Rule Induction on Word Problems as a

Function of Learner Characteristics

and Task Variables

Nancy Steacy
B.A., University of British Columbia, 1974

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of

Psychological Foundations

ACCEPTED
FACULTY OF GRADUATE STUDIES
[redacted]
DEAN
DATE October 6, 1987

We accept this thesis as conforming
to the required standard

[redacted]

Daniel G. Bachor, Ph.D.

[redacted]

Beverly A. Timmons, Ed.D.

[redacted]

James H. Vance, Ph.D.

[redacted]

Lloyd O. Ollila, Ph.D.

UNIVERSITY OF VICTORIA

July 1987

Supervisor: Daniel G. Bachor, Ph.D.

ABSTRACT

The rule induction rate and ability of Grade 6 students on word problems were assessed. Two classes (N = 57) in a suburban elementary school were administered 144 word problems in three sets. The problems were all isomorphic to each other and, if recognized as such, could be solved using the same procedural rule. The problems varied in whether or not they contained extraneous information and, when present, in the location of that information within the problem context. They also varied in the complexity of the sets involved and in the language used to establish that complexity. Lastly, the problems contained all real words, nonsense nouns, or nonsense nouns and verbs as a way of manipulating the level of language familiarity.

The subjects were grouped according to their performance on measures of three learner characteristics, academic self-concept, test anxiety, and memory span, and according to gender for one comparison. Subjects performed differentially on the total problem set in terms of both rule induction rate and ability. Individual performances over the three problem sets were observed to follow one of three different patterns wherein performance either was maintained at an uniformly high level, improved dramatically, or was uniformly low. The influence of learner characteristics seems to have over-powered any effects of the task variable manipulations. Two separate performance groups were

differentiated by the memory span measure. These groups remained distinct regardless of changes in the task. The word problem performance of subjects with low academic self-concept was almost always differentiated from that of all other subjects. The performances of high and low test anxiety groups were sometimes distinct, though neither were ever significantly different from a middle group.

Finally, memory span was a reasonable predictor of rule induction rate, and consequently, its predictive power decreased over the three problem sets. Academic self-concept was a moderate predictor of rule induction ability over the three trial tasks. Level of test anxiety was essentially unrelated to rule induction ability or rate on the isomorphic word problems.

Examiners:

[redacted]

Daniel G. Bachor, Ph.D.

[redacted]

Beverly A. Timmons, Ed.D.

[redacted]

James H. Vance, Ph.D.

[redacted]

Lloyd O. Ollila, Ph.D.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I wish to acknowledge the continuing support of my thesis committee, especially my supervisor, Dan Bachor.

I am also indebted to the teachers and students of Fairburn Elementary School who co-operated in the research study.

I want to extend my gratitude to my family, Robert, Erin, and Sean, who tolerated with good humour my preoccupation with my work.

Finally, I would like to thank my parents who taught me that learning is its own reward.

Chapter 1

INTRODUCTION TO THE PROBLEM

"Problem solving" is a catch phrase in education in the 1980's and is often heralded as one of the most important goals of education (Branca, 1980; Cyert, 1980). This recognition accorded problem solving is not new (Dewey, 1910; Polya, 1945, 1957; Thorndike, 1922), although its prominence in the post "back-to-the-basics" era is being asserted with renewed vigour across curricula, but particularly in the area of instruction and learning in mathematics. Whereas there is a general unanimity regarding the importance of problem solving in pedagogy, a universally accepted definition of what constitutes problem solving is not so readily forthcoming nor is a consensus on how to teach it. Similarly, a comprehensive and cohesive theory (or theories) of problem solving on which to base instruction and curricula also is lacking. The problem solving research, while voluminous, is very much an amorphous mass, some might even say "mess" (Lester, 1980, 1983; Reed, 1982; Suydam, 1980).

A large portion of the problem solving research has employed "laboratory" tasks as opposed to "classroom" tasks. Research using the latter, while less frequent, has ranged across school subject areas with much of it having had a mathematical focus. Within this mathematical context one of the classroom-like tasks examined most often has been the infamous word or story problem, anathema to student and teacher alike (Blakenship & Lovitt, 1976; Hoy, 1984; Sharma, 1981).

Problems, problem solving, and rule induction

Researchers in both the classroom and clinical setting, employing math word problems or otherwise, have recognized that whether a specific task actually constitutes a problem is particular to the individual attempting the task, and that its status can change for that individual over time. Dewey (1910) suggested early in the history of problem solving research that a problem is an individual-specific situation initially causing "perplexity" or "confusion" and containing elements which were unanticipated or unfamiliar. A recently elaborated definition by Lester (1983) illustrates the complex nature of what constitutes a problem for a given individual or group at a given time.

> A "problem" is a task for which:
>
> 1. the individual or group confronting it wants or needs to find a solution,
> 2. there is not a readily accessible procedure that guarantees or completely determines the solution, and
> 3. the individual or group must make an attempt to find a solution (pp. 231-232).

Lester's first point emphasized the role of motivation in problem solving. His second consideration stressed that implicit in problem solving is an initial state of failure, a realization which runs counter to many educators' concern that any failure may be harmful to students' self-concepts. The third component of the definition recognized the fundamental part that effort plays in successful problem solving. When all three of Lester's points are considered together, one realizes that "a great deal of information is needed about each student to create real problem solving experiences" and that "a problem for one person may not be a problem for another" (p. 232). Individual differences such as prior knowledge, past experience, intellectual or other necessary abilities, and/or level of interest may influence whether tasks or situations are construed as problems.

A task requires problem solving then, when it is perceived as novel; that is, when the individual does not have an immediately available method (strategy or rule) for deriving the answer (Gagné, 1977). Indeed, a task for which a rule is readily available is an "exercise" in the conscious application of a known rule in a known context, as opposed to a problem. Similarly, if a rule is not only immediately available but is automatically invoked, the task is mere "rote drill" (Cawley, Fitzmaurice, Shaw, Kahn & Bates, 1979). Like exercises and rote drill activities, problem solving also relies on previously learned rules but requires (a) that such rules be combined in a new and creative manner resulting in problem solution or (b) that a rule or rules learned earlier be applied in a new context again resulting in problem solution (Gagné, 1977).

Simon and Lea (1974) have differentiated between problem solving and rule induction. They have suggested that problem solving is limited to the activity involved in finding the solution to a specific task. Rule induction, though, would seem to be involved in the transferring of the newly-generated rule from that specific task to a class of tasks. Once abstracted, the new rule or new contextual relevance of old rules may be applied to any problem the individual recognizes as having the same essential elements as the initial task.

Given that a problem consists of a task for which the individual does not have a readily available solution-strategy, problem solving essentially involves finding the task specific rule(s) and transforming that task from a problem to an exercise. Rule induction involves recognizing or expanding a class of problems for which the specific rule is generally applicable, effectively transforming the entire class of problems into an exercise. In both problem solving and rule induction such transformations would seem to be at least in part a function of the individual's perception of the task, which would in turn appear to be influenced by both task factors and individual-specific learner characteristics brought to the problem solving situation. That is, ease of solution (specific or generalized) is a complex interactive variable, or set of variables, based on the nature of both the problem and the problem-solver.

Task factors influencing problem solving on arithmetic word problems

In the research which has concentrated on problem solving in the context of arithmetic word problems there has been an attempt to identify a number of specific task factors which function as key determinants of problem difficulty. However, very few of these variables have been identified as unequivocally influential. That the research findings when compared across studies have been frequently discrepant is not surprising given the variety of word problems used, the levels and types of students tested, and the specific variables manipulated (Suydam, 1980). The research has been less than systematic and often carried out in the absence of a well-delineated theoretical model.

In task factor analysis the focus has been, not unexpectedly, on the dual contexts, linguistic and arithmetic, of math word problems, that is, on factors affecting readability and computational difficulty. Some specific language variables found to contribute to a story problem's readability for at least some populations are (a) vocabulary level (Caldwell & Goldin, 1979; Jerman & Rees, 1972), (b) syntactic complexity (Larsen, Parker & Trenholme, 1979), (c) length of text (Jerman & Rees, 1972), and (d) the familiarity of the language (Sharma, 1982).

The nature of the computations required to solve a word problem would seem to be directly implicated in establishing the level of problem difficulty. Thus, it is not surprising that a number of computational variables have been identified as being possibly influential. Such variables include the following: (a) type of operation involved; that is, the basic operations appear to progress in difficulty with addition being easier than subtraction, which is easier than multiplication and with division being the most difficult (Carpenter, Corbitt, Kepner, Lindquist & Reys, 1980); (b) computational level, that is, the size of the numbers involved and whether regrouping is required (Cawley et al., 1979); and (c) number of operations required; that is, multi-step problems are more difficult than single-step problems (Carpenter et al., 1980).

Arithmetic word problems have been construed as having a translation and a computation phase (Goldin & Caldwell, 1979; Paige & Simon, 1966). During translation problems are read and the appropriate relationship between the elements is conceptualized, resulting typically in a number sentence which then is solved during the computation phase. Whereas, readability factors certainly would seem to affect problem translation, there would appear to be another group of task factors operational during this first phase. These are translation phase factors which influence problem difficulty after initial reading, factors which affect the individual's logical analysis and conceptualization of the problem (Bachor, 1987). Representative of such variables is extraneous information which must during conceptualization be identified as unnecessary and discarded. The existence of this third class of task factors (or dichotomy within the factors affecting the translation phase) suggests that though necessary it is not sufficient to be able to read the words and do the calculations involved. The successful problem solver must be able to logically reconstruct in some meaningful way the essential elements of the task (Riley, Greeno & Heller, 1983). Other factors which may influence logical analysis of word problems include (a) the order in which the information is presented within the problem (Greene, 1974; Jerman & Rees, 1972), (b) the relationships between sets and subsets, for example, superordinate-subordinate classifications either given or assumed to be prior knowledge of the problem solver (Cawley et al., 1979), (c) whether the relationship between given elements in such that its translation results in a linear canonical number sentence (Carpenter & Moser, 1982; Rosenthal & Resnick, 1974), (d) the presence of "cue" or "distractor" words, which either legitimately indicate the appropriate arithmetic operation or suggest an inappropriate operation (Cawley et al., 1979; Goodstein, 1981), and (e) the subject and/or object set complexity (Cawley et al., 1979).

Whereas task factors contained in word problems have been widely investigated (Jerman & Rees, 1972), albeit not always systematically, there has been less research into the interaction of such factors with characteristics of problem-solvers to determine if specific task factors differentially define problem difficulty across specific learner characteristics. Thus, a consideration of the learner characteristics influencing performance on word problems and how such subject variables interact with task variables to determine performance would seem to be imperative to the understanding of the word problem conundrum.

Some learner characteristics assumed to influence performance on math word problems

Learner characteristics, those individual attributes the student brings to the test or learning situation, are numerous and diverse, and include such variables as gender, age, socio-economic background, IQ, scholastic achievement, experience and motivation (Gage & Berliner, 1979; Lester, 1980, 1983). Of the learner characteristics thought to influence math word problem solving, those most frequently addressed in the research literature are reading and computational abilities, and general level of intellectual functioning (Balow, 1964; Goodstein, 1981). However, teachers have been long puzzled and frustrated by students with adequate reading and computational skills who still experience difficulties with math word problems (Hoy, 1984). Consequently though these characteristics have been identified in the research as contributing significantly to an individual's performance, they are unable to account for all the performance variance between individuals. Nonetheless, individual differences in other cognitive and affective characteristics which also reasonably might be assumed to influence performance on arithmetic word problems have typically received only cursory attention. Three such possible performance correlates are information processing capacity, academic self-concept and test anxiety.

Information processing capacity. It has been demonstrated that an individual has a limited processing capacity which determines the amount of information which can be consciously acted upon at any given time (Bachelder, 1977a, 1977b; Bachor, 1976; Case, 1978b, 1985; Miller, 1956, Pascual-Leone, 1970). This limited capacity has been equated with attentional span and storage space in short term memory (Travers, 1982). The efficiency of this processing capability has been found to be in part developmentally-determined, increasing with age until at least mid-adolescence (Case, 1974, 1978b, 1985; Chi, 1976; Dempster, 1982; Pascual-Leone, 1970; Pascual-Leone, Goodman, Subelman & Ammon, 1978). It has been suggested that the functional level of this capacity may be constrained by other individual-specific cognitive and affective characteristics [e.g., anxiety (Case, 1974; Pascual-Leone, 1970)].

The specific structure, functioning and development of this limited capacity has been the subject of considerable research, theorizing and debate (Case, 1978b, 1985; Case, Kurland & Goldberg, 1982; Chi, 1976; Huttenlocher & Burke, 1976; Pascual-Leone, 1970, Pascual-Leone, Goodman, Subelman & Ammon, 1978). However, it is generally accepted that the individual has a demonstrable attentional span (typically assessed through digit or word span tests) and that the theoretical controversies and inconsistencies do not affect the practical implications of the construct in terms of the individual's problem solving ability. If the attentional demands of a task exceed the attentional energy the individual can apply to the task, successful completion of the task will be difficult or perhaps unlikely (Bachor, 1976; Case, 1978a). Therefore, functional information processing capacity may be expected to effectively constrain performance on certain types of arithmetic word problems by certain individuals.

Academic self-concept. Self-concept has been viewed as a mediational factor in an individual's behaviour and in his/her perception of that behaviour (Boersma & Chapman, 1979; Wells & Marwell, 1976), yet due to difficulties in operationalizing the multifaceted-construct and the use of poorly constructed assessment instruments the relationship between self-perception and behaviour has been hard to establish empirically (Shavelson, Huber & Stanton, 1976). Similarly, despite observations that children's attitudes toward themselves as learners appear to be related to their academic performance, when global measures of self-concept have been correlated with measures of academic achievement, the logically-conceived relationship has failed to be supported unequivocally (Boersma & Chapman, 1979). Still, Bloom (1976) suggested that students' perceptions of their academic abilities may affect the effort they expend on school tasks and their willingness to persist in the face of difficulties. Task outcomes then, are not only subject to the individual's cognitive abilities, but also may be influenced by his or her beliefs about the nature of those abilities; beliefs which may or may not be realistic. Therefore, an individual's academic self-concept may be expected to influence the functional limits of problem solving ability (Brookover & Gotlieb, 1964).

Test anxiety. Test anxiety has been defined as "a set of responses to a class of stimuli that have been associated in the individual's experience of evaluation or testing" (Sieber, 1980, p. 18) or as "an unpleasant feeling or emotional state that has physiological and behavioural concomitants and that is experienced in formal testing or other evaluative situations" (Dusek, 1980, p. 85). It has been equated with the individual's perception of the evaluative nature of the task situation and not with the real presence or absence of evaluation which appears to trigger test anxiety (Sieber, 1980).

High test-anxious students have exhibited less on-task behaviour and express more task-irrelevant egocentric statements than low test-anxious students (Dusek, 1980; Zatz & Chassin, 1983). Geen (1980) has suggested that test anxiety affects cue utilization; that is, the ability to scan and assess the task components. Given the constraints of a limited attentional capacity and the self-orientation typical of test-anxious individuals, Geen has proposed that much of that capacity is given over to task-inappropriate self-referential cognitions, precluding maximal utilization of task cues. Not surprisingly then, high-anxious subjects consistently have recalled fewer items on digit span tests than low-anxious subjects (Mueller, 1980). As with academic self-concept, test anxiety level may lessen an individual's ability to effectively use an already limited processing capacity.

Summary

Problem solving is gaining new pre-eminence across education curricula, particularly in mathematics, where the one common problem solving activity is the word problem. Problem solving is the process wherein the individual finds a means based on previous learning to a solution which is not readily apparent. An extension or corollary of problem solving is rule induction wherein the individual is able to generalize a solution strategy to other problems recognized as having certain similar relevant elements. Both problem solving and its corollary, rule induction, would appear to be functions of the task factors of the problem(s) and characteristics of the problem-solver.

Structurally, word problems have been examined in terms of three sets of task factors which might reasonably be expected to influence the level of difficulty: (a) language factors which affect the initial reading of the text, (b) logical analysis factors which influence the construction of appropriate relationships between relevant elements, and (c) mathematical factors which affect performance after the reading and conceptual reconstruction of the problem. This last set of variables is operational during the computation phase. It would appear that language and logical analysis variables are implicit to the translation phase of the solution-finding process. Thus, because the computational demands of story problems have an obvious direct influence on problem difficulty while the effects of some translation stage variables have been less obvious, and because of the temporal primacy in the solution-finding process of translation phase variables, the task factors examined in this study are at the language and logical analysis level. The computational demands of the problems employed are reduced and consciously controlled.

Similarly, because a student's response to any task factor would appear to be a function of individual-specific affective and cognitive factors brought to the problem solving situation, it also would seem reasonable to examine several learner characteristics and how they may be related to performance as specific task factors are manipulated and to the ability to induce an appropriate rule. What constitutes a problem for whom and for how long?

Research Questions

The purpose in this study, then, is essentially four-fold and the following general research questions will be addressed:

1. Will Grade 6 students induce rules differentially on word problems?

2. How will the performance of Grade 6 students on several affective and cognitive measures be related to rule induction on word problems?

3. Will task factor manipulations differentially affect Grade 6 students' rule induction as a function of differences in learner characteristics?

4. Will learner characteristics, other than reading and computational ability and general intelligence level(i.e., memory span, academic self-concept, and test-anxiety) be reasonable predictors of rule induction on word problems?

Chapter 2

REVIEW OF THE LITERATURE

Given the breadth and depth of the several research areas which have been integrated in this thesis, only those theoretical positions and empirical studies with direct implications are considered. In particular, the author acknowledges that the literature pertaining to general problem solving is vast and consequently no attempt is made to include many disparate aspects of the research included within the broad parameters of that field. To establish a basic framework, however, some perspectives on problem solving are briefly elaborated as are some parameters assumed to influence rule induction across analogous tasks. The focus then is narrowed to consider problem solving in mathematics, especially the arithmetic word problems. Task factors, learner characteristics, and process variables which in the literature have been found to influence performance on word problems are examined. Similarly, the potential influence of some task and subject variables hitherto uninvestigated in this context or less than comprehensively so are postulated. Finally, some typologies for classifying word problems are considered with specific reference to how they categorize analogous problems and allow for investigations of rule induction.

Some Basic Theoretical Perspectives on Problem Solving

Historically, problem solving has been construed differentially. Gagné (1977), whose hierarchical learning theory might be viewed, at least in part, as evolving from an earlier behaviourist perspective, has suggested that problem solving is the most complex type of human learning. He has seen problem solving as the process of arriving at a solution which consists of a new combination or rearrangement of previously learned concepts and rules. Through problem solving, the individual can derive new ideas independently.

Much of the problem solving research has emmanated from a cognitive perspective in which the learner is seen as an active participant in solving problems. Two polarized perspectives are representative of the variations within the cognitive viewpoint: a developmental stage approach and an information processing approach.

Stage theorists have proposed that all behaviour is developmental in nature. How an individual attempts to solve a problem and the type of problems he can successfully solve is an indication of his cognitive-developmental stage (Piaget, 1952). Each stage is defined by certain characteristic problem solving behaviours (Ginsburg & Oper, 1979). Inappropriate or inefficient problem solving behaviours may reflect a lack of cognitive maturity rather than a potential for creativity or inventiveness.

Information processing theorists generally have proposed a sequential model in which incoming data are analyzed and synthesized in steps. Some such models have used a computer analogy where the "program" allows the individual to deal adaptively with the demands of the situation (Parill-Bumstein, 1981). All such models have emphasized the processes involved in the selection, encoding, storage and retrieval of information and have viewed problem solving within such a context (Newell & Simon, 1972; Rumelhart, 1977).

In a recently developed cognitive theory, Pascual-Leone (1970; Pascual-Leone, Goodman, Ammon & Subelman, 1978) and Case (1974; 1985) have proposed a functional process model within a broad developmental structure in an attempt to account for differences in learning behaviour between individuals and between domains for a given individual. Certain aspects of this theory will be elaborated on in the discussion of memory span later in this chapter.

Problem solving and Rule Induction Across Analogous Tasks

If problem solving can be characterized as that which we do when we don't immediately know what to do, then rule induction might be construed as recognizing when we do know what to do and doing it. Thus, in a discovery learning situation problem solving and rule induction would seem to be temporally linked, with problem solving antecedent to rule induction. The first time an example of a particular task type is met it will at least temporarily constitute a problem. Further meetings with other instances of the task type may or may not constitute problems depending on the individual's identification of the task type. According to Egan and Greeno (1974)

> a subject begins the task lacking certain critical knowledge. During the performance of the task, the subject may acquire new knowledge structures....The knowledge structures are stored in working memory. Subjects abstract the rules by observing one or more examples of the rules in use (p. 43).

In the absence of direct instruction (i.e., being told the procedural rule) successful problem solving is necessary but not sufficient for rule induction to occur. There is no assurance that either direct instruction or successful problem solving relative to a particular task will result in that successful performance being transferred to other tasks within the class of tasks (Freeze, 1986; Gagné, 1977).

Rule induction, then, has been seen as a two-step process, dependent upon the ability to recognize a given task or tasks as analogous to a task for which one knows how to proceed, followed by application of that known procedure. Using puzzle problems (e.g., "the missionaries and cannibals"), Gick and Holyoak (1980) found that identifying a task as analogous to an already solved task and the transfer of that solution procedure do not necessarily follow in a lock-step fashion. They have suggested that how the individual represents the tasks in memory, both conceptually and procedurally, and the "semantic distance" between the analogous tasks may influence both the recognition of similarities and the application of the appropriate rule. Gentner (1982) has proposed that an analogous status across tasks is determined by similarities in the structural relationships between problem elements or objects and not by the attributes of the elements or objects per se. However, similarities in element or object attributes, that is, commonalities in problem context, frequently may be construed as establishing analogous situations and can result in misapplication of a procedural rule. Not surprisingly, Silver (1981) found that "good" and "poor" problem solvers differed significantly in their ability to recall the structure of previously solved problems, with poor problem solvers often recalling only contextually-based information. This would seem to support Gick and Holyoak's (1980) contention that how a task is encoded in memory may indeed affect the transfer of learning from one task to another task in the same class. Rule induction as it may specifically apply to word problems is considered later in the discussion of word problem typologies.

Mathematical Problem Solving and Word Problems

Problem solving in education commonly has been associated with mathematics or other subject areas with a large mathematical component (e.g., physics). Here, the research has been somewhat atheoretical and unsystematic (Lester, 1980, 1983; Suydam, 1980). Theory and its validation through research, has failed to keep pace with a commonly perceived educational goal: "Learning to solve problems is the principal reason for studying mathematics" [National Council of the Supervisors of Mathematics (NCSM) , 1977, p. 27]. Suydam (1980) has suggested that word problems originated to fill the gap, "because educators could find no other plausible way to communicate the applicability of the skills and ideas of mathematics lessons to everyday life" (p. 35). She also has acknowledged that many sets of word problems are not problems at all but rather exercises in applying a rote algorithm or basic facts, that is, application of the appropriate rule occurs but for the wrong reasons educationally. Students are able to apply a procedural rule due to the format of materials and the predictability of a curriculum. For example, instruction in multi-digit subtraction with regrouping is typically followed by word problems all of which require multi-digit subtraction with regrouping. The analogous status of the word problems is guaranteed and does not have to be induced by the students.

Three sets of interrelated variables have been commonly identified as contributing to word problem solving performance and by implication to rule induction across word problems: (a) task characteristics associated with the problem's format and structure; (b) learner (subject) characteristics, individual differences students bring to the problem solving situation; and (c) process variables, the specific strategies used during task execution. What is not so clear is the nature of the complex relationships between the three groups of factors and thus far, researchers have failed to account satisfactorily for the complexities involved using existing problem solving models (Vander Zanden & Pace, 1984). In the following sections these variables are discussed individually and at greater length.

Task Variables

Within the context of story problems, research attempts have been made to analyze the relative difficulty associated with a number of task variables(i.e., with the physical formatting of the problem information). The work of Jerman (Jerman & Rees, 1972; Jerman, 1973) is typical of this approach which employs regression analyses and factor analysis to determine the relative influence of various task factors. Using such procedures some language, logical deductive, and computational variables have been identified as accounting for large portions of variance in problem difficulty.

The readability level of a word problem appears to have a significant influence on its solvability. In their study of subjects in Grades 4 and 6, Caldwell and Goldin (1979) found that the vocabulary level of the words incorporated into story problems had a significant effect on performance. Similar findings have been reported for the same grade levels by Linville (1976) and Paul, Nibbelink and Hoover (1986). Jerman and Rees (1972) reported essentially the same results in their multiple regression analyses across age groups. Such findings are logical in that vocabulary level is a determinant of comprehension and without accurate comprehension of the problem situation, solution finding likely will be negatively affected.

Also not surprising, the syntactic complexity of the text has been identified as affecting problem readability at least for some populations (Larsen, Parker & Trenholme, 1978; Linville, 1976). When vocabulary and mathematical computational level were controlled for, highly complex syntactic structures adversely affected the performance of eighth graders (Wheeler & McNutt, 1978). Similarly, the length of the word problem has been demonstrated to influence word problem difficulty (Jerman & Rees, 1973). Finally, the familiarity of the language used to establish problem context has been identified as affecting students' performance (Sharma, 1981). Even when the actual words used to relate the problem are "readable" for the target population if the context portrayed is not familiar, solution-finding would appear to be impeded. Unfamiliar contexts then might be expected to influence students' ability to identify analogous problems. In the present study the effects of language familiarity on performance and rule induction are examined.

Variations in logical-deductive task factors, factors which influence conceptual understanding even when readability level is appropriate, have been demonstrated to affect word problem difficulty. Verbal cues such as "altogether," "left," and "more" have been found to facilitate performance (Goodstein, 1981;

Sharma, 1981) sometimes permitting students to calculate a correct answer without reading or comprehending the entire text (Cawley et al., 1979). Verbal cues, then, may invite rote application of a procedural rule. Students also have been found to overgeneralize the use of procedural rules in the presence of such cue words, even though the words may not be functioning in that cueing capacity (e.g., "left" may refer to direction as well as to a quantity remaining) and their inclusion is intended to be misleading (Blakenship & Lovitt, 1976).

"Canonical" representations of set relationships (e.g., a + b = ? and a - b = ?) have been identified as less difficult than "noncanonical" representations (e.g., a + ? = c and ? - b = c) (Carpenter & Moser, 1983; Rosenthal & Resnick, 1974). Typically, noncanonical relationships are the result of indefinite quantifiers in the word problem text (e.g., "some boys" or "a number of marbles"). Cawley et al. (1979) and Bachor, Steacy and Freeze (1986) have used the terms "direct" and "indirect" to differentiate between these two problem types. Only canonical or direct problems are considered in this research study.

The presence of extraneous information, that is, additional sets superfluous to problem solution, similarly has been found to affect performance on word problems (Blakenship & Lovitt, 1976; Carpenter, Corbitt, Kepner, Lindquist & Reys, 1980), as has the location of that information within the problem text (Greeno, 1974; Jerman & Rees, 1972). Neither has had consistently profound effects across populations and the influence of the latter has not proven unequivocally for any population (Bilsky & Judd,1986; Englert, Culatta & Horn, 1986; Goodstein, 1972; Muth, 1984; West, Lee & Anderson, 1969). The effects of both the presence and the relative placement of extraneous information are examined further in this study.

Cawley et al. (1979) have suggested that another task variable which might influence problem conceptualization is the complexity of the sets involved. They have postulated the existence of four different set complexity conditions in arithmetic word problems; that is, problems may be defined as containing simple or complex subject sets and simple or complex object sets resulting in four possible permutations. If a set is complex, the problem solver may be required to combine the elemental sets into some larger inclusive set. The language used to establish set complexity is another possibly influential variable (Bachor & Steacy, 1987). Complexity may be achieved by reclassifying subordinate sets under a given superordinate or by grouping the elemental sets into a compound set using the co-ordinate conjunction "and." While hypothesized to influence problem difficulty, this concept of set complexity has remained largely uninvestigated. Therefore, the effects of set complexity and of variations in the language used to create complex sets are considered in this study.

The final category of task factors identified as contributing to ease of problem-solution are computational variables. It would seem inherently obvious that the arithmetic demands of any word problem would influence directly its level of difficulty and there is much research to indicate that this is indeed the case (e.g., Carpenter et al., 1980). Since the computational demands of word problems are not under consideration in the present study, this body of research will not be elaborated on here.

Learner Characteristics

It already has been suggested that word problem solving performance, and by extension, rule induction across analogous problems are likely complex interactive processes between task and subject variables. The individual's cognitive capacities, his or her perception of the task and of himself or herself as problem-solver, and other psychological or physiological states would seem to put real constraints on that performance. Of the numerous learner characteristics possibly implicated in the solving of arithmetic word problems, three have dominated in the literature: (a) level of intellectual functioning, (b) reading ability, and (c) computational ability.

In a number of studies over several decades, it has been demonstrated that mentally handicapped and average learners perform differently on story problems (Bilsky & Judd, 1986; Cawley & Goodman, 1969; Cruikshank, 1948; Goodstein, 1972). Differential responses to task factor manipulations (e.g., the presence of extraneous information) by IQ groups also have been reported by such researchers, supporting the contention that problem solving is interactive in nature.

Research into the role of reading ability in word problem solving has yielded less than definitive results, in spite of the fact that reading is undeniably involved early in the problem solving process. However, in the majority of such research, significant correlations between general reading ability and performance on word problems have been found. For example, Treacy (1944) reported overall effects for reading ability on problem solving scores for Grade 7 students. However, he did not find significant results for all reading subskills, suggesting that, at least for this grade level, certain aspects of reading ability may be more directly implicated in problem solving than others. Balow (1964), too, found significant correlations between reading ability and performance on word problems for sixth graders. After partialling-out the effects of IQ, however, the correlation was greatly reduced, causing Balow to suggest that the relationship between reading and problem solving ability may be due to the high correlation of each with general intelligence and that no direct cause and effect relationship between the two abilities exists. Finally, from the results of numerous studies, it has emerged that computational skills and word problem solving ability are highly correlated (Balow, 1964; Carpenter et al., 1980; Kilpatrick, 1978).

The influence of any number of other, perhaps not so task-specific, subject variables on performance on word problems remains essentially unexplored. It was with this in mind that memory span, test anxiety, and academic self-concept were included as independent variables in the present study. These three subject variables were specifically chosen from a range of possible variables because they have been implicated theoretically and empirically as determinants of functional information processing capacity. Given that the effects of these learner characteristics have rarely, if ever, been investigated in the specific context of arithmetic word problems, the ensuing discussion will focus on the constructs themselves in an attempt to postulate their possible effects on word problem performance.

Memory Span. The concept of memory span and its assumed role in cognitive functioning has had a long and prominent history attested to by the emphasis given to measures of span in the assessment of general intelligence. If one accepts that the individual has a limited processing capacity which can be equated with working memory and the amount of information which can be actively operated on therein, one has postulated a basic stricture on task performance. If one also accepts that such a capacity is in some way developmentally-determined, then the complexity of tasks which can be successfully completed should increase with age. Much theorizing and research has centered on these two postulates (e.g., Case, 1974a, 1974b, 1978, 1985; Case, Kurland & Goldberg, 1982; Chi, 1978; Chi & Klahr, 1975; Pascual-Leone, 1970; Pascual-Leone, Goodman, Subelman & Ammon, 1978) with broad agreement on their general accuracy and a great deal of debate on what is actually limited and what develops with age, the memory space itself or the operational efficiency with which it is used.

Of particular interest here is the theory of Pascual-Leone (1970; Pascual-Lone, Goodman, Ammon & Subelman, 1978) and Case (1978; 1985) in which they have as part of a larger model hypothesized and produced empirical evidence in support of the existence of a quantitative construct, M-space, which appears to increase developmentally. Essentially M-space is the maximum number of discrete pieces of information or schemes which can be controlled or integrated at one time. Though initially proceeding from the same theoretical stance introduced by Pascual-Leone (1970), it is likely that Pascual-Leone and Case presently would disagree on the developmental determinants of M-space (Case, 1985), debating between a developmentally increasing absolute storage capacity and a developmentally increasing operational efficiency which frees more storage space. Nonetheless, such agreements would not appear to affect the practical educational significance of a limited and quantifiable processing capacity which may effectively determine the complexity and of the tasks a given individual can successfully complete.

Regardless of theoretical differences, Pascual-Leone (1970) and Case (1974b) have reported the M-space levels and corresponding developmental substages given in Table 1. There is some debate over the accuracy of the lower and upper ends of the scale (Case, 1985), but for school-age children these broad parameters appear to hold with reasonable consistency. In reporting such levels, however, both Pascual-Leone and Case would allow that the functional level of M-space is subject to a number of individual-specific variables. They have suggested that the scheme or repertoire of schemes which may become activated in response to a specific task can be "boosted" by a number of organismic factors. The scheme which is boosted may be task relevant or irrelevant, efficient or otherwise; thus, they have accounted for individual differences in task performance. Within this theoretical perspective, then, performance on arithmetic word problems as on other tasks would appear to be under some generalized developmental constraints which may be further influenced by individual-specific task responses. If the amount of information necessary for word problem solution exceeds an individual's functionally-determined processing capacity, successful performance on that problem is unlikely. Similarly, if the individual is unable to solve individual analogous problems successfully, it becomes virtually impossible for that individual to compare and recognize the similar structural relationships of the problems which are necessary for rule induction to occur. Ironically, however, once rule induction has occurred the demands on working memory of any problem within the entire class of problems may be significantly reduced.

Table 1

Developmental Ranges for M-space

Age	Developmental Substage	M-space
3-4	early preoperations	e + 1
5-6	late preoperations	e + 2
7-8	early concrete operations	e + 3
9-10	middle concrete operations	e + 4
11-12	late concrete--early formal operations	e + 5
13-14	middle formal operations	e + 6
15-16	late formal operations	e + 7

Adapted from Case, 1974b, p. 549.

Note. e = the theoretical construct which represents the amount of mental energy needed to activate automatic schemes.

Two individual-specific variables, academic self-concept and test anxiety, which may boost task appropriate or inappropriate schemes and thus contribute to functional M-space level, are discussed next.

Academic Self-concept. Shavelson, Hubner and Stanton (1976), have theorized that self-concept is a "multi-faceted" construct which appears to become differentiated with age. The pre-schooler has a rather global concept of himself/herself, whereas the adolescent has both a general concept and a multi-dimensional view of self relative to the roles assumed (e.g., academic, social, or physical roles). Some aspects of self-concept then become more role and situationally determined as the individual matures. Global or specific, self-concept would appeared to be a composite of how we see ourselves generally or situationally and our perception of how significant others see us (Marsh & Shavelson, 1985). Thus, external feedback about our behaviour is an important source of concept development. Self-concept also would seem to be strongly associated with other affective characteristics such as locus of control and the attribution of successes and failures to luck or to effort (Schunk, 1985). Persons with low self-concept are generally more likely to view themselves as ineffectual and to give up in the face of difficulty.

Academic self-concept has been posited as one facet of the more global construct. It's emergence appears to be a phenomenon of the middle school years, perhaps resulting from accumulated experience in the academic situation (Marsh & Shavelson, 1985; Purkey, 1970). While a relationship between academic self-concept and academic achievement has been found consistently across studies (Boersma & Chapman, 1979; Kistner, Haskett, White & Robbins, 1987; Shavelson, Hubner & Stanton, 1982; Winne, Woodlands & Wong, 1982), the specific nature of the relationship has been hotly debated (Potterbaum, Keith & Ehly, 1986). Is the relationship causal and if so, what causes what? Or is the observed relationship the result of some third variable with which both academic self-concept and academic performance correlate highly? Regardless of theoretical position, these researchers have determined that persons identified as having poor academic self-concepts do generally perform differently on academic tasks than do their peers with higher levels of self-concept and conversely, that poor performers are likely to have poor self-concepts.

Summers (1984) also found that academic self-concept and test anxiety correlate highly with each other, in that individuals with low academic self-concepts frequently are high-test anxious. Such a relationship between perceived academic ability and anxiety level when that ability is being evaluated would seem logical. It also has been determined that students' perceptions of their own performance and teachers' perceptions of that performance do not differ significantly (Boersma & Chapman, 1979), suggesting that students' judgment of their academic abilities may be reasonably accurate, whether that judgment is the result or cause of that performance.

Academic self-concept, then, may be either a reflection of real performance constraints or a constraining influence in itself. Regardless, given the demonstrated relationship between academic achievement and self-concept one might expect that students with poorer academic self-concept should generally perform more poorly on word problems than students with better self-concepts. Similarly, given the relationship between perceived abilities and causal attribution, students with low self-concepts may, if the word problems are perceived as difficult, feel their efforts are ineffectual and literally "give up." Without sustained effort rule induction is unlikely to occur.

Test Anxiety. Anxiety has been viewed as both potentially facilitative and debilitative depending on the anxiety level itself, the demands of the specific situation and the individual involved. In the Yerkes-Dodson Law (1908) it is held that increases in arousal (anxiety) result in improved performance up to an optimal level (Travers, 1982). Increases in arousal beyond that point have been found to inhibit performance (Heinrich & Spielberger, 1982). Test anxiety has been seen as a special case of this general anxiety construct and can be construed as the level of psychological and physiological arousal in the face of perceived evaluation (Dusek, 1980). Test anxiety then is not a phenomenon confined to tests per se, but may affect performance in any situation in which the individual's behaviour is thought to be under scrutiny.

Working from a Hullian Drive Theory perspective (wherein drive, arousal or anxiety is the result of the competing levels in the needs of the individual and the intensity of noxious stimulation) Heinrich and Spielberger (1982) have postulated the following relationships between anxiety level and learning:

1. For simple or easy learning tasks in which correct response are dominant and competing error tendencies are minimal, the performance of high-anxious subjects will be superior to that of low-anxious subjects.
2. For difficult learning tasks, in which competing error tendencies are strong relative to correct responses, high drive will activate these error tendencies, and the performance of high-anxious subjects will be inferior to that of low-anxious subjects.
3. For tasks of intermediate difficulty the stage of learning must be taken into account. High anxiety will be detrimental to performance early in learning

> when the strength of correct responses is weak relative to competing error tendencies. Later in learning, high anxiety will begin to facilitate performance as correct responses are strengthened and error tendencies are extinguished (p. 146).

In this context the error tendencies reflect the degree to which the response has become habituated. The third postulate would seem to have direct implications for the effects of anxiety on rule induction. High anxiety may be debilitative only in the early stages of a rule induction task and may eventually become noninterfering or even facilitative as rule abstraction occurs, i.e., as an habituated response becomes generalized reducing error tendencies.

S. B. Sarason and his colleagues (Sarason, Davidson, Lighthall, Waite & Ruebush, 1960) have studied extensively test anxiety in elementary school students. A special measure, the Test Anxiety Scale for Children, was devised to facilitate their research and is still the standard instrument for assessing levels of test anxiety in such children. They found that high anxious students perform more poorly on academic tasks and emit more negative egocentric cognitions concerning their abilities to cope with the task than do low anxious students. These authors have indicated that the level of anxiety in individual children may be a reflection of their early familial experiences, the way in which their parents cope with their own arousal situations, the way in which the child's affective behaviour is responded to, and the critical and evaluative nature of the home environment. Children, then, arrive at their formal school with a predisposition, based on prior experience, to interpret a wide range of cues as evaluative or non-evaluative, and to respond to those cues which are interpreted as evaluative with certain established coping behaviours which may be appropriate or inappropriate. This would suggest, as Spielberger (1966) has emphasized, that the evaluative nature of a task is dependent upon the individual's prior experience, that tasks are perceived within a personal context.

In 1971 Wine proposed a cognitive-attentional interpretation of the anxiety interference model. High anxiety negatively impacts on task performance by directing cognitive attention away from the task and focussing it inward on self-evaluation and deprecation. Anxiety serves to interfere in the utilization of relevant task cues by usurping attention for self-centered rather than task-centered cognitions.

The concept of cue utilization was based heavily on Easterbrook's (1959) hypothesis in which he has stated that "emotional arousal consistently narrows the range of our utilization in task performance" (Wine, 1971, p. 97), and that arousal can be facilitating or debilitating depending upon the necessity of peripheral cues to successful task resolution. Consistent with this hypothesis was Geen's (1980) summary of research in which the presence of task relevant and irrelevant cues has been manipulated and the effects of such manipulations on the performance of extreme anxiety groups were assessed. While performance of high anxious individuals was generally inferior to low anxious individuals, it also was affected less by cue manipulation than low anxious individuals' performance, which was facilitated by the addition of relevant cues and debilitated by the presence of irrelevant cues. This last finding may have implications for students' performance on word problems with extraneous information.

Wine (1980; 1982) more recently has expanded the attentional interference model of test anxiety to include the concept of an individually-determined limited information processing capacity which necessitates a limited attentional energy construct. If an individual has a specific amount of attentional energy to deploy at any given time and this attention can be task-directed or diverted away from the task by self-oriented cognitions, task performance may be affected by the direction of attentional deployment. The more complex the task (level of task complexity is also individual-specific), the more debilitating the effects of misappropriated attention. Consistent with some earlier research such a view also would allow that anxiety, because of its energizing nature, can be facilitative as long as the attentional energy consumed by the anxiety-produced self-referential non-task oriented cognitions plus the attentional demands of the task do not exceed the attentional capacity of the individual. Beyond that point, anxiety interferes with and narrows the perception of task cues and thus inhibits performance.

Test anxiety then might be expected to have a complex effect on students' performance on arithmetic word problems and rule abstraction across problems with similar structural relationships. First, the effects of a certain level of test anxiety may be specific to the individual student and to the demands of the problem. Second, a certain level of anxiety may differentially affect performance over time. Finally, extremes in anxiety level may produce differential effects across task factor manipulations. To the best of the author's knowledge only one study of arithmetic word problem solving has used test anxiety as the independent variable (West, Lee & Anderson, 1969). The performance of 58 Grade 6 and 8 subjects on the Test Anxiety Scale for Children (Sarason et al., 1960) was compared with their ability to select relevant from irrelevant information. From the results it was inferred that anxiety level interacted with the presence of extraneous information, increasing problem difficulty for high anxious subjects.

Process Variables

The third focus of word problem research has been the processes or strategies individuals use during problem solving. These process variables might best be construed as manifestations of the interactions between the two research foci already considered, task factors and learner characteristics. The strategies individuals deploy during problem solving would seem to reflect their perceptions and conceptualizations of the demands of the task, their cognitive capabilities, their perceptions of those capabilities, and their motivation to complete the task. Process variables would not appear to exist independent of the task or the learner attempting the task. From the process variable research, however, may come some insights into rule induction behaviour across analogous word problems(i.e., in "typing" problems we are identifying a group of problems for which a generalized procedural rule may be appropriate). This issue is discussed further in the following section in which word problem typologies are examined.

Classifying Math Word Problems

Because math word problems are difficult for many students and specific problems are more difficult than others, several attempts have been made recently to distinguish word problem types. The essentially independent, though parallel research of Carpenter and Moser (1982; 1983) and Greeno and Riley

(Greeno, 1980; Riley, Greeno & Heller, 1983) in which they have focussed on simple word problems involving addition and subtraction is seminal in this area. The major premise of their work are that children choose strategies on the basis of their conceptual understanding (i.e., their representation of the problem) and that their representations of problems largely reflect the semantic structure of those problems. Riley et al. (1983) have defined "semantic structure" as "conceptual knowledge about increases, decreases, combinations, and comparisons involving sets of objects" (p. 159), that is, the ways sets can be related. In Appendix A the categories for classifying word problems as delineated by Riley and her associates are illustrated. Carpenter and Moser's system is essentially the same save for slight changes in nomenclature. Both classification systems categorize word problems on the basis of the logical conceptual understanding established by the semantic relationships between the sets within a given problem.

Despite the fact that for the mature problem solver all 16 problem types, found in Appendix A, would be solved using simple canonical number sentences ($a + b = ?$ and $a - b = ?$), the two research groups have determined that certain solution strategies typically are employed by children for each of the conceptual categories. For example, for change (result set unknown) problems when given manipulatives (blocks) to aid in solution, children most often use a "separating-from" strategy. However, for change (change set unknown) problems children typically use an "adding-on" strategy; while for compare (difference set unknown) problems, children frequently use a "matching" strategy.

A hierarchy of problem difficulty within these classification systems also has been identified (Carpenter & Moser, 1982, 1983; Riley, Greeno & Heller, 1983). It would appear that some semantic relationships are more difficult to conceptualize than others even at this low level of computational requirement. Four levels of difficulty have been indicated by the research results. They are, from easiest to hardest, as illustrated in Table 2.

Table 2

Hierarchy of difficulty within Riley, Greeno, and Heller's classification of word problem types

Level	Problem Type
1	Change (result unknown)
	Combine (combined value unknown)
2	Change (change set unknown)
	Compare (difference unknown)
3	Combine (subset unknown)
	Change (start set unknown)
	Compare (compared quantity unknown)
4	Compare (referent set unknown)

(Results for equalize problems remain inconclusive.)

Adapted from Thompson and Hendrickson (1986, p. 24).

Carpenter and Moser, who have conducted their research with students in Grades 1 to 3, have found that by Grade 3 the hierarchy of difficulty is not so apparent as students become more able to conceptualize the semantic relations involved.

Although using children's conceptual understanding as the basis for establishing verbal problem types would appear presently to be the prevalent classifying procedure, it is not the only one. Caldwell and Goldin (1979), for example, have attempted, without obtaining significant results, to classify problems according to their assumed cognitive-developmental demands and the developmental stage of the students who can successfully solve them. Nor is the conceptual understanding classification system without its limitations, in that learner characteristics and task variables are essentially ignored. This last concern is a major one and has been acknowledged by Carpenter and Moser (1983), who grant that "differences in wording or context may result in systematic differences that are unrelated to the (semantic) structure of the problem or to how children solve it" (p. 19). Similarly, the differentiations in strategy noted for primary students may not be present in students at the intermediate level or higher. Conceptual understanding of word problems likely changes with age resulting in an extension of problem-type boundaries and a reduction in the actual number of earlier perceived types.

Yet another method of classifying problems, proposed by Reed (1987), allows us to focus on and reintegrate the concepts of problem solving and rule induction in the context of word problems. Reed's problem categories are not mutually exclusive from the other classification systems discussed and can in fact be combined with them to further elucidate the process of rule abstraction on analogous problems. In several studies, Reed (1987; Reed, Dempster & Ettinger, 1985) has examined university students' ability to recognize similarities between word problems. He has suggested that word problems may be related in one of four ways in terms of their story contexts and solution procedures. First, and perhaps most obviously, they may be unrelated, bearing no resemblance contextually or procedurally. Second, problems may be "similar," sharing a common context but requiring different solution-finding procedures. Third, they may be "isomorphic," having different story contexts but using the same procedural rule in solution-finding; that is, "the equations needed to solve these problems are structurally identifiable" (p. 124). Finally, word problems may be "equivalent," both similar contextually and isomorphic procedurally.

Like Gentner (1982), Reed (1987) has concluded that it is the corresponding structual relationships between sets or objects rather than the attributes of the objects themselves, which determines the analogous status of problems regardless of their contexts. If they are correct in this assumption, it would seem that the concept of rule induction would be most applicable to isomorphic problems. Because these problems by definition can be solved using the same relational rule, it is across such problems that rule abstraction may occur; that is, the "general properties of a class of solutions are inferred by solving a small number of problems in the class" (Egan & Greeno, 1974, p. 79). Before rules can be transferred across contexts, it is necessary that problems containing isomorphic relationships be identified through "mapping." Mapping refers to the process of assigning every member in a relational set to a corresponding member in another relational set, to establish relational congruency. If recognition of analogous relationships and the transfer of an appropriate rule occur; isomorphic problems in effect may become non-problems as students no longer need to formulate a problem solving rule, but have one readily available.

Reed (1987) has suggested that two other factors, problem "transparency" and a level of "base specificity" in the learner, contribute to the identification of isomorphic word problems. The concept of "transparency" would appear to be concerned with task factors, that is, changes in problem context or format, which might impede or facilitate the identification of analogous problems. Similarly, the concept of a basic level of understanding, "base specificity," necessary for problem classification would seem to allow that there are individual differences which may interfere with or preclude that classification process and, by extension, rule induction. From the results of his studies, Reed has concluded that novice problem solvers, those lacking a base specificity, are more likely to type problems according to similarities in context than to similarities in relational structure. These findings are consistent with Silver's (1981) earlier conclusions regarding novice and expert problem solvers. Reed also found that the more transparent (similar in context and format) isomorphic problems are, the more likely they are to be identified as requiring the same solution procedure which is consistent with Gick and Holyoak's findings (1980) regarding the "semantic distance" between isomorphic problems.

In summary, then, the investigation of problem solving in the context of arithmetic word problems has emanated from the three different perspectives associated with more general investigations of problem solving (Lester, 1980, 1983): (a) the isolation of task factors which may affect problem difficulty (Jerman & Rees, 1972; Linville, 1970); (b) the identification of some characteristics of good problem-solvers (Krutetskii, 1976; Muth, 1984); and (c) the examination of the conceptual processes learners use in defining the semantic relationships found in specific types of problems (Carpenter & Moser, 1982; De Corte, Vershchaffel & De Win, 1985; Riley, Greeno & Heller, 1983). The common focus inherent in all of the above orientations, is the attempt to illuminate variables which influence problem solving behaviour in the context of specific novel tasks. A corollary of this research, at least implicitly, would seem to be a concern with the variables that affect the abstraction of rules which may be induced during problem solving, and which may effectively increase the learner's capacity to solve other similar problems (Egan & Greeno, 1974; Gagne, 1977; Gagne & Briggs, 1974).

In reviewing the literature two major tenets have emerged. First, a better approach to the study of word problems may be an integrative one which recognizes the wide range of variables possibly implicated in word problem solving and the complexity of their potential interactions. Second, an important goal for word problem solving researchers is the identification of variables which can potentially affect students' ability to classify problems according to conceptual type and to induce procedural rules appropriate for specific types. It was within these theoretical parameters that the present study was conceived and undertaken.

Chapter 3

METHODS

Restatement of the Problem

Despite the recent emphasis on arithmetic word problems in the classroom and a corresponding increase in arithmetic word problem research, a comprehensive description of the relationships between problem structure, learner attributes and word problem solving and rule induction ability has failed to emerge. Many task variables have been suggested as possible sources of problem difficulty and some have been empirically verified though not always unequivocally or consistently. The influence of learner variables (other than reading, computation, and general intellectual ability) on performance outcome has received less attention. The concept of rule induction in the context of story problems similarly has been the object of little scrutiny. This neglect may be reflected in the frustration that word problems generate in the classroom. A need for continued research to identify influential task and learner variables and the nature of their relationship to word problem solving and rule induction ability and rate is indicated. A better theoretical understanding of the relationships involved may contribute to the creation of more effective instructional strategies and improved student performance. It is to this end that the present study was dedicated: A clarification of the influence of some task and learner characteristics on students' rule induction ability and rate as measured on their performance on a set of isomorphic word problems.

Design

The research design for the study was descriptive and correlational. Differences in rule induction rate and ability for the entire sample were investigated. Similarly, the influence of several learner characteristics on rule induction rate and ability was examined. The relationship between students' performance on learner characteristic measures and their performance on word problems as task factors were manipulated was considered. Finally, the predictive nature of the learner characteristics relative to word problem performance and rule induction rate and ability was explored.

Variables

The following task variables were included in the study: (a) the presence and location of extraneous information within the problem text; (b) the complexity of the subject and object sets involved; (c) the grouping of complex sets versus the reclassification of complex sets; and (d) the familiarity of the language which the word problems were constructed. Each of these variables will be clearly operationalized in the discussion of problem generation and illustrated with the presentation of prototypes. Several subject variables were included in the study: (a) information processing capacity or memory span; (b) level of academic self-concept; (c) level of test anxiety; and for some but not all comparisons (d) gender.

Subjects and Sampling Procedure

The subjects were the students (N=57) in two Grade 6 classes in a Victoria, British Columbia elementary school which serves an essentially middle-class suburban neighbourhood. Male students outnumbered female students in the two classes (31 males; 26 females). At the time of data collection subjects ranged in age from 137 months to 157 months. Mean age was 143.13 months (approximately 11 years, 11 months) with a SD of $\pm$ 4.39 months. Subjects' participation was contingent upon receipt of parental permission. It was recognized at the proposal stage that the use of intact classes would affect the generalizability of the results.

For some analyses the subjects were further divided into subgroups on the bases of their performances on the learner characteristic measures discussed in the "Instruments" section. The procedures for assignment to groups are discussed later when they become pertinent to the results under consideration.

Instruments

Word Problems. Subjects' rule induction rate and ability as determined by their performance on isomorphic word problems was measured on a researcher-designed set of questions. The total number of problems was 144, which was divided into three administration-facilitating counterbalanced sets of 48 randomized problems each. Problem-generation followed a model proposed by Cawley et al. (1979) and further refined and elaborated by Bachor (1985) and Bachor, Steacy and Freeze (1986). This particular approach to generating items allowed for the controlled manipulation of the task variables under investigation.

A discussion of the parameters guiding problem generation and of the incorporated task variables now follows and illustrative problem types are presented.

During problem generation the math word problems were held constant as follows:

1. All problems were isomorphic to each other; i.e., had the same relational structure between sets and, therefore, could be solved using the same procedural rule.

2. All problems required single digit addition. All addends were larger than 1 and smaller than 10, all correct sums were between 4 and 18, inclusive. This constraint was adhered to because the relationship of computation ability to problem solving and rule induction ability was not under investigation in the present study. All subjects had demonstrated criterion-level basic addition skills making the computational requirements of the word problems essentially a non-issue in ascertaining task difficulty.

3. All problems were text implicit (Bachor, 1985); that is, all information necessary to problem solution was supplied, but integration of information across problem statements was required.

4. All problems were direct (Bachor et al., 1986; Cawley et al., 1979); that is, the implied operation is indeed the appropriate operation. Problems contained neither indefinite quantifiers (e.g., "some") nor misleading cue words.

5. All problems were syntactically similar. Problem information statements followed a simple sentence structure of the subject-verb-object pattern. All problems concluded with interrogative statements following the "How many objects did the subject(s) verb?" pattern. Half the problems required

reclassification of set information and thus included a reclassification statement in the form: "Subordinate a and subordinate b are superordinate x" (e.g., "Roses and daisies are flowers."). Finally, no modifiers were used in any problems.

6. All extraneous information was set relevant as opposed to set irrelevant (Bachor, 1987; Bachor & Steacy, 1987; Bachor et al., 1986). Two problems which illustrate the difference between these two types of extraneous information are now presented. The following word problem is an example of a problem containing set relevant information:

> The sailor owned 2 boats. The doctor owned 5 boats. The teacher saw 6 boats. How many boats did the sailor, doctor and teacher own?

The third problem statement is extraneous to problem-solution, but is relevant to the complex subject set of "sailor, doctor, and teacher" and also to the simple object set "boats." It is only irrelevant to the action set "owned". In comparison the following is an example of a problem containing set irrelevant information:

> Jane read 8 stories. Jane read 3 poems. The dog chased 3 rabbits. How many stories and poems did Jane read?

The third problem statement is again extraneous to problem-solution, but is also irrelevant to the subject set "Jane," the action set "read," and the complex object set "stories" and "poems." All extraneous information in the present study was of the type found in the first example; that is, it was, while unnecessary, in fact inappropriate to the problem-solution, related to the subject and object sets under consideration. The discussion and examples of set complexity below will further elucidate the relationship.

According to the task variables under investigation the word problems were varied as follows:

1. Extraneous information. There were four levels of this variable. Some problems contained no extraneous information (NEI), for example:

> The boy has 8 balls. The boy has 6 bats. How many balls and bats does the boy have?

For some problems containing extraneous information, that information was presented in the first statement (EI1), for example:

> Ben broke 8 cookies. Jill ate 5 cookies. Chris ate 2 cookies. How many cookies did Ben, Jill and Chris eat?

In other problems the extraneous information was presented in the second statement (EI2), for example:

> The girl rode 4 horses. The man saw 7 horses. The boy rode 6 horses. How many horses did the girl, man and boy ride?

In still other problems the extraneous information was presented in the third statement (EI3), for example:

> Mr. Ross bought 3 houses. Mrs. Jones bought 9 houses. Mr. Smith painted 4 houses. How many houses did Mr. Ross, Mrs. Jones and Mr. Smith buy?

2. Set complexity. This variable had three levels. Some problems contained a complex subject and a simple object (CSSO), for example:

> The doctor picked 6 apples. The nurse picked 4 apples. How many apples did the doctor and nurse pick?

Other problems contained a simple subject and a complex object (SSCO), for example:

> The princess baked 7 tarts. The princess baked 8 cookies. How many tarts and cookies did the princess bake?

Yet other problems contained both a complex subject and a complex object (CSCO), for example:

> Bob scared 4 robins. Bill scared 9 sparrows. Bob and Bill are boys. Robins and sparrows are birds. How many birds did the boys scare?

3. Grouping of sets versus reclassification of sets. According to this two-level variable all complex sets were either combined using compound set language or renamed using subordinate/superordinate set language. The following is an example of a problem requiring grouping (GRP); that is, using compound set language:

> Mary has 2 cats. Bob has 4 cats. How many cats do Mary and Bob have?

In contrast, the following is the parallel problem requiring reclassification (RC); that is, using subordinate/superordinate set language:

> Dave ate 6 candies. Jeff ate 8 candies. Dave and Jeff are boys. How many candies did the boys eat?

4. Language familiarity. Three parallel sets of problems were generated in the manipulation of language familiarity making this a three-level variable. For some word problems all words used were real words (REAL), (e.g., all problems presented in the previous examples). The vocabulary level for these problems was

no higher than Grade 3/4. For other problems phonetically-correct one-syllable nonsense words were substituted for all nouns (NN), for example:

> Mog has 5 zorns. Blit has 4 zorns. How many zorns did Mog and Blit have?

Finally, in other problems phonetically-correct one-syllable nonsense words were substituted for all nouns and verbs (all content words) (NNV); for example:

> Poom wamped 3 glits. Poom wamped 8 blips. Glits and blips are vocks. How many vocks did Poom wamp?

All nonsense words displayed the appropriate grammatical markers; that is, "_s" denoted plural nouns and "_ed" denoted verbs in the past tense.

Each word problem simultaneously represented one level of each of the four task variables. For example, the following problem is described as having extraneous information in the first statement (EI1), having a complex subject and simple object (CSSO), having nonsense nouns (NN), and requiring grouping (GRP).

> Nom wanted 9 gorps. Flan ate 6 gorps. Lonk ate 3 gorps. How many gorps did Nom, Flan and Lonk eat?

Similarly, the following problem has extraneous information in the third statement (EI3), has a simple subject and a complex object (SSCO), uses real words (REAL), and requires reclassification (RC).

> The child picked 9 tulips. The child picked 2 roses. The child smelled 7 daisies. Tulips, roses and daisies are flowers. How many flowers did the child pick?

The word problems were hand-scored by the researcher as were all the instruments. Each individual problem solution was required to meet two criteria before it was counted as "correct." The first criterion was based on the correct choice of the numbers appropriate to problem-solution as opposed to the accurateness of the actual computation; that is, for all problems students were required, despite the simple nature of the computations involved, to show their work in the space provided and it was on the basis of the numbers extracted from the problem that this first criterion was judged. Therefore, careless errors in computation were not allowed to confound the results. The second requirement for a problem to be considered correct was based on the selection of the appropriate object set. Students were instructed not to answer with an isolated numeral but to include the appropriate noun set (e.g., "9 apples" or "They had 9 apples" as opposed to simply "9"). Prior to beginning the first problem set students were trained on practice items to use the work space and to answer with what they considered the appropriate object set. To the best of the researcher's judgment all students understood the task requirements before administration of Problem Set 1. Instructions on how to answer were repeated prior to administering Problem Sets 2 and 3.

Based on the above scoring procedure rule induction rate and ability were inferred from subjects' performance on the three problem sets. Differential pass rules of 100%, 90% and 80% were established and students' level of mastery was assessed and compared on the three problem sets. Rule induction rate was construed as the relative problem set on which criterion was reached. (Criterion was set at 90% correct, but only after a post hoc analysis of the subjects' overall performance according to the pass rules was conducted.) Rule induction ability was equated with the attainment of mastery (90% correct) regardless of the problem set on which it occurred.

The same rule for choosing the correct numbers and the accurate noun set was appropriate across all problem types. The extraneous number set always was found in the statement containing the different verb; that is, information could be excluded on the basis of a dissimilar action set. The correct noun set always could be found in the interrogative statement following the words "How many." The somewhat ambiguous nature of the word problems was recognized and the scoring procedure was deliberately flexible to allow for several possible interpretations. Thus, for "grouping" problems containing extraneous information and complex objects (36 of 144 problems) it was possible to eliminate one of the three members of the compound set on the basis of the "verb" rule used to identify extraneous information. Inclusion or exclusion of this extraneous member of the compound set did not affect correctness. Similarly, "reclassification" problems with complex objects could be answered using the given superordinate or by reiterating the subordinates (54 of 144 problems), again, both ways of answering were acceptable. It was assumed that a certain rate and level of mastery could be identified and equated with rule induction.

Counting Span Test. Subjects' information processing capacity was determined by performance on the Counting Span Test (Case, Kurland & Goldberg, 1982), which is essentially two measures, counting speed and counting span, the latter being of greater interst in the present study. Due to time constraints only the students in one of the two classes (n=34) were administered these measures. The speed test is a measure of "operational efficiency" (Case, 1985; Case et al., 1982) while the span test is technically an M-space measure and not a test of short-term memory span (Case et al., 1982; Pascual-Leone, 1970). M-space tests require transformation of information prior to retrieval and not just straight recall (Case et al., 1982). Case (1982; 1985) has found the two measures, counting speed and counting span, to be highly correlated. Both tests are individually administered and require the subject to count dots on cards. Operational speed is calculated from the total time taken to accurately count all the dots on eight cards (50 dots), yielding a speed in milliseconds. M-space is calculated by determining the span of counting operations which can be stored and recalled. On a total of 30 trials, the subject is required to count and store dots per card on a number of cards ranging from one to six before retrieving the sequence of counting products. M-space is the largest set on which the subject can recall the counting products of the complete card set on at least three of the five trials administered, and thus resultant scores can range from one to six. (See Appendix B for sample counting span trials.)

Despite the work of Case and his colleagues, no formal reliability or validity data is available for either measure. However, Lennox (1985) found M-space to be a reasonable predictor of reading ability in elementary students (Grades 3 to 5).

<u>Students' Perception of Ability Scale</u>. Level of self-concept in the academic situation was measured on the <u>Students' Perception of Ability Scale</u> (SPAS) (Boersma & Chapman, 1977). The commercially-available scale consists of 70 items to be answered "yes" and "no". It yields a full scale score and six subscale scores (General ability, arithmetic, reading/spelling, penmanship/neatness, school satisfaction, and confidence). Whether an answer of "yes" or "no" indicates high or low academic self-concept is item-specific and a scoring key is provided by the publisher.

The SPAS would appear to be reliable. Boersma and Chapman have reported a full scale Cronbach's alpha of .915, with subtest alphas between .686 and .855. Full scale test-retest reliability has been reported at .834, with subscale values ranging from .714 to .824. They also have suggested that moderate correlations with several academic achievement measures indicates construct validity.

Test Anxiety Scale for Children. Subjects' level of test anxiety was measured on the Test Anxiety Scale for Children (TASC) (Sarason, Davidson, Lighthall, Waite & Ruebush, 1960). The TASC consists of 30 forced-choice items read by the examiner to which subjects respond by circling "yes" or "no" on the supplied answer sheet. Level of test anxiety is the number of "yes" responses (Dusek, 1980). (See Appendix C for TASC questions and sample answer sheet.)

Because highly defensive children often will not admit to feelings of anxiety, the Defensiveness Scale for Children (DSC) (Sarason, Hill & Zimbardo, 1964) was administered as well as the TASC. Consisting of 38 examiner-read, "yes" - "no" items, including an 11 item lie scale, the subjects' level of defensiveness if the number of "no" responses. Those subjects exhibiting extreme defensiveness can be eliminated, if need be, from the sample (Dusek, 1980). In this case three subjects scoring two or more S.D.'s above the mean on the DSC were eliminated from the data analyses involving the test anxiety variable. (See Appendix D for DSC questions and sample answer sheet.)

Widely used, both the TASC and DSC, also would seem to be reliable. Ruebush (1963) has reported TASC internal consistency ranging from .82 to .89 and test-retest reliability from .71 to .82. For the DSC he has reported internal consistency at .82, but has given no test-retest reliability coefficient.

Procedures

The data was collected in two formats, large group (entire class) and individual subject administration. All instruments were administered by the researcher following the delivery of standardized introductory and instructional protocols appropriate to the particular instrument. Protocols may be found in Appendix E. The three sets of word problems were given on consecutive days and in the class situation. The two Grade 6 classes completed the same problem set on the same day. Problems were presented in booklet form, four problems to the page. (See Appendix F for sample pages from test booklets.) Each session began with several practice problems. No time constraints were imposed and all subjects were able to complete each of the three sets unrushed within one hour.

The SPAS, TASC, and DSC, in that order, were administered, one daily, prior to each of the three problem sets. The SPAS comes in booklet form from the publisher and subjects can read each item silently and mark the appropriate response. The TASC and DSC, as mentioned earlier, are designed to be examiner-read with subjects circling "yes" and "no" for each item on the answer sheets provided. Each of these instruments took approximately 15 minutes to complete.

During the week following the completion of the other instruments the two measures included in the <u>Counting Span Test</u> were administered individually in a private setting to a subset of the sample ($\underline{n}$=34). Each individual administration took approximately 30 minutes. Upon conclusion of data collection both classes received a brief explanation of the nature of the project and were invited to question the researcher. The entire body of data was collected within a two week period. To prevent researcher bias no data was examined until the data collection was complete.

Hypotheses

The following hypotheses were generated prior to conducting the research study.

H1 Subjects will perform differentially on a word problem rule induction task in two ways:

H1a Some subjects will fail to induce an appropriate rule.

H1b Rule induction will occur at different rates for different subjects.

H2 Rule induction ability and rate will not differ significantly as a function of gender.

H2a Males and females will show no significant differences in their ability to induce an appropriate rule.

H2b Males and females will show no significant differences in their rate of rule induction.

H3 When subjects are assigned to sub-groups on the basis of performance on learner characteristic measures, rule induction will occur differentially for sub-groups within a given characteristic.

H3a Memory span sub-groups will perform differentially across the problem sets, with subjects having a larger memory span out-performing those with a lower span.

H3b Academic self-concept sub-groups will perform differentially across the problem sets, with subjects having a high academic self-concept out-performing those with a low self-concept.

H3c Test anxiety sub-groups will perform differentially across the problem sets with subjects who are low test anxious out-performing those who are high anxious.

H4 When subjects are assigned to learner characteristic sub-groups, the relationship between the performance of those sub-groups will change as a function of task factor manipulations.

H4a The relationship between performance levels of memory span sub-groups will vary across manipulations in the presence and placement of extraneous information, the set complexity, the set language, and the level of language familiarity.

H4b The relationship between performance levels of academic self-concept sub-groups will vary across manipulations in the presence and placement of extraneous information, the set complexity, the set language, and the level of language familiarity.

H4c The relationship between performance levels of test anxiety sub-groups will vary across manipulations in the presence and placement of extraneous information, the set complexity, the set language, and the level of language familiarity.

H5 When considered together, memory span, academic self-concept, and test-anxiety levels will account for variations in total performance across subjects and in their performance on each of the three problem sets.

Chapter 4

RESULTS

Because of the ceiling effect inherent in a rule induction paradigm and also the effects of the wide dispersion of subjects within some learner characteristic groups for some dependent variables in the present study, analyzing the data was not straightforward. First, the ceiling effect produced badly skewed distributions which a log-linear transformation of the raw data did not correct (Atkinson, 1985; Bartlett, 1947; Wetherill, 1981). Second, the large differences in group variances resulted in the homogeneity of variance assumptions being failed for the multivariate analyses. Descriptive statistics (means and standard deviations) for the dependent and independent variables are given in Appendix G, Table 3. How these data analysis problems were addressed will be discussed as the procedures become relevant to the variables under discussion. All statistical analyses were conducted using the SPSS-x: Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner & Bent, 1985).

Reliabilities

Before any analyses relative to the hypotheses were conducted, odd-even reliability for subjects' performance each on the problem sets were calculated. Only odd-even measures of reliability were considered logical in that differences in split half and alternate form scores were anticipated. The odd-even reliabilities for each of the three problem Sets were as follows: (a) for Problem Set 1, .94, (b) for Problem set 2, .93, and (c) for Problem Set 3, .96.

Two measures of internal consistency also were calculated for subjects' performance on the Students' Perception of Ability Scale, The Test Anxiety Scale for Children, and the Defensive Scale for Children. The SPAS had an odd-even reliability of .88 and a split-half reliability of .84. Odd-even and split-half reliabilities for the TASC were .80 and .75, respectively. For the DSC, the odd-even reliability was .82 and the split half was .79. All six correlations were consistent with those obtained by the instrument developers. No internal reliabilities were conducted for the Counting Span Test as it did not seem reasonable to do so given the nature of the measure. In the CST, trials are randomized and only 5 sets of trials are administered. Thus, given the format, computing measures of internal consistency would not seem logical.

Rule Induction Performance on Word Problems For Entire Sample

Given that all problems were randomized and were relationally isomorphic regardless of task factor manipulations and that all but one subject maintained or improved performance over the three problem sets, subjects' scores on the final set were taken as indicative of rule induction occurrence over the three administrations. It was assumed that rule induction should result in perfect or near perfect performance; that is, if a subject had abstracted a procedural rule he would apply it uniformly across problems. Therefore, for the purpose of this study, four performance groups were defined as those students scoring (a) 100%, (b) 90-99%, (c) 80-89%, or (d) less than 80%. In establishing such performance intervals it was hypothesized that subjects would perform differentially and assumed that students in the 100% and 90-99% groups likely had recognized the isomorphic nature of the problems and had identified and applied an appropriate rule consistently across items. Subjects obtaining 90-99% correct were included in this mastery group in the belief that even good problem solvers make occasional random errors. It also was assumed a priori that if subjects scored 80-89% correct they may not have recognized the analogous relational structure of all the problems and may not have developed a generalized procedural rule, although it was difficult to speculate about the rule induction status of this group. Finally, it was assumed that those subjects who responded correctly to less than 80% of the items on the third problem set probably had failed to identify the isomorphic status of the problems, and consequently, had not induced or applied an appropriate rule. The hypothesis of differential performance and rule induction ability was supported by the range in students' scores (from a low of 17% correct to a high of 100%) on the third problem set. Given that the most common score (mode) on this set was 48 (100%) correct and that the standard error of measurement associated with any score was ± 1.75, it was possible with a pre-set level of confidence to identify subjects whose "true" score could not have been 48. When intervals around the observed scores of 8 subjects scoring 42 or less were established using three standard errors of measurement ($p < .01$), the intervals failed to include the possibility that their true score was 48. Using the performance groups to describe the dispersion of scores of the 57 subjects, 28 scored 100%, 17 scored 90-99%, 2 scored 80-89%, and 10 scored less than 80%. Thus, while the majority of subjects (45 scored 90% correct or better) would appear to have recognized the analogous structural relationship of the problems, there was a small though distinct number of subjects (10 or perhaps 12 depending on how one characterizes the 2 individuals in the 80-89% correct group) who after the three exposures still may have failed to realize that the word problems were all isomorphic. Therefore, given the failure of some subjects to reach the performance criterion, Hypothesis 1a, that some subjects would fail to induce a rule over the 144 analogous word problems was accepted.

It was also hypothesized that rule induction group membership would change over time, i.e., that rule induction would not occur simultaneously for all subjects who did ultimately achieve mastery. In Table 4 the actual number of subjects in each performance group for each problem set (PS1, PS2 and PS3) are provided. In Figures 1 to 3 the same information is presented graphically. An analysis of these data revealed significant overall effects for group membership over time, x^2 (6, $\underline{N}$ = 57) = 46.00, ($\underline{p}$ < .001). From this result and the tracing of individual subjects' performances over the three days came two important observations. First, there were essentially only two performance groups, consisting of subjects scoring 90% or above and those scoring less than 80% correct on any of the three problem sets. The 80-89% correct group was virtually non-existent for all three problem sets suggesting that indeed at any point there were two distinct and separate performance groups, those subjects with and those subjects without an appropriate rule applied consistently across contexts. Second, subjects' rule induction ability over the three problem sets could be characterized in one of three ways: (a) those subjects who immediately or very quickly recognized the analogous nature of the questions (i.e., scored 90% or better all

Table 4

Number of subjects by percentage of items correct for three problem sets

Percentage of Items Correct	Problem Set PS1	PS2	PS3
100%	16	22	28
90-99%	15	19	17
80-89%	4	3	2
<80%	22	13	10

N=57

problem sets) and were, therefore, faced with an exercise as opposed to a problem solving situation (n=34), (b) those subjects who failed to recognize the isomorphic basis of the problems initially, but had done so by the final problem set (i.e., scored less than 80% on the first problem set, but 90% or better on the last set), (n=14), and (c) those subjects who for the duration of the three-part task failed to establish the similar relational structure of the problems (i.e., always scored less than 80%) and, therefore, may have approached each question as a new problem, frequently unsuccessfully (n=10). This characterization of subjects' performance over time failed to account for the behaviour of the two subjects, those achieving in the 80-89% correct groups on the final set. Conceivably, some subjects in this 80-89% group and the less than 80% correct group would have joined those who over time developed a generalized rule if the number of trials had been increased. Also, the second group, the "improvers," could further be subdivided on the basis

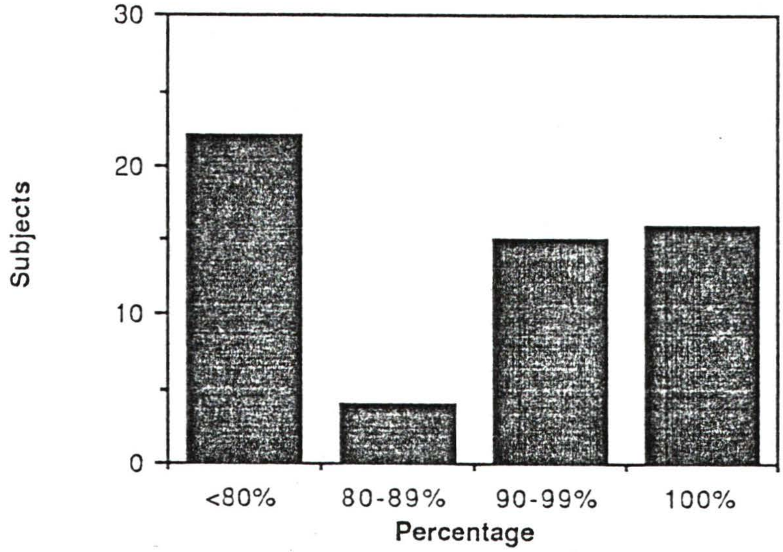


Figure 1. Subjects Grouped by Percentage Correct on Problem Set One.

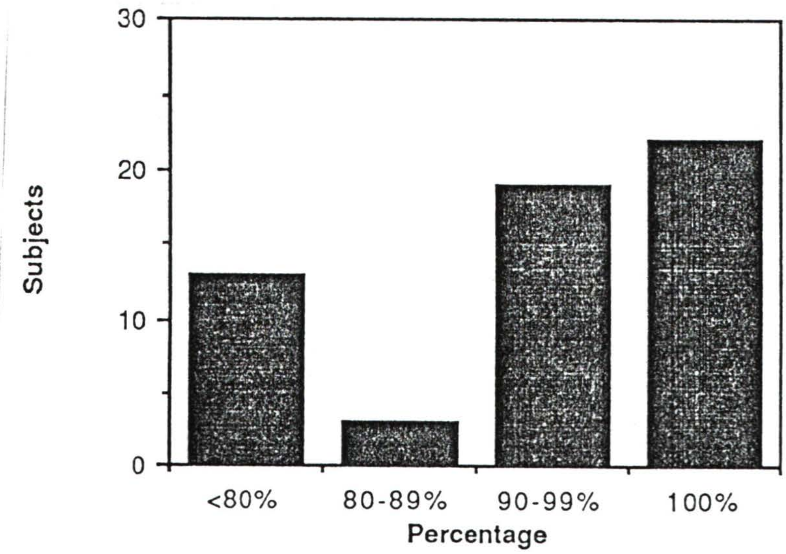


Figure 2. Subjects Grouped by Percentage Correct on Problem Set Two.

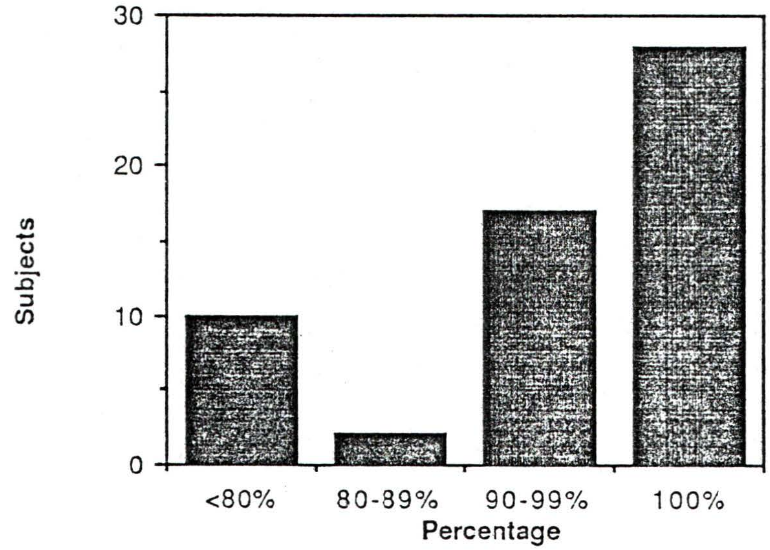


Figure 3. Subjects Grouped by Percentage Correct on Problem Set Three.

of their performance on PS2, that is, the length of exposure necessary for mastery. There was a group who lacked mastery initially, but reached the 90% or better criterion by PS2 ($\underline{n}$=9). Another group did not achieve mastery until the final problem set ($\underline{n}$=5). On the basis of these observations, Hypothesis 1b, that students induce a rule at differential rates, also was accepted.

Interestingly, a post hoc analysis of subjects' split-half performance on PS1 (the first 24 problems compared to the last 24 problems in the set) similarly supported this assumption of differential induction rates. These data are reported in Table 5. An analysis of these data revealed significant differences in group membership over the split-halves, $x^2(3, \underline{N} = 57) = 46.37$, ($\underline{p} < .001$). Again, too, the 80-89% correct group for both halves was small or non-existent relative to the other performance groups.

Table 5

Number of subjects by percentage of items correct over split halves of Problem Set One

Percentage of Items Correct	PSI 1st Half	PSI 2nd Half
100%	18	26
90-99%	14	7
80-89%	0	4
<80%	25	20

$\underline{N}$=57

Rule Induction Performance by Gender

It was hypothesized that rule induction ability and rate would not be differentiated by gender. The assumption regarding non-differentiated ability across males and females was supported on the basis of subjects performance on PS3, $\underline{t}(55) = -0.27$, $p > .20$. However, an informal analysis of the gender data across the three problem sets has suggested differences in rate of induction. In

Table 6 the number of subjects by gender and performance group for each of the problem sets are given. It can be noted that if groups are collapsed into those scoring 90% correct or above and those scoring below 90%, as in Table 7, the female subjects who eventually score 90% or above do so by PS2, whereas that group continues to grow for male subjects over the three problem sets. Collapsing the data in such a manner can be justified on the basis of the assumptions underlying the initial creation of the performance groups, that is, that perfect or near perfect performance (e.g., a score of 90% or better) would be considered indicative of rule induction occurrence. Therefore, it would appear that, whereas males and females ultimately performed at the same level, they achieved that level of performance at different rates. Thus, Hypothesis 2a and the assumption of no gender differences in ultimate rule induction was supported. However, Hypothesis 2b and the assumption of no such differences in rate of induction was not supported by the results.

Table 6

Number of subjects by gender and performance group for three problem sets

Percentage of Items Correct	PS1		PS2		PS3	
	F	M	F	M	F	M
100%	10	6	13	9	15	13
90-99%	7	8	8	11	6	11
80-89%	0	4	1	2	1	1
<80%	9	13	4	9	4	6

_n_(females)=26 _n_(males)=31

Table 7

Number of subjects by gender scoring 90% correct or above and below 90% for the three problem sets

Percentage of Items Correct	PS1 F	PS1 M	PS2 F	PS2 M	PS3 F	PS3 M
$\leq$90%	17	14	21	20	21	24
>90%	9	17	5	11	5	7

n(females)=26 n(males)=31

Performance by Learner Characteristic Groups

Before discussing the analyses involving learning characteristic groups, two issues are addressed briefly. First, the procedures by which subjects were assigned to groups are delineated. Second, the rationales for the basic statistical approaches are introduced. Following these discussions, an hypothesis by hypothesis, reporting of the results is presented.

Assignment of Subjects to Groups. Subjects were assigned to groups on the basis of their performance on three learner characteristic measure, the CST, SPAS, and TASC. The fourth such measure, the DSC, was administered only to exclude highly defensive subjects from the test anxiety analyses. By virtue of their performance on the Counting Span Test, subjects were assigned to memory span or M-space groups. On the CST they demonstrated M-spaces of two (MSP2) (n = 3), three (MSP3) (n = 11), four (MSP4) (n = 12), or five (MSP5) (n = 8). Such a memory span range in the age group under consideration would seem to be consistent with the results obtained by Case and his colleagues (1982).

Three equal-sized academic self-perception groups (n = 19) were generated from subjects' SPAS scores. The low-scoring third of the subjects were assumed to have low self-perception (ASP1), the next third were assumed to have average self-perception (ASP2), and the final third were considered to have high academic self-perception (ASP3). In forming the ASP groups if the cut-off score between two groups was shared by two more subjects, the statistical program randomly divided those subjects between the groups.

Similarly, subjects were assigned to three equal-sized test anxiety groups (n = 18) on the basis of their performance on the TASC. The reduced n was the result of the exclusion three subjects who scored more than 2 SD above the sample mean on the DSC. The one-third of the subjects scoring highest on the TASC were assumed to be high anxious (TA1), the next third, mid anxious (TA2), and the low-scoring third, low anxious (TA3). Again, for a shared score at a cut-off point, groups were formed by random assignment of those subjects to either the higher or lower group in the same manner in which the ASP groups were created.

Basic Approaches to Data Analyses. Given the theoretical nature of the M-space construct, (i.e., that each level is discrete from every other level) only non-parametric analyses were used to examine these data. Also, given that there are no commonly-used multivariate non-parametric analogs, all analyses originated as univariate procedures on performance by M-space groups (Results given in Appendix G, Table 8) and were followed by individual group comparisons as indicated. Some nonstatistical observations also were made.

Overall analyses of variance for academic self-perception and test anxiety groups over total performance, the individual problem sets, and the task variables were completed both parametrically on the raw and the transformed data and then non-parametrically because the homogeneity of variance assumption was not met. Regardless of procedure used, the results were found to be significant almost without exception. (See Tables 9 and 10 for the parametric results.) Due to the failure to meet the homogeneity of variance assumptions, however, post hoc analyses using the pooled variance estimate were not conducted. Therefore, to further investigate the specific relationships between groups, two procedures were applied to the data. First, the variances at the 95% confidence interval were compared for each group over each level of the dependent variables, allowing the relationship between probable group means to be observed. (See Tables 11 to 15 in Appendix G for the confidence intervals.) Second, $\underline{t}$-tests were run using the separate variance estimate procedure in which it is assumed that homogeneity of variance has not been met. To reduce the probability of spurious results due to completing multiple $\underline{t}$-tests two confidence levels were set a priori at $\underline{p} < .01$ and $\underline{p} < .001$. Again, some nonstatistical observations were made.

Table 9

Multivariate analyses for task variables by learner characteristic groups

Task Variable	df	ASP Groups F	TA Groups F
Extraneous Information	8	2.86**	2.78**
Set Complexity	6	3.68**	1.54*
Set Language	4	5.83+	3.06*
Language Familiarity	6	3.80**	2.23*
Problem Sets	6	5.38+	2.19*

*p<.05 **p<.01 +p<.001

Table 10

Univariate analyses of levels of the dependent variables by learner characteristic

	Variable Levels	ASP Groups F	TA Groups F
Extraneous Information	NEI	8.38**	4.38
	E11	10.39**	5.96*
	E12	12.86**	6.15*
	E13	10.73**	4.50
Set Complexity	CSSO	10.33**	4.38
	SSCO	11.30**	5.23*
	CSCO	12.46**	6.90*
Set Language	GRP	11.10**	6.25*
	RC	11.35**	4.59
Language Familiarity	REAL	8.29**	4.53
	NN	10.39**	5.03*
	NNV	12.29**	5.52*
Problem Sets	PS1	8.55**	5.02*
	PS2	15.45**	5.02*
	PS3	6.86*	5.09*
Total Performance		10.48**	5.19*

df = (2,54) *p<.01 **p<.001

Problem Solving and Rule Induction Performance by Learner Characteristic Groups

Memory Span. As described earlier a subset of the sample (n=34) demonstrated memory spans of 2 (n=3), 3 (n=11), 4 (n=12), and 5(n=8). The performance levels of the four MSP groups across the three problems sets is given in Table 16. Chi-squares, for each of the problem sets were significant as follows: (a) for PS1, $X^2(9, n=34) = 20.44$, ($p < .02$), (b) for PS2, $X^2(9, n=34) = 26.74$, ($p < .002$), and (c) for PS3, $X^2(9, n=34) = 18.82$, ($p < .03$). However, these results must be interpreted cautiously as the assumption of expected frequency per cell was often violated. Nonetheless, a significant interaction between memory span and performance level on the word problems for all sets is suggested, though the specific source of the interaction is not revealed by the procedure.

A less formal and nonstatistical approach to these data would seem to suggest that memory span was more related to performance on the initial problem set, than on the second and third sets. All 3 subjects in the MSP2 group scored less than 80% on PS1 and 9 of 11 subjects in the MSP3 group scored less than 90%. However, of the 20 subjects with memory spans of 4 or 5 all but three scored 90% correct or better. By PS3 these contrasts in performance had moderated somewhat and the 90% correct demarcation no longer seemed to have differentiated MSP2 and 3 subjects from MSP4 and 5 subjects with such clarity. Many, though not all, subjects in the two lower memory span groups did reach the 90% criterion for rule induction on the final problem sets. Thus, Hypothesis 3a would seem to have been supported, though it is recognized that memory span may not affect rule induction ability if length of exposure to the task is not

Table 16

Number of subjects by memory span groups and performance level

Percentage of Items Correct	MSP2[a]	MSP3[b]	MSP4[c]	MSP5[d]
Problem Set One				
100%			6	4
90-99%		2	4	3
89-89%		2	1	
<80%	3	7	1	1
Problem Set Two				
100%		1	7	6
90-99%	1	5	5	2
89-89%	1			
<80%	1	5		
Problem Set Three				
100%		1	8	7
90-99%	2	6	3	1
80-89%		1	1	
<80%	1	3		

[a]*n* = 3 [b]*n* = 11 [c]*n* = 12 [d]*n* = 8

constrained. It does, however, appear to differentiate individuals on the basis of rate of induction, where larger memory spans may be equated with more rapid induction.

The results of Kruskall-Wallis ANOVA's computed on performance scores by MSP groups for each of the problem sets can be seen in Appendix G, Table 8. From these results it is apparent that the memory span groups did function differently on all sets. These data also would suggest that for these word problems rather than four memory span groups as differentiated by the CST there were actually only two groups. The first consisted of subjects with MSP of two or three and the second subject with MSP of four or five. Individual group comparisons supported this observation. With one exception ordinally adjacent groups were never significantly different from each on any problem set. Comparisons between MSP3 and MSP4, the only exception, were always significantly different, as were all comparisons involving non-adjacent groups. Post hoc analyses of the data using the combined span groups, MSP2/3 and MSP4/5, as the independent variable produced all significant comparisons. Thus, Hypothesis 3a as originally stated was rejected. The contention that memory span would differentiate performance on word problems did find some support, but only for combined lower (MSP2/3) and higher (MSP4/5) memory span groups.

Academic self-concept. The performance levels of subjects in the three equal-n academic self-perception groups were compared for each of the three problem sets. Table 17 contains those data. Again chi-squares revealed significant interactions between groups and performance levels as follows: (a) for PS1, $X^2(6, N=57) = 19.76$, ($p < .004$), (b) for PS2, $X^2(6, N=57) = 28.80$, ($p < .001$), and (c) for PS3, $X^2(6, \underline{N}=57) = 15.02$, ($\underline{p} < .03$). Again these results must be viewed cautiously due to the violation of expected frequencies within some cells and

Table 17

Number of subjects by academic self-concept group and performance level

Percentage of Items Correct	ASP1[a]	ASP2[a]	ASP3[a]
		Problem Set One	
100%		6	10
90-99%	4	6	5
80-89%	1	2	1
<80%	14	5	3
		Problem Set Two	
100%	1	8	13
90-99%	5	8	6
80-89%	3		
<80%	10	3	
		Problem Set Three	
100%	5	10	13
90-99%	5	6	6
89-89%	1	1	
<80%	8	2	

[a] $\underline{n}$ = 19.

because in the procedure used the major source of the interaction is not indicated.

Therefore, as with memory span, a more informal approach to the data seemed revealing. Viewing the tables directly would suggest that academic self-concept appears to have been less strongly associated with rate of rule induction than with ability to induce rules, as the low academic self-concept (ASP1) group still split virtually evenly on either side of the 90% correct criterion for mastery on PS3. However, ASP1 subjects, generally, were both less likely and slower to reach criterion.

A multivariate analysis of performance on the three problem sets by the academic self-perception groups revealed significant overall effects. (See Table 9.) Univariate analyses for performance by groups on each of the problem sets were also significant, as shown in Table 10. Individual comparisons of ASP groups for PS1, PS2, and PS3 (using $\underline{t}$-tests with separate variance estimates) revealed significant differences between ASP1 (low self-concept) and ASP3 (high self-concept) for PS1 and PS2 ($\underline{p} < .001$) and also for PS3 ($\underline{p} < .01$). For ASP1 and ASP2 (moderate self-concept) significant differences ($\underline{p} < .01$) were found for PS1 and PS2. On PS3, however, differences between these two groups, according to the a priori setting of confidence levels, were non-significant, $\underline{t}$ (21.79) = -2.27, $\underline{p}$ = .033. For ASP2 and ASP3 comparisons revealed no real differences in performance on any problem set, with all $\underline{t}$-values having $\underline{p}$-levels greater than .01.

Thus, on the basis of these analyses Hypothesis 3b, that academic self-perception groups would perform differentially across the problem sets was rejected, though it did find support in the performance of some groups on some problem sets. Low and high self-concept groups always performed distinctly with marked differences in performance at any point in time (PS1, PS2, or PS3). In rule induction ability over the 144 word problems low and moderate self-concept groups were different on PS1 and PS2, but not PS3 suggesting differences in rule induction rate between these two groups but perhaps not ultimate ability. Moderate and high groups were never distinct, performing similarly in terms of rule induction rate and ability.

Test anxiety. The number of subjects for each of three equal-n test anxiety groups by four performance levels on the three problem sets were compared. This data is presented in Table 18. Here again chi-squares were calculated and significant interactions between groups and performance levels found as follows: (a) for PS1, $X^2(6, \underline{n}=18) = 13.32$, ($\underline{p} < .04$), (b) for PS2, $X^2(6, \underline{n}=18) = 13.16$, ($\underline{p} < .05$), and (c) for PS3, $X^2(6, \underline{n}=18) = 16.34$, ($\underline{p} < .02$). Here, too, the same caveats regarding interpretation expressed in the two previous sections apply.

Once again a non-statistical approach to the data was illuminating. Low anxiety (TA1) would appear to have been strongly associated with relatively good performance on all problem sets and thus with early rule induction occurrence. Moderate levels of test anxiety (TA2) may have been only weakly associated with performance level as evidenced by the dispersion of TA2 subjects over the performance levels, particularly on PS1. The relatively static performance of TA2 subjects over the problem sets, especially on PS2 and 3 would suggest a stable relationship between moderate test anxiety and rule induction over time though the influence for specific individuals may have varied. High anxiety (TA3) would

Table 18

Number of subjects by test anxiety group and performance level

Percentage of Items Correct	TA1[a]	TA2[a]	TA3[a]
		Problem Set One	
100%	9	4	3
90-99%	6	5	4
80-89%	1	3	
<80%	3	7	12
		Problem Set Two	
100%	12	6	4
90-99%	6	8	5
80-89%		1	2
<80%	1	4	8
		Problem Set Three	
100%	15	7	6
90-99%	4	8	5
80-89%			2
<80%		4	6

[a] $\underline{n}$ = 18

appear to have influenced significantly performance at least initially, with 12 of 19 TA3 subjects scoring less than 80% on PS1. On PS2 and PS3 the effects of high anxiety appear to have become more individual-specific.

A multivariate analysis of performance on the three problem sets by the three test anxiety groups revealed significant overall effects, though not at the highly significant level associated with academic self-perception. (See Table 9.) Univariate analyses of groups by individual problem sets were similarly significant, as shown in Table 10. Again, individual group comparisons for each problem set were calculated using t-tests with separate variance estimates. The only comparison which was significant at $p<.01$ was that between the TA1 (low anxious) group and the TA3 (high anxious) group on PS1, which supports the contention concerning the early influence of high levels of test anxiety. All other group comparisons on all problem sets revealed no distinct differences in group performances. Therefore, Hypothesis 3c, that test anxiety groups would perform differentially across the problems was unsubstantiated except for the specific case mentioned above and was rejected. Though low test anxiety was associated with rapid rule induction, level of test anxiety could not differentiate rule induction rate or ability for subjects in high or moderate anxiety groups.

Task Variable Manipulations by Learner Characteristic Groups

Memory Span. Kruskal-Wallis One-Way ANOVAs on each task variable level (four extraneous information levels, three set complexity levels, two set language levels, and three language familiarity levels as shown in Appendix G, Table 8) by memory span group showed significant overall differences between groups for every level. Mann-Whitney U-tests comparing every group individually with every other group revealed a consistent pattern of significant and nonsignificant relationships. The following pairs of MSP groups were always significantly different ($p<.01$) for every task variable level: MSP2 and 4, MSP2 and 5, and MSP3 and 5. Comparisons between MSP2 and 3 and MSP4 and 5 were not significant. The only remaining comparison, MSP3 with MSP4, was always significant except for on word problems containing no extraneous information (NEI), $U = 28.5$, $p = .012$ (corrected for ties), given the a priori standards for significance set at $p < .01$.

In these results only minimal support was found for Hypothesis 4a, that the relationships between the performance of memory span groups would change as a function of variations in task factors. Thus, it was rejected as initially stated. Differences in group relationships for specific task factor levels frequently would appear to have been due more to general performance differences than to task variable manipulations except for the one case noted above. Differences and similarities in group performances seem to exist regardless of the variables in the word problems, with essentially only two distinct memory span groups, MSP2/3 and MSP4/5. This echoes the observation made earlier in the analyses of the performance of MSP groups over the problem sets.

Academic self-perception. Multivariate analyses of variance for the four task variables (extraneous information, set complexity, set language, and language familiarity) by academic self-perception groups resulted in significant overall effects (Wilks' Lambda) as shown in Table 9. Univariate analyses for each level of the four task variables were conducted also (Table 10), revealing overall groups effects for every level.

Individual ASP group comparisons (*t*-tests using separate variance estimate) for each task factor level suggested a consistent pattern in the relationship between group performance, largely regardless of task variations. ASP1 (low self-concept) and ASP3 (high self-concept) groups were always different on all levels of the dependent variables ($p<.001$ in all cases except for problems containing REAL language where the comparisons were significant at $p<.01$), suggesting a general performance difference between these two groups. ASP1 and ASP2 (moderate self-concept) groups almost always performed differently ($p<.01$), except for on problems with no extraneous information (NEI) $t(32.66) = -2.41$, $p = .022$, and on problems using REAL language, $t(23.25) = -2.59$, $p = .016$. The performance comparisons for ASP2 and ASP3 groups were never significantly different ($p>.01$ in all cases), irrespective of any task factor variations.

These results gave only minimal support to Hypothesis 4b, that the relationship between the performance of academic self-perceptions groups would vary differentially across task factor manipulations. Thus, the hypothesis as originally postulated was rejected. Significant differences were found for some groups on some task factor levels, but not between all groups or all levels. However, here again, actual group relationships on any given level were most likely due to general performance differences or similarities and rarely to variations in the task. Only the relationship between ASP1 (low) and ASP2 (moderate) groups may have varied in response to task variable manipulations. Another observation would suggest that on these word problems the performance of the low ASP group was virtually always different from either of the other ASP groups; that is, low ASP would seem to have been strongly associated with a relatively generalized low level of performance. Membership in the other ASP groups did not have a similar power to differentiate between average and high levels of performance on the word problems.

Test Anxiety. Multivariate analyses of the performance on the levels of each of the task variables by test anxiety groups were all significant at $p<.05$ or better (Table 9). Univariate analyses for each task factor level by groups revealed several levels for which overall group differences were non-signficant. (See Table 10.) These included problems with no extraneous information (NEI), problems with a complex subject but simple object (CSSO), problems which reclassified subordinate sets under a superordinate set (RC) and problems using all real words (REAL). Thus, the influence of test anxiety level may have varied with task factor manipulations.

Individual group comparisons (again t-tests with separate variance estimates) were calcualted for those task variable levels for which the univariate analyses were significant. Few individual comparisons between groups revealed significant performance differences ($p<.01$). Only comparisons between TA1 (low anxiety) group and TA3 (high anxiety) group were significant, with the low anxious group outperforming the high anxious group and then not for every task factor level. The performance of these two groups was not significantly different for problems in which the extraneous information was presented in the third statement (EI3), $t(20.28) = 2.39$, $p = .027$. They were different ($p<.01$) (a) on problems in which the extraneous information was presented first (EI1) or second (EI2), (b) on problems with a simple subject and complex object (SSCO) and also those with both a complex subject and object (CSCO), (c) on problems with nonsense nouns (NN) and those with nonsense nouns and verbs (NNV), and finally (d) on problems in which elements of complex sets were grouped together using the co-ordinate conjunction "and" (GRP). No comparisons for any other two groups were ever significant.

Thus, Hypothesis 4c, that the relationship between the performance of test anxiety groups would vary differentially across task factor manipulations received more support than either 4a or 4b. Again, however, that support was not unconditional, though level of test anxiety would seem to have interacted more with the task variations than did either memory span level or academic self-perception level, and may have contributed less to generalized differences in performance. Significant differences frequently were observed between the performance low anxious (TA1) group and the high anxious group (TA3). Having moderate levels of test anxiety (TA2) though, appears to have had little influence on general performance or on interactions with task factor manipulations. Subjects with moderate test anxiety seem to have performed on an individually-specific basis.

The Predictive Value of Three Learner Characteristics

It was hypothesized that memory span, academic self-concept, and test anxiety levels in combination would be reasonable predictors of performance on a rule induction task consisting of 144 arithmetic word problems. The hypothesis was subdivided to consider the overall predictive nature of the independent variables and also their predictative ability relative to each of the problem sets (PS1, PS2, and PS3). Consequently, four stepwise multiple regression analyses were computed entering the three learner characteristics as possible predictor variables and performance on the appropriate word problems as the criterion variable. As memory span data was calculated for only 34 subjects and not the entire sample of 57, only the scores for those subjects for whom a complete set of learner characteristic data were available were included in the regression analyses.

Using performance scores on PS1 as the predictor variable, MSP level accounted for the greatest amount of shared variance and thus loaded first. Adding ASP level to the regression equation significantly increased the amount of variance accounted for by the other two learner characteristics. TA level, however, was not added as doing so would not increase the percentage of variance significantly. (See Table 19.)

There was a change in the loading order of the predictor variables when scores on PS2 were used as the criterion variable. Here ASP level accounted for the largest amount of shared variance. MSP level, which loaded second, further increased the value of the adjusted R^2 significantly. Again, however, TA was not included in the regression equation as it would not account for a significant amount of variance not already accounted for by self-concept and memory span. (See Table 20.)

The regression analysis in which performance on PS3 was used the criterion variable produced only one significant predictor variable, ASP level. Neither MSP or TA level could increase the percentage of shared variance accounted for in any significant way. (See Table 21.)

Table 19

Multiple stepwise regression analyses on PS1 by learner characteristics

Variable entered on Step 1 - MSP

Multiple R	0.560		Percentage of variance accounted for: 29.3%
R Square	0.314	F=14.65**	
Adjusted R Square	0.293		

Variable entered on Step 2 - ASPGRP

Multiple R	0.632		Additional percentage of variance accounted for: 6.7%
R Square	0.399	F=10.29**	
Adjusted R Square	0.360		

**p<.001

The results of the multivariate regression analysis, in which the criterion variable was the total performance scores across all 144 word problems were, not surprisingly, very similar to the results obtained for performance on PS2. ASP level accounted for most of the shared variance. Adding MSP level further increased the amount of variance significantly. Yet again, adding TA level would not have increased the predictive power of the equation and thus test anxiety was not included in the stepwise regression analysis. (See Table 22.)

The adjusted R^2's in the four analyses after all significant predictor variables had been loaded, varied from a low of .228 for PS3 and a high of .411 for total performance suggesting a moderate predicative ability for some of the

Table 20

Multiple stepwise regression analyses on PS2 by learner characteristics

Variable entered on Step 1 - ASPGRP

Multiple R	0.585		Percentage of variance accounted for: 32.2%
R Square	0.342	F=16.65**	

Variable entered on Step 2 - MSP

Multiple R	0.662		Additional percentage of variance accounted for: 8.0%
R Square	0.439	F=12.11**	
Adjusted R Square	0.402		

**p<.001

Table 21

Multiple stepwise regression analyses on PS3 by learner characteristics

Variable entered on Step 1 - ASPGRP

Multiple R	0.501		Percentage of variance accounted for: 22.8%
R Square	0.251	F=10.72*	
Adjusted R Square	0.228		

*p<.01

Table 22

Multiple stepwise regression analyses on total performance by learner characteristics

Variable entered on Step 1 - ASPGRP

Multiple R	0.560		Percentage of variance
R Square	0.314	F=14.66**	accounted for: 29.3%
Adjusted R Square	0.293		

Variable entered in Step 2 - MSP

Multiple R	0.669		Additional percentage
R Square	0.447	F=12.52**	of variance accounted for:
Adjusted R Square	0.411		11.8%

**p<.001

learner characteristic variables on some portions of the task. Academic self-concept was always implicated in a predictive capacity, with memory span becoming less predictive over time. Test anxiety was never a predictive factor.

Summary of Results

Significant differences in rule induction ability and rate were found across the sample. At least 10 subjects failed to abstract an appropriate rule according to the imposed criterion, (i.e., scored less than 80% correct on the third and final problem set). Three rule induction rate groups were identified; those subjects who immediately or very rapidly reached mastery (90% correct or better), those who did not have a rule initially but eventually did reach mastery, and those who failed to find an appropriate procedural rule. Although differences in rule induction rate by gender were found, with females' induction rate exceeding that of males, no gender differences in ultimate rule induction ability over the three problem sets were found.

Memory span differentiated performance on the three problem sets in a decreasingly powerful way; that is, it may have affected rule induction rate more than ultimate ability on the specific word problem task. For the specific word problem task used in the study, subjects effectively fell into two groups whose performance was virtually always differentiated: those with a memory span of two or three and those with a memory span of four or five. Post hoc analyses on the combined groups verified this conclusion. This pattern of performance by the combined memory span groups was maintained regardless of task factor manipulations except for word problems containing no extraneous information where performance was not significantly differentiated.

Academic self-perception was associated more strongly with rule induction ability than rate of induction, particularly for those subjects with poor self-concepts, who almost always performed more poorly than subjects with average or good self-concepts. This pattern of differentiated behaviour by the low academic self-perception group was found for two of the three problem sets and for most task variable manipulations. The low academic self-perception group performed similarly to the moderate self-perception group only on the final problem set, on word problems containing no extraneous information and on problems using only real words. The low group always remained differentiated from the high group.

The influence of test anxiety was less easily characterized, although high levels of test anxiety would seem to have influenced subjects' performance on the initial problem set, where the low anxious group outperformed the high anxious group. All other group comparisons across problem sets were nonsignificant. Of the three learner characteristics, test anxiety was the most responsive to task factor manipulations in that there was no generally predictable group performance patterns which existed independent of the task variable levels. Any significant differences on given task factor levels were found only between low and high anxious groups. Moderate levels of test anxiety seems to have had no consistent group relationship to performance on the word problems.

Two of the three learner characteristics, academic self-concept and memory span were moderately effective as combined predictors of performance. The power of academic self-concept as a predictor was maintained over the three problem sets, whereas memory span's predictive ability decreased over time.

Statement of Limitations

The preciseness and the generalizability of the results reported here may be affected by several theoretical and practical limitations which should be acknowledged. First, in operationalizing the concept of rule induction it was equated with level of performance. Thus, rule induction occurrence was inferred. The inference was based on the assumption that rule induction inextricably linked to another more observable behaviour, accuracy of performance. Therefore, the validity of the interpretations concerning rule induction rely on the validity of the underlying assumption. Second, the study suffers from the problems inherent in a correlational design. Cause and effect relationships between variables cannot be assumed. Interpretations must always allow for the possibility that any correlations between variables are due to mutual correlations with a third unidentified variable and do not reflect direct influence. Third, the convenient nature of the sampling procedure must be construed as affecting the generalizability of the results. The sample under study may not have been typical of any other or a more random group of Grade 6 students. Fourth, while using the nonsense words did create novel tasks and reduced the confound of individual differences in vocabulary development, it also was a source of some levity which may have affected the seriousness with which subjects approached the word problems. Similarly, the knowledge that this was a research project and the results would not in any way affect subjects' academic standing may have influenced the motivational level with which the problems were approached. Fifth, the limited number of trials may have resulted in somewhat premature conclusions which may or may not have been borne out over a larger number of trials. Finally, the data analyses suffered from several problems already alluded to, (i.e., some badly skewed distributions and widely discrepant variances between groups).

Chapter 5

DISCUSSION

Rule induction in the context of word problems was investigated as an extension of the solution-finding process in which procedural rules are generalized across a class of word problems identified as isomorphic. The purpose in the study was to (a) compare Grade 6 students' ability to induce an appropriate procedural rule on 144 isomorphic word problems presented in three sets of 48 problems each and the rate at which a rule is induced, (b) further examine their rule induction ability and rate in terms of differences in gender, memory span, academic self-perception and test anxiety, (c) assess the influence of some task variables on students' performance relative to their level of memory span, academic self-perception and test anxiety, and (d) determine the predictive power of these combined learner characteristics in relation to students' performance on the word problems. The exploratory nature of the study is recognized and the results are acknowledged as less than definitive. The results, however, may increase our understanding of the process of rule induction in the context of word problems and some constraints on that process. Similarly, some relevant research and practical eductional implications may be derived from the obtained results.

Interpretation of the Results

Rule induction ability and rate. Based on their levels of performance on the third problem set, the Grade 6 students included in the study appear to have varied in their ability to induce a rule over the 144 analogous problems, as 20% of subjects failed to reach 90% criterion. From an examination of students' performance levels on all three problem sets it would seem that for those students who did induce a rule there were variations in their rate of induction. The number of subjects in the 90% correct or better performance group continued to increase substantially within the first problem set and over the three problem sets. Thus, it is suggested that Grade 6 students may differ both in their rule induction ability and rate. As noted earlier, it is conceivable that with an increased number of trials some members of the non-mastery group on Problem Set 3 may have eventually induced a rule. Therefore, differences in the final level of performance may be more the result of the experimental design than of any real differences.

From these data it was observed that at no time was there a significant number of subjects achieving at the 80-89% level. Subjects' performance on all problem sets and on both halves of the first set fell into two distinct and separate groups, those who scored 90% or better and those scoring less than 80%. This may indicate that the occurrence of rule induction is associated with a marked, rather than a gradual improvement in performance. The mastery criterion for this particular task was 90% correct, somewhat higher than some proponents of mastery learning might have suggested (Bloom, 1976). Rule induction, then, was associated with near perfect performance, in that having an appropriate procedural rule may have reduced the difficulty of the specific word problems greatly, increasing the likelihood of accurate solution-finding; that is, the problems had become exercises.

From an examination of the influences of four learner characteristics (gender, memory span, academic self-concept, and test anxiety) some possible correlates of rule induction ability and rate may be identified.

No differences in final rule induction ability were found by gender. Males and females were equally capable of inducing a rule and reaching mastery, as determined by their non-differentiated performance on Problem Set 3. There were, however, gender differences in the rate of rule induction. Females' performance plateaued more quickly than males; that is, those female subjects who did achieve mastery did so by the second problem set, with their performance remaining stable over the third set. Males' performance never plateaued, but continued to improve over the three problem sets. Thus, differential rates by gender are suggested. From a slightly different perspective, however, this gender data may be interpreted to suggest that females achieve their ultimate stable performance level, whatever that may be, more rapidly than males, but that the performance of the latter group continues to improve. Therefore, again, it must be acknowledged that the conclusion that no differences in induction ability existed may be a function of the actual number of trials administered. Either interpretation though, gives support to the conclusion that these Grade 6 students did exhibit differential induction rates by gender on the word problems. The gender differences may be explained by some other unidentified variable or interaction of variables associated with the task (e.g., whether the students construed the problem as essentially a reading or an arithmetic task may have influenced the observed differences in the performance of females and males).

When the rule induction data were examined relative to subjects' memory spans, some interesting patterns appeared to emerge. Memory span was strongly associated with performance on Problem Set 1, less so on Problem Set 2, and even less so on Problem Set 3. Larger memory spans (i.e., of four or five) were highly associated with mastery performance regardless of problem set. Smaller spans (two or three) are strongly related to lack of mastery on the first problem set, but are increasingly less influential on the succeeding two sets. Thus, for subjects having a relatively low span, size of span would seem to have influenced their starting level of performance but was unrelated to the level they ultimately achieved. More than half the subjects with a memory span of three or less had reached criterion (90%) by the final problem set. Memory span, then, is perhaps more related to rate of induction than to absolute achievement. This conclusion is consistent with the literature which suggests that smaller functional memory spans do not preclude learning, but rather make it more difficult and slower (Bachor, 1976; Case, 1978).

From the memory span data on rule induction performance came another consistent observation. For performance on the word problem task there were effectively only two memory span groups, a lower group (those with spans of two or three) and a higher group (those with spans of four or five). This differentiation between groups may imply either that the general memory demands of the word problems or that the demand of the strategy most students were applying during problem solving was four. Since it was impossible, because of the research design, to ascertain the specific nature of the memory demand of the problems or the strategy most typically used, such an interpretation must remain conjecture.

A somewhat different pattern was observed when the rule induction data were interpreted relative to the subjects' level of academic self-perception. Relatively stable performances by groups across the three problem sets were observed. This led to the conclusion that self-perception may be more significantly associated with determining or reflecting the level of rule induction ability than with the rate of induction. Whether academic self-concept was a cause or a manifestation of rule induction ability on the word problems remains undetermined.

Individual group comparisons revealed that the low self-perception group performed consistently and significantly more poorly than the high self-perception group across problem sets. The performance of low and average self-perception group were similarly related across the first two problem sets. On Problem Set 3, however, the performance of the low group approached that of the average group. The low self-concept group, then, was associated with an almost generalized poorer performance relative to the other groups. Low academic self-concept also would seem to be associated with relatively slow increments in performance over time, again suggesting a greater link to rule induction ability than to rate of induction.

No consistent pattern emerged when the rule induction data were assessed in terms of the subjects' level of test anxiety. The only individual group difference in performance existed between the low and high anxiety groups and only on the first problem set. This was in agreement with Heinrich and

Spielberger's (1982) contention that on tasks of intermediate difficulty high anxiety is only detrimental to performance early in the learning process. Level of test anxiety bore only individual-specific relationships to rule induction ability or rate. No generalized relationship relative to performance level across the word problems existed, an observation which would again find some support in the test anxiety literature (Sieber, 1980; Tobias, 1979).

The Predictive Nature of the Learner Characteristics. On total performance over the 144 word problems academic self-perception proved to be the best predictor of performance. When self-perception and memory span were considered together they became moderately powerful predictors accounting for 41% of the variance in subjects' performance on the word problems. Test anxiety was unable to account for variations in total performance above that already accounted for by academic self-perception and memory span.

More important to the rule induction paradigm were the results of the multiple regression analyses for each of the problem sets. For Problem Set 1 memory span was the most significant predictor with academic self-perception loading second in the regression equation. Test anxiety was not a significant factor in predicting performance on this or any other set. For Problem Set 2 self-perception was the most significant predictor with the addition of memory span increasing the predictive power of the equation. On Problem Set 3 only academic self-concept was a significant predictor.

These data may be interpreted as suggesting that memory span level was more influential in determining rate of rule induction than an ultimate level of performance on the word problems. Academic self-concept, however, may have been associated more highly with achievement level. Finally, the influence of test anxiety level would seem to be more individual-specific and not associated with a generalized induction ability or rate.

Not surprisingly, these regression analyses results confirmed the earlier interpretations of the rule induction data relative to the individual learner characteristics (memory span, academic self-concept, and test anxiety). Memory span may determine rate of learning, but not what can be learned if time constraints are not imposed. The complexity of this particular rule induction task resulted in slower rates of induction for subjects with lower spans but did not preclude induction. Academic self-concept, whether a reflection or a determinant of ability, may be significantly related to manifested ability level. Academic self-concept, then, may be strongly associated with what will be learned. Finally, as has been suggested in the literature (Heinrich & Spielberger, 1982; Spielberger, 1966), the influence of test anxiety may be more responsive to interactions between specific task demands and the individual's ability relative to the task than to some overall learning rate or level of achievement.

<u>The Response of Learner Characteristic Groups to Task Variable Manipulation</u>. Despite the isomorphic relational structure of the word problems, some task variables which were hypothesized to be possible influences on problem difficulty were manipulated. It was anticipated that the groups for each learner characteristic (memory span, academic self-perception, and test anxiety) would perform differentially as the variables were manipulated (i.e., that certain task variables might be identified as impeding rule induction for some groups). This was not, however, generally the case.

For memory span, differences between individual groups were either always significant or never significant regardless of the task variable levels. This pattern had only one exception. When the group with a memory span of three was compared to the group with a span of four they performed differently on all task variable levels except on problems containing no extraneous information on which they performed similarly. This exception would appear to be logical theoretically, in that those problems containing no extraneous information would seem to have made less demands on working memory in that they had one fewer pieces of information to be processed than did problems with extraneous information. Other than the case just noted, the differences in performance between individuals with spans greater than three and those with spans less than four were generalized and overwhelmed any statistically significant relationships in performance which might have been the result of task variable manipulations.

The response of academic self-perception groups to task factor variations was similar to that of the memory span groups. Here, too, most comparisons followed a consistent pattern revealing significant differences between the low self-perception group and both other groups and no differences in performance between the moderate and high self-perception groups. Two exceptions to this generalized pattern existed: The performance of low and moderate academic self-perception groups were not different on problems containing no extraneous information or on problems containing only real words. If academic self-concept is a reflection of some real ability level, these exceptions can be explained if both of these problem types can be construed as requiring less ability than do other types. If academic self-concept is a determinant of ability level, these exceptions may have occurred because these problems were perceived by all subjects as relatively easy and within their ability. But, again, the effects of task variations on group comparisons was over-powered by the more generalized differences in group performances.

Test anxiety level seems to have been somewhat more responsive to task variable manipulations and no overwhelming performance relationships between groups emerged. All three test anxiety groups (low, moderate and high) performed similarly on problems with no extraneous information, problems with a complex subject and simple object, problems in which subordinate sets were reclassified under a given superordinate, and problems containing only real words. It already has been suggested that problems with no extraneous information and those using only real words may have been perceived as easier and/or actually were easier than problems containing other levels of the same task variable. The same argument can be made for the remaining two cases relative to the other levels of the respective task variables. Problems with a complex subject and a simple object are the only problems in which the simple set is the quantified set, which may have made such problems easier than those in which the quanitified set was a complex set. In problems in which sets were reclassified under a superordinate, some of the necessary information was actually "chunked" (Miller, 1956), which may have effectively reduced problem difficulty for some subjects, if they took advantage of this externally imposed organizational information. Thus, under task conditions which are perceived as easy or are in reality easy for individual subjects, level of anxiety may be irrelevant.

While the mid anxious group was never differentiated from either of the other test anxiety groups, the low and high test anxiety groups did perform differently for all other variable levels with one exception. These two groups did perform similarly on problems in which the extraneous information was presented in the third statement, just prior to the posing of the question. This result perhaps may be explained in terms of the contiguity of the two essential information statements (Goodstein, 1981) and that they were processed before the extraneous information was encountered, though this conjecture is somewhat speculative. However, if one had speculated a priori on the basis of theory and previous research, precisely which task variable levels might have proved easiest, the levels nondifferentiated by performance across test anxiety groups would have been the most likely choices.

For two of the three learner characteristics considered, memory span and academic self-concept, the power of the level of the characteristic to differentiate performance on the word problems was pervasive. The conclusion that the effects of learner differences may supercede the effects of task variations is consistent with the findings in some other areas of learning (e.g., reading comprehension and questioning research (Peterson & Swing, 1983; Raphael & Pearson, 1985; Raphael & Wonnacott, 1986) . To use Reed's (1987) terminology "base specificity" may have been more relevant than degree of problem "transparency" in determining the recognition of the isomorphic status of the word problems and the induction and application of a generalized procedural rule. Only when the performance data were examined in terms of test anxiety level did problem "transparency" appear to become an issue in performance.

Implications for Future Research

Whereas the concept of rule induction has been explored from several different perspectives, rarely has it been examined in the context of classroom-like tasks, such as word problems. Therefore, there would seem to be a need in the rule induction research to mesh the theoretical base with the practical educational situation. This study was an attempt to address that need in some small way.

Future research in the area could pursue a variety of directions, only a few of which are suggested. First, a similar study with an extended number of trials might address the issue of what happens to that group of students who had not abstracted a rule after three trials. Second, the present study could be replicated with students at other grade levels to assess some developmental trends in rule induction ability. Do differences in induction rate and ability increase or decrease with age? Third, other learner characteristics (e.g., cognitive style factors) could be used as the independent variables on a rule induction task involving word problems. Thus, further insights into the nature of the individual differences which affect rule abstraction could be identified. Fourth, other task variables which might affect the "transparency" of the isomorphic status of a class of word problems, could be manipulated. Were the results in the present study produced by choosing the wrong variables for manipulation? Finally, several other research designs might be employed to further investigate rule induction in the context of word problems. Reed's (1987) approach in which university students "typed" algebra word problems, might be used with arithmetic word problems, to evaluate the identification process involved in rule induction in the absence of rule application. Through the use of "think aloud" protocols obtained while individual students solve isomorphic word problems, rule induction might become an observable behaviour and not merely inferred on the basis of correct responses.

Whatever the specific focus in future word problem solving research, task, subject, and process variables must be integrated if a reasonably accurate picture of the complex activity is to emerge. In addition, attention should be given to the role of rule induction in the facilitation of the problem solving process and to how it functions in effectively transforming word problems into exercises.

Educational Implications

From the results of the present study come several implications for the discovery learning approach to word problems in the classroom. First, such a learning approach may be effective for most students but seemingly not for all. Some individuals who failed to reach criterion on the task as it existed may have benefited from feedback during their performance. Others may only have reached criterion after direct instruction in the appropriate procedural rule. Second, given enough exposure to a word problem task, a student's initial inability to perform to criterion may be unrelated to his/her final level of perfomance. Starting off relatively poorly does not preclude rule abstraction from ultimately occurring. Thus, differential rates of learning must be accounted for in the instructional design. Third, characteristics, other than reading and arithmetic ability, which students bring to the word problem solving situation may influence directly what is learned and how guickly it is learned. Therefore, it might be helpful to assess those students having difficulties with word problems in terms of more than their reading and arithmetic ability if we are to understand the nature of their difficulties. Fourth, given that most students can induce appropriate rules, the instructional habit of following exercises in double-digit multiplication with word problems all of which require double-digit multiplication in the solution-finding procedure, would seem counterproductive and fails to give students the opportunity to use those inductive skills in solving word problems. Word problems probably should not be presented already "typed", but rather exemplars of multiple types should be presented for comparison and analyses, allowing students to discover the relational and procedural differences and similarities which define the parameters of an isomorphic class of word problems.

Conclusion

In much of the word problem research the emphasis has been on the identification of problem types. The nature of specific typologies and the guidelines used to establish problem types have reflected the theoretical bias of the particular researcher(s). Classifying word problems would also seem to be a valuable skill for students. If problem typing is based on the identification of isomorphic problems, that is, those with the same relational structure, it may encourage the abstraction of generalized procedural rules which then can be applied across the class of analogous problems. Transforming classes of problems into exercises via rule induction would seem to be both a logical and natural extension of the problem solving process. As was found in this study, the majority of students probably can induce appropriate procedural rules for isomorphic problems, though the rate at which they do so may vary. Both the ability to induce rules and the rate at which they are induced may be influenced more by some characteristics of the problem-solvers than by variations in the surface structure of the word problems. These influential learner characteristics may include more than reading and computational ability. Further investigation of the interaction between subject and task variables may clarify this relationship and help to explain students' performance on word problems in the classroom.

REFERENCES

Atkinson, A. C. (1985). Plots, transformations and regressions. Oxford: Clarendon Press.

Bachelder, B. L., & Denny, M. R. (1977a). A theory of intelligence: I. Span and the complexity of stimulus control. Intelligence, 1, 127-150.

Bacheler, B. C., & Denny, M. R. (1977b). A theory of intelligence: II. The role of span in a variety of intellectual tasks. Intelligence, 1, 237-256.

Bachor, D. G. (1976). Information processing capacity and teachability of low achieving students. Unpublished doctoral dissertation. University of Toronto. (O.I.S.E.)

Bachor, D. G. (1979). Suggestions for modifications in testing low-achieving adolescents. Journal of Special Education, 13(4), 443-452.

Bachor, D. G. (1985). Questions for and by the learning disabled student. In J. F. Cawley (Ed.), Mathematics for the learning disabled (pp. 183-208). Rockville, MD: Aspen Corporation.

Bachor, D. G. (1987, July). Towards a taxonomy of word problems. Paper presented at the Annual Meeting of the International Group for the Psychology of Mathematics Education, Montreal, Quebec.

Bachor, D. G., Steacy, N., & Freeze, D. R. (1986, June). Factors contributing to students' failure of word problems: Some preliminary results. Paper presented at the Annual Conference of the Canadian Society for Studies in Education, Winnipeg, Manitoba.

Bachor, D. G., & Steacy, N. (1987, June). The importance of language shifts and extraneous information in determining word problem difficulty. Paper presented at the Annual Conference of the Canadian Society for Studies in Education, Hamilton, Ontario.

Balow, I. H. (1964). Reading and computation ability as determinants of problem solving. Arithmetic Teacher, 12, 18-22.

Bartlett, M. S. (1947). The use of transformations. Biometrics, 3, 39-52.

Bilsky, L. H., & Judd, T. (1986). Sources of difficulty in the solution of verbal arithmetic problems by mentally retarded and non-retarded individuals. American Journal of Mental Deficiency, 90, 395-402.

Blakenship, C. A., & Lovitt, T. C. (1976). Story problems: Merely confusing or downright befuddling? Journal for Research in Mathematics Education, 7, 290-298.

Bloom, B. S. (1976). Human characteristics and school learning. New York: McGraw-Hill.

Boersma, F. J., & Chapman, J. W. (1977). Students' Perception of Ability Scale. Edmonton, Alta: PsiCan Consulting Ltd., University of Alberta.

Branca, N. A. (1980). Problem solving as a goal process and basic skill. In S. Krulik (Ed.), Problem solving in school mathematics: 1980 Yearbook (pp. 3-8). Reston, VI: NCTM.

Brookover, W. B., & Gotlieb, D. (1964). Sociology of education. New York: American Book Co.

Caldwell, J. H., & Goldin, G. A. (1979). Variables affecting problem difficulty in elementary school mathematics. Journal for Research in Mathematics Education, 10, 323-336.

Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Lindquist, M. M., & Reys, R. E. (1980). Solving verbal problems: Results and implications from national assessments. Arithmetic Teacher, 28, 8-12.

Carpenter, T. P., & Moser, J. M. (1983). The acquisition of addition and subtraction concepts. In R. Lesh & M. Landhu (Eds.), Acquisition of mathematics concepts and processes (pp. 7-44). New York: Academic Press.

Case, R. (1974a). Structures and strictures, some functional limitations on the course of cognitive growth. Cognitive Psychology, 6, 544-573.

Case, R. (1978a). A developmentally based theory and technology of instruction. Review of Educational Research, 48, 439-469.

Case, R. (1978b). Intellectual development from birth to adulthood: A neo-Piagetian interpretation. In R. S. Seigler (Ed.), Children's thinking: What develops? (pp. 37-72). Hillsdale, NJ: Erlbaum.

Case, R. (1985). Intellectual development: Birth to adulthood. New York: Academic Press.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. Journal of Experimental Child Psychology, 33, 386-404.

Cawley, J. F., Fitzmaurice, A. M., Shaw, R. A., Kahn, H., & Bates, H. (1979). Math word problems: Suggestions for LD students. Learning Disabilities Quarterly, 2, 35-41.

Chi, M. T. H. (1976). Short term memory limitations in children: Capacity or processing deficits? Memory and Cognition, 23, 266-281.

Chi, M. T. H., & Klahr, D. (1975). Span and rate of apprehension in children and adults. Journal of Experimental Child Psychology, 19, 434-439.

Cruickshank, W. M. (1948). Arithmetic ability of mentally retarded children: Ability to differentiate extraneous material from needed arithmetic facts. Journal of Educational Research, 42, 161-170.

Cyert, R. M. (1980). Problem solving and educational policy. In D. Tuma & F. Rief (Eds.), Problem solving and education (pp. 3-8). Hillsdale, NJ: Lawrence Erlbaum.

De Corte, E., Verschchaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's representations and solutions. Journal of Educational Psychology, 77, 460-470.

Dempster, F. N. (1978). Memory span and short term memory: A developmental study. Journal of Experimental Child Psychology, 26, 419-431.

Dewey, J. (1910). How we think. Boston: D. C. Heath & Co.

Dusek, J. B. (1980). The development of test anxiety in children. In I. G. Sarason (Ed.), Test anxiety: Theory, research, and applications (pp. 87-110). Hillsdale, NJ: Erlbaum.

Dweck, C. S., & Wortman, C. B. (1982). Learned helplessness, anxiety, and achievement motivation. In H. W. Krohne & L. Laux (Eds.), Achievement, stress, and anxiety (pp. 92-125). Washington, D. C.: Hemisphere.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 66, 183-201.

Egan, D. E., & Greeno, J. G. (1974). Theory of rule induction: Knowledge acquired in concept learning, serial pattern learning, and problem solving. In L. W. Gregg (Ed.), Knowledge and cognition (pp. 43-104). Potomac, MD: Erlbaum.

Englert, C. S., Culatta, B. E., & Horn, D. G. (1987). Influence of irrelevant information in addition word problems on problem solving. Learning Disability Quarterly, 10, 29-36.

Freeze, D. R. (1986). A comparison of cognitive and behavioral programs to promote the transfer of learning across educational settings for four learning disabled children. Unpublished masters thesis, University of Victoria, Victoria, B. C.

Gage, N. L., & Berliner, D. C. (1979). Educational psychology (2nd ed.). Boston: Houghton Mifflin.

Gagne, R. M. (1977). The conditions of learning (3rd ed.). New York: Holt, Rinehart and Winston.

Gagne, R. M., & Briggs, L. J. (1974). Principles of instructional design. New York: Holt, Rinehart and Winston.

Geen, R. G. (1980). Test anxiety and cue utilization. In I. G. Sarason (Ed.), Test anxiety: Theory, research, and applications (pp. 43-61). Hillsdale, NJ: Erlbaum.

Gentner, D. (1982). Are scientific analogies metaphors? In D. S. Miall (Ed.), Metaphor: Problems and perspectives (pp.106-132). Brighton, England: Harvester Press.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. Cognitive Psychology, 12, 306-355.

Ginsburg, H., & Opper, S. (1979). Piaget's theory of intellectual development. Englewood Cliffs, NJ: Prentice-Hall.

Goodstein, H. A. (1981). Are the errors we see the true errors? Error analysis in verbal problem solving. Topics in Learning & Learning Disabilities, 1(3), 31-45.

Goodstein, H. A., Cawley, J. F., & Helfgott, J. (1971). Verbal problem solving among educable mentally retarded. American Journal of Mental Deficiency, 76, 238-241.

Greeno, J. G. (1974). Processes of learning and comprehension. In L. W. Gregg (Ed.), Knowledge and cognition. Potomac, MD: Lawrence Erlbaum.

Greeno, J. G. (1980). Analysis of understanding in problem solving. In R. Kluwe & H. Spada (Eds.), Developmental models of thinking (pp. 199-212). New York: Academic Press.

Heinrich, D. L., & Spielberger, C. D. (1982). Anxiety and complex learning. In H. W. Krohne & L. Laux (Eds.), Achievement, stress and anxiety (pp. 145-165). Washington, DC: Hemisphere.

Hoy, C. (1984). Story problems: Vehicles to develop thinking skills. Academic Therapy, 19, 567-572.

Jerman, M. (1973). Problem length as a structural variable in verbal arithmetic problems. Educational Studies in Mathematics, 5, 109-123.

Jerman, M., & Rees, R. (1972). Predicting the relative difficulty of verbal arithmetic problems. Educational Studies in Mathematics, 4, 306-323.

Jordan, J. (1978). Why Johnny can't visualize. Vector, 19(4), 59-63.

Kilpatrick, J. (1978). Research on problem solving in mathematics. School Science and Mathematics, 78, 189-192.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. Psychological Review, 92, 109-129.

Kistner, J., Haskett, M., White, K., & Robbins, F. (1987). Perceived competence and self-worth of LD and normally achieving students. Learning Disability Quarterly, 10, 37-44.

Krulik, S., & Rudnick, J. A. (1982). Teaching problem solving to preservice teachers. Arithmetic Teacher, 29(6), 42-45.

Krutetskii, V. A. (1976). The psychology of mathematical abilities in school children (J. Teller, Trans.). Chicago: University of Chicago Press.

Larson, S. C., Parker, R., & Trenholme, B. (1978). The effects of syntactic complexity upon arithmetic performance. Learning Disability Quarterly, 1(4), 80-85.

Lennox, N. D. (1985). Remedial reading: An investigation of the neurological impress method, the repeated readings method, and a combined neurological impress-repeated readings method. Unpublished masters thesis, University of Victoria, Victoria, B. C.

Lester, F. K. (1980). Research on mathematical problem solving. In R. J. Shumway (Ed.), Research in mathematics education (pp. 286-323). Reston, VI: National Council of Teachers of Mathematics.

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

Linville, W. J. (1976). Syntax, vocabulary, and the verbal arithmetic problem. School Science and Mathematics, 76, 152-158.

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical nature. Educational Psychologist, 20, 107-125.

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

Mueller, J. H. (1980). Test anxiety and the encoding and retrieval of information. In I. G. Sarason (Ed.), Test anxiety: Theory, research, and applications (pp. 63-86). Hillsdale, NJ: Erlbaum.

Muth, K. D. (1984). Solving arithmetic word problems: Role of reading and computational skills. Journal of Educational Psychology, 76, 205-210.

National Council of the Supervisors of Mathematics. (1977). Position paper on basic mathematic skills. Arithmetic Teacher, 25(1), 19-22.

Newell, A.,& & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. (1985). SPSS-x: Statistical package for the social sciences. New York: McGraw Hill.

Parill-Burnstein, M. (1981). Problem solving and learning disabilities: An information processing approach. New York: Grune & Stratton.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's development stages. Acta Psychologica, 32, 301-345.

Pascual-Leone, J. (1978). Compounds, confounds, and models in developmental psychology: A reply to Trabasso and Foellinger. Journal of Experimental Child Psychology, 26, 18-40.

Pascual-Leone, J., Goodman, D., Subelman, I., & Ammon, P. (1978) Piagetian theory and neo-Piagetian analysis as psychological guides in education. In J. M. Gallagher & J. A. Easley (Eds.), Knowledge and Development, Vol II: Piaget in Education (pp. 243-289). New York: Plenum Press.

Paul, D. J., Nibbelink, W. H., & Hoover, H. D. (1986). The effects of adjusting readability on the difficulty of mathematics story problems. Journal for Research in Mathematics Education, 17, 163-171.

Peterson, P. L., Swing, S. R. (1983). Problems in classroom implementation of cognitive strategy instruction. In M. Pressley & J. Levin (Eds.), Cognitive strategy research: Educational applications (pp. 267-284). New York: Academic Press.

Piaget, J. (1952). The origins of intelligence in children. New York: W. W. Norton & Company.

Polya, G. (1957). How to solve it (2nd ed.). Garden City, NY: Doubleday.

Potterbaum, S. M., Keith, T. Z., & Ehly, S. W. (1986). Is there a causal relation between self-concept and academic achievement? Journal of Educational Research, 79, 140-144.

Purkey, W. W. (1970). Self concept and school achievement. Englewood Cliffs, NJ: Prentice-Hall.

Raphael, T. E., & Pearson, P. D. (1985). Increasing students' awareness of sources of information for answering questions. American Educational Research Journal, 22, 217-235.

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., Dempster, A., & Ettinger, M. (1985). The 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 problme-solving ability in arithmetic. In H. P. Ginsburg (Ed.), The development of mathematical thinking (pp. 153-196). New York: Academic Press.

Rosenthal, D. J. A., & Resnick, L. B. (1974). Children's solution processes in arithmetic word problems. Journal of Educational Psychology, 66, 817-825.

Rumelhart, D. E. (1972). Introduction to Human Information Processing. New York: John Wiley & Sons.

Sarason, S. B., Davidson, K. S., LIghthall, F. F., Waite, R. R., & Ruebuch, B. K. (1960). Anxiety in elementary school children. New York: John Wiley & Sons.

Sarason, S. B., Hill, K. T., & Zimbardo, P. G. (1964). A longitudinal study of the relation of test anxiety to performance on intelligence and achievement tests. Monographs of the Society for Research in Child Development, 29, #7.

Schoenfeld, A. H., & Hermann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memeory and Cognition, 5, 484-494.

Schunk, D. H. (1985). Self-efficacy and classroom learning. Psychology in the Schools, 22, 208-223.

Sharma, M. C. (1981). Using word problems to aid language and reading comprehension. Topics in Learning and Learning Disabilities, 1(3), 61-71.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. Review of Educational Research, 46, 407-441.

Sieber, J. E. (1980). Defining test anxiety: Problems and approaches. In I. G. Sarason (Ed.), Test anxiety: Theory, research, and applications (pp. 15-40). Hillsdale, NJ: Erlbaum.

Silver, E. A. (1981). Recall of mathematical problem information: Solving related problems. Journal of Research in Mathematics Education, 12, 54-64.

Summers, G. C. (1984). Learned helplessness, test anxiety and learning disabilities: An assessment with intermediate-age children. Unpublished masters thesis, University of Victoria, Victoria, B. C.

Suydam, M. N. (1980). Untangling clues from research on problem solving. In S. Krulik (Ed.), Problem solving in school mathematics: 1980 Yearbook. Reston, VI: NCTM.

Thorndike, E. L. (1922). Psychology of arithmetic. New York: Macmillan.

Tobias, S. (1979). Anxiety research in educational psychology. Journal of Educational Psychology, 72, 573-582.

Travers, R. M. (1982). Essentials of learning. New York: Macmillan.

Treacy, J. P. (1944). The relationship of reading skills to the ability to solve arithmetic problems. Journal of Educational Research, 38, 86-96.

Vander Zanden, J. W., & Pace, A. J. (1984). Education psychology: In theory and practice (2nd ed.). New York: Random House.

Wells, L. E., & Marwell, G. (1976). Self-esteem: Its conceptualization and measurement. Beverley Hills, CA: Sage.

West, C. K., Lee, J. F., & Anderson, T. H. (1969). The influence of test anxiety on the selection of relevant from irrelevant information. Journal of Educational Research, 63, 51-52.

Wetherill, G. B. (1981). Intermediate statistical methods. New York: Chapman and Hall.

Wine, J. D. (1971). Test anxiety and direction of attention. Psychological Bulletin, 76, 92-104.

Wine, J. D. (1980). Cognitive-attentional theory of test anxiety. In I. G. Sarason (Ed.), Test anxiety: Theory, research, and applications (pp. 349-385). Hillsdale, NJ: Lawrence Erlbaum.

Winne, P. H., Woolands, M. J., & Wong, B. (1982). Comparability of self-concept among learning disabled, normal and gifted students. Journal of Learning Disabilities, 15, 470-475.

Zatz, S., & Chassin, L. (1983). Cognitions of test anxious children. Journal of Consulting and Clinical Psychology, 51, 526-534.

APPENDIX A

WORD PROBLEM CLASSIFICATIONS (RILEY, GREENO & HELLER)

Action	Static
CHANGE	COMBINE
Result unknown	Combine value unknown
1. Joe had 3 marbles. Then Tom gave him 5 more marbles. How many marbles does Joe have now?	1. Joe has 3 marbles. Tom has 5 marbles. How many marbles do they have altogether?
2. Joe had 8 marbles. Then he gave 5 marbles to Tom. How many marbles does Joe have now?	Subset unknown
	2. Joe and Tom have 8 marbles altogether. Joe has 3 marbles. How many marbles does Tom have?
Change Unknown	COMPARE
3. Joe had 3 marbles. Then Tom gave him some more marbles. Now Joe has 8 marbles. How many marbles did Tom give him?	Difference unknown
	1. Joe has 8 marbles. Tom has 5 marbles. How many marbles does Joe have more than Tom?
4. Joe had 8 marbles. Then he gave some marbles to Tom. Now Joe has 3 marbles.	2. Joe has 8 marbles.

How many marbles did
he give to Tom?

Start unknown

5. Joe had some marbles.
Then Tom gave him 5
more marbles.
Now Joe has 8 marbles.
How many marbles did
Joe have in the
beginning?

6. Joe had some marbles
Then he gave 5 marbles
to Tom.
Now Joe has 3 marbles.
How many marbles did Joe
have in the beginning?

EQUALIZING

1. Joe has 3 marbles.
Tom has 8 marbles.
What could Joe do to
have as many marbles
as Tom?
(How many marbles
does Joes Need to
have as many as Tom?)

2. Joe has 8 marbles.
Tom has 3 marbles.
What could Joe do to have
as many marbles as Tom?

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

Compared quality unknown

3. Joe has 3 marbles.
Tom has 5 more marbles
than Joe.
How many marbles does
Tom have?

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

Referent unknown

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

6. Joe has 3 marbles.
He has 5 marbles less
than Tom.
How many marbles does
Tom have?

From Riley, Greeno & Heller (1983), p. 160.

APPENDIX B

OVERVIEW OF CST, SAMPLE TRIALS, AND RECORD FORM

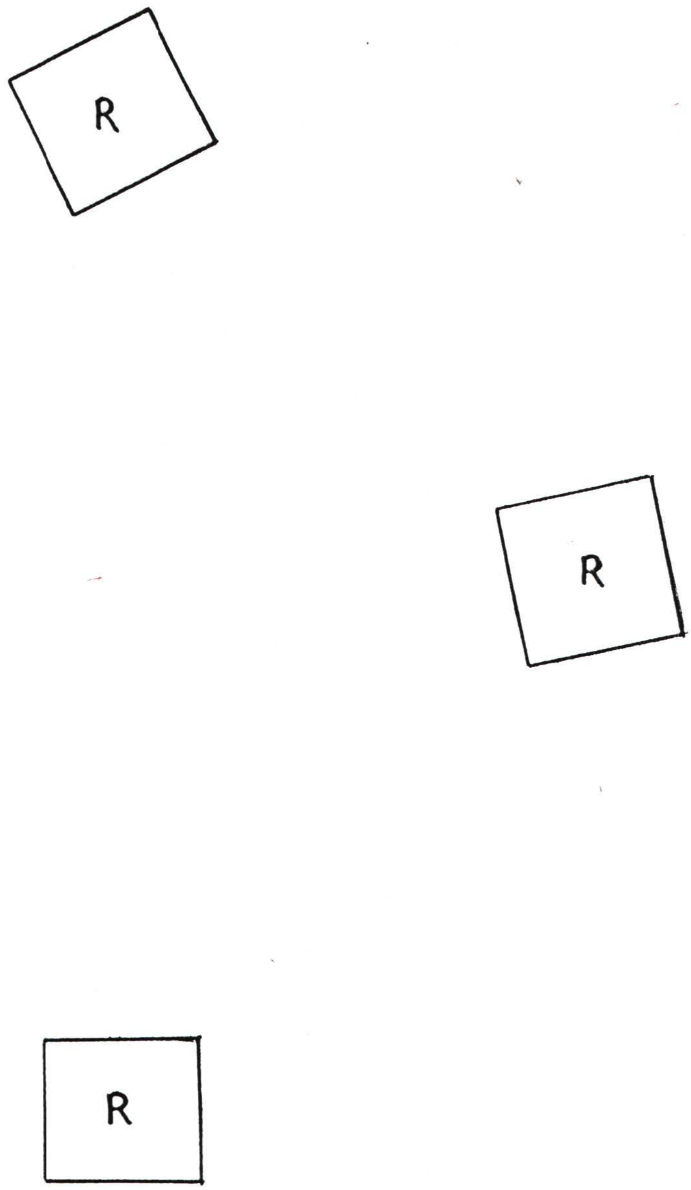
(Adapted from Case, Kurland, and Goldberg, 1982)

The Counting Span Test is designed to assess memory span or operational memory across the age range from four years to adulthood. Subjects are presented with sets of cards, 21.5 cm by 28 cm. On each card a number of orange distractor squares are interspersed with red squares. Each square is 3 cm square. The subjects' task is to count the red squares on each of the cards and then, after all the cards in the set have been counted, to recall the number of red squares that appeared on each in the order they appeared. The full test consists of 114 cards (in 5 practice trials and 30 test sets). The largest set size for which the subject can recall the ordered set with 60% accuracy (i.e., 3 out of 5 trials of that set size) is taken as his/her memory span.

A copy of one practice trial and one test trial is included as a sample. An R indicates the red square and an unmarked square indicates an orange one.

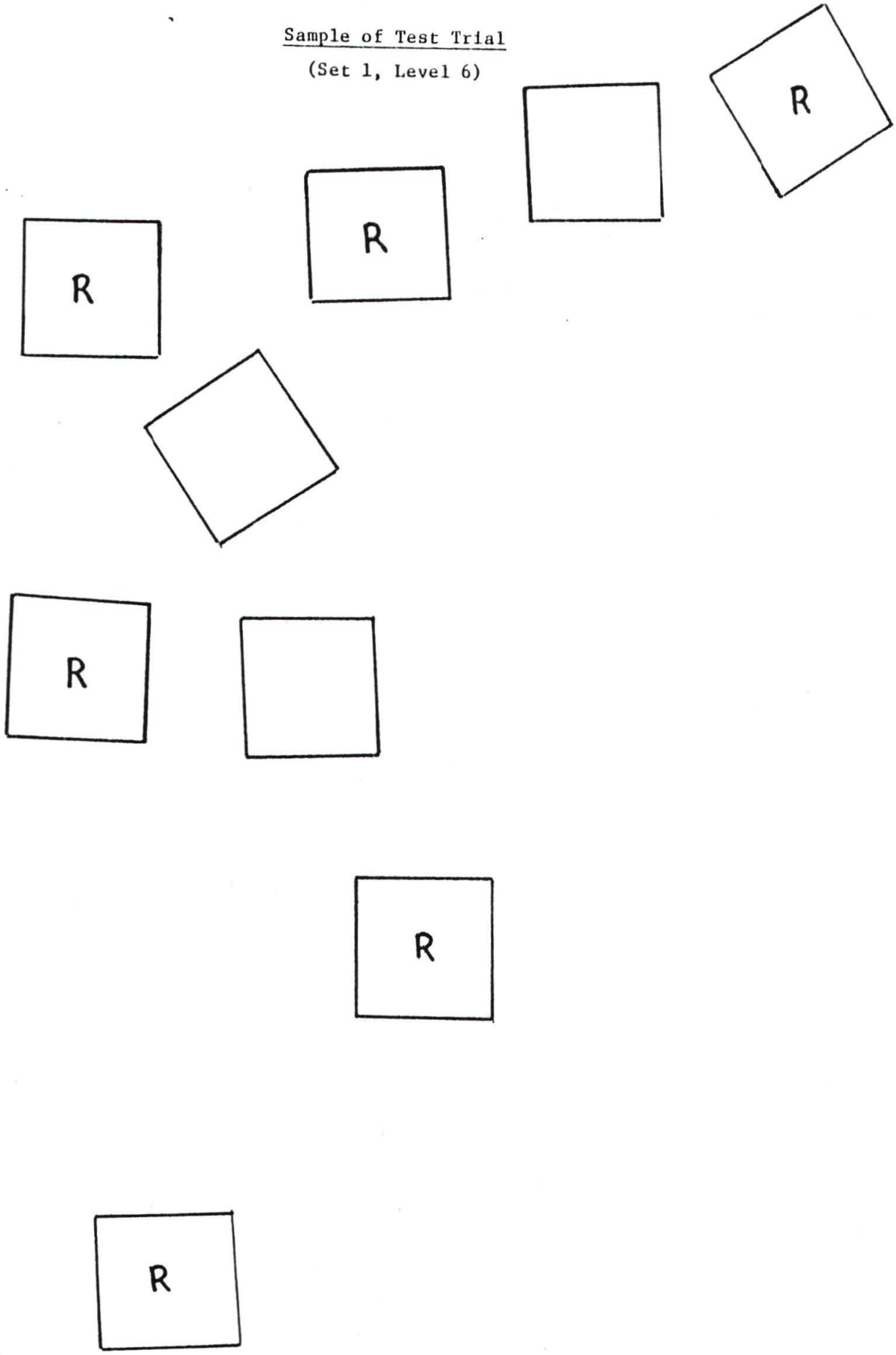
Sample Practice Trial

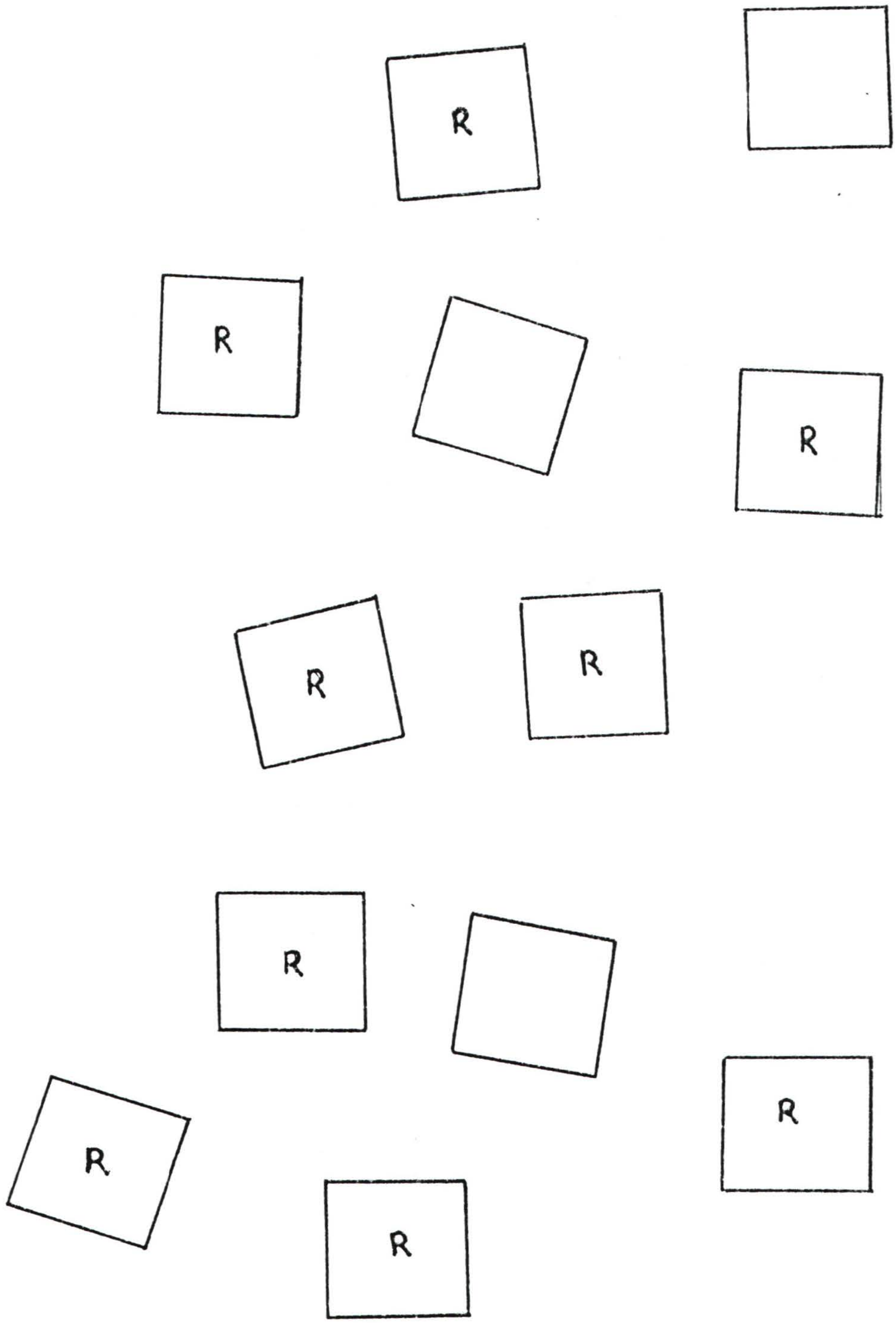
(Training Phase A)

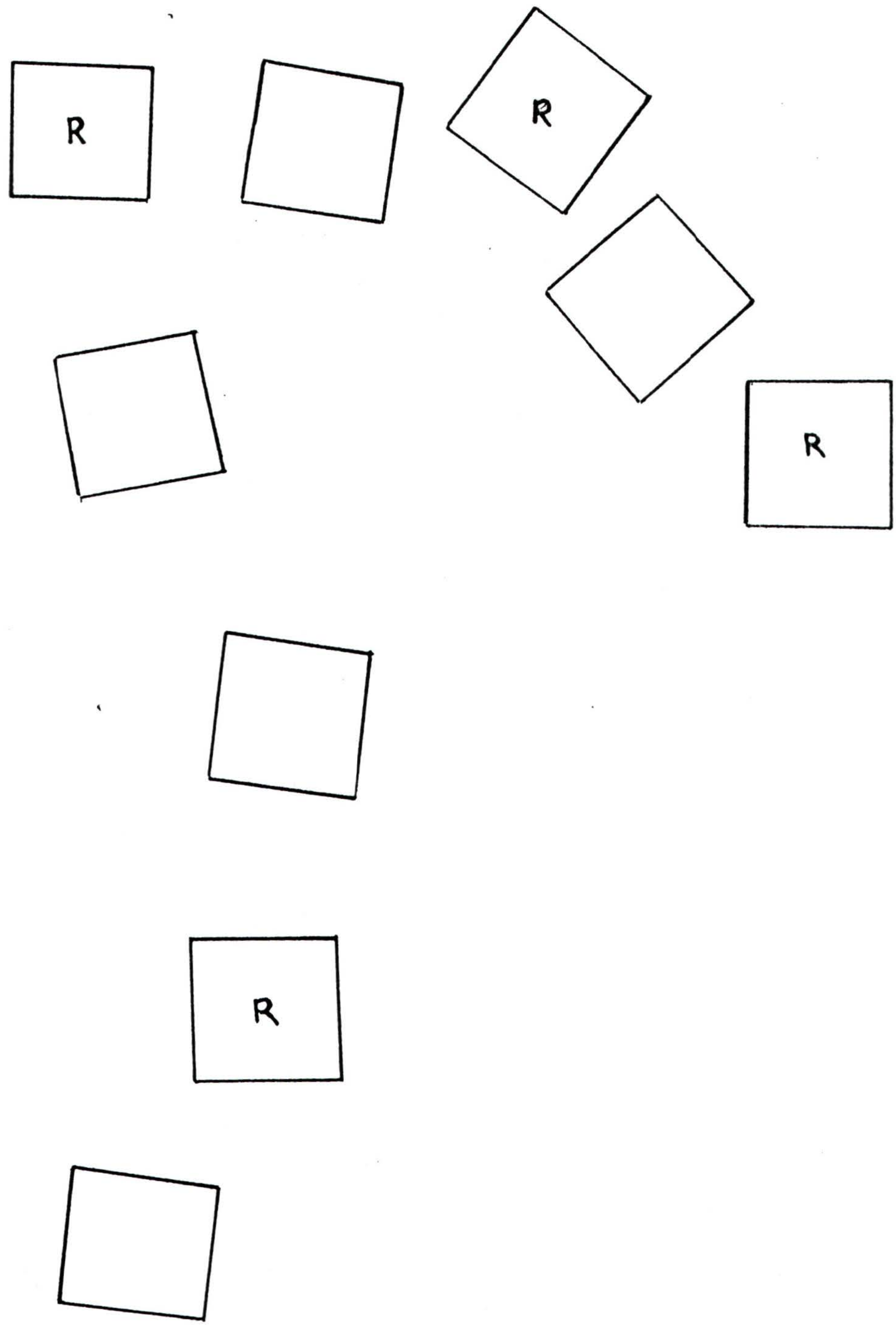


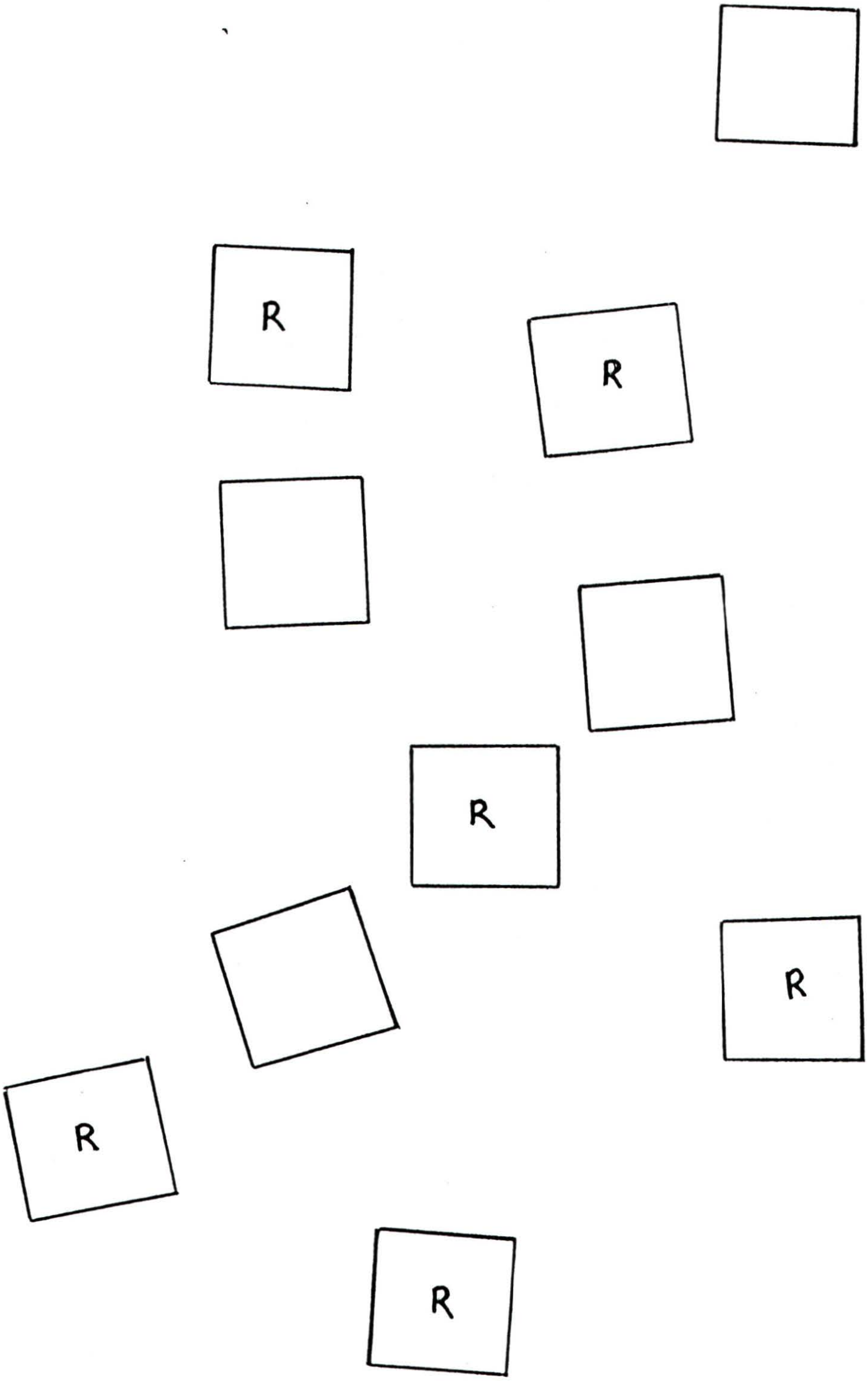
Sample of Test Trial

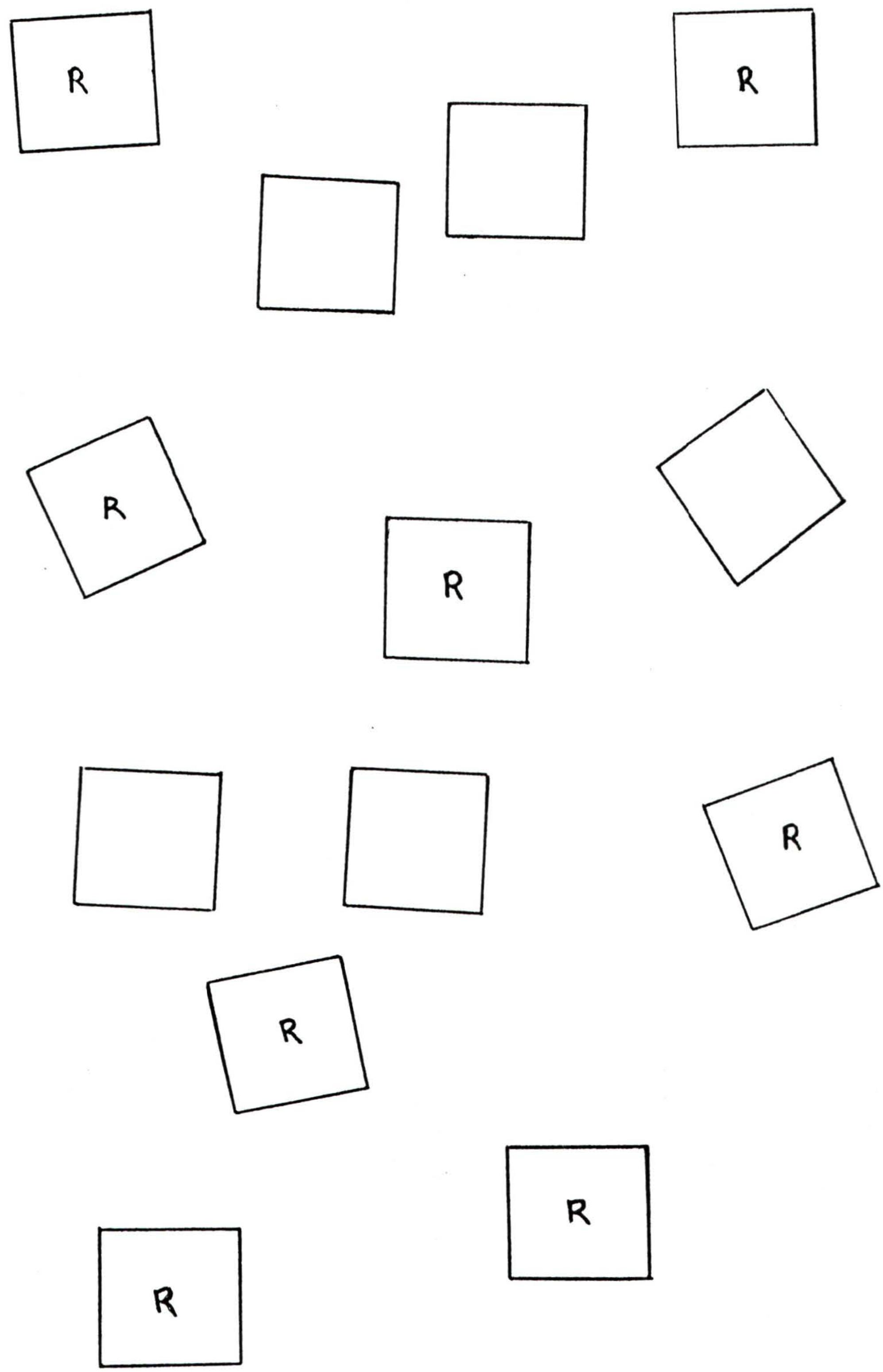
(Set 1, Level 6)

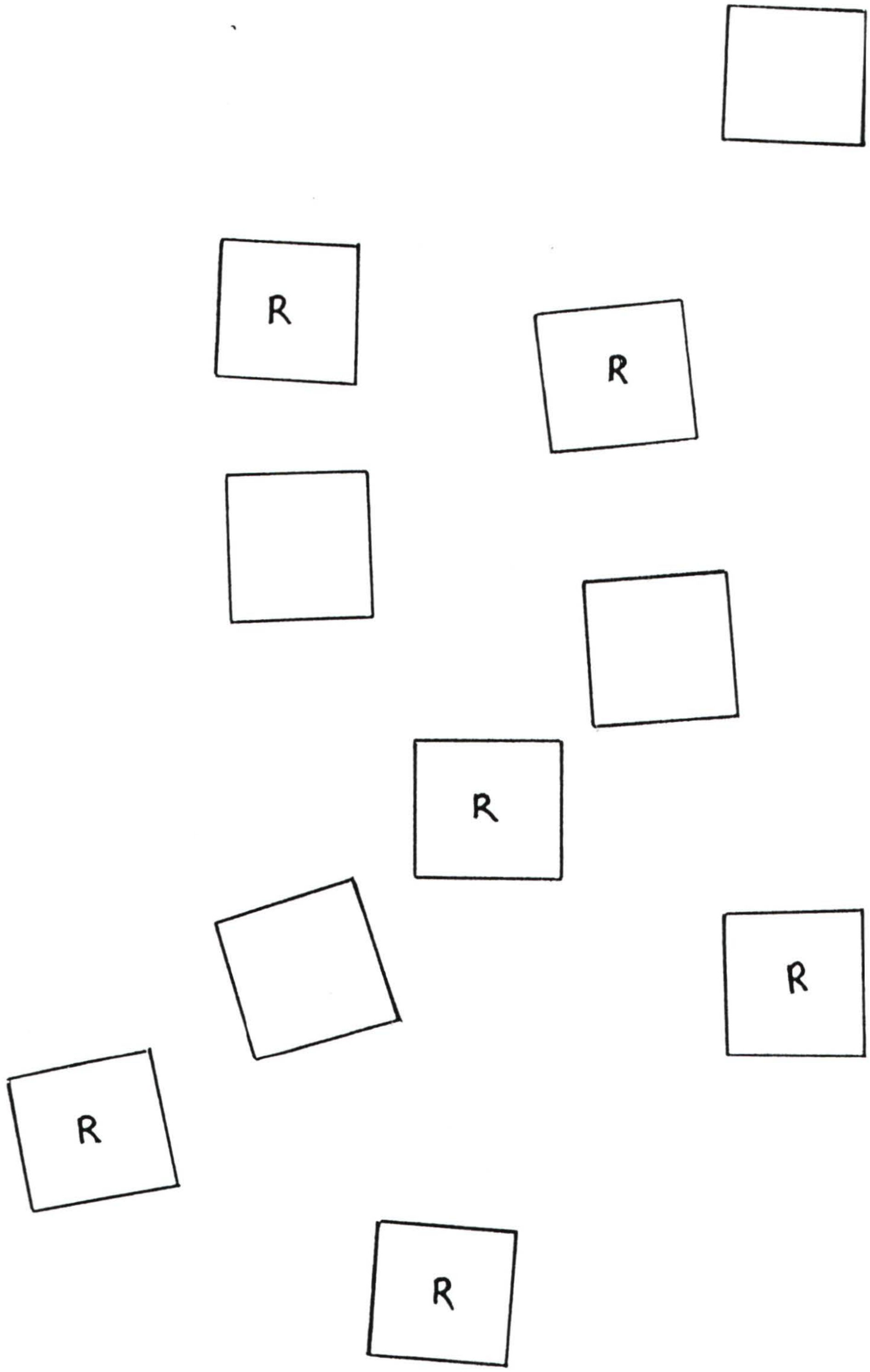



R
R
R
R
R
R
R
R
R


R
R
R
R


R
R
R
R
R
R


R
R
R
R
R
R
R
R


R
R
R
R
R
R

CARD COUNTING TASK

Answer Sheet

Name: ______________________________

Age: ______________________________
(in months)

TRAINING PHASE

(A)	3	
(B)	3,5	
(C)	6	
(D)	4,8,6	
(E)	6,3	

TESTING PHASE

SET	LEVEL	ITEM	STUDENT RESPONSE
1	1	6	
	6	6,3,8,7,4,9	
	3	7,4,8	
	4	4,9,6,5	
	2	7,4	
	5	7,4,6,3,8	
2	4	6,4,3,8	
	6	4,2,5,8,3,7	
	3	3,6,4	
	2	5,3	
	5	4,7,6,9,5	
	1	8	
3	2	4,7	
	4	5,9,7,3	
	1	3	

SET	LEVEL	ITEM	STUDENT RESPONSE
	5	5,7,4,8,3	
	3	7,9,6	
	6	6,5,8,3,9,4	
4	4	7,5,4,8	
	3	8,3,6	
	6	5,3,6,8,4,7	
	2	4,6	
	5	6,3,7,2,5	
	1	7	
5	6	6,4,7,5,8,3	
	3	5,7,4	
	5	8,5,3,7,9	
	1	4	
	4	6,3,8,5	
	2	6,3	

APPENDIX C

TASC QUESTIONS AND ANSWER SHEET

(S.B. Sarason, Davidson, Lighthall, Waite & Ruebush, 1960.)

Items are read by the examiner and students respond "yes" or "no" on the supplied answer sheet.

1. Do you worry when the teacher says that he/she is going to ask you questions to find out how much your know?
2. Do you worry about being promoted, that is, passing from this grade to the next grade at the end of the year?
3. When the teacher asks you to get up in front of the class and read aloud, are you afraid that you are going to make some bad mistakes?
4. When the teacher says that he/she is going to call upon some boys and girls in the class to do arithmetic problems, do you hope that he/she will call upon someone else and not on you?
5. Do you sometimes dream at night that you are in school and cannot answer the teacher's questions?
6. When the teacher says that he/she is going to find out how much you have learned, does your heart begin to beat faster?
7. When the teacher is teaching you about arithmetic, do you feel tha tthe other children in the class understand him/her better than you?

8. When you are in bed at night, do you sometimes worry about how you are going to do in class the next day?

9. When the teacher asks you to write on the blackboard in front of the class, does the hand you write with sometimes shake a little?

10. When the teacher is teaching you about reading, do you feel that other children in class understand him/her better than you?

11. Do you think you worry more about school than other children?

12. When you're at home and you are thinking about your arithmetic lessons for the next day, do you become afraid that you will get the answers wrong when the teacher calls upon you?

13. If you are sick and miss school, do you worry that you will do more poorly in your school work than other children when you return to school?

14. Do you sometimes dream at night that other boys and girls in your class can do things you cannot do?

15. When you are home and you are thinking about your reading lesson for the next day, do you worry that you will do poorly on the lesson?

16. When the teacher says that he/she is going to find out how much you have learned, do you get a funny feeling in your stomach?

17. If you did very poorly when the teacher called on you, would you probably feel like crying even though you would try not to cry?

18. Do you sometimes dream at night that the teacher is angry because you do not know your lessons?

19. Are you afraid of school tests?

20. Do you worry a lot <u>before</u> you take a test?

21. Do you worry a lot <u>while</u> you are taking a test?

22. After you have taken a test do you worry about how well you did on the test?

23. Do you sometimes dream at night that you did poorly on a test you had in school that day?

24. When you are taking a test, does the hand you write with shake a little?

25. When the teacher says that he/she is going to give the class a test, do you become afraid that you will do poorly?

26. When you are taking a hard test, do you forget some things you knew very well before you started taking the test?

27. Do you wish a lot of times that you didn't worry so much about tests?

28. When the teacher says that he/she is going to give the class a test, do you get a nervous or funny feeling?

29. While you are taking a test do you usually think you are doing poorly?

30. While you are on your way to school, do you sometimes worry that the teacher may give the class a test?

<u>TASC</u> Answer Sheet

Name ____________________

Teacher _________________

1.	yes	no	16.	yes	no
2.	yes	no	17.	yes	no
3.	yes	no	18.	yes	no
4.	yes	no	19.	yes	no
5.	yes	no	20.	yes	no
6.	yes	no	21.	yes	no
7.	yes	no	22.	yes	no
8.	yes	no	23.	yes	no
9.	yes	no	24.	yes	no
10.	yes	no	25.	yes	no
11.	yes	no	26.	yes	no
12.	yes	no	27.	yes	no
13.	yes	no	28.	yes	no
14.	yes	no	29.	yes	no
15.	yes	no	30.	yes	no

APPENDIX D

DSC QUESTIONS AND ANSWER SHEET

(S.B. Sarason, Hill & Zimbardo, 1964.)

Items are read by the examiner and students respond "yes" or "no" on the supplied answer sheet.

* 1. Do you love to play sports best of all?

* 2. Should girls be just as brave as boys?

** 3. Do you ever worry about knowing your lessons?

4. Do you sometimes dream about things you don't like to talk about?

5. Are you sometimes afraid of getting into arguments?

6. When someone scolds you does it make you feel badly?

** 7. Do you ever worry about what people think of you?

8. When you get mad do you ever tell anyone else about it?

9. Do you sometimes feel like hurting someone?

** 10. Do you ever worry that you won't be able to do something that you want to do?

11. Do you like to play in the snow?

12. Are you sorry for some of the things you have done?

13. When one of your friends won't play with you, do you feel badly?

** 14. When you were younger, were you ever scared of anything?

15. When someone makes you mad, do you ever tell them about it?

16. Do you feel cross and grouchy sometimes?

17. Are there some people that you don't like?

** 18. Have you ever been afraid of getting hurt?

19. Since you started school, have you ever felt like crying?

20. Do you feel it's important to think about how you can get people to like you?

21. Do you like to go to the teach in the summertime?

** 22. Do you ever worry about something bad happening to someone you know?

23. Sometimes when you get mad, do you smash something?

24. When you hurt somebody's feelings, does it make you feel badly?

25. Do you wish your teacher paid more attention to you?

** 26. Do you ever worry about what is going to happen?

27. Do you sometimes have arguments with your mother and father?

** 28. Are you ever unhappy?

29. Are there some things you just don't like to talk about?

30. If you think someone doesn't lik you, does it bother you?

31. Do you like to go on trips with your mother and father?

** 32. Has anyone ever been able to scare you?

33. Dpo you feel terrible if you break something which belongs to somebody else?

34. Do you lose your temper sometimes?

** 35. Have you ever had a scary dream?

36. When you are worried about something, do you like to talk about it?

37. Does it bother you if the teacher chooses someone else instead of you to do something for her (or him)?

** 38. Do you every worry?

39. When you've done something wrong, is it hard for you to say you're sorry?

40. Is it hard for you to tell someone you're scared?

---

* Filler items not scored but included to help control for "warm-up" effects.

** Items of the Lie Scale for Children.

DSC Answer Sheet

Name ____________________

Teacher ____________________

1.	yes	no	21.	yes	no
2.	yes	no	22.	yes	no
3.	yes	no	23.	yes	no
4.	yes	no	24.	yes	no
5.	yes	no	25.	yes	no
6.	yes	no	26.	yes	no
7.	yes	no	27.	yes	no
8.	yes	no	28.	yes	no
9.	yes	no	29.	yes	no
10.	yes	no	30.	yes	no
11.	yes	no	31.	yes	no
12.	yes	no	32.	yes	no
13.	yes	no	33.	yes	no
14.	yes	no	34.	yes	no
15.	yes	no	35.	yes	no
16.	yes	no	36.	yes	no
17.	yes	no	37.	yes	no
18.	yes	no	38.	yes	no
19.	yes	no	39.	yes	no
20.	yes	no	40.	yes	no

APPENDIX E

PROTOCOLS

Protocols for the First Morning of Data Collection

Hi! I'm sure I'm pretty familiar to most of you by now. (The author had worked on several other research projects in the school.) My name is Nancy Steacy and I am a graduate student at the University. My area of study is Educational Psychology which means I'm interested in how people learn.

Each morning we will begin by answering a questionnaire which will ask you how you feel about a lot of things. Then there will be a set of word problems to complete, but I think you'll discover they may be different than those problems you are used to answering.

(To the subjects who were administered the CST the following also was said:)

Over the next two weeks I will also be asking each of you to come down to the "__________" room by yourself and we will play a memory game. We will draw up a timetable and put it on the board with Mr. "__________"'s permission.

(To all subjects:)

I appreciate that you have agreed to take part in my project. I also want you to know that your results on any of the tests or questionnaires we are about to do are between you and me. It is very important to my research results that you answer the questionnaires honestly. Please don't worry about how your neighbour or best friend is answering or how you think I might want you to answer. I want to know how you really feel. It is also important that on the word problems (and memory game) you try to do your best. Any questions?

Before we start does anyone need a pencil?

Administration of SPAS

The first questionnaire we're going to complete asks how you feel about school. Remember that a questionnaire is not a test and that there are no "right" or "wrong" answers; that you are being asked how you feel. **(SPAS booklets were passed out.)**

Please fill in the information on the cover. Read the directions for marking your answers. Don't open the booklet until we are all ready. Everyone ready? Okay turn the page. Let's look at the directions at the top of the page. They say "__________" **(The researcher read them out loud.)**

Let's do #1 together. It says: "__________" **(The researcher read the first item out loud.)** Think about it and circle either "yes" or "no" depending on whether it describes you. Please make your circles neat enough so I won't have difficulty trying to decide what your answer is. Now you read #2 to yourselves and circle "yes" or "no". Everyone understand what they are supposed to do? There are several statements which talk about "printing", it might be more helpful if you think of "writing" instead. Now you may complete the rest of the questionnaire. Please read each item carefully before you answer. Put your hand up if you don't understand any of the statements. When you are finished, close the booklet on the top of your desk and I will pick it up.

Administration of Problem Set One

Now on to the word problems, but as I said I think you will soon see these are problems with a difference; and that difference is in the words used in them. Some problems contain real words which are all familiar to you. Other problems use a combination of real words and nonsense words. Still other problems contain mostly nonsense words. However, all the problems can be solved.

(Problem Set One was passed out.) Please fill in the information on the cover and when that is completed please read the sample questions. Do not open the booklet until everyone is ready.

I would like everyone to look at sample question #1, it says "__________" **(The researcher read it out loud. See Appendix F for actual sample items.)** This is how I want you to answer. **(The researcher wrote on the board "15 pencils and pens".)** It is very important that you include "pencils and pens". And even though you may be able to do many of the calculations in your head, please put "6+9=15" in your work space. I need to see which numbers and which operation you used for my research.

Now look at question #2. Here you see an example of some of the nonsense words used, it says "__________" **(The researcher read it.)** What would the answer be? What will you write in the work space? What will you write in the answer space? **(Selected subjects responded orally to these questions.)**

Does everyone understand what they are supposed to do? Read each question carefully. Show your work for every question and answer with more than a number. When you finish, close your booklet on the top of your desk. I will pick it up. If you finish before some of your classmates, please carry on very quietly with your own work. You may now begin.

Protocols Given the Second Morning of Data Collection

Administration of TASC

This morning I am going to ask you some questions. These questions do not have right or wrong answers. They are questions about some things which might worry you. Unlike yesterday's questionnaire, I will be reading these items to you. Again, you will be asked to circle "yes" or "no". Here are the answer sheets. **(The answer sheets were distributed.)** Please fill in your name and your teacher's name. Everyone done that?

Listen carefully to each question and circle either "yes" or "no" beside the number of that item. Remember people think and feel differently. The person next to you may circle different answers than you do. Your answers will depend on how you think and feel.

If you do not understand any question raise your hand and I will explain it. Let's start now. Question #1 is... **(The researcher read it out loud.)** Circle "yes" or "no". Everyone ready to go on. Question #2 is... **(All questions were delivered orally by the researcher.)** Question #3 is... and so on.

(After question 18 and before 19 the research said:)

In the following questions the word "test" is used. What I mean by "test" is any time the teacher asks you to do something to find out how much you know or how much you have learned. You could be asked to answer questions on paper, on the blackboard or orally. Do you understand what I mean by "test"? It is any time the teacher asks you to do something to find out how much you know. **(The remainder of the TASC items were then read.)**

Administration of Problem Set Two

Now, you will complete the second set of word problems. They are similar to yesterday's problems. **(Problem Set Two was distributed to subjects.)** Again, please fill in the information on the front cover. Everyone ready to continue?

Let's review how I want you to answer the problems. Remember, please don't answer with just a number like "15". It is important that you tell me 15 "what", such as "15 apples and oranges" or "15 blorgs", whatever you think is appropriate. Also, remember to do your work in the work space even though I know that you are able to do it in your head. Any questions? You may now begin.

Protocols given on the Third Morning of Data Collection

This morning we will answer the last questionnaire and do the last set of problems. This morning's questionnaire is similar to yesterday's in that I will be reading the items to you. Again, please circle "yes" or "no" depending upon whether you agree or disagree with the question. Here are the answer sheets. **(The answer sheets were distributed.)** Before we begin please fill in the information at the top of the sheet. Everyone ready?

Again, listen carefully to each item and circle either "yes" or "no" beside the number of that item. Please remember to answer as honestly as you can. Let's begin. **(The researcher then read the DSC items out loud.)**

Administration of Problem Set Three

And now we are on to the final set of word problems. **(Problem Set Three was distributed.)** Please fill in the information on the front cover once again. Everyone finished?

Do you recall the two things I want you to remember about doing the word problems? **(The researcher solicited responses from several subjects.)** Did everyone remember what "______" and "______" just told us about how you should answer the questions using more than a number, and that you are to show your work in the work space? Great! You now may begin.

(After completion of Problem Set Three those subjects who were not going to be administered the CST were thanked again for participating. The researcher's gratitude was expressed individually to those subjects who did subsequently complete the CST. After all the data collection was complete all subjects were briefly apprised of the nature of the study and were invited to question the researcher.)

Protocol Given to Individual Subjects Prior to the Administration of the Counting Span Test

Practice Phase

We're going to play a memory game in which you will count and remember things. **(Subject was shown card from Training Phase A. See record form in Appendix B.)** See this card. I'm going to show you a lot of cards like this one. What I want you to do is take your finger and count each of the red squares out loud. **(The researcher demonstrated touching each square and counting out loud.)** Now you do it. **(The subject counted.)** Now, I'm going to cover up this card **(The researcher placed a blank card directly over card just counted.)** and all you have to do is tell me how many red squares were on the card. How many? **(If the subject responded incorrectly the procedure was repeated.)**

Now let's try another one. **(Training Phase C.)** I want you to count the squares just like you did last time, but count only the red squares not the orange ones, like this. **(The researcher demonstrated touching each red square and counting out loud.)** Now you try. **(The subject counted and then the researcher covered the card.)** How many red squares were on it? Right, that's the way we do it. **(The procedure was repeated if the subject was incorrect.)** Now let's try another. **(Training Phase B.)** First count this card. **(The researcher made sure that each red square and only the red squares were touched and counted out loud.)** Remember how many red squares were on that card and count this one. **(The researcher placed the second card directly over the first.)** Now I'll cover them both up. **(The researcher covered the second card with a blank card.)** and you tell me how many were on the first card, and then how many were on the second card.

(The process was repeated if necessary.) Sometimes there will be as many as 4, 5 or 6 cards to count before you tell them back to me. **(Training Phases D and E were then presented.)**

Counting Span Test Items

Now we are ready to try some more cards. Remember to count carefully out loud, touching each red square as you count it. You can count as fast or slow as you want, whatever speed is best for you. Try to tell me the number of red squares on each card in the order in which you see them. **(The 30 trials comprising the CST were then presented.)**

APPENDIX F

SAMPLE PAGES FROM PS1, PS2, AND PS3

(Showing Practice Items and Formatting of Test Items)

NAME: ______________________________

TEACHER: ___________________________

DATE: ______________________________

PART I

SAMPLE QUESTIONS

--Work space--------

1. Bob bought 6 pencils.
   Bob bought 9 pens. How
   many pencils and pens did
   Bob buy?

Answer ____________________________

____

--Work space--------

2. Mog owned 4 nops. Mog
   owned 7 nops. How many nops
   did Mog own?

Answer--

---Work space-----

17. Blick berved 8 pids. Meap berved 3 jates. Hamp tonked 6 shruts. Blick, Meap, and Hamp are tives. Pids, jates, and shruts are sulls. How many sulls did the tives berve?

Answer ______________________________

____

---Work space-----

18. Blip jonned 3 gorps. Golt jonned 2 gorps. How many gorps did Blip and Golt jon?

Answer ______________________________

____

---Work space-----

19 Min kissed 6 veeds. Flamp hit 7 veeds. Gade hit 3 veeds. How many veeds did Min, Flamp, and Gade hit?

Answer______________________________

____

---Work space-----

20. Jason had 4 stallions. Jason had 4 mares. Stallions and mares are horses. How many horses did Jason have?

Answer ______________________________

____

--------------------------------------Work space-----

1. Brap fummed 4 rimps.
Gom nigged 7 plints.
Smide fummed 2 bools. How many rimps, plints, and bools did Brap, Gom, and Smide fum?

Answer ____________________

____

--------------------------------------Work space-----

2. Fred collected 3 stamps.
Kelly collected 9 coins.
Greg lost 8 stickers. How many stamps, coins, and stickers did Fred, Kelly, and Greg collect?

Answer ____________________

____

--------------------------------------Work space-----

3. Ann mailed 3 letters. Tony mailed 5 letters. Frank received 7 letters. How many letters did Ann, Tony, and Frank mail?

Answer ____________________

____

--------------------------------------Work space-----

4. The bloss matched 5 sligs.
The fant touched 4 sligs.
The mirst touched 3 sligs.
Blosses, fants, and mirsts are dobs. How many sligs did the dobs touch?

Answer ____________________

____

Page 7, PS3

--Work space-----

25. Jim lost 6 dimes. Sarah found 8 nickles. Maria lost 6 quarters. How many dimes, nickles, and quarters did Jim, Sarah, and Maria lose?

Answer ____________________________

____

--Work space-----

26. The phock saw 7 vards. The gurf counted 4 vards. The slout counted 6 vards. How many vards did the phock, gurf, and slout count?

Answer ____________________________

____

--Work space-----

27. The strape hussed 3 cronks. The nox hussed 3 cronks. The chand pobbed 6 cronks. How many cronks did the strape, nox, and chand huss?

Answer ____________________________

____

--Work space-----

28. The otter ate 6 clams. The otter,ate 4 fish. The otter saw 2 shrimp. How many clams, fish and shrimp did the otter eat?

Answer ____________________________

____

APPENDIX G

TABLES

Table 3

Means and standard deviations for dependent and independent measures

Independent Measures	X	SD
SPAS[a]	47.95	11.50
TASC[b]	9.33	6.00
DSC[c]	13.14	5.49
Dependent Measures		
PS1[d]	39.39	10.68
PS2[d]	42.70	8.95
PS3[d]	43.72	9.32
Total Performance[e]	41.94	9.68

Note. N = 57 (except for TASC, where n = 54).

Descriptive statistics for the Counting Span Test were not claculated as it did not seem to be logical to do so given the discrete nature of the construct and the limited range of possible scores (from 2 to 5).

[a]SPAS = Students' Perception of Ability Skill. Number of items = 70. Higher scores indicate higher levels of self-concept.

[b]TASC = Test Anxiety Scale for Children Number of items = 30. Higher scores indicate higher levels of test anxiety.

[c]DSC = Defensiveness Scale for Children Number of items = 40. Higher scores indicate higher levels of defensiveness.

[d]PS1 = Problem Set One; PS2 = Problem Set Two; PS3 = Problem Set Three. Maximum score for any problem set = 48.

[e]Maximum score = 144.

Table 8

Kruskal-Wallis ANOVAs of levels of the task variables and problem sets by M-space groups

Task Variable Level	M-space Group	n	Mean Rank	Chi-square[a]
Extraneous Information				
NEI	MSP2	3	8.00	
	MSP3	11	11.50	
	MSP4	12	21.38	
	MSP5	8	23.50	13.94*
EI1	MSP2	3	8.00	
	MSP3	11	8.73	
	MSP4	12	22.92	
	MSP5	8	25.00	20.86+
EI2	MSP2	3	9.33	
	MSP3	11	8.59	
	MSP4	12	23.75	
	MSP5	8	12.33	19.53**
EI3	MSP2	3	7.17	
	MSP3	11	10.45	
	MSP4	12	23.00	
	MSP5	8	22.81	15.83*

Table 8 Cont'd.

Task Variable Level	M-space Group	n	Mean Rank	Chi-square[a]
Sex Complexity				
CSSO	MSP2	3	6.00	
	MSP3	11	8.68	
	MSP4	12	24.21	
	MSP5	8	23.88	22.04+
SSCO	MSP2	3	6.67	
	MSP3	11	8.91	
	MSP4	12	23.67	
	MSP5	8	24.13	21.80+
CSCO	MSP2	3	9.00	
	MSP3	11	10.09	
	MSP4	12	21.50	
	MSP5	8	24.88	16.00*
Set Language				
GRP	MSP2	3	7.17	
	MSP3	11	9.23	
	MSP4	12	23.38	
	MSP5	8	23.94	18.79**

Table 8 Cont'd.

Task Variable Level	M-space Group	n	Mean Rank	Chi-square[a]
RC	MSP2	3	7.33	
	MSP3	11	8.64	
	MSP4	12	23.42	
	MSP5	8	24.63	21.50+
Language Familiarity				
REAL	MSP2	3	4.50	
	MSP3	11	10.36	
	MSP4	12	23.08	
	MSP5	8	23.8`	18.59**
NN	MSP2	3	10.00	
	MSP3	11	8.45	
	MSP4	12	24.08	
	MSP5	8	22.88	
NNV	MMP2	3	7.83	
	MSP3	11	8.64	
	MSP4	12	23.38	
	MSP5	8	24.50	21.29+

Table 8 Cont'd.

Task Variable Level	M-space Group	n	Mean Rank	Chi-square[a]
Problem Sets				
PS1	MSP2	3	9.00	
	MSP3	11	9.05	
	MSP4	12	23.42	
	MSP5	8	23.44	17.70**
PS2	MSP2	3	6.83	
	MSP3	11	10.32	
	MSP4	12	22.50	
	MSP5	8	23.88	16.86**
PS3	MSP2	3	9.83	
	MSP3	11	9.50	
	MSP4	12	22.00	
	MSP5	8	24.63	17.30**

N=34

Note. NEI = no extraneous information; EI1 = extraneous information first; EI2 = extraneous information second, EI3 = extraneous information third.

CSSO = complex subject/simple object; SSCO = simple subject/complex object; CSCO = complex subject/- complex object.

GRP = sets grouped together with "and"; RC = sets reclassified under a superordinate.

REAL = all real words; NN = nonsense nounse; NNV = nonsense nouns and verbs.

PS1 = problem set one; PS2 = problem set two; PS3 = problem set three.

[a]Chi-squares are corrected for ties.

*$p<.01$.

**$p<.001$.

+$p<.001$.

Table 11

Extraneous information by learner characteristic

Groups by Subject Variable		n	Presence or order of presenting extraneous information: NEI	EI1	EI2	EI3
			Academic self-perception			
ASP1		19				
	M		31.58	25.00	23.90	25.42
	SD		4.41	10.30	10.07	10.42
	CI95		29.45-33.71	20.04-29.96	19.04-28.75	20.40-30.44
ASP2		19				
	M		34.58	32.63	32.53	33.21
	SD		3.72	5.64	5.39	3.75
	CI95		33.05-36.11	29.91-35.35	29.94-35.11	31.40-35.02
ASP3		19				
	M		35.63	34.74	34.37	34.58
	SD		.76	2.49	2.89	2.67
	CI95		35.27-36.00	33.54-35.94	32.98-35.76	33.29-35.87
			Test anxiety			
TA1		18				
	M		35.79	34.84	34.42	34.16
	SD		.54	2.14	2.57	2.71
	CI95		35.53-36.05	33.81-35.87	33.18-35.66	32.85-35.46

Table 11 cont'd.

Groups by Subject Variable		n	Presence or order of presenting extraneous information: NEI	EI1	EI2	EI3
TA2		18				
	M		33.11	30.95	30.42	31.79
	SD		4.56	7.12	6.91	5.93
	C195		30.91-35.30	27.52-34.38	27.09-33.75	28.93-34.65
TA3		18				
	M		32.90	26.58	25.95	27.26
	SD		3.59	10.39	10.60	10.62
	CI95		31.17-34.62	21.57-31.59	20.84-31.06	22.15-32.38

Note. Maximum score = 36.

ASP1 = low academic self-perception group; ASP2 = mid academic self-perception group; ASP3 = high academic self-perception group; TA1 = low test anxiety group; TA2 = mid test anxiety group; TA3 = high test anxiety group.

NEI = no extraneous information; EI1 = extraneous information first; EI2 = extraneous information second; EI3 = extraneous information third.

Table 12

Set complexity by learner characteristic

Groups by Subject Variables		n	Level of Set Complexity CSSO	SSCO	CSCO
			Academic self-perception		
ASP1		19			
	M		37.11	35.84	33.16
	SD		10.16	11.64	12.07
	CI95		32.21-42.00	30.23-41.45	27.34-38.97
ASP2		19			
	M		44.63	44.68	43.53
	SD		4.35	5.90	8.06
	CI95		42.54-46.73	41.84-47.53	39.64-47.42
ASP3		19			
	M		46.21	46.95	46.37
	SD		2.92	1.87	3.27
	CI95		44.81-47.62	46.05-47.85	44.79-47.94
			Test anxiety		
TA1		18			
	M		45.95	46.68	46.74
	SD		2.86	2.19	2.18
	CI95		44.57-47.33	45.63-47.74	45.69-47.79

Table 12 cont'd.

Groups by Subject Variables		n	Level of Set Complexity CSSO	SSCO	CSCO
TA2		18			
	M		42.95	42.79	40.74
	SD		5.91	8.19	10.60
	CI95		40.10-45.80	38.84-46.74	35.89-45.59
TA3		18			
	M		39.00	38.00	35.58
	SD		10.60	11.60	12.31
	CI95		33.95-44.16	32.41-43.59	29.65-41.51

Note. Maximum score = 48.

ASP1 = low academic self-perception group; ASP2 = mid academic self-perception group; ASP3 = high academic self-perception group; TA1 = low test anxiety group; TA2 = mid test anxiety group; TA3 = high test anxiety group.

CSSO = complex subject/simple object; SSCO = simple subject/complex object; CSCO = complex subject/- complex object.

Table 13

Set language by learner characteristic

Groups by Subject Variables		n	Method of Establishing Set Complexity: GRP	RC
	Academic self-perception			
ASP1		19		
	M		50.84	54.84
	SD		17.34	17.31
	CI95		42.84-59.20	46.50-63.19
ASP2		19		
	M		64.32	68.26
	SD		11.76	7.01
	CI95		58.65-69.98	64.89-71.64
ASP3		19		
	M		69.11	70.74
	SD		4.63	3.25
	CI95		66.87-71.34	68.91-72.04
	Test Anxiety			
TA1		18		
	M		69.00	70.26
	SD		4.19	3.25
	CI95		66.98-71.02	68.70-71.83

Table 13 cont'd.

Groups by Subject Variables		n	Method of Establishing Set Complexity GRP	RC
TA2	18			
	M		61.47	
	SD		13.01	11.62
	CI95		44.89-62.69	50.20-66.64
TA3		18		
	M		53.79	58.42
	SD		18.46	17.05
	CI95		44.89-62.69	50.20-66.64

Note. Maximum score = 72.

ASP1 = low academic self-perception group; ASP2 = mid academic self-perception group; ASP3 = high academic self-perception group; TA1 = low test anxiety group; TA2 = mid test anxiety group; TA3 = high test anxiety group.

GRP = sets grouped together with "and"; RC = sets reclassified under a superordinate.

Table 14

Language familiarity by learner characteristic

Groups by Subject Variables		n	REAL	NN	NNV
			Academic Self-perception		
ASP1		19			
	M		39.53	34.16	31.37
	SD		9.51	12.43	14.01
	CI95		34.94-44.11	28.16-40.15	24.62-38.12
ASP2		19			
	M		45.58	43.68	42.90
	SD		3.68	6.67	8.75
	CI95		43.81-47.35	40.47-46.90	38.68-47.11
ASP3		19			
	M		46.90	45.90	46.32
	SD		1.52	3.78	3.43
	CI95		46.16-47.63	44.07-47.72	44.66-47.97
			Test Anxiety		
TA1		18			
	M		46.63	46.00	46.11
	SD		2.19	3.07	3.21
	CI95		45.56-47.69	44.52-47.48	44.56-47.65

Table 14 Cont'd.

Groups by Subject Variables		n	REAL	NN	NNV
TA2		18			
	M		44.95	41.05	39.90
	SD		3.84	9.70	11.53
	CI95		43.10-46.80	36.38-45.73	34.34-45.45
TA3		18			
	M		40.43	36.68	34.58
	SD		9.87	12.09	14.16
	CI95		35.66-45.18	30.86-42.51	27.76-41.40

Level of Language Familiarity: REAL, NN, NNV.

Note. Maximum score = 48.

ASP1 = low academic self-perception group; ASP2 = mid academic self-perception group; ASP3 = high academic self-perception group; TA1 = low test anxiety group; TA2 = mid test anxiety group; TA3 = high test anxiety group.

REAL = all real words; NN = nonsense nouns; NNV = nonsense nouns and verbs.

Table 15

Problem set by learner characteristic

Groups by Subject Variables		n	PS1	PS2	PS3
			Academic self-perception		
ASP1		19			
	M		32.26	35.26	38.00
	SD		11.18	11.76	13.83
	CI95		26.88-37.65	29.60-40.93	31.33-44.67
ASP2		19			
	M		41.53	45.32	45.58
	SD		9.99	4.28	4.51
	CI95		36.71-46.34	43.25-47.38	43.40-47.75
ASP3		19			
	M		44.37	47.53	47.58
	SD		6.51	.96	.69
	CI95		42.23-47.51	47.06-47.99	47.25-47.91
			Test anxiety		
TA1		18			
	M		44.95	46.74	47.68
	SD		4.35	2.64	.67
	CI95		42.85-47.05	45.46-48.02	47.36-48.01

Table 15 cont'd.

Groups by Subject Variables		n	Problem Set PS1	PS2	PS3
TA2		18			
	M		38.16	43.11	44.63
	SD		12.16	8.12	5.29
	CI95		32.30-44.02	39.16-47.02	42.08-47.18
TA3		18			
	M		35.05	38.26	38.84
	SD		11.43	11.69	13.97
	CI95		29.54-40.56	32.63-43.90	32.11-45.57

Note. Maximum score = 48.

ASP1 = low academic self-perception group; ASP2 = mid academic self-perception group; ASP3 = high academic self-perception group; TA1 = low test anxiety group; TA2 = mid test anxiety group; TA3 = high test anxiety group.

PS1 = problem set one; PS2 = problem set two; PS3 = problem set three.

VITA

Surname: Steacy   Given Names: Nancy

Place of Birth: Victoria, B.C.   Date of Birth: July 26, 1951

Educational Institutions Attended, with Dates of Entering and Leaving:

University of British Columbia	1970	to	1974
		to	
		to	
		to	

Degrees Awarded, with Dates and Names of Institutions:

B.A.	1974	University of British Columbia

Awards:

PARTIAL COPYRIGHT LICENSE

Title of Thesis:

RULE INDUCTION ON WORD PROBLEMS AS A
FUNCTION OF LEARNER CHARACTERISTICS
AND TASK VARIABLES

Author: [redacted]

Nancy Steacy

August 28, 1987
