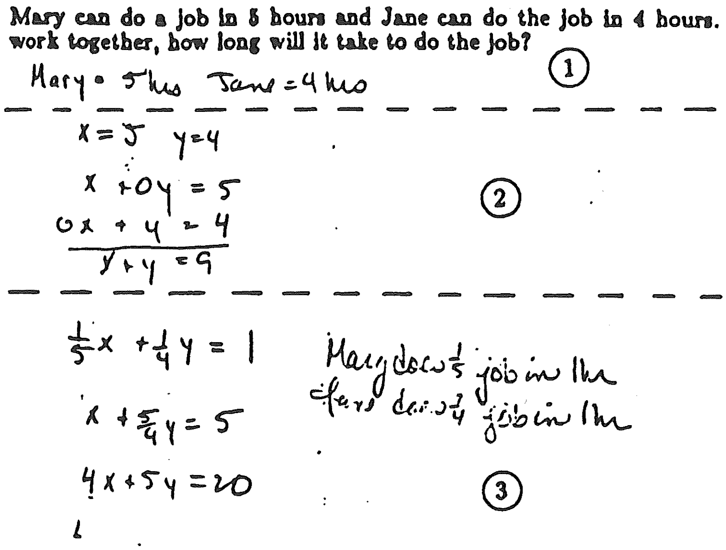

Figure 12: Protocol of subject w23 on the wta problem.

$$880 = \frac{160}{T} \qquad T = \frac{880}{160}$$

- *variable errors:* we observed two types of errors related to the concept of variable. In "switch errors", the meaning of a variable changes in the course of problem solving. In "label errors", subjects are using variables as labels for quantities[16]. For instance, in the return-trip problem (mrt), subject m10 writes the equation $1B + 8W = 6hrs$, explained as "for every 1 hour on the bus, it takes 8 hours to get back".

- *arithmetic errors:* such as $\frac{880}{160} = \frac{11}{4}$ (which subject m20 recovers from after verification, using the scaling strategy mentioned earlier).

---

of the final solution. In fact, this manipulative error results in recovery from a previous conceptual error, suggesting that use of the formalism is often subordinated to semantic intentions.

[16]Although infrequent, label variable errors are surprising among students with substantial mathematical training.

- *impasse:* the subject reaches a point where s/he cannot continue with the current method. A good example of impasse is provided by episode 3 of Figure 6, where a correct simulation by hourly increments cannot reach the non-integer solution and overshoots it, thus causing a switch to another method.
- *lost:* the subject reaches a point where s/he cannot determine how to proceed, as in episode 3 of Figure 12;
- *final solution:* the subject reaches a result and presents it as solution to the problem;
- *final solution wrong:* the subject realizes or believes that the solution presented is incorrect.

The preceding discussion of our scoring taxonomy, of necessity, gives an overly linear picture of our operational approach to written problem-solving protocols. In fact, most of the judgements described above were usually made quickly (from 5 to 20 minutes per protocol) and with little dissension among the scorers. By force, each category was rated with at least 75% agreement over four scorers; many categories approached unanimous agreement. In the analyses which follow, higher-level conceptual categories will be constructed for descriptive purposes by carefully combining atomic category judgements. Thus we will be able to speak of subjects having a "final episode" or a series of adjacent episodes during which model-based reasoning is used. Beyond the results presented in a moment, we expect the set of scored protocols to provide a rich dataset for continuing analysis.

3.4 Results

Although analyses of these data are continuing, we will present results which give an overview of problem solving processes evident in the protocols, and give specific attention to findings which at this point appear promising or interesting. In presenting these results, we make distinctions between subjects' "problem solving attempts," the episodic structure of those attempts, and their "final solutions" offered in the written protocol. By problem solving attempts, we mean all of the behaviors evident in the written protocol, which may include several distinct episodes. By episode

Table 1: Percentage of subjects with a scored category during their solution attempts.

Problem	MOD	MRT	WTA	WC
Content				
Algebra	82.4	85.9	71.8	63.5
Model	30.6	22.4	35.3	47.1
Ratio	17.6	14.1	15.3	42.4
Procedure	0.0	1.2	21.2	0.0
Units	3.5	1.2	1.2	1.2
Annotation	7.1	15.3	21.2	29.4
Diagram	69.4	36.5	8.2	9.4
Purpose				
Comprehension	84.7	64.7	57.6	60.0
Solution attempt	100.0	100.0	100.0	100.0
Verification	28.2	20.0	7.1	20.0
Transitions				
Solution	97.6	75.3	85.9	97.6
Impasse	9.4	10.6	7.1	4.7
Lost	4.7	21.2	15.3	3.5
Wrong	16.5	38.8	25.9	16.5
Errors				
Omission	7.1	21.2	23.5	11.8
Commission	17.6	49.4	42.4	14.1
Arithmetic	9.4	4.7	3.5	2.4
Algebra	5.9	8.2	8.2	0.0
Variable	1.2	5.9	14.1	2.4

Errors in scored episodes.

k. Conceptual errors of omission and commission increase for the more difficult problems (*mrt* and *wta*), and appear much more frequently than manipulative errors (arithmetic, algebraic or variable errors) on all problems.

While these findings are only preliminary, several interesting patterns emerge. First, subjects' written annotations are not composed solely of material generated while performing algebraic transformations. Instead, many subjects appear to use various forms of direct situational reasoning, which we have termed **model-based reasoning**, conducted within the confines of their understanding of a story posed by the problem text. Second, although most subjects do present a solution in some form, their efforts do not appear as a smooth progression towards a quantitative solution. Rather, their problem solving efforts are often interrupted by varied conceptual difficulties which must be repaired before a solution is found. Finally, while manipulation errors within the mathematical formalism do occur, they are overshadowed by conceptual errors of omission or commission as a primary source of problem solving difficulty[17]. Each of these issues will be examined more closely in following sections.

3.4.2 Episodic structure: final episodes and model-based reasoning

Examination of the written protocols clearly shows that subjects undertake a variety of problem solving activities when attempting to solve these problems, particularly when they encounter difficulties in reaching a solution. However, the previous findings speak only to the *presence* of various conditions in subject's problem solving efforts. By our scoring, subjects averaged approximately 2.5 scored episodes per problem solving effort, with some protocols presenting evidence for as many as 10 distinct episodes. We will

[17] "Errors" of omission are difficult to interpret in this context, since within an episode a subject may not yet have inferred a necessary constraint. These omitted constraints may be incorporated during later problem solving efforts.

Table 3: Final episodes: content and errors by correctness.

Problem	MOD			MRT			WTA			WC		
Outcome[k]	C	I	N	C	I	N	C	I	N	C	I	N
n	77	6	2	44	15	26	52	21	12	78	5	2
Content												
Algebra	58	6	0	36	8	20	43	5	7	44	2	1
Model	3	0	0	4	2	6	2	1	2	12	1	0
Ratio	13	0	2	4	3	0	5	3	2	22	1	1
Procedure	0	0	0	0	0	0	1	11	1	0	0	0
Units	2	0	0	0	0	0	0	0	0	0	0	0
Not scored	1	0	0	0	2	0	1	1	0	0	1	0
Errors												
Omission	0	1	0	0	4	6	0	8	4	0	2	0
Commission	1	5	0	0	10	10	1	19	6	1	2	0
Arithmetic	4	0	0	1	1	0	1	2	0	0	1	0
Algebraic	3	1	0	0	2	2	1	3	0	0	0	0
Variable	0	1	0	0	2	0	2	2	1	2	0	0

[k]C = correct; I = incorrect; N = no solution.

Relations between solution outcome and rated categories.

Table 3 shows content and error categories for the problem solving episode during which a final outcome occurred. For episode **content**, an individual subject received a single category score, so cell frequencies sum to give appropriate marginal totals. For a few subjects (1, 2, 2, and 1 subjects across problems) there was insufficient information in the written protocol to enable unambiguous scoring; hence the content of their final episodes was not scored, and they are excluded from our content analysis. For episode **errors**, an individual subject may achieve a correct solution but still demonstrate an error. Furthermore, an individual may have been scored as having several types of errors. Hence, cell entries may not always coincide with marginal subject totals.

Several aspects of final episode content were interesting and consistent with findings for overall solution attempts:

d. With the exception of problem *mod*, conceptual errors are more prevalent than manipulation errors. This is particularly true of the more difficult problems (*mrt* and *wta*).

e. Contrasting subjects who achieved a correct solution with those who did not (incorrect or no solution given), it is clear that subjects with a correct solution had strictly fewer conceptual errors (1:6, 0:30, 1:37 and 1:4 across problems). Subjects reached a correct solution in these cases as a result of offsetting manipulative errors which, perhaps fortuitously, "corrected" their conceptual errors.

f. Manipulative errors, while generally more prevalent among subjects who did not achieve a correct solution, were also observed among those giving a correct solution. Across problems, there were 7, 1, 4 and 2 manipulative errors in the final solution episodes of subjects who achieved a correct solution. These manipulative errors were typically corrected within that final episode to allow for a correct solution.

g. On the *wta* problem, use of an averaging procedure contributes substantially to the number of observed or conceptual errors. Further analyses of the content of conceptual errors are ongoing.

One interpretation of these results might be that manipulative errors are less frequent but more recoverable than conceptual errors. That is, subjects who make an error during a problem solving episode are more likely to recover from that error if it stems from arithmetic or algebraic manipulation than if it is a result of misunderstanding the quantitative structure of the problem. Given our scoring of episodes and conceptual errors, however, this conclusion is not supported by the contrast reported above. Further analysis is required to gain an adequate description of how conceptual errors are corrected across multiple episodes. At present, it appears that the most serious errors among this group of relatively competent problem solvers occur at the level of conceptual understanding rather than the level of formal, manipulation skills.

Table 4: Precursors to situational reasoning: errors and transitional status in a previous episode by purpose of a model-based episode.

Problem	MOD			MRT			WTA			WC		
n	26			19			30			40		
Purpose[k]	C	S	V	C	S	V	C	S	V	C	S	V
No preceding episode	7	1	0	1	4	0	17	4	0	10	2	0
No errors in preceding episode												
Subgoal	3	9	0	0	6	0	2	2	1	10	7	0
Found solution	0	0	0	0	0	0	0	0	0	0	4	2
Lost, impasse	1	0	0	1	0	0	0	2	0	0	1	0
Wrong	0	0	0	0	1	0	0	0	0	0	0	0
Errors in preceding episode												
Subgoal	1	0	0	0	0	0	0	0	0	0	1	0
Found solution	0	0	0	0	0	0	0	0	0	1	0	0
Lost, impasse	0	0	0	0	0	0	0	0	0	0	0	0
Wrong	2	2	0	1	5	0	0	2	0	1	1	0

[k]C = comprehension; S = solution attempt; V = verification.

f. Subjects "abandon[22]" a prior, error-free episode infrequently and when doing so show no particular pattern in their subsequent use of model-based reasoning.

Model-based reasoning episodes preceded by an episode with errors are less frequent than those discussed above, but fall into similar categories. Relatively few subjects have preceding errors yet are unaware of those errors (achieve a subgoal or final solution). This occurs only 3 times across all subjects and all problems. Subjects who are aware of their preceding error decide that they are wrong (there are no impasse or lost transitions).

g. Of those who have determined they are wrong, subsequent model-based reasoning is used either for comprehension or as an attempt to find a solution. Those attempting a solution are in the majority on problems *mrt* and *wta.*

These findings, although based on a subset of the overall sample and highly speculative, suggest an interpretation of the purpose of model-based reasoning. Model-based reasoning is undertaken for one of four basic reasons: as a *preparatory comprehension* strategy when the model-based episode is either the first problem solving activity attempted or follows other comprehension episodes, as a *solution* strategy when subjects feel they are on track, as an *evidence gathering* strategy when a solution has been found previously (this is infrequent), or as a *recovery* strategy when subjects suspect that their comprehension or solution efforts may be "off track."

Apart from inferring subjects' reasons for undertaking model-based reasoning, we would also like to characterize the efficacy of this reasoning strategy. As with the analyses above, interactions between episodes are complex and difficult to extract from our coding of written protocols. In an initial attempt at assessing efficacy, however, we will simply examine the occurrence of any errors within episodes, excluding errors of omission as explained previously. Table 5 shows the relationship between errors during a preceding episode (when there is one) and errors within the model-based reasoning episode.

[22]Lost, impasse or wrong as a transition out of the preceding episode

structure of the problem during later reasoning episodes. Subjects also attempt to find solutions directly through model-based reasoning, generally without introducing errors. Alternately, after encountering an error during previous problem solving activities, subjects may be able to recover through the use of model-based reasoning. Finally, model-based reasoning may play a confirmatory role when subjects have identified important problem constraints or a possible solution.

3.5 Discussion and further analyses

The findings presented above are offered as a preliminary exploration of "competent" algebra story problem solving. By choosing the term competent, we hope to contrast the problem solving behaviors we have observed in this sample against images of "expertise" in problem solving which are often portrayed in the literature. Rather than smoothly executing a set of highly practiced skills, many subjects in our sample appear to *construct* solutions to a presented algebra story problems. These constructions often proceed with some difficulty and include reasoning activities only partly connected to algebraic or arithmetic formalisms.

The latter finding is important given our initial hypotheses about knowledge sources which might support the elaborative construction of a solution-enabling problem representation. Our observation of reasoning strategies outside the traditional algebraic formalism gives indirect support for the hypothesis that reasoning about the situational context of a problem can serve as a justification for assembling quantitative constraints that may eventually lead to a correct solution. Unfortunately we see little overt, performance-level evidence for significant positive or negative transfer between true or false isomorphic pairings.

Many aspects of these protocols are under further study. First, despite little gross evidence of transfer effects from our ordering manipulation over problem pairs, approximately 10% of our subjects gave evidence for some form of transfer during their solution attempts. Many of these cases were instances of negative transfer that either directly violated the quantitative and situational structure of the target problem[24] or demonstrated a failure

[24]For example, subject W08 on problem *wc* incorrectly attempts to add working rates

about how these constraints might be assembled during problem solving.

b. The **situational structure** of a restricted space of algebra story problems is briefly explored. We argue that abstract situational materials not only form the basis for "stories" used as vehicles for algebra problems, but also provide a language of justifications which subjects may use to assemble quantitative representations described above.

Turning to the exploratory study of problem solving, we can summarize as follows:

c. A descriptive **vocabulary of problem solving behaviors** is developed and used to score a sizable set of written problem solving protocols. This vocabulary is built around the idea of problem solving episodes, each of which plays a tactical role in a larger solution attempt.

d. Viewed as a collection of episodes, subjects' overall solution attempts are surprising in several respects:

   1. Solution efforts are not composed entirely of reasoning episodes conducted within arithmetic and algebraic formalisms. Instead, subjects appear to use a variety of reasoning tactics including what we term "model-based reasoning." In model-based reasoning, the subject simulates or runs a model of the situational context of the problem by choosing successive values along a variable dimension (e.g., time) of that context.
   2. Solution attempts cannot be characterized as a smooth progression of problem solving steps within a uniform representational space. Instead, subjects detect inconsistencies in their understanding of the problem and may use a variety of activities to recover from those inconsistencies.
   3. Conceptual errors occur when subjects omit necessary constraints or introduce (commit) errorful constraints in their understanding of the problem. Manipulative errors, in contrast, are generally misapplications of arithmetic or algebraic operations, or incorrect use of variables. In this sample of "competent" problem

case, the model-based episode is usually error free, suggesting that reasoning within the situational context of the problem may help subjects to recover from existing problem solving errors.

Further analyses of these data are underway, but the image of "competent" problem solving that emerges is surprising in many respects. Despite demonstrable achievements in mathematics, many of our subjects appear to have considerable difficulty in finding solutions for typical compound motion and work problems. Furthermore, much of the material we have gathered gives strong evidence for reasoning strategies that lie at least partially outside algebraic and arithmetic formalisms. We hypothesize that these reasoning strategies, particularly what we call "model-based reasoning," play an important role in subjects' elaborative construction of solution-enabling representations of algebra story problems.

5 References

Dellarosa D. (1985) Abstraction of problem-type schemata through problem comparison. TR – 146, Institute of Cognitive Sciences, University of Colorado.

Greeno J.G. (1983) Conceptual entities. In D. Gentner and A.L. Stevens (Eds.) *Mental Models*. Lawrence Erlbaum Associates: Hillsdale, New Jersey, 227–252.

Greeno J.G. (1985) Advancing cognitive science through development of advanced instructional systems. Paper presented at AERA meetings, Chicago.

Greeno J.G., Brown J.S., Foss C., Shalin V., Bee N.V., Lewis M.W. and Vitolo T.M. (1986) Cognitive principles of problem solving and instruction. Report 41, Berkeley Cognitive Science Report Series.

Hinsley D., Hayes R. and Simon H. (1979) From words to equations: meaning and representation in algebra word problems. In P.