DOCUMENT RESUME

ED 388 711 TM 024 158

AUTHOR	Sebrechts, Marc M.; And Others
TITLE	Toward a Cognitive Basis for Quantitative Ability Measures.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-93-22
PUB DATE	Mar 93
NOTE	78p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Algebra; *Cognitive Processes; Constructed Response; Construct Validity; Difficulty Level; Equations (Mathematics); High Achievement; Higher Education; Performance; Prediction; Problem Solving; Thinking Skills; *Undergraduate Students; *Word Problems (Mathematics)
IDENTIFIERS	*Graduate Record Examinations; *Mathematical Thinking; Scholastic Aptitude Test

ABSTRACT

The construct validity of algebra word problems for measuring quantitative reasoning was examined, focusing on an analysis of problem attributes and on the analysis of constructed-response solutions. Constructed-response solutions to 20 problems from the Graduate Record Examinations (GRE) General Test were collected from 51 undergraduates. Regression analyses indicated that models including factors such as the need to apply algebraic concepts, problem complexity, and problem content could account for 37% to 62% of the variance in problem difficulty. Four classes of strategies were identified for constructed response problems: equation formulation, ratio setup, simulation, and other (unsystematic) approaches. Higher achieving students used equation strategies more and unsystematic approaches less than lower achieving examinees. Problem conception errors were the best predictor of performance on the constructed-response problems and the Scholastic Aptitude Test mathematics test (SAT-M). In contrast, procedural errors contributed to the prediction of performance on the constructed-response problems but not to standing on the SAT-M. Overall, these results provide support for the construct validity of GRE algebra word problems and of SAT-M as measures of quantitative reasoning. Appendix A provides some sample problems, and Appendix B is a table of attribute codings. (Contains 10 tables, 1 figure, and 41 references.) (Author/SLD)

RR-93-22

RESEARCH REPORT

TOWARD A COGNITIVE BASIS FOR QUANTITATIVE ABILITY MEASURES

Marc M. Sebrechts
Mary Enright
Randy Elliot Bennett
Kathleen Martin

Educational Testing Service
Princeton, New Jersey
March 1993

2

Toward a Cognitive Basis for Quantitative Ability Measures

Marc M. Sebrechts

Mary Enright

Randy Elliot Bennett

Kathleen Martin

Abstract

The construct validity of algebra word problems for measuring quantitative reasoning was examined from two perspectives, one focusing on an analysis of problem attributes and the other on the analysis of constructed-response solutions. Twenty problems that had appeared on the Graduate Record Examinations General Test were investigated. Constructed-response solutions to these problems were collected from 51 undergraduates. Regression analyses of problem attributes indicated that models including factors such as the need to apply algebraic concepts, problem complexity, and problem content could account for 37% to 62% of the variance in problem difficulty. With respect to constructed-response solutions, four classes of strategies were identified: equation formulation, ra' ɔ setup, simulation, and other (unsystematic) approaches. Higher achieving students used equation strategies more and unsystematic approaches less than lower achieving examinees. Examinees' errors were classified into eight principal categories. Problem conception errors were the best predictor of performance on the constructed-response problems and on SAT-M. In contrast, procedural errors contributed to the prediction of performance on the constructed-response problems but not to standing on SAT-M. Overall, these results provide support for the construct validity of GRE algebra word problems and of SAT-M as measures of quantitative reasoning. A preliminary theoretical framework for describing performance on algebra word problems is proposed and its usefulness for more systematic design of tests is discussed.

Acknowledgements

We wish to thank the many individuals who contributed to this research. Linda Redmond and Sara Chittum helped to categorize strategies and errors. We benefited greatly from Janice Scheuneman's expertise in analyzing problem attributes, which she generously shared with us. Beverly Whittington provided advice and suggestions based on her experience in developing tests of quantitative ability. Eleanore De Young, Altamese Sherrell, and Ruth Yoder all helped with the preparation of the manuscript. Financial support for this research was provided by the Program Research Planning Council.

Toward a Cognitive Basis for Quantitative Ability Measures

One function of validation research is to accumulate a body of empirical evidence that supports certain interpretations of test scores while discounting alternative interpretations. A second, perhaps less appreciated function, is to foster the development of theoretical accounts of performance that can serve as a basis for more principled test design (Messick, 1989). This latter function has become increasingly important given innovations such as computer adaptive testing and the growing demand for tests that provide more detailed, instructionally relevant information. Cognitive analysis of aptitude tests is a method well-suited to serving both functions of construct validation. In addition to producing evidence relevant to what interpretations of test scores are and are not justified, cognitive analysis is useful in uncovering hypothetical constructs that may underlie and influence performance. These hypothetical constructs can be subjected subsequently to further empirical tests and, if supported, can be used to design items and scoring rubrics.

In the present study, our first goal was to explore the degree to which algebra word problems, which frequently appear on standardized tests, are valid indicators of quantitative reasoning ability. Problems were analyzed from two perspectives. One focused on identifying what problem attributes contributed to problem difficulty. The other focused on describing performance in terms of examinees' problem-solving strategies and errors. These perspectives were coupled because it is increasingly recognized that neither ability nor item difficulty are unidimensional *psychologically* although they may be so *psychometrically* (Embretson, 1983; Snow & Lohman, 1989). Performance is a result of an interaction between an individual and a problem and needs to be understood in light of both the knowledge and skills the individual brings to the situation and the nature of the demands imposed by the problem. Individuals who get a particular problem wrong may do so for a variety of reasons. Similarly, problems that are equal in difficulty are not necessarily difficult because of identical factors. Describing the varied factors that contribute to problem difficulty and to proficient performance is an important way of evaluating the construct validity of quantitative ability tests such as those included on the Graduate Record Examinations (GRE) or the Scholastic Aptitude Test (SAT). In addition, a more detailed understanding of the characteristics of problems and performance is critical if tests are to be used to provide more descriptive or diagnostic information to test users.

A second goal of the present study was to describe and evaluate a preliminary theoretical account of problem solving in a small domain (algebra word problems) that can support detailed descriptions of individual performance. This account describes how problem characteristics and cognitive processes interact in ways that can influence individual strategies and errors as well as item difficulty and overall proficiency. The next section describes a general theoretical model which includes the relevant cognitive architecture that mediates problem solving. This is followed by a description of how the cognitive processes in this model, together with problem characteristics, serve as the basis for mathematical problem solving.

<u>Theoretical Model</u>

Most current theories of human cognition assume that problem solving reflects a limited set of capacities. Assuming normal sensory and motor functioning, differences in problem solving can be attributed to the interaction between a problem-solving context and a central cognitive processor that manipulates information in memory. According to this view (Anderson, 1983; Card, Moran, & Newell, 1983; Newell, 1990) the principle cognitive activity underlying problem solving consists in modifying working memory based on the cognitive processor's "recognize-act" procedures. The problem context introduces information into working memory. The current contents of working memory are then compared to possible action sequences in long-term memory by associative links. When a match is found or "recognized," the "action" in long-term memory is carried out and the contents of working memory are updated accordingly. The contents of working memory can be used to generate external changes or to record a solution, and the external context can in turn change the status of working memory. Complex behavior results from the combination of a large number of such memory transformations, using the current state to find associated states and generate new actions.

This model suggests two central cognitive components that will influence mathematical problem solution: processing ability and memory contents. Processing abilities reflect the speed and capacity of information transfer in working memory. Generally, working memory is thought to handle approximately seven distinct chunks of information at one time (Miller, 1956). Without sustained attention, items will remain active for only a few seconds. As a consequence, any effective problem solving must work within these constraints. If too much information must be retained in working memory, or if it must be processed too quickly, the problem cannot be solved. Working memory estimates are generally fairly stable for a wide range of tasks. However, there are differences in what constitutes an informational memory unit or chunk. Increased expertise results in more compact representation, and thus expands the functional capacity of working memory.

The second major factor underlying problem solving is the contents of long-term memory, or content knowledge. This includes knowledge of linguistic structure and meaning, of mathematical procedures and standard problem types, of general problem-solving routines and heuristics, and of objects and events in the world. Problem solving success will depend on the amounts and kinds of knowledge available.

The utility of that knowledge will depend, of course, on retrieval which is influenced both by problem context and a problem solver's expertise. Various aspects of the problem context such as the phrasing of a problem or the problem content play a role in the extent to which the information in memory will be located and utilized. For example, a problem statement is more likely to match a solution structure in memory if the problem statement's surface structure fits directly with a known problem structure in memory. In addition, in the course of problem solving, additional information may be generated and stored externally (e.g., equations, tables, diagrams), modifying the external problem context. The combination of the external and internal representations then provides a framework within which the problem is solved.

The influence of problem context and the nature of memory retrieval also vary with expertise. Experts appear to have their knowledge organized in ways that make it possible to detect patterns and form coherent representations more quickly (Polya, 1973; Glaser, 1991). More proficient students appear to have categories of problems represented in memory, and they are able to use that categorization as the basis for retrieving a solution process (Hinsley, Hayes, & Simon, 1977). Experts are more likely to use the problem (deep) structure as the basis for a search of memory, whereas novices tend to use problem (surface) features (Glaser, 1991). In addition, greater proficiency appears to be associated with improved strategies for developing qualitative models of a situation rather than proceeding directly with quantitative manipulations (Paige & Simon, 1966).

Cognitive Architecture and Mathematical Problem Solving

The model of human cognition outlined suggests that there are two primary sources of problem-solving difficulty: working memory and long-term memory. Here we consider how that general cognitive model can account for problem solving in mathematics. Working memory capacity limits will help determine how problem information can be effectively structured. Although there is some evidence of a relationship between working memory capacity and mathematical proficiency (Hiebert, Carpenter, & Moser, 1982), most research suggests that working memory capacity is fairly consistent across individuals with varying mathematical ability (Spiegel & Bryant, 1978; Briars, 1983). As noted above, however, the functional capacity of working memory can vary with expertise. The quadratic equation, for example, may represent a number of separate chunks for a novice, whereas it may be a single chunk for a more proficient mathematician.

Long-term memory, in contrast, is assumed to have no capacity limits. Its role in problem solving is determined by content. Success in problem solving will depend on the currently retrievable contents of long-term memory, including knowledge of specific mathematical relations as well as general problem-solving strategies.

Solving algebra word problems requires that individuals retrieve from long-term memory both everyday and more specialized mathematical knowledge. This knowledge is manipulated in working memory in order to comprehend the problem situation and construct an integrated model or representation. This representation is then used to formulate a solution plan which in turn draws on other mathematical skills for execution. This approach to mathematical problem solving can be summarized in terms of four general cognitive activities: problem translation, problem integration, solution planning and monitoring, and solution execution (Mayer, Larkin, & Kadane, 1984; Mayer, 1987). The remainder of this section suggests how the properties of working and long-term memory function jointly with item attributes to influence each of these four activities.

Problem translation. When reading a problem, students must use linguistic knowledge to interpret or translate what is being stated, and to restate the givens and goals in their own terms. They must recognize the meaning of specific measures such as "meter" or "yard" as well as more complex terms such as "compound interest." In addition, they often must use a wide range of factual and common-sense knowledge, such as the assumption that a train traveling in a certain direction is moving in a straight line. While overall linguistic complexity (i.e., length and linguistic structure of the problem statement) is known to be associated with problem difficulty (Barnett, 1984), the contribution of the underlying mathematical complexity to this relationship has not been fully evaluated. However, there are other, more specific factors that are likely to affect problem difficulty. For example, there is evidence that people have more difficulty interpreting relational statements such as "John is six years older than Sue" than assignment statements such as "John is nine years old" (Mayer, 1982; Loftus & Suppes, 1972). When children are asked to repeat problems after listening to them, they often ignore the relational aspects, so that "Bill has seven more apples than Sue" would be remembered as "Bill has seven apples" (Riley, Greeno, &

Heller, 1983). Another source of difficulty is the need to notice significant details such as differences in measurement units (i.e., minutes vs. hours) among variables, or between the givens and the required answer. Domain familiarity can also influence comprehension: If a problem statement includes terms or concepts that are not stored explicitly in long-term memory, translation becomes more difficult, if not impossible. Finally, problem solvers may have to distinguish which information in the problem statement is relevant and irrelevant to the problem solution.

Problem integration. Underlying algebra word problems are one or more alternative mathematical structures that embody relationships (mathematical operations) among the problem elements and specify paths from the givens to the goals. Typically, most of the problem elements, the final goal, and some constraints are explicit in the problem statement and identifying them is a matter of translation or comprehension. However, other constraints and the relationships among the problem elements are often implicit in the problem situation and much of the challenge in solving algebra word problems resides in uncovering these implicit relationships and constraints and organizing them into a larger structure.

The simplest way in which this integrated representation can be achieved is by the triggering of a previously stored solution strategy (schema) from long-term memory. This schema can then serve as the structure for the problem solution.

Not surprisingly, the presence of such schemata is highly dependent on prior training and experience. Mayer (1981) analyzed algebra word problems from secondary school algebra texts and found that these problems could be classified into eight families based on the "story line" and source formulas such as "distance x rate = time" or "interest rate x principle = dividend." He further identified approximately 100 subcategories of problem types within these families and found that the frequency of occurrence of these problem subcategories varied from 4 to 25 per 1000. In a subsequent study, Mayer (1982) found that these frequencies were positively correlated with probability of recalling a previously read problem. Furthermore, recall errors demonstrated a tendency to convert low frequency problems into similar high frequency ones. Mayer concluded that the presence of problem-solving schemata is a function of the amount of attention they receive in high school. Developing problem expertise is, in part, a function of improved representations in memory through example and practice.

Hinsley, Hayes, and Simon (1977) suggested that competent problem solvers used problem categorization as an important means to select a solution schema. They found that competent students can categorize problems into problem types after hearing only a few words of the problem statement. One subject, for instance, after hearing the words, "A river steamer..." said "It's going to be one of those river things with upstream, downstream, and still water. You are going to compare times upstream and downstream--or if the time is constant, it will be the distance" (p.97). This ability to sort problems quickly is useful and is probably one of the hallmarks of expertise (Glaser, 1991).

This rapid triggering of schemata can also be a liability. Students often learn schemata linked to particular contexts. For example, one "expert" high school algebra schema is for "river problems," describing movement of a boat with and against a current (VanLehn, 1989). However, this type of problem schema is dependent on surface features and may therefore be overly specific. If students use the surface cues for retrieving schemata they are likely to make errors when the problem context is different: A problem involving a plane flying with a tail wind or head wind might not be seen as having mathematical structure similar to "river problems." In contrast, successful problem solvers are more likely than unsuccessful ones to categorize mathematics problems on the basis of structural rather than surface features (Silver, 1979). The inappropriate level at which problem types are stored in memory may account for the widely reported difficulty students have in new problem contexts despite substar ial success within a given context.

If the problem statement does not directly elicit a specific schema from long-term memory, the problem solver must compose a representation that captures the underlying quantitative structure. Although story lines can provide some clues, similar story lines may reflect very differen quantitative structures (Mayer, 1982). In order to capture this relevant quantitative structure independent of the surface context, a number of network notations have been developed (Hall, Kibler, Wenger, & Truxaw, 1989; Reed, 1987; Reed, Dempster, & Ettinger, 1985; Shalin & Bee, 1985). For example, Shalin and Bee analyzed the quantitative structure of word problems in terms of elements, relations, and structures. Elements include (a) extensives or primary quantities, (b) intensives such as rates that map two extensives to each other, (c) differences or additive comparisons of two quantities, and (d) factors that involve multiplicative comparisons of quantities. Many word problems consist of one or more triads of elements combined in additive or multiplicative relationships. One of the relationships Shalin and Bee described, a multiplicative relationship among one intensive and two extensives, is typical of many algebra word problems such as those involving speed, interest, costs, and work.

For complex problems that involve more that one triad, problem structure describes the way that these triads are linked. Shalin and Bee found that many two-step word problems could be classified as an exemplar of one of three linked structures--hierarchy, shared-whole, shared-part--and that these problem structures had an effect on problem difficulty. An illustration of a structural representation for the following word problem is given in Figure 1:

> "A person invests $10,000, some at 8 percent per year and some at 10 percent per year. The annual income from this investment is $870. How much was invested at 8 percent?"

As seen in Figure 1, this problem can be described as having four elemental triads or substructures related through hierarchical and shared-part linkages.

Insert Figure 1 about here

Although this underlying structure is not meant to serve as a description of the student's cognitive activity, it can help to explain differences in performance. For example, because of working memory capacity limitations, the more elements and relationships there are, the more difficult a problem is likely to be. However, knowledge about basic, complementary mathematical relationships among elements such as "distance, rate, and time" or "dividends, interest, and principal" should help individuals to group or chunk subparts of a problem, and this knowledge may vary in accessibility for different problem classes. Integrating these chunks into a larger structure requires recognizing constraints that are operating in the problem situation.

A constraint in the above problem is that the income from the two investments adds up to the total income.

Problems can also be characterized by a situational structure (Hall, Kibler, Wenger, & Truxaw, 1989) which preserves certain kinds of information such as semantic and spatial relations that are lost in the quantitative abstraction. A situational analysis may underlie solutions that emphasize model-based reasoning and problem simulation rather than the algebraic abstraction. For example, the above problem can be solved by recognizing that complementary portions of the $10,000 investment are put in at each rate and that the annual income will increase as a greater portion of the money is placed at the higher (10%) rate. A table of possible solutions with gradual convergence can then be constructed.

Although quantitative and situational structures do not constitute psychological models, they do specify structural constraints that influence possible representations. These representations can in turn influence solution strategies. For example, problems with concrete temporal intervals, such as the one above, are amenable to a situational analysis with discrete temporal units. As a consequence, it is feasible to use a representation that models successive time intervals in order to converge on a solution. In contrast, problems that specify a value in "abstract" rather than concrete increments (e.g., "*y* seconds" in Problem #3 or "*K* kilometers" in Problem #15, see Appendix A) do not yield a straightforward situational mapping; such problems therefore are less likely to be solved using a modeling/simulation approach.

Solution planning and monitoring. For any given problem, there are multiple approaches to a solution. The chosen approach will depend on how the problem has been translated and whether or not an appropriate schema has been retrieved. In addition, the approach will depend on the kinds of strategic knowledge a student has stored in memory. For example, a student may try decomposition in which a problem is broken down into constituent parts; he or she may use backward reasoning in which a goal state is specified, components of the target equation are set as subgoals, and subgoals are in turn broken into finer subgoals until a solution can be achieved by substitution or simple computation. While a plan prescribes the actions that should lead to problem solution, the effectiveness of the plan and the accuracy with which actions are executed need to be evaluated or monitored as the plan is implemented.

Although planning has been shown to aid problem solving (Schoenfeld, 1979), empirical studies indicate that its use by students is very limited (Branca, Adams, & Silver, 1980; Briars, 1983). More competent problem-solvers tend to identify an appropriate schema which requires a specific solution plan, to monitor the success of their progress as they work on their solution, and to verify the result. Less competent problem solvers often approach a problem at a superficial level, without adequate planning. Lester (1983) found that poor problem solvers often use keywords as the basis for determining which operators to apply, and then apply the chosen operators directly to the numbers in the problem statement. Less competent problem solvers likewise tend to assume that there is one rather direct solution and, as a consequence, fail to monitor and evaluate their solutions as they are working (Briars, 1983).

Because planning and monitoring are superordinate to, and integrated with other aspects of problem solving, few factors can be identified that have a unique impact on specific problem-solving activities. One exception may be the nature of the problem goal which can be classified as either a quantity or as an expression containing a variable. The presence of a variable in a goal would seem to make it more difficult to evaluate the reasonableness of a result. For example, finding that a car was traveling 600 miles an hour should alert an individual to a possible error in the solution. However, an answer of x/y miles per hour cannot be so readily evaluated.

Solution execution. Once a sequence of steps has been planned, the solution must be implemented by executing those steps. In general, this consists of a series of computations as well as symbolic manipulations. Typically, more competent students will plan the whole solution before execution, whereas the less competent students may execute portions of an incomplete set of plans before formulating further steps. In either case, the difficulty of these steps will again depend on the extent to which they are already stored in memory and can be simply retrieved.

For most word problems, some simple procedures (such as addition of two single-digit numbers) can generally be solved by rote; the relevant "known facts" are retrieved from memory as needed (Fuson, 1982). More complex procedures, however, may require a sequence of steps that present multiple opportunities for making errors. There are probably at least two sources for errors in execution. First, the student may retrieve a procedure that is not specified in enough detail. For example, he or she may move a variable by simply removing a value from one side of the equation and subtracting it from the other, a procedure that works for addition, but not for subtraction. Second, although a student may have the correct procedure in long-term memory, an error may be produced by the incorrect execution of an operation from working memory. For example, a decimal point may be placed inappropriately, or a value may be changed inadvertently during transcription.

In the ideal case, translation, integration, planning and monitoring, and execution tend to proceed in a linear manner without error or inefficiency. However, the more common case includes substantial movement between the various stages. A solution, for example, is partially executed and a difficulty is encountered. The student might then go back to read the problem again, and generate a new translation. VanLehn (1988) has suggested that the difficulties in a solution strategy or the "impasses" are critical to an understanding of student solutions. Students do not follow a linear path, but rather when they encounter a difficulty, they search for a "repair" that generates a solution within the new context they have constructed in their solution effort.

Empirical Analysis

This paper uses the theoretical model described above to account for performance on a set of algebra word problems. More specifically, an attempt is made to provide a preliminary cognitive rationale for word-problem performance on the GRE General Tests's quantitative section. To that end, detailed analyses are performed on student solutions to identify strategies and errors in a set of typical word problems. These analyses are placed within the context of the way in which cognitive processes interact with item attributes to account for performance.

Method

Materials

All word problems (excluding quantitative comparison problems) were identified on disclosed versions of the GRE General Test quantitative sections administered from 1985 to 1989. This problem pool was composed of 75 problems that were classified into categories of Equation, Probability, Distance = Rate x Time, Interest, Work, and Graduated Rate by an ETS test developer. Twenty problems, shown in Appendix A, were selected from the four latter categories. These 20 problems had been administered in a multiple-choice format on the GRE. In the present study they were administered in a constructed-response format. Problem difficulty, as measured by equated delta values for different samples of examinees who took the GRE, is shown in Table 1. (Equated delta is a transformation of percent correct with a mean of 13 and a standard deviation of 4, and is comparable across samples of examinees. In contrast with percent correct, equated delta and difficulty are positively related: the more difficult a problem, the higher the equated delta). The mean equated delta was 13.48, indicating that this set of problems is of average difficulty.

Insert Table 1 about here

Problem representations. As noted earlier, there are numerous alternative representations that can be formulated for any given problem statement. For purposes of our analysis, we used two alternative representations that capture essential problem information in a parsimonious form for each of the twenty problems. The first of these consisted of one or more equations that incorporated all the relevant values and variables from the problem statement, any implicit values or other necessary variables, and the relationships among these values and variables as concisely as possible. (These equation-based representations for all twenty problems are included in Appendix A.) These equations were developed by two of the authors in consultation with another individual with expertise in developing quantitative items for standardized tests. The second form of representation was a diagram based on Shalin and Bee's (1985) analysis of word problems that lends itself to qualitative descriptions of aspects of the problems' structure. Examples of both forms of representation are illustrated in Figure 1.

Item attributes. After the problem representations were agreed upon, each problem was coded for the attributes shown in Table 2. These attributes were selected based on a review of previous research on problem characteristics in mathematics (cf. Goldin & McClintock, 1984; Mayer, 1981, 1982; Scheuneman, in preparation; Shalin & Bee, 1985), as well as discussions among the project staff. These attributes attempted to capture the essential properties of the problem statement as well as critical features of each of the two alternative problem representations.

Insert Table 2 about here

Subjects. Fifty-one undergraduate students at The Catholic University of America were recruited as part of a course requirement in Introductory Psychology. Thirty-three students were female and 18 were male. Four students were in the Nursing School, 10 in the School of Engineering and Architecture, and 37 were in the School of Arts and Sciences; of this last group, 25 had selected majors--17 in humanities, 3 in math or science, 2 in the Social Sciences and 3 in Business. SAT scores were available for forty-one of the subjects. In those cases in which students had taken the test more than once, the highest score was used. For these students, the mean SAT-M score was 578 with a standard deviation of 91; the mean SAT-V score was

506 with a standard deviation of 74. In contrast, mean scores for all examinees in 1987, the year most subjects would have taken the SAT, were 476 (SD=122) for SAT-M and 430 (SD=111) for SAT-V (College Board, 1987).

Procedure

All problems were presented in open-ended format, with no answer options available. After completion of consent forms, students received general instructions together with a set of twelve word problems which they solved at their own pace. When they had completed those problems, students were given twelve algebraic equations to solve which tapped the requisite symbolic and computational skills needed to solve the word problems; these procedural problems are not discussed further here. Finally, after completing the equations, the students were given an additional eight word problems. The entire session lasted from one to two hours depending on the individual. No supplementary materials or calculators were available during the session.

Scoring of solutions

Each problem was analyzed for the correctness of the end result and for the presence of any errors in the solution. A solution was considered "correct" if the answer was exact, if it was more precise than requested, if it differed from the desired answer only by precision beyond the first decimal place, or if the answer was correct given a plausible alternative reading of the problem stem. For example, one problem (#1) asked for the number of "complete steps" required. The answer "115.2" was not in complete steps, but was otherwise accurate and therefore was considered correct. In determining the presence of errors, the presumption was in favor of the student. If the end result was correct, no error was assigned unless it could be clearly established how the error resulted in a correct end result. There was a negligible number of such solutions.

Analysis of strategies and errors

Each solution was analyzed for the strategies used and errors made. Since most students appeared to use one dominant approach, the analysis considered only the final strategy for each problem. An initial taxonomy of strategies and errors was developed by one of the authors based on the actual solutions and test developer suggestions. This taxonomy was expanded as necessary by two research assistants who reanalyzed each problem. A subset of problems was then analyzed again by one of the authors. Any discrepancies or questions were resolved by joint analysis with the research assistants. Thirty responses with their analyses were then presented to a test developer for final confirmation of the structure of the taxonomy.

Results

Responses were available from nearly all subjects for all problems. Of the 1020 solutions (51 subjects X 20 problems) data was missing for 25 responses. In 6 cases, it was not possible to tell if the subject had unintentionally skipped the problem. For the remaining 19 cases, the omissions were clearly intentional and were treated as errors.

Percent correct for each problem is shown in Table 3. Mean performance across subjects was 42% correct (S.E.= 3.5), and varied across subjects from 0% to 85% on the 20 problems. Only one quarter of the subjects got at least 50% of the problems correct and only three subjects got at least 75% of the problems correct.

Insert Table 3 about here

Performance in this study was highly correlated with standard measures of problem difficulty and student proficiency. The results from the sample of 51 subjects showed a difficulty ordering very similar to that obtained when these problems were administered nationally in multiple-choice format. Problem difficulty as measured by percent correct for our constructed-response sample correlated -0.87 with the GRE equated deltas based on national administrations.

For the forty-one subjects for whom SAT scores were available, SAT scores were positively correlated with percent correct (Pearson $\underline{r}$ with SAT-M = 0.74; with SAT-V = 0.53). Using a least squares simple linear multiple regression, it was found that the stronger relationship was with mathematical ability ($b = 1.33$; $t(38) = 5.51$, $p < .01$) although there was also a reliable effect of verbal ability ($b = .63$, $t(38) = 2.12$, $p < .05$); the adjusted R^2 for both variables together was 0.58.

Item Attributes

Preliminary analysis was concerned with identifying attributes that occurred with reasonable frequency in this small set of items and determining if these attributes were related to item difficulty. Two measures of item difficulty were used: the percent correct obtained in the current study (constructed-response format) and the equated delta from the GRE administrations (multiple-choice format). The fact that only 20 problems were available limited the sensitivity of the analyses of item attributes. Dichotomous variables were included in the analysis only if there was a minimum of five instances of the less-frequent value for the set of 20 problems. This resulted in the exclusion of some potentially interesting attributes. For example, having to express an answer in terms of a variable rather than a quantity seemed to make problems particularly difficult (Problems #3, #13, and #15), but given the small number of instances, the specific contribution of this variable could not be evaluated.

In Table 4, first order correlations between measures of problem difficulty and item attributes are presented. (The coding for each problem on these attributes is presented in Appendix B.) These attributes are grouped into categories on the basis of both logical analysis and intercorrelations among the attributes. The first category, labeled "Algebra," consisted of only one attribute, the appearance of a variable more than once in the equation-based representation. This attribute tended to have low to moderate correlations with other attributes (-.03 to .43) and was a particularly good predictor of the difficulty of constructed-response problems.

Insert Table 4 about here

The finding that the "Algebra" attribute was related to difficulty is important. Many problems in our set were "non-algebraic" in that they could be represented using only one variable and solved through a series of arithmetic operations without any need to manipulate variables. Other problems, however, required either more than one variable or the repetition of a variable in the representation. For these problems, examinees had to engage in some minimal algebraic thinking such as representing one variable in terms of another or expressing an answer in terms of a variable. Thus, the application of algebraic concepts, which requires abstraction from a context and fosters generalization and formalization of the problem situation, characterized some of the more difficult problems. According to the Curriculum and Evaluation Standards of the National Council of Teachers of Mathematics (1989), the ability to represent quantitative situations with expressions that include variable quantities is a central competency that should be developed in the high school years if not earlier. The NCTM standards state that an "understanding of algebraic representation is a prerequisite to further formal work in virtually all mathematical subjects" (p. 150). That college students--who should be proficient in representing problems algebraically--found these questions relatively difficult is disturbing.

The next category in Table 4 consists of four attributes that are measures of the complexity of either the equation-based or the structural representations of the problems. These attributes had high intercorrelations with each other (.72-.97) and were good predictors of difficulty, as might be expected given an assumption of limited cognitive capacity. Level of nesting was an especially strong predictor of difficulty in the GRE sample.

A third group of attributes was categorized as content attributes and included the need to convert measures of time and the presence of metric measures or money in the problem statement. The intercorrelations among this group of variables were moderately high and sometimes negative (-.54 to .54). Content categories tended to be mutually exclusive: Problems that were about money seldom included metric measures or measures of time. The time conversion attribute logically could have been categorized as a complexity attribute in that it requires additional arithmetic operations; however, it was included in the content category because of its association with other content variables which also sometimes required additional arithmetic operations (percents to decimals or fractions). The content attributes were among the best predictors of difficulty in the college sample, although time conversions and metric measures were associated with increased difficulty, whereas money was associated with decreased difficulty, perhaps reflecting relative familiarity with using the relevant units.

Finally, linguistic attributes, including whether there were specific relational words in the problem statement and the frequency of various kinds of propositions, are grouped together in Table 4. However, the intercorrelations of these attributes with each other were low to moderate (.03 to .67) and similar to their correlations with complexity attributes (-.05 to .70). The correlations of these attributes with problem difficulty also were low to moderate.

Multiple regression analyses. Multiple regression analyses were carried out to evaluate the independent contribution of the various attributes to difficulty. Based on both theoretical considerations and evaluation of the first-order correlations, a 2-variable model including the "algebra" attribute and a complexity attribute, "level of nesting," was selected for evaluation. Content and linguistic attributes were added to this model individually to see if they accounted for any additional variance. The addition of linguistic attributes did not produce significant changes in R^2 and are not discussed further. Summaries of the 2-variable and 3-variable models which included content attributes are presented in Table 5 and in Table 6 for the college sample and the GRE samples respectively. The 2-variable model accounted for 32% and 41% of the variance in the two measures of problem difficulty. Both the time conversion and the money content variables produced significant increases in R^2 for the prediction of percent correct in the college sample. However, the fact that the content attributes in these problems are confounded makes it difficult to interpret the effects of these attributes unambiguously. Are "money" problems easier to solve because people are more familiar with them (and can retrieve appropriate schema more easily) or are they easier because they do not involve time conversions or working with unfamiliar metric measures? For the prediction of equated delta, only the addition of the money content variable produced a significant increase in R^2.

Insert Tables 5 and 6 about here

In sum, a 3-variable model taking into account the need to apply algebraic concepts, problem complexity, and some aspect of content accounted for between 37% to 62% of the variance in two independent estimates of problem difficulty in this set of 20 problems. Because these estimates of item difficulty were obtained under very different circumstances, including differences in subject cohort, motivation, and problem format, the influence of these attributes seems robust. Furthermore, the finding that these attributes, rather than linguistic ones, are related to problem difficulty, is supportive of the construct validity of these problems as measures of quantitative reasoning. However, because of the small number of problems in the set, the confounding of attributes, and the low frequency of some potentially important attributes, the results of these analyses must be viewed as suggestive. In future research, problems should be designed to vary item attributes systematically so that their independent contributions can be evaluated.

Strategies

Strategy taxonomy. Subjects used a wide range of approaches to the constructed-response problems, although there was substantially more consistency for correct than incorrect solutions. Of some 600 incorrect solutions, 75% resulted in unique final answers given by only one subject, indicating that the numerical end-result alone is not an adequate means to diagnose errors in the solution process.

Three broad classes of strategy were evident: equation formulation, ratio setups, and simulations. Equation approaches were those in which the students attempted to establish and work through the solution primarily by using a set of equations, with or without explicit variables. Ratio setups used a specific set of equations that followed the canonical form $a/b = c/d$, with one of the terms representing the unknown. Simulations were strategies in which the students "modeled" the situation by indicating the state at different iterations on one variable, usually time. The frequency of occurrence of each of these approaches is shown in Table 7.

Insert Table 7 about here

The three primary strategies accounted for 99% of correct solutions and 68% of the incorrect solutions. Of the total solutions, 2% were not attempted and were therefore considered incorrect (labeled "N" for "No Work Shown" in Table 7). Students used strategies that were not easily classified in 17% of their solutions, almost all of which were incorrect. These alternative strategies (classified as "Other" in Table 7) included guessing and trial-and-error. The use of these strategies varied markedly with problem difficulty; 84% of "other" strategies were for the more difficult half of the problems, as defined by equated delta.

Equation strategies accounted for 65% of solutions overall and 81% of correct solutions. This strategy was used more frequently than the ratio or simulation strategies for 18 of the 20 problems. Most typically these solutions were constructed from relatively simple, learned formulae, such as d = rt.

Although almost all of the problems could be solved using a single equation with appropriate conversions, relatively few subjects followed that strategy. This was particularly noticeable with problems which required the statement of one term of the equation in terms of another. For example, in Problem #4, if the time of the first train were represented as *x*, the time of the second train would be x-2. In Problem #5, if the amount invested at 8% were represented as *y*, the amount at 20% would be "$10,000 - *y*." However, only 14 of the subjects used this approach on at least one of these two problems. Surprisingly, only 4 subjects used the procedure on both problems, suggesting that the context was important in procedure selection.

More commonly, subjects attempted to decompose the problem into its constituents. Most solutions employed the strategy of isolating the smallest possible step and incrementally adding to those steps. This approach appears to be followed even in cases in which this approach is not most efficient. For example, Problem #6 asks how long it would take to double the principal of $750 given an annual interest rate of 5%. For that problem, 90% of the forty-two subjects who had an identifiable strategy included the specific dollar amount in their solution, the vast majority determining the yearly return of $37.50 and then dividing $750 by that amount to arrive at a solution of 20 years. In this case, however, the specific dollar amounts could have been ignored since, for any principal amount, it would take 20 years to reach 100% at 5% per year.

Ratio setup was the preferred strategy for Problem #1 and constituted at least 10% of the solution strategies in five other cases (Problems #2, #3, #8, #9, and #15). These problems can be identified as those in which the structure of the problem maps reasonably directly onto a ratio. The canonical form of such problems is: If quantity-A of unit-1 is generated in quantity-B of unit-2, what quantity-C of unit-1 will be generated in quantity-D of unit-2? Problem #1 has the simplest and most direct mapping from text to ratio and is the one for which this strategy is most successful. When a time transformation (Problems #3, #8, and #15), a variable (#3) or a constant (#15) is introduced, the strategy becomes less frequent and less successful.

Finally, simulation was the preferred strategy for problem #5; students estimated the amount invested at 8% and the amount at 10%, determined the outcome, and then revised their estimates. Simulations also constituted the strategy for over 10% of the solutions in three other instances (Problems #4, #10, and #11). In each case, the student attempted to determine the target values for each increment of time until a solution was achieved. For example, a student would indicate the distance from the station for each of two trains at each hour (Problem #4).

Situational determinants. Based on the large percentage of "standard" strategies associated with correct solutions, it appears that, whenever possible, subjects tried one of a very limited number of known procedures.

The key to using prior knowledge is being able to retrieve the appropriate "schematic" elements and to combine those components to construct a satisfactory representation. Forming such a representation, however, depends on several factors including the surface context of the problem and its linguistic structure.

The surface context serves as one clue to solution strategies. For example, one of two approaches was usually applied to the train catch-up problem (#4). Of the 40 subjects who had an identifiable strategy, 50% used equations that took into account the discrepancy between the trains while 42% set up some form of table for a simulation of the two train states. Surface context, however, can also elicit the wrong approach. On Problem #6, which involved an interest bearing account, many students assumed that the problem concerned compound interest, although only simple interest was required.

The linguistic structure of the problem stem can also induce certain strategies. For example, problem #1 maps directly onto a ratio strategy. For that problem, a ratio approach accounts for 57% of correct solutions. The linguistic structure can also interfere with problem solution when it appears to match the quantitative structure, but does not. For example, Problem #2 appears to have a ratio structure that can be expressed as "time-1:speed-1::time-2:speed-2." In this case, the assumption is wrong since the ratios as formulated are not equivalent.

If neither the surface context nor the linguistic structure serves as adequate retrieval cues for solution methods, as was the case for many of the problems, then it is necessary to abstract from the context. Subjects must then use heuristics and more general problem-solving approaches to try to construct a representation. Subjects could do this reasonably well for problems with concrete scenarios, but showed substantial difficulty when the problem introduced variables or constants (Problems #3, #13, and #15). The more abstract formulation appears to have prevented subjects from employing simple schemata that they otherwise were quite capable of using. For example, although the underlying solution structure is very similar for Problems #1 and #3, subjects had 74% correct on Problem #1 and only 22% on Problem #3. This difficulty seems to stem from two sources, the use of variables *x* and *y* instead of numerical values and the introduction of a time conversion. In fact, the presence of more than one variable in a problem and the presumed need to use algebraic rather than simple arithmetic concepts had a particularly detrimental effect on students' ability to develop a coherent approach to problem solving. The correlation of this problem attribute and the log of the frequency of "other strategies" was .64 ($p<.002$), indicating that the presence of variables led students to use other (mostly ineffective) solution methods.

In addition, subjects appeared to use simulation when the equation solution to a problem would require the use of multiple instances of a variable (Problems #4 and #5) or a variable nested in a multipart equation (Problem #11). Presumably if subjects cannot generate the more complex equation structures, they use a simulation strategy.

Strategies and subject proficiency. Although equation/schema strategies were most prominent, there were individual differences in strategy selection. Most noticeably, the more proficient subjects tended to use the equation strategy more than the less proficient subjects. Use of equation strategies increased with better performance ($r_{pbis} = 0.64$; $t\ (49) = 5.82$, $p<.01$) and with higher SAT-M scores ($r_{pbis} = 0.68$; $t\ (39) = 6.48$, $p<.01$). Ratio and simulation strategies were used roughly equally across ability levels, whereas strategies that could not be categorized (including guessing and trial and error) decreased with performance ($\underline{r}_{pbis} = -0.65$; $\underline{t}$ (49) = 5.96, $\underline{p}<.01$) and SAT-M ($\underline{r}_{pbis} = -0.59$; $\underline{t}$ (39) = 5.09, $\underline{p}<.01$).

Errors

Error Taxonomy. As described above, each solution was individually analyzed and each error was categorized by at least two raters. The taxonomy attempted to provide as much detail as possible without making unnecessary inferences about the subjects' performance. Using this approach, 17 types of errors were identified in the subjects' solutions. These error types were grouped into eight major error categories as shown in Table 8. (No measure of reliability of error types or categories is provided here; the taxonomy is intended only as an initial framework.) In cases in which no specific source could be established for an error, that error was classified as a guess (Table 8: I.D.)

Insert Table 8 about here

Errors and problem attributes. The frequency of occurrence of major error types for each of the 20 problems ordered by problem difficulty is shown in Table 9. Roughly one-third of the errors stem from problems in which subjects fail to develop a representation that captures the problem solution structure; these errors are labeled Problem Conception Errors (Table 9: I). Thirteen percent of these problem conception errors are cases in which the student solves for something other than what was requested in the problem (I.A); almost half of this 13% was associated with one problem (#9). That problem involves three separate devices working independently but simultaneously to drain a tank; many students interpret the problem as three separate devices which are not working together and thus produce three separate values in their answers. Another twenty percent of the errors in this category consist of misusing the givens or in making additional assumptions that are not part of the givens (I.B). Finally, two-thirds of the problem conception errors are instances in which the student did not complete the solution or guessed (I.C. and I.D). In these instances, students have no clear sense of how to proceed. This kind of error is especially evident for problems that (a) utilize variables instead of specific quantities (Problems #13 and #18), (b) have multiple variables in the canonical quantitative representation (#5), or (c) have a deep level of nesting (#19).

Insert Table 9 about here

The relationship between item attributes and problem conception errors was explored through correlational analyses. Log transformations of the error data were performed because of skewed distributions. As might be expected from the analysis of item difficulty, both the algebra and complexity attributes were positively related to the log of the frequency of such errors. However, these two classes of attributes were associated with different classes of conceptual errors. The correlations between the four complexity attributes and "problems-with-givens" errors ranged from .47 to .58 (p's <.05). Problems that are complex, as defined in this study, typically have more givens, and thus more opportunities for information to be misread or misinterpreted. The algebra attribute, however, correlated .62 (p <.05) with failures to complete problems. Again, the need to work with variables rather than quantities appears to have a devastating effect on the ability to develop a coherent approach to problem solution and led to incomplete solutions.

As might be expected, other types of errors were associated with specific item attributes but occurred only when the problem required certain procedures. Thus, weight errors (II.C) occurred only for Problems #13 and #16, clock and elapsed-time errors (II.D) only for #4 and #14, and constants and variable errors (V) almost exclusively for Problems #3, #5, #13, and #15. In addition, almost 90% of conversion errors (VI) were associated with Problems #3, #8, #14, #15, and #16.

In contrast, errors in labeling units (VII) and carrying out procedures (VIII) appeared to be distributed relatively evenly across problems with two exceptions. On Problem #1, subjects often failed to carry out the procedure to provide an answer in "complete steps." On Problem #14, subjects frequently failed to carry out their computations to the required two decimal places, and in this instance, that slip affected the outcome.

<u>Strategies and errors</u>. As shown in Table 7, roughly half of the solutions using any one of the three major strategies (equation, ratio, simulation) resulted in an error, whereas virtually all of the solutions achieved through strategies not falling into one of these groupings produced an error. Most categories of error occurred with each type of solution strategy, although the opportunity for certain errors was substantially less for ratio and simulation strategies because they were used so much less frequently than equations. Schema (II.A), Plan Setup (II.B), and Plan Execution Errors (II.E) were most common in problems that emphasized the

equation strategy. Two categories of error, ratio (III) and simulation (IV), could occur only in those cases in which the related strategy was employed.

Errors and subject proficiency. Each major error category was inversely related to constructed-response performance and SAT-M score as shown in Table 10. The strength of the contributions of each category was determined by a least-squares stepwise regression in which percent correct for 51 subjects and SAT-M for 41 subjects were predicted from the number of errors in each of the eight error categories. Error categories were entered in order of the greatest zero-order correlation with the predicted variable. For constructed-response performance, five of the eight categories (all except ratio setup, simulation/modeling, and units) contributed significantly to the regression, with problem conception contributing the greatest percentage of the variance followed by procedural errors.

Insert Table 10 about here

For SAT-M, only one error category, problem conception, contributed significant variance in the stepwise regression. Problem conception errors decrease gradually with proficiency. Whereas specific procedural errors were significant (but weak) indicators of how an individual did on the constructed-response problems, they did not predict the general quantitative reasoning proficiency measured by SAT-M.

Discussion

Construct validity of algebra word problems

One goal of this research was to gather evidence in support of the construct validity of GRE algebra word problems as measures of quantitative reasoning ability from two perspectives. The first perspective focused on the analysis of problem attributes and the second on the analysis of students' solution strategies and errors. We found that problem attributes such as the need to apply algebraic concepts, problem complexity, and problem content were important determinants of difficulty in this small set of problems. Furthermore, the mathematical complexity of the problems was found to be strongly associated with difficulty whereas linguistic complexity was not.

When solution processes were analyzed, more proficient students tended to show a greater use of systematic equation strategies and a less frequent use of highly idiosyncratic strategies. This may reflect the fact that the more proficient student has better problem-solving techniques, and that proficiency is more likely to result from known techniques than from generation of a completely novel solution strategy.

Performance on constructed-response problems in this study was highly correlated with SAT-M scores, suggesting that our analysis and the standardized test scores are capturing similar underlying abilities. It is therefore plausible that the characteristics we have described can be applied to the analysis of SAT-M scores. Greater proficiency, as measured by SAT-M, resulted in fewer errors in each of the eight error categories analyzed. Both constructed-response performance and SAT-M were most strongly related to differences in the number of conceptual errors. Procedural errors played a lesser role in specifying constructed-response performance and showed no significant relationship with SAT-M. In sum, the major difficulty on these problems appears to be in appropriate conceptualization and planning of a solution, *not* in executing the steps or working out the equations.

These data provide preliminary support for the view that SAT-M represents a measure of quantitative reasoning rather than computational ability. Most of the problems used here, which are reasonably representative of GRE discrete quantitative word problems, can be solved using a limited factual base and a collection of general quantitative problem-solving skills. They do not require retrieval from memory of facts, such as the conversions of kilometers to miles, nor do they demand retrieval of complex formulae. Performance differences among individuals in this context appeared to depend on the ability to represent the problem, to recognize general algorithms, to utilize heuristics, and to construct solutions.

Two other findings illustrate the *psychological* multidimensionality of algebra word problems. First, problem attributes can be combined to predict problem difficulty just as error types can be combined to account for proficiency. Secondly, different problem attributes have differential effects on students' strategies and errors. Thus, construct validity problems might arise when examinees are tested on different sets of problems, as in computer adaptive testing (Embretson & Wetzel, 1987). If one examinee receives problems that are difficult because of the need to formulate an algebraic representation while another examinee is tested on problems that are difficult because of structural complexity, the same abilities may not be brought to bear by both examinees. Finally, these findings suggest how more informative descriptions of student performance might lead to more systematic problem designs. More systematic problem designs would be fostered by a theoretical framework linking problem attributes with individual differences in performance, as we discuss below.

Integration of the theoretical framework with the study results

The data reported here fit reasonably well with a cognitive processing model based on memory search. As students try to construct a solution representation, they follow a limited set of strategies to achieve a successful solution. They may retrieve a previously known "schematic" solution, construct an equation based on problem components, employ a ratio, or generate a solut·.n by simulating changes in various properties over time.

The selection of a strategy appears to depend upon the mapping between the problem stem and representations in memory. Some problems will trigger familiar plans for certain students (e.g., a train catch-up problem as in Problem #4); others will map to a ratio (e.g., when a known distance and time are followed by an unknown distance and a second time as in Problem #1); still others will suggest a simulation (e.g., when two conditions are varying over time until a target is reached, as in Problem #5).

Some of the more successful students write out formulae or immediately retrieve an equation. For some problems, this is a straightforward direct mapping. As problems become more complicated, however, equation specification becomes a multi-step process. Solutions are constructed from partial schemata which are modified to fit the specific problem, and individual equations are constructed to fit meaningful substructures of the problem. For example, for Problem #6, it is possible to write a complete equation of the form "Start Time + (Distance-1 / Rate-1 + (Total Distance - Distance-1) / Rate-2) = Finish Time" and then fill in the values by appropriate substitution, but that approach is rare. Instead, students tend to determine separately the time for the first part of the trip, perhaps by retrieving a generalized $d = rt$ formula and then transforming it to yield $t = d/r$. The second distance is then determined, followed by the time for that distance. The separate times are combined, and finally, the elapsed time is added to the starting clock time. As the number of components to manipulate (conversions, weightings, constants, etc.) increases, students have more difficulty maintaining that information in memory, and, as a consequence, separating out situational components for which they have known procedures. Problem solving appears successful in these circumstances only if the problem can be decomposed into segments that are manageable within memory constraints.

Problems also become more difficult as the solution procedure becomes increasingly abstracted from the problem stem. This abstraction can result from either the situational or the quantitative structure of the problem. Students appear to use their understanding of the problem situation to help mediate the decomposition of the problem and the formation of an associated representation. Some problems do not possess a clear, concrete situational structure; that is, it is difficult to form a specific model of the described situation. This is especially the case with the introduction of constants or variables as in Problems #3 and #15. In those cases, none of the simulation strategies are successful, and the students try a large number of "other" strategies, again almost always unsuccessfully. Students also construct solutions based on a learned set of quantitative structures, and problem difficulty is therefore related to the underlying quantitative structure. In our sample, this structure is especially complicated for problems that require an "algebraic" equation in which a variable is used to represent more than one instance of the concept, as in Problems #4 and #5. In those cases, there are a substantial number of simulations and "other" strategies, presumably because of the complexity of the appropriate equation formulation. Again, the frequency of non-standard strategies and the accompanying errors increase dramatically with problem difficulty.

Since almost none of the "other" strategies generated successful solutions, it appears to be the case that these "other" strategies (including trial and error) are employed primarily when none of the three primary strategies can be used. These "other" strategies frequently involve manipulation of various subparts of the given information in the absence of a coherent solution plan. In a number of cases, subjects make some attempt to solve the problem, but then indicate that they simply do not know how to proceed.

These results are consistent with a model of sequential memory access in which a student retrieves the most specific solution strategy available from memory. If a problem-specific solution is available, it is retrieved and applied. If not, the student may try to reformulate the problem, primarily by decomposition, to match known solution strategies. Of course, even in these retrieval scenarios, the student may need to reconstruct or model the situation in a way that provides an adequate mapping to representations in memory. The more indirect the relationship between the problem statement and the desired solution strategy--that is, the greater the abstraction--the harder the problem.

If these more specific retrieval strategies are unsuccessful, the student may utilize more general simulation or model building procedures. If none of these approaches works, then students attempt to manipulate the given information in order to generate a plausible solution.

Evaluation of the framework and future directions

The cognitive framework presented in this study was useful for identifying, organizing, and evaluating problem attributes and processing characteristics that underlie performance on a set of algebra word problems. Given the increasing emphasis on open-ended problems, it is especially important to establish how such problems measure proficiency. The data here suggest that, for our limited sample, proficiency based on constructed-responses is highly correlated with proficiency based on standardized multiple-choice tests. Similar results have been reported by Bridgeman (1992). At the same time, there are differences in the importance of attributes and types of errors in these contexts. For example, specific problem content (e.g., money) was significantly correlated with constructed-response performance but not with a standardized measure of multiple-choice difficulty (equated delta). Likewise, computational errors played a more substantial role in success on constructed-response problems than they did in SAT-M scores. (These SAT scores were obtained on test forms that included only multiple-choice items.) Whereas constructed responses may provide a better way to assess different aspects of proficiency, such responses may also be more subject to the specifics of item construction. Interestingly, our data support SAT-M as a measure of quantitative reasoning, suggesting that open-ended responses may introduce other cognitive factors into the assessment. Multiple-choice items may reduce the impact of procedural and computational errors and increase the construct validity of the test as a measure of quantitative reasoning (Katz, Friedman, & Bennett, 1993). Expanding on this cognitive analysis will be an important part of any response to the call for more open-ended problems in assessment. As part of that goal, there are many ways in which the current framework can be improved so as to provide a more principled basis for problem development and scoring.

First, it is necessary to examine the attributes studied here in a more systematic way with a larger sample. This examination would require constructing specific items which show systematic variation on the relevant attributes and gathering data to test hypotheses about the effects of those attributes on performance. For example, the presence of a variable in the final answer appeared to be a major factor in problem difficulty, but there were insufficient instances to test that relationship. In other cases, attributes were confounded, so it was difficult to clearly isolate the independent contributions of attributes.

Secondly, it is important to consider the generality of the sources of problem difficulty described here. Our sample included 27% of the total word problems from the disclosed forms of the GRE-Q over a period of five years and 62% of the word problems in the four major categories that we analyzed. It therefore seems reasonable to suggest that our descriptions provide a good characterization of a substantial portion of GRE-Q word problems. It is less clear how these results would generalize to other types of problems. For example, problems that require more algebraic manipulation (e.g., $4 x 2 - 4xy + y^2 = 0$) would be expected to result in more procedural errors. Likewise if the problems were to require specific content knowledge, such as the formula for the area of a circle, then content-based errors would be likely to increase.

Third, the strategies described here account for only a limited portion of the data; they provided only major groupings for the final strategy selected by the students. A more detailed analysis would be needed to describe the patterns of solution construction and repairs that led to success or failure. For example, it is possible that proficiency differences are related to the sequence of strategies employed or the way in which impasses are handled.

Finally, the framework described here needs to be extended in order to provide more detailed understanding of the process of representation and integration. In the present study, the phase of integration and representation played an important role in performance. We suggested that success in this phase reflected individual differences in search and retrieval from long-term memory; those students who have appropriate representations of the relevant type of problem in memory perform better. Thus improved problem representation and integration may be, at least in part, a function of prior experience with relevant problems, and such prior experience may also be responsible for some of the individual differences in SAT-M. Nevertheless, most of the problems we considered did not require any detailed knowledge of facts, and could be solved with limited general problem-solving skills, especially in the absence of any strong time constraints. However, despite the fact that many students appeared to have the requisite quantitative skills, they often failed to use them successfully in certain contexts. To more fully understand these difficulties, we need to develop a clearer understanding of the relationship between underlying mathematical structure, item attributes (e.g., the introduction of variables), and solution difficulty.

One hypothesis is that reasoning occurs within specific domains and that systematic change of problem context will change the way in which problems are addressed. Given a familiar environment, general reasoning resources could be more readily applied. A baseball fan could compute a range of batting averages, but might have difficulty with determining on-time airline records that required similar quantitative reasoning skills. An alternative hypothesis is that there is something about the abstractive properties independent of domain that makes certain problems hard. Similar probability problems should be approximately equivalent in difficulty, according to this view. In order to compare these alternatives, it would be necessary to define a set of problems with similar underlying quantitative structures, but variable contexts.

A structural representational system might be particularly useful (cf. Hall et al., 1989; Shalin & Bee, 1944) as a principled basis for designing problems and for analyzing responses. Problems can be designed so that structure varies within content area and is made comparable across different contexts (concrete as well as abstract). Then, the effects of these dimensions can be evaluated independently. Student responses could be analyzed in terms of their mastery of elemental structures and their ability to integrate elemental structures into various higher-order ones in different contexts. Student errors could be taxonomized with respect to the level of structure at which errors occur. Other important aspects of the analyses of student responses would involve describing the strategies they use to represent problems in different contexts, their flexibility in applying different strategies, and the efficiency of the strategies and problem decompositions they use.

References

Anderson, J.R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Barnett, J. (1984). The study of syntax variables. In G. A. Golden & C. E. McClintock (Eds.), Task variables in mathematical problem solving, (pp. 23-68). Philadelphia, PA: Franklin Press.

Branca, N. A., Adams, V. M., & Silver, E. A. (1980). Problem solving processes of ten and eleven year olds. Berkeley, CA: University of California Press.

Briars, D. J. (1983). An information-processing analysis of mathematical ability. In Individual differences in cognition (pp. 181-204). Academic Press.

Bridgeman, B. (1992). A comparison of quantitative questions in open-ended and multiple-choice formats. Journal of Educational Measurement, 29, 253-271.

Card, S. K., Moran, T. P., & Newell, A. (1983). The psychology of human computer interaction. Hillsdale, NJ: Lawrence Erlbaum Associates.

College Board. (1987). 1987 Profile of SAT and Achievement Test Takers. New York: The College Board.

Embretson, S. (Whitely). (1983). Construct validity: Construct representation versus nomothetic span. Psychological Bulletin, 93, 179-197.

Embretson, S. E. & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension tests. Applied Psychological Measurement, 11, 175-193.

Fuson, K. C. (1982). An analysis of the counting-on solution procedure in addition. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), Addition and subtraction: A cognitive perspective (pp. 67-81). Hillsdale, NJ: Lawrence Erlbaum Associates.

Glaser, R. (1991). Expertise and assessment. In M. C. Wittrock & E. L. Baker (Eds.), Testing and cognition (pp. 17-30). Englewood Cliffs, NJ: Prentice Hall.

Goldin, G. A. & McClintock, C. E. (Eds.). (1984). Task variables in mathematical problem solving. Philadelphia, PA: Franklin Institute Press.

Hall, R., Kibler, D., Wenger, E., & Truxaw, C. (1989). Exploring the episodic structure of algebra story problem solving. Cognition and Instruction, 223-283.

Hiebert, J., Carpenter, T. P., & Moser, J. M. (1982). Cognitive development and children's solutions to verbal arithmetic problems. Journal for Research in Mathematics Education, 83-98.

Hinsley, D. A., Hayes, J. R., & Simon, H. A. (1977). From words to equations--Meaning and representation in algebra word problems. In M. A. Just & P. A. Carpenter (Eds.), Cognitive processes in comprehension (pp. 89-106). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Katz, I. R., Friedman, D. E., & Bennett, R. E. (1993). Examinees' strategy differences in solving stem-equivalent constructed-response versus multiple-choice SAT-Mathemtical items. Princeton, NJ: Educational Testing Service.

Krutetskii, V. A. (1976). The psychology of mathematical abilities in schoolchildren. Chicago: University of Chicago Press.

Lester, F. K. (1983). Trends and issues in mathematical problem solving research. In R. Lesh & M. Landau (Eds.), Acquisition of mathematics concepts and processes (pp. 229-261). New York: Academic Press.

Loftus, E. F., & Suppes, P. (1972). Structural variables that determine problem solving difficulty in computer-assisted instruction. Journal of Educational Psychology, 63, 521-42.

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

Mayer, R. E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories, and templates. Instructional Science, 10, 135-175.

Mayer, R. E., Larkin, J. H., & Kadane, J. B. (1984). A cognitive analysis of mathematical problem-solving ability. In R. J. Sternberg (Eds.), Advances in the psychology of human intelligence (pp. 231-273). Hillsdale, NJ: Lawrence Erlbaum Associates.

Messick, S. (1989). Validity. In Robert L. Linn (Ed.), Educational Measurement (3rd ed.), (pp. 13-104). New York: American Council on Education/Macmillan.

Miller, G. A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing informatior. Psychological Review, 63, 81-97.

National Council of Teachers of Mathematics (1989). Curriculum and Evaluation Standards for School Mathematics, Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics (1988). The ideas of algebra, K-12. Reston, VA: National Council of Teachers of Mathematics.

Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.

Paige, J. M., & Simon, H. A. (1966). Cognitive processes in solving algebra word problems. In B. Kleinmutz (Eds.), Problem solving (pp. 51-119). New York: Wiley.

Polya, G. (1973). Induction and analogy in mathematics. Princeton, NJ: Princeton University Press.

Reed, S. K. (1987). A structure-mapping model for word problems. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 124-139.

Reed, S. K., Demptser, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. Journal of Experimental Psychology, Learning, Memory, and Cognition, 11, 106-125.

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

Scheuneman, J. & Grima, A. (in preparation). Characteristics of quantitative items associated with differential performance for female and black examinees. Princeton, NJ: Educational Testing Service.

Schoenfeld, A. H. (1987). Cognitive science and mathematics education: An overview. In A. H. Schoenfeld (Eds.), Cognitive science and mathematics education (pp. 1-31). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schoenfeld, A. H. (1979). Explicit heuristic training as a variable in problem-solving performance. Journal for Research in Mathematics Education, 173-187.

Shalin, V., & Bee, N. V. (1985). Analysis of the semantic structure of a domain of word problems (Technical Report No. APS-20). Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Silver, E. A. (1979). Student perceptions of relatedness among mathematical verbal problems. Journal for Research in Mathematics Education, 10, 195-210.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In Robert L. Linn (Ed.), Educational measurement (3rd ed.) (pp. 263-331). New York: American Council on Education/Macmillan.

Spiegel, M. R., & Bryant, N. D. (1978). Is speed of processing information related to intelligence and achievement? Journal of Educational Psychology, 904-910.

VanLehn, K. (1988). Toward a theory of impasse-driven learning. In H. Mandl & A. Lesgold (Eds.), Learning issues for intelligent tutoring systems (pp. 19-41). New York: Springer Verlag.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), Foundations of cognitive science (pp. 527-579). Cambridge, MA: MIT Press.

Table 1

Problem Descriptions

Problem #	Descriptive Key	Test Development Classification	Equated Delta
1	16 Steps	Dist. = Rate x Time	8.20
2	Train Return	Dist. = Rate x Time	10.80
3	*xy* Rate	Dist. = Rate x Time	13.70
4	Train Catch-up	Dist. = Rate x Time	15.40
5	8/10% Invest	Interest	14.30
6	$750 Interest	Interest	13.00
7	Down Payment	Interest	10.40
8	Brads	Work	12.00
9	"R,S,T Machines"	Work	13.80
10	Lawyer	Graduated Rate	9.90
11	Bricklayer	Graduated Rate	12.70
12	Electricity	Graduated Rate	9.80
13	*h* Hour Trip	Dist. = Rate x Time	15.60
14	600 Mile	Dist. = Rate x Time	13.50
15	*k* Kilometers	Dist. = Rate x Time	16.00
16	Bike Ride	Dist. = Rate x Time	16.60
17	4 Machines	Work	14.80
18	Workers V & W	Work	15.40
19	Carpenter	Work	19.80
20	Drain Tank	Work	13.80

Note. Equated Delta is a linear transformation of percent correct with a mean of 13 and a standard deviation of 4.

Table 2

Attributes of Algebra Word Problems

Attributes Based on the Problem Statement

Givens -
the number of quantities, factors, and variables that were explicit in the problem statement

Numeric and content characteristics -
whether the problem statement included whole numbers, fractions, decimals, percents, rates, measures of time, distance, volume, or money

Linguistic characteristics -
counts of the number of arguments, predicates and modifiers in the problem statement as well as whether any relational words that express a quantitative relation between quantities or variables such as twice, older than, later than, appeared in the problem statement

Distracting information -
presence of letter labels (Train T) that might be confused with variable labels, or of quantitative information that is irrelevant to the solution of the problem

Attributes Based on Problem Representation

Equation-based Attributes

Implicit values-
values necessary to represent the problem as an equation that were not given in the problem statement

Variables in the equation -
the number of times variables appear in the equation

Operations -
the number of mathematical operations in the equation

Conversions -
whether scale conversions such as minutes to hours or percent to decimals or fractions are required

Nest level -
the number of levels of parentheses and/or the number of equations in the representation

Table 2 con't

Attributes of Algebra Word Problems

Answer form -
whether expected answer was a quantity or an expression containing a variable

Structural Attributes (based on Shalin & Bee, 1985)

Elementary structures -
the number of lower-order groupings of quantities and/or variables and whether they involved additive relations among extensives (primary quantities), additive relations among intensives (rates), multiplicative relations among extensives and intensives, and multiplicative relations among extensives and factors

Higher-order structures -
the number of higher-order relations among elementary structures and whether they involved a hierarchy, a shared whole, a shared part, or a shared rate (a special case of a shared part)

Table 3

Summary of Correct Responses by Problem Across Subjects

Problem #	Number of Students Attempting a Solution	Number of Students with Correct Solutions	Percent of Students with Correct Solutions
1	50	37	74
2	51	26	51
3	50	11	22
4	51	17	33
5	51	19	37
6	49	33	67
7	51	42	82
8	51	29	57
9	51	24	47
10	51	38	75
11	50	30	60
12	51	38	75
13	51	5	10
14	51	7	14
15	51	6	12
16	51	5	10
17	51	15	29
18	51	11	22
19	50	3	06
20	51	25	49

Table 4

Correlations Between Problem Attributes and

Measures of Problem Difficulty (n=20)

	Examinee Sample	
Attributes	College Sample	GRE Samples
Algebra		
More than 1 variable in the representation	-.54*	.42+
Complexity		
Level of nesting	-.40+	.60**
No. of operations in the representation	-.31	.46*
No. of elemental structures	-.22	.34
No. of higher order structures	-.27	.41+
Content		
Time conversions	-.39+	.14
Metric measures	-.38+	.35
Money	.57**	-.36
Linguistic		
Relational Word	-.08	.28
No. of arguments	-.17	.14
No. of predicates	-.24	.32
No. of connectives	-.10	.16
No. of modifiers	-.22	.32

Note. Performance in the college sample is expressed as percent correct. The GRE samples' performance is in terms of equated delta.

$^{**}p<.01$. $^{*}p<.05$. $^{+}p<.10$.

Table 5

Prediction of Problem Difficulty (Percent Correct) for College Sample:

Estimated Regression Parameters and R^2 Values (n=20)

	Alternative Models			
	1	2	3	4
Intercept	56.16	71.48	62.97	49.17
Beta for Attributes				
More than 1 variable	-.48*	-.39*	-.38+	-.42**
Level of nesting	-.32+	-.50**	-.37+	-.36*
Time conversions		-.50**		
Metric measures			-.30	
Money				.54**
df	(2,17)	(3,16)	(3,16)	(3,16)
R^2	.39	.60	.47	.68
Adjusted R^2	.32	.54	.37	.62
Significance of change in R^2		<.01	n.s.	<.001

<u>Note</u>. Adjusted R^2 is corrected for the number of variables in the model.

*p<.05. **p<.01. +p<.10.

Table 6

Prediction of Problem Difficulty (Equated Delta) for GRE Sample:

Estimated Regression Parameters and R^2 Values (n=20)

	Alternative Models			
	1	2	3	4
Intercept	11.19	10.34	10.73	11.92
Beta for Attributes				
More than 1 variable	.33+	.28	.22	.29+
Level of nesting	.55**	.60**	.60**	.57**
Time conversions		.32+		
Metric measures			.34+	
Money				-.36*
df	(2,17)	(3,16)	(3,16)	(3,16)
R^2	.47	.56	.58	.60
Adjusted R^2	.41	.48	.50	.53
Significance of change in R^2		<.10	<.10	<.05

Note. Adjusted R^2 is corrected for the number of variables in the model.

$^*p<.05$. $^{**}p<.01$. $^+p<.10$.

Table 7

Percent of Correct and Incorrect Solutions for Each Strategy Category

	CORRECT					INCORRECT						OVERALL					
Problem #	E	R	S	O	TOT	E	R	S	O	N	TOT	E	R	S	O	N	n
1	30	42	0	2	74	12	12	0	2	0	26	42	54	0	4	0	50
2	47	4	0	0	51	20	22	6	2	0	49	67	25	6	2	0	51
3	14	6	0	2	22	42	8	4	24	0	78	56	14	4	26	0	50
4	22	0	12	0	33	24	2	20	20	2	67	45	2	31	20	2	51
5	18	0	20	0	37	16	0	20	24	4	63	33	0	39	24	4	51
6	61	2	4	0	67	16	2	0	10	4	33	78	4	4	10	4	49
7	78	4	0	0	82	14	2	0	2	0	18	92	6	0	2	0	51
8	53	4	0	0	57	27	14	2	0	0	43	80	18	2	0	0	51
9	37	10	0	0	47	39	6	0	6	2	53	76	16	0	6	2	51
10	65	0	10	0	75	22	0	4	0	0	25	86	0	14	0	0	51
11	38	0	22	0	60	16	0	22	2	0	40	54	0	44	2	0	50
12	75	0	0	0	75	12	8	0	6	0	25	86	8	0	6	0	51
13	10	0	0	0	10	49	0	2	39	0	90	59	0	2	39	0	51
14	14	0	0	0	14	82	0	0	4	0	86	96	0	0	4	0	51
15	8	4	0	0	12	39	18	2	25	4	88	47	22	2	25	4	51
16	10	0	0	0	10	67	2	0	22	0	90	76	2	0	22	0	51
17	29	0	0	0	29	47	0	0	14	10	71	76	0	0	14	10	51
18	16	0	4	2	22	22	0	0	53	4	78	37	0	4	55	4	51
19	4	2	0	0	6	10	4	0	74	6	94	14	6	0	74	6	50
20	47	2	0	0	49	45	0	0	4	2	51	92	2	0	4	2	51
Column Means	34	4	4	0	42	31	5	4	17	2	58	65	9	8	17	2	

E - Equation/Schema R - Ratio S - Simulation/Modeling O - Other/Unidentified N - No Work/Skip

Table 8

List of Principal Error Types

I. Problem Conception

A. Misconceptions of Problem Objective

- Solutions provide information other than what is requested in the problem statement.

- Example: Problem #9. The student indicates how much water will be drained rather than how much will be left.

B. Problems with Givens

- Information concerning the given information in the problem is misread or misinterpreted. Includes making assumptions not in the problem statement.

- Example. Problem #13. "The first third...took twice as long" is interpreted to mean that each remaining third took twice as long, or equivalently that the first third took "half" as long.

C. Failure to Complete

- Work is terminated before a solution is reached, taken to indicate that the student either misconstrues the problem or is unable to generate a strategy for continuing.

D. Guess

- The student appears to have no specific direction; the failure to achieve a correct solution could not be ascribed to any other specific error type. In some instances students write that they are making a guess.

II. Equation Setup

A. Scheme Failure

- Solutions include use of the correct concepts, but in incorrect relations. Major equations include $d=rt$, $i=rp$, $balance=rp-deposit$. Errors also include equations that demonstrate misconceptions of "rate" and failing to differentiate events that are concurrent and sequential.

- Example: Problem #4. Rather than setting equal two simultaneous distances ($r1t1 = r2t2$), the distances are added ($r1t1 + r2t2 = x$).

Table 8 con't

B. Plan Setup Errors

- Errors in formulating a set of equations that reflect an error in correct structure although they do not appear to be a part of a coherent schema. This includes a few instances of misconceptions about equations (e.g. two unknowns in one equation), confusing rate with time, incorrectly distributing partial values, and confusing concurrent and successive work episodes.

- Example. Problem #19. "If A would take 6 days to complete a job and A and B take 4 days together, then B takes 1/2 of 4 or 2 days, which is 1/3rd the rate of A.

C. "Weight" Errors

- Either the rate or time associated with some events is not appropriately weighted. Subjects frequently assume that a simple average of two rates will give the appropriate overall rate regardless of the time at each rate.

- Example. Problem #16. If the rate to school is 0.12 kilometers per minute and the return trip rate is 0.24 kpm, then the average speed is (0.12 + 0.24)/2 = 0.18 kpm.

D. Clock and Elapsed Time

- These are difficulties with the use of clock time or the treatment of clock and elapsed time as the same.

- Example. Problem #4. Answer given as relative rather than clock time.

E. Plan Execution Errors

- These are errors that reflect a correct plan but a failure in some part of its implementation. The errors, however, are not merely computational.

- Example. Problem #17. Provide an answer using only 1 machine instead of 4.

III. Ratio Setup

- Solutions use matching of units (a/b = c/d), but misorder the terms or include incorrect items. Common errors, include problems with the need to invert the labels for ratio equivalence.

- Example. Problem 2: Subjects set up the equivalence 60 mph/3.5 hr = 50 mph/x hr instead of the correct version, 60/50 = x/3.5.

Table 8 con't

IV. Simulation/Modeling

- Solutions "model" the situation typically by an explicit or implicit table indicating incremental changes in the state of problem variables. Errors resulted from incorrect estimates of steps in the simulation or matching of incompatible units.

- Example. Problem 2:
   50 miles ->1 hour;
  +50 miles ->2 hours;
  +50 miles ->3 hours;
  +50 miles ->4 hours;
  +10 miles ->4 hours 10 minutes

V. Constants and Variables

A. "Constant" Errors

- Constants are ignored, given a specific unjustified value, or used in an inappropriate way.

- Example. Problem #3. The answer is given as "x/60" without inclusion of "y."

B. "Variable" Errors

- Variables are ignored, the same variable is used for more than one meaning, or unnecessary variables are added.

- Example. Problem #5. The variable 'x' is used to represent both the amount at 10% and the amount at 8%: ".08x + .10x = $870."

VI. Conversion

- Units are transformed improperly. This occurs most commonly with time units, but also includes percent to decimal conversion.

- Example: Problem 2. "4.10 hrs" is converted to "4 hours 10 minutes."

VII. Units

- Units are ignored, improperly combined, or misapplied. In some cases, a solution has the correct value, but then is given an incorrect label.

Table 8 con't

VIII. Procedural Difficulties

A. Computational

- Errors in arithmetic operations of addition, subtraction, multiplication, and division as well as errors in precision.

B. Transcription

- Errors are made in copying information from one part of the problem to the next.

Table 9

Frequency of Each Type of Error Across Subjects for Problems Ordered by Percent Correct (n = 51)

Error Category	Problem Number																				Cumulative Frequency
	7	10	12	1	6	11	8	2	20	9	5	4	17	3	18	14	15	13	16	19	
I. Problem Conception	1	9	8	4	7	4	4	1	7	25	18	19	6	16	25	8	19	29	13	31	254
A. Misconceive Objective								1		16	1		1	2	3		9	1			34
B. Problems With Givens		9		2		2	1		1	4		3	3	2	1	1	1	7	4	9	50
C. Failure to Complete			3	2	4	1			2	3	12	8	2	9	8	6	7	17	6	8	98
D. Guess	1		5		3	1	3		4	2	5	8		3	13	1	2	4	3	14	72
II. Equation Setup	6			2	5	3	1	4	10	2	1	16	25	4	12	13	8	14	38	10	174
A. Schema Problems	6			2	5			4	10			4		4		3	8	5	19		70
B. Plan Setup Errors										2	1	2	1		12			2		10	30
C. "Weight" Errors																		7	19		26
D. Clock and Elapsed Time												10				10					20
E. Plan Execution Errors						3	1						24								28
III. Ratio Setup			4	1			2	15		1			1				2				26
IV. Simulation/Modelling		1				7		3				5									16
V. Constants and Variable						1		1	1		13	1		14			15	11			57
A. "Constant" Errors								1						13			15	11			40
B. Variable						1			1		13	1		1							17
VI. Conversion	2				1	2	13	6						16		26	20		12		98
VII. Units				1		2	1	2					3	2	2		4		4		21
VIII. Procedural		3	3	9	5	6	10	7	7	3	1	2	3		2	30	6	2	14		113
A. Computational		3	2	8	5	5	10	5	7	2	1	2	3		2	26	6	2	13		102
B. Transcription			1	1		1		2		1						4			1		11
Column Totals	9	13	15	17	18	25	31	39	25	31	33	43	38	52	41	77	74	56	81	41	759
Percent Correct	82	75	75	74	67	60	57	51	49	47	37	33	29	22	22	14	12	10	10	6	

Table 10

Relation of Error Categories to Performance on this Sample Set and Proficiency as Measured by SAT-M

	Constructed-Response Performance			SAT-M Proficiency		
Error Category	Simple Correlation	R-Squared	BETA	Simple Correlation	R-squared	BETA
I. Problem Conception	-0.72	0.51	-0.70**	-0.68	0.45	-0.61**
II. Equation Setup	-0.39	0.14	-0.23**	-0.15	-0.004	-0.11
III. Ratio Setup	-0.37	0.12	-0.08	-0.40	0.14	-0.17
IV. Simulation/Modeling	-0.30	0.07	-0.09	-0.43	0.17	-0.19
V. Constants and Variable	-0.21	0.02	-0.16**	-0.08	-0.02	-0.07
VI. Conversion	-0.34	0.10	-0.22**	-0.31	0.07	-0.11
VII. Units	-0.30	0.07	-0.07	-0.24	0.03	-0.01
VIII. Procedural	-0.30	0.07	-0.24**	-0.04	-0.02	-0.07
Multiple Regression		0.80			0.45	
All Eight Categories		0.82			0.53	

Note. R^2 for individual categories are from Simple Least-Square Linear Regression. R^2 for the multiple regression, adjusted for the number of variables in the model, and Beta coefficients are from Stepwise Multiple Regression.

$**p<.01$

Figure 1
Structural Representations for Algebra Word Problems

Problem 5, Equated Delta = 14.3

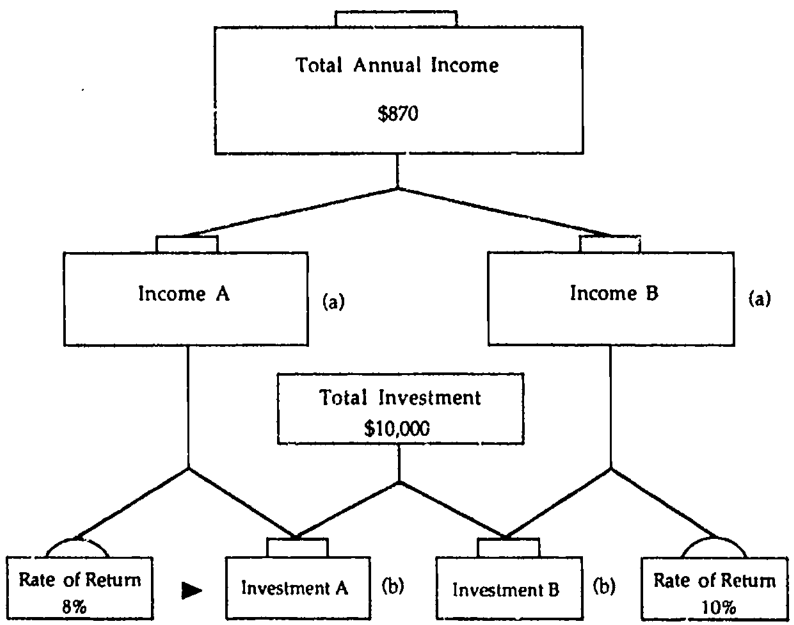
Text

A person invests $10,000, some at 8 percent and some at 10 percent per year. The annual income from this investment is $870. How much was invested at 8 percent?

Equation-based representation

$(8\%)x + (10\%)(\$10{,}000 - x) = \$870$

Structural representation (based on Shalin & Bee, 1985)



[bracket] = Extensives
[arc] = Intensives
(a) = Heirarchical Link
(b) = Shared Part Link
▶ = Goal

Problem 6, Equated Delta = 13.0

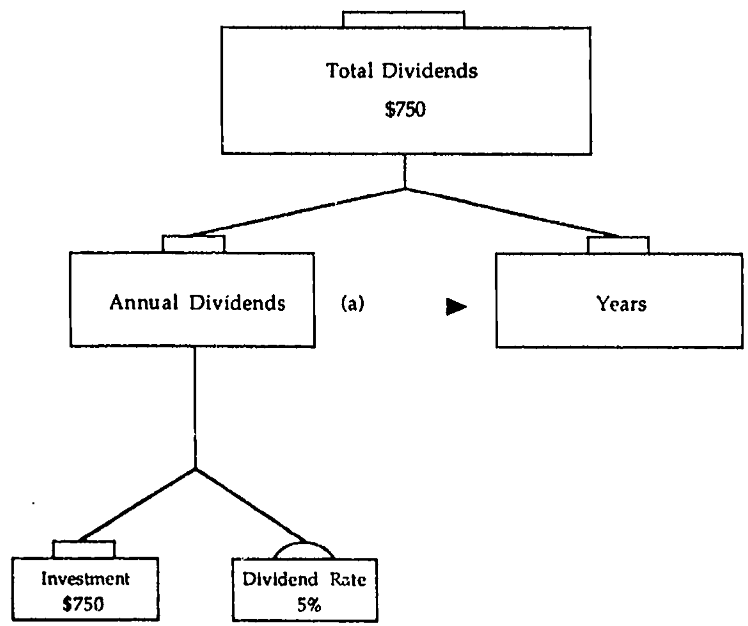
Text

Money in a certain investment fund earns an annual dividend of 5 percent of the original investment. In how many years will an initial investment of $750 earn total dividends equal to the original investment?

Equation-based representation

$$\frac{750}{750 \times (5\%)} = x$$

Structural representation (based on Shalin & Bee, 1985)



[bracket] = Extensives
[arc] = Intensives
(a) = Heirarchical Link
▶ = Goal

Appendix A

GRE Algebra Word Problems and

Equation-based Representations

Appendix A
GRE Algebra Word Problems and Equation-based Representations

1. When walking, a certain person takes 16 complete steps in 10 seconds. At this rate, how many complete steps does the person take in 72 seconds?

$$\frac{16 \text{ steps} \times 72 \text{ sec}}{10 \text{ sec}} = x \text{ steps}$$

2. A train travels from City X to City Y in 3 hours and 30 minutes at an average speed of 60 miles per hour. If the train returns at an average speed of 50 miles per hour, how long does the return trip take?

$$\frac{\{3' 30''\}^{*} \times 60 \text{ mph}}{50 \text{ mph}} = x \text{ hrs}$$

3. If a certain object has been moving at the constant rate of x meters per minute, how many meters has the object moved in the last y seconds?

$$\{x \text{ mpm}\} \times y \text{ sec} = z \text{ meters}$$

4. At 9:00 a.m. train T left the train station and two hours later train S left the same station on a parallel track. If train T averaged 60 kilometers per hour and train S averaged 75 kilometers per hour until S passed T, at what time did S pass T?

$$60 \text{ mph} \times t \text{ hrs} = 75 \text{ mph} \times (t - 2) \text{ hrs}$$
$$9 \text{ am} + t \text{ hrs} = \{x\} \text{ o'clock}$$

5. A person invests $10,000, some at 8 percent per year and some at 10 percent per year. The annual income from this investment is $870. How much was invested at 8 percent?

$$\{8\%\}\ \$x + \{10\%\}\ (\$10{,}000 - \$x) = \$870$$

6. Money in a certain investment fund earns an annual dividend of 5 percent of the original investment. In how many years will an initial investment of $750 earn total dividends equal to the original investments?

$$\frac{\$750}{\$750 \times \{5\%\}} = x \text{ years}$$

7. A buyer must make a 15 percent down payment on the house she is purchasing for $32,000. If she has already made a $500 deposit, how much more will she need for the down payment?

$$(\{15\%\} \times \$32{,}000) - \$500 = \$x$$

*Brackets { } are used to indicate the need for a conversion such as between scales of time or percents to decimals/fractions.

8. At the rate of 36 brads per 10 seconds, how many brads does a machine produce per hour?

$$\frac{36 \text{ brads} \times \{\{1 \text{ hr}\}\}}{10 \text{ sec}} = x \text{ brads}$$

9. Three devices, R, S, and T, each working by itself at a constant rate, require 4, 8, and 12 hours, respectively, to drain 240 gallons of water. If there are 240 gallons of water in a tank, how many gallons of water will be left in the tank after each of the three devices has worked by itself for exactly 1/2 hour?

$$240 \text{ gal} - \tfrac{1}{2} \text{ hr} \left(\frac{240 \text{ gal}}{4 \text{ hr}} + \frac{240 \text{ gal}}{8 \text{ hr}} + \frac{240 \text{ gal}}{12 \text{ hr}}\right) = x \text{ gal}$$

10. A lawyer charges $100 for the first hour of service and $75 for each additional hour. A bill of $625 represents how many hours of the lawyer's services?

$$\$100 + \$75(x \text{ hrs} - 1) = \$625$$

11. A bricklayer receives $6 per hour for a 7-hour day and 1 1/2 times his regular hourly rate of pay for time in excess of 7 hours during a single day. If he received $54 for a single day's work, how long had he worked that day?

$$(\$6 \times 7 \text{ hr}) + (1\tfrac{1}{2} \times \$6 \times (x - 7)\text{hrs}) = \$54$$

12. Electricity to operate a window air conditioner costs 3 cents per hour and to operate an attic fan 1.2 cents per hour. How many hours would the attic fan have to be in operation for the cost to be the same as the cost to operate the air conditioner for 8 hours?

$$\frac{3 \text{ ¢ph} \times 8 \text{ hr}}{1.2 \text{ ¢ph}} = x \text{ hrs}$$

13. The first third of a 75-mile trip took twice as long as the rest of the trip. If the first third took h hours, what was the average speed, in miles per hour, for the whole trip?

$$2t = h \text{ hrs}$$

$$\frac{75 \text{ miles}}{(h \text{ hrs} + t)} = x \text{ mph}$$

14. On a 600-mile motor trip, Bill averaged 45 miles per hour for the first 285 miles and 50 miles per hour for the remainder of the trip. If he started at 7:00 a.m., at what time did he finish the trip?

$$7 \text{ am} + \frac{285 \text{ miles}}{45 \text{ mph}} + \frac{(600 - 285) \text{ miles}}{50 \text{ mph}} = \{x\} \text{ o'clock}$$

15. A car is travelling at an average speed of 80 kilometers per hour. On the average, how many <u>seconds</u> does it take the car to travel k kilometers?

$$\left\{\left\{\frac{k \text{ km}}{80 \text{ kph}}\right\}\right\} = x \text{ sec}$$

16. Jane takes 50 minutes to ride her bike from home to school. She takes 25 minutes to ride her bike home along the same route. It is 6 kilometers in each direction. What is her <u>average speed</u> in kilometers per hour for the round trip?

$$\frac{(6 + 6) \text{ km}}{\{(50 + 25) \text{ min}\}} = x \text{ kmph}$$

17. The 4 machines at company K operated simultaneously and at a constant rate r to fill 5/8 of a production order for wrenches. The remaining wrenches needed to fill the order were produced in 3 hours with all 4 machines running at the same constant rate r. The cost of operating each machine for one hour was \$22.00. What was the total cost of machine operation for this order?

$$\frac{(1 - 5/8) \text{ order}}{3 \text{ hrs}} = \frac{1 \text{ order}}{t \text{ hrs}}$$

$$\$22 \text{ per hr per machine} \times 4 \text{ machines} \times t \text{ hrs} = \$x$$

18. Worker W produces n units in 5 hours. Workers V and W, working independently but at the same time, produce n units in 2 hours. How long would it take V alone to produce n units?

$$n \text{ units} \,/\, \left(\frac{n \text{ units}}{2 \text{ hrs}} - \frac{n \text{ units}}{5 \text{ hrs}}\right) = x \text{ hrs}$$

19. A carpenter worked alone for 1 day on a job that would take him 6 more days to finish. He and another carpenter completed the job in 4 more days. How many days would it have taken the second carpenter to do the complete job working alone?

$$\left(1 \text{ job} - \left((4 + 1) \text{ days} \times \frac{1 \text{ job}}{(6 + 1) \text{ days}}\right)\right) \times \frac{1 \text{ job}}{x \text{ days}} = 4 \text{ days}$$

20. How many minutes will it take to fill a 2,000-cubic-centimeter tank if water flows into the tank at the rate of 20 cubic centimeters per minute and is pumped out at the rate of 4 cubic centimeters per minute?

$$\frac{2000 \text{ cc}}{(20 - 4) \text{ ccpm}} = x \text{ min}$$

Appendix B

Coding of Attributes for Problems in Set

Appendix B

Coding of Attributes for Problems in Set

	Problem																			
Attributes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
<u>Algebra</u>																				
More than 1 variable in the representation	0	0	1	1	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0
<u>Complexity</u>																				
Level of nesting	0	0	0	2	1	0	1	0	1	1	2	0	2	1	0	1	2	1	3	1
No. of operations in the representation	2	2	1	4	4	2	2	2	7	3	5	2	3	5	1	3	5	4	7	2
No. of elemental structures	2	2	1	4	4	2	2	2	6	4	5	2	3	5	1	3	5	4	5	2
No. of higher order structures	1	1	0	3	4	1	1	1	6	4	6	1	3	5	0	2	5	4	7	1
<u>Content</u>																				
Time conversions	0	1	1	1	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0
Metric measures	0	0	1	1	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1
Money	0	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	1	0	0	0
<u>Linguistic</u>																				
Relational Word	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0
No. of arguments	4	11	7	10	6	8	6	5	9	6	8	8	8	6	7	11	11	7	4	6
No. of predicates	4	8	4	9	4	4	6	3	8	6	7	8	6	5	5	10	14	8	8	7
No. of connections	3	9	4	11	7	5	6	3	15	4	9	6	6	10	3	6	10	5	6	8
No. of modifiers	3	3	3	3	1	7	0	0	3	2	5	3	5	3	1	4	8	3	5	4