INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.
- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.
- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17"x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.

Order Number 8719155

Generalization via multiple exemplar training and two methods of rule training: Do rules help?

Feinberg, Hal, Ph.D.

The University of North Carolina at Greensboro, 1987

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs or pages ✓
2. Colored illustrations, paper or print ✓
3. Photographs with dark background ____
4. Illustrations are poor copy ____
5. Pages with black marks, not original copy ____
6. Print shows through as there is text on both sides of page ____
7. Indistinct, broken or small print on several pages ✓
8. Print exceeds margin requirements ____
9. Tightly bound copy with print lost in spine ____
10. Computer printout pages with indistinct print ____
11. Page(s) ________ lacking when material received, and not available from school or author.
12. Page(s) ________ seem to be missing in numbering only as text follows.
13. Two pages numbered ____. Text follows.
14. Curling and wrinkled pages ____
15. Dissertation contains pages with print at a slant, filmed as received ✓
16. Other ________________________________

University
Microfilms
International

GENERALIZATION VIA MULTIPLE EXEMPLAR TRAINING AND TWO METHODS OF RULE TRAINING: DO RULES HELP?

by

Hal Feinberg

A Dissertation submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
of Doctor of Philosophy

Greensboro
1987

Approved by

Dissertation Adviser

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of the Graduate School at the University of North Carolina at Greensboro.

Dissertation Adviser Richard N. Roberts

Committee Members P. Scott Lawrence

Anthony J. DeCasper

August 28, 1986
Date of Acceptance by Committee

August 27, 1986
Date of Final Oral Examination

FEINBERG, HAL, Ph.D. Generalization via Multiple Exemplar Training and Two Methods of Rule Training: Do Rules Help? (1987) Directed by: Dr. Richard N. Roberts. 254 pp.

The failure to find generalization in many applied studies with children has provided impetus for the development of both the self-instructional training (SIT) paradigm and the metacognitive training paradigm. Process research is lacking in both paradigms, given the array of training components within which verbalization training is typically embedded.

The present study asked this fundamental question: Was generalization enhanced by training children to verbalize task requirements as they engaged a task series, relative to training multiple exemplars alone? Another question followed: If rule training did facilitate generalization, were the effects attributable to the training of rules per se, or to the problem-solving behavior specified by the rules?

Thirty-eight preschoolers were trained with four sets of matrix completion tasks in a pre-post design. The multiple exemplar group (ME) received minimal instructions and feedback across training tasks, as did all groups. The rule training condition (RT) additionally required children to verbalize a rule and then perform specified problem-solving responses. The problem-solving control group (PSC) isolated the impact of the rule training per se, by requiring problem-solving responses alone. Rule discovery training (RD) encouraged the child to verbalize task

requirements.

Results indicated that only the maintenance items revealed group differences on an unprompted posttest. Groups RT and PSC were comparable to each other and superior to group ME.

On generalization items of the prompted posttest, groups RT and PSC were comparable to each other, and superior to groups RD and ME, which also were comparable. The problem-solving responses taught to RT and PSC children may have mediated generalization. Problem-solving responses increased in frequency from unprompted to prompted posttest, and were predictive of correct answers on a trial-to-trial basis.

Rules taught to RT children did not appear to function as readily emitted responses. RT children emitted rules infrequently at unprompted posttest, and did no more problem-solving than PSC children. The prompts of the prompted posttest appeared necessary to do what it was hoped the rule training would do: promote high frequencies of problem-solving responses.

ACKNOWLEDGEMENTS

I wish to thank my committee chairman, Dr. Richard N. Roberts, for his considerable advice and support during the years this work was in progress. Thanks also to my committee members for their input in the early stages of this project. They are: Dr. Anthony DeCasper; Dr. Gary Hoover; Dr. P. Scott Lawrence, and; Dr. Richard Shull.

Special thanks must go to the Center for the Development of Early Education at Kamehameha Schools, which provided a wealth of resources, both human and material, that enabled my research. Individuals at Kamehameha Schools who played a role in the execution of my dissertation were truly too numerous to mention. A few who stood out included: Dr. Roland Tharp, who offered valuable conceptual input; Cynthia Hinze and Nina Mattie, who were exceptional trainers, and; Dr. Burt Furuta, who made computer analyses happen.

Also appreciated was the conceptual and pragmatic input of Dr. Ann Brown and Dr. Joseph Campione. Finally, thanks go to Betty Ona and the KCAA preschools for their graciousness in permitting me to work with their children.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Overview

The generality of the effects of experimental intervention upon nontargeted as well as targeted responses has become the focus of substantial empirical and theoretical work in the applied arena. Until recently, generalization was viewed merely as a failure to discriminate. But data generated by various theoretical paradigms and covering a wide range of tasks, behaviors, settings, and population samples, have often revealed little or no generalization across surprisingly similar contexts. These findings encouraged some writers to suggest that generalization was best considered a theoretically important process to be studied in its own right (Stokes & Baer, 1977; Kirschenbaum and Tomarken, 1982).

Stokes and Baer (1977) categorized the procedures that have been used to obtain generalization in an effort to offer a preliminary technology of generalization. Most often cited were methods of "train and hope," and sequential modification, unsystematic and restrictive methods that did not truly contribute to a technology of the programming of generalization. Although Stokes and Baer and Dick and Roberts (1982) called for the teaching of antecedent verbal mediators as a logically promising means of inducing generalization, work of this nature has for the most part been neglected in the applied analysis of behavior in favor of more conventional reinforcement paradigms that targeted solution responses.

Failure to obtain substantial generalization has not been confined to studies in behavior analysis, however. Data generated from the study of the ontogeny of memory have yielded parallel results. For example, the responses of young mentally retarded and normal preschool children do not generalize broadly as compared with the behavior of normal or older children. These difficulties appeared on testing tasks that were slightly different structurally, had slightly different response requirements, were administered in different settings, or were administered by different persons. If, however, training included feedback about the value of the mnemonic or strategy monitoring skills, learning disabled, retarded, and very young children sometimes applied mnemonic strategies across situations and tasks. These results led some researchers to conclude that a deficit in awareness of appropriate strategy usage, rather than a strategy deficit, was responsible for poor generalization. For writers such as Flavell (Flavell & Wellman, 1977) and Brown (1975), memory awareness and monitoring are critically important skills which tap into executive control processes crucial to generalized strategy application.

Brown (1979) recently called for memory training which explicitly teaches children to ask themselves questions about the nature of task demands, to monitor their success as they apply task strategies, and to assess the quality of their performance in the hope that such broadly worded verbalizations will maximize the chances of generalized effects. Brown acknowledged the similarity of this training to many self-instructional training (SIT) packages within the cognitive behavior modification (CBM) paradigm. SIT derived its impetus from the failure of conventional behavioral interventions to yield generalized effects.

Although several SIT studies or SIT-like studies produced response generalization across settings (e.g., Meichenbaum & Goodman, 1971; Bornstein & Quevillon, 1976) and across tasks or responses (Palkes, Stewart, and Kahana, 1968; Palkes, Stewart, and Friedman, 1972), others found no such generalization across settings (Friedling & O'Leary, 1979), nor across tasks or responses (Douglas, Parry, Martin, & Gaston, 1977; Lovitt & Curtiss, 1968; Robin, Armel, & O'Leary, 1975). It is important to note that particularly in this paradigm, the relationship between training and generalization tasks in many of these studies was not systematically varied on logical grounds. Rather, generalization was sought in some instances between tasks with virtually no overlapping response requirements (e.g., analogue academic to reading).

The approaches of both the metacognition and CBM paradigms suffer in their implicit theoretical or procedural assumptions. In the case of traditional memory research, it is unwarranted to assume that the verbalization of supposedly metastrategic rules gain access to higher-order and qualitatively different executive processes than do task-specific rules.

While CBM does not ascribe special theoretical status to generally phrased rules or self-questioning strategies, those who have adopted the approach have typically been equally remiss in assuming that the rules will exercise functional control over behavior. The inconsistent outcome results which plagued the SIT literature may partially be a result of this assumption. To this writer's knowledge, virtually no CBM studies have explicitly manipulated children's learning histories with rules.

The present approach ascribed no special status to abstractly worded rules, rules which avoided the inclusion of task-specific referents, or rules which specified self-checking or self-monitoring responses. While some rules may specify and control a larger or different set of responses than others as a function of prior learning, there is no need to infer a qualitatively different process. Thus, it is not necessary to assume that if children fail to spontaneously (in the absence of explicit prompts) behave in accord with newly-learned task-specific strategies on a transfer task, they will behave in accord with newly-learned metastrategies on that transfer task. In each case it may be necessary to ensure, at least initially, that verbal rules exercise functional control over behavior.

Categorizing instructed verbalizations in terms of their specification (or lack of specification) of response-reinforcer contingencies or problem-solving responses is superior to many labels commonly employed to describe different experimental instructions such as concrete and conceptual (Kendall & Wilcox, 1980), or general and specific. Only by being precise about rules for identifying and classifying verbalizations can the effects of intervention themselves be generalized to other population samples or contexts. Descriptions of instructed verbalizations as general and specific (e.g., Schleser, Cohen, & Thackwray, 1983) are particularly troublesome as they appear to describe the effects of training rather than the occasion for the effects, and hence may easily encourage investigators to offer circular accounts of generalized effects. It is a present thesis that different rules are neither general nor specific: they may exercise varying degrees of control over responding as a function of the responses and contingencies they specify, the conditions under which they are emitted, and the child's history with these and similar verbalizations.

In the present study, generalization was examined across tasks that were logically related. Drawing from Becker (1971) and Carnine and Becker's (1982) discussions of generalized concept learning, while the examination of generalization and derivation of functional response classes is a matter of the evaluation of experimental investigation, the programming of generalization is as well a matter of logical considerations. The logical considerations made by an experimenter during his investigation of generalization deserve greater attention. So too, does the possibility of further extending generalization via the experimental manipulation of verbal antecedents. Finally, in agreement with Meichenbaum & Asarnow (1979), perhaps failures to spontaneously employ verbal mediators in learning or generalization are best viewed not as deficiencies within the child to produce or mediate, but as instructional deficiencies on the part of the experimenter. Indeed it is the experimenter who is responsible for programming an environment that will ultimately maximize the probability of learning and generalized learning.

The present study examined the influence of rules upon the generality of training effects. Most basically, it addressed the effects of programming a sequence of tasks upon performance on task instances both within the training task range and without. The tasks differed systematically from each other along an objectively specifiable continuum. The question was to what extent did rule training promote learning (including acquisition, maintenance, and generalization) over and above the programming of multiple exemplars (ME). Of interest was a comparison of two methods of engaging young children in rule emission. The first method was to prompt, in a noncoercive manner, verbalization of task rules as children engaged the tasks. The second method was to directly model and instruct children's verbalization of task rules. The rules specified problem-solving responses and the contingency between such responses and task outcome. The present study further examined this second method by experimentally distinguishing the training of children in the use of rules from the training of behavior the rules specified.

Generalization

This discussion of generalization begins with a definition of terms. For Stokes and Baer (1977) generalization was said to be the occurrence of relevant behavior, "under different non-training conditions without the scheduling of the same event in those conditions as had been scheduled in the training conditions (p. 350)." The non-training conditions may be across subjects, settings, people, behaviors, and/or time. Stokes and Baer wrote,

> Thus, generalization may be claimed when no extratraining manipulations are needed for

> extratraining changes; or may be claimed when some extra manipulations are necessary, but their cost or extent is clearly less than that of the direct intervention (p. 350).

The present author endorsed this definition of generalization and concerned himself primarily with generalization across tasks (in which both task stimuli and response requirements varied between training and testing) and generalization across time.

Stokes and Baer categorized techniques designed to assess, or implement generalization. Of greatest relevance to the present study were the methods of train and hope, sequential modification, sufficient exemplars, and mediate generalization. In the "train and hope" method, which characterized nearly half of the experiments they reviewed, generalization across one or more dimensions was noted but not actively programmed. In sequential modification, if generalization did not occur, intervention was explicitly programmed for that setting, person, behavior, or time. One illustration of this method cited by Stokes and Baer was that of Meichenbaum, Bowers, and Ross (1968) who found an absence of behavior changes from an afternoon intervention period to the morning period in a classroom for adolescent offenders. Generalization to the morning period required that the same intervention be applied in that setting.

Another illustration of sequential modification was provided in a study by Lovitt and Curtiss (1968), who found that successful intervention for one-digit subtraction problems did not generalize to two-digit subtraction problems. Nor did the same intervention, after successful application to two-digit mathematics problems generalize to mathematics problems of a slightly more complex format. The definition of this method (the application of intervention to each condition in which generalization is desired) itself implied an absence of generalization. If intervention must address every situation in which generalization is desired, then no generalization has taken place.

In Stokes and Baer's method of training sufficient exemplars, sufficient examples of the same generalization lesson are taught until the induction is formed (i.e., until generalization occurs sufficiently to satisfy the problem posed). Note that the difference between this method and sequential modification is typically a measurement difference and not a procedural one. Sequential modification is usually concerned with generalization to only a few untrained conditions, and after its completion, generalization to other untrained circumstances often remains unexamined. With the training of sufficient exemplars, as Stokes and Baer pointed out,

> Generalization to untrained stimulus conditions and to untrained responses is programmed by the training of <u>sufficient</u> exemplars (rather than all) of these stimulus conditions or responses (p. 355).

Just as the procedural definition for sequential modification implies that generalization to untrained conditions has not occurred, the procedural definition for training sufficient exemplars implies that generalization has been obtained.

Training multiple exemplars and looking for and measuring changes in untrained situations has been shown to be a successful means of producing generalization, particularly when the exemplars are stimulus exemplars and the stimuli are experimenters. For example, Stokes, Baer, and Jackson (1974) found that the training of retarded children's greeting responses by two experimenters (but not one) was sufficient to produce generalization of the response to over 20 members of the institution staff who had not participated in the response training. The efficacy of training response exemplars or stimulus exemplars in a form other than experimenter exemplars is theoretically promising. But the training of multiple exemplars have not been subjected sufficiently to experimental investigation, except perhaps in the realm of generalized concept learning (Becker, 1971; Carnine & Becker, 1982). Two exceptions follow.

Baer, Peterson, and Sherman (1967) examined generalization across response exemplars. They found that reinforcing a subset of children's motor imitations of puppets produced imitative responses that had never previously been trained or reinforced. Solnick and Baer (1984) highlighted the importance of empirical evaluation of functional response classes. These writers reasoned that such evaluations might guide attempts at training sufficient exemplars and contribute to their efficacy. Solnick and Baer monitored preschoolers' performance on five formats of number-numeral correspondence problems. They found that intervention in perhaps one, two, or at most three problem formats ensured generalization to the remaining untrained formats. Stokes and Baer lament the paucity of research on training multiple exemplars.

It should be noted that the overwhelming majority of studies in the behavior analysis literature subsumed under the generalization categories described above as well as others (i.e., introduce to natural maintaining contingencies, use indiscriminable contingencies) intervened via manipulation of response-reinforcer contingencies. The systematic manipulation of antecedents, and in particular verbal antecedents was rarely explored as a means of inducing generalization. This is surprising, as self-verbalizations constitute responses that can be emitted relatively independent of the immediate environmental conditions. Verbal responses may have stimulus control functions. Thus, the potential for generalization is promising. As Skinner (1966) wrote,

> It is much easier to construct useful discriminative stimuli in verbal form. [Verbal responses are] easily recalled and capable of being executed anywhere. The verbal response makes it easier to learn to discriminate...to retain the discrimination over a period of time...to respond appropriately when the original discrimination is forgotten (p.231).

Similarly, Stokes and Baer (1977) wrote,

> Language is a response, of course; it is also, equally obviously, a stimulus to the speaker as well as to the listener. Thus it meets perfectly the logic of a salient common stimulus, to be carried from any training setting to any generalization setting that the child may ever enter (p.361).

Stokes and Baer explicitly cited mediate generalization as a means of securing generalization. For these writers, mediated generalization required establishing a response as part of the new learning that was likely to be utilized in other problems as well, and constituted sufficient commonality between the original learning and the new problem to result in generalization. The most commonly used mediator is language, but self-control and self-management procedures also exemplify it.

Stokes and Baer cited Risley and Hart (1968) as an example of an analysis of mediated generalization. These writers initially found that contingent reinforcement for four-year olds' reports (true or untrue) of play with a particular item, increased the occurrence of such reports, but did not influence actual play with the item. In a next phase, reinforcement was contingent upon true report only (the children had to both play with item X and report it to earn reward). In most instances, this procedure produced correspondence between saying and doing with play behavior increasing to meet the occurrence of its report.

In a second experiment, the reinforce content--reinforce correspondence sequence was repeated over a series of five additional items or activities. By the time the children encountered the third or fourth item, reinforcing content alone yielded substantial increases in the occurrence of nonverbal as well as verbal behavior. Risley and Hart concluded that generalization had been demonstrated over the course of the items, as verbal behavior began to control non-verbal behavior, such that saying would lead to doing.

Two comments about this study as it bears on generalization deserve mention. First, even in the phases of the study in which generalization was purported to occur, reinforcement contingencies (i.e., intervention) for content were still in effect. When the reinforcement contingency was later switched to a new item, reports and usage of the previous item approached baseline. According to Stokes and Baer, generalization may be claimed when no extratraining manipulations are needed for extratraining changes, or when some extra manipulations are needed but their cost or extent is clearly less than that of the direct intervention. Risley and Hart's manipulations thus met this criterion, as over the course of training, reinforcing content alone yielded results that were initially obtainable only by reinforcing both content and behavior, a definite savings. However, this constituted a more liberal example of generalization than typically encountered in self-instructional training paradigms, where prompts and contingencies for self-verbalization are discontinued during testing.

Second, differential reinforcement of saying led to doing in novel item instances not solely as a function of contingencies applied to corresponding verbal and nonverbal behavior. A likely critical element to the generalization that was obtained was the programming of multiple verbal-nonverbal response exemplars. Though not emphasized by the writers, the multiple exemplar component was most likely necessary for generalized effects.

Lovitt and Curtiss (1968) also trained a verbal response to bridge the learning from one set of conditions to another. Lovitt and Curtiss offered the observation that in comparison with the usual formal evaluation of response-reinforcer contingency manipulations, teachers often manipulate antecedent events in the form of instructions or mnemonics, and evaluations of such instructions are generally casual. Their intervention consisted of simply instructing and modeling the verbalization of arithmetic problems before putting down an answer. During baseline, there was no such verbalization demand. For all three formats of single digit, double digit, and complex format subtraction problems, this intervention substantially reduced error rate and increased accuracy rate. Further, when baselines were reinstated and the child was discouraged from verbalizing, the behavioral effects of intervention (accuracy and error rates) maintained.

However, generalization from problem type to problem type was not evident. That is, as mentioned earlier, high accuracy rates and low error rates did not transfer from one-digit subtraction problems to two-digit subtraction problems, nor from two-digit subtraction problems to one-digit subtraction problems of a more complex format, despite the similarity of the problems. Thus, the intervention which immediately targeted problem verbalization, produced no generalization across highly similar tasks. The effect of verbalization generalized over time, though it might have been the case that reinforcing solution responses would have led to comparable maintenance effects.

Lovitt and Curtiss did not explicitly address this question of the comparative effects of reinforcing solution responses versus manipulating verbal antecedents, but experimenters who did were Grimm, Bijou, & Parsons (1973). This study employed an operant problem-solving model to help two boys learn the concept of number. The arrangement of a one-to-one learning context, in which tutors provided continuous reinforcement for correct responses (outcomes) failed to increase accuracy rates above baseline for either child. Subsequently, training was initiated in which the covert, early part of the response chain was made overt (i.e., verbally identifying the written numerals, verbally counting the symbols, and pointing to symbols while counting). Social reinforcement for each class of chaining response was eliminated over the course of training, until the only overt response the child was explicitly reinforced for was the circling solution response.

Training produced high rates of accuracy which remained high after training was faded out. This study's importance lay in its explicit demonstration of the value of training problem-solving techniques versus response-outcome consequation in acquiring the concept of number. Implicitly, it highlighted the usefulness and versatility of language as a means of targeting typically covert problem solving responses that occur early in the response chain. This targeting of early, covert responses facilitated acquisition. But how then, might the training of verbal mediators facilitate generalization? Dick and Roberts (1982) addressed just this question with respect to response generalization. They wrote that,

> Two behaviors may be topographically related because both share highly relevant components. The keystone behavior is defined as that shared component. Once taught, it serves as an antecedent which affects the subject's skill level for a number of different responses. A program which teaches antecedent keystone behaviors may hold more promise for response generalization than programs which consequate terminal responses (p.3).

It might be noted at this point that the limits of generalization may be topographically determined, but need not necessarily be so. Solnick and Baer described the variety of task formats that they presented to children as exemplars of the concept of number-numeral correspondence, and implied that what may have functionally "held" response classes together were their shared problem-solving response requirements. For example, Solnick & Baer suggested that their preschoolers' performances in formats 4 and 5 (versus 1 through 3) appeared as a response class because in both cases the solution involved counting while simultaneously holding or attending to the target numeral. In other words, a greater memory requirement was present in formats 4 and 5 than in 1 through 3.

For the learning of tasks and concepts, then, where various task instances present both structural differences and response requirement differences, training verbal antecedent mediators may serve to cue common response requirements that occur relatively early in attentional and problem-solving phases. This might pave the way for a greater degree of response generalization than that which typically results from an operant reinforcement paradigm. Training verbal mediators is a method for obtaining generalization that has been infrequently explored within the applied analysis of behavior.

A review of the behavior analysis literature as it addressed generalization, suggested that the procedures adopted by investigators were typically either unsystematic (e.g., train and hope) or failed to constitute valid instances of generalization (e.g., sequential modification). The method of training sufficient exemplars, both stimulus and response, appeared quite effective though its demonstration within this literature was confined largely to multiple stimulus exemplars with experimenters as stimuli.

The training of verbal mediators and verbal problem-solving responses yielded dramatic effects upon response acquisition and maintenance. Though such methods have great theoretical potential for generalization, they only rarely addressed the generalization issue. One study (Risley & Hart, 1968) that did address the issue did not permit a separate evaluation of the effects of verbal mediation from that of multiple exemplars upon the generalization that resulted (a difficulty not uncommon to generalization studies born of other theoretical paradigms as well, to be discussed below). This kind of evaluation might prove quite useful from both conceptual and applied perspectives.

Cognitive Behavior Modification

Bornstein and Quevillon (1976) noted that frequently employed operant procedures such as response cost, contingency management, and timeout have been shown to improve the behavior of hyperactive and disruptive children. However, these procedures often produced limited effects. Some procedures failed to affect the behavior of a percentage of the subjects treated (Kazdin, 1973) or failed to yield training effects over time or across untrained conditions (e.g., Bornstein & Hamilton, 1975; O'Leary & Kent, 1973).

Such results, along with Luria's (1961) model of the development of verbal self-regulation, and behavioral formulations of self-control, encouraged a variety of self-instructional training (SIT) studies with children. To those working in the CBM paradigm, training children in verbal mediational skills was likely to produce a set of self-control skills that the child might apply in a wide range of situations. Meichenbaum and Goodman's (1971) oft-cited study is prototypical of the SIT studies (in terms of both procedures and outcomes) and thus is reviewed in some detail.

Citing the verbal mediation literature, Meichenbaum and Goodman suggested that a training program designed to improve task performance and engender self-control should provide explicit training in the comprehension of the task, the spontaneous production of mediators, and the use of such mediators to control nonverbal behavior. In their initial experiment, impulsive second grade children were exposed to either cognitive training or to one of two control procedures, applied to tasks tapping a range of skills from sensorimotor to complex problem-solving. Cognitive training featured a sequence of steps which progressed toward increased independence: the trainer initially performed a task, talking aloud while the child watched; eventually the child performed the task while verbalizing covertly. The verbalizations which the trainer modeled and the child subsequently used included:

> (a) questions about the nature and demands of the task so as to compensate for a possible comprehension deficiency; (b) answers to these questions in the form of cognitive rehearsal and planning in order to overcome any possible production deficiency; (c) self-instructions in the form of self-guidance while performing the task in order to overcome any possible mediation deficiency; and (d) self-reinforcement (Meichenbaum & Goodman, 1971, p. 117).

The results indicated that the self-instruction (cognitive training) group improved relative to control groups at one-month follow-up on a variety of analogue measures including the Picture Arrangement subtest of the WISC and response latency on the MFF, but not on MFF errors, Block Design, or Coding subtests. This resulted despite the similarity of training and testing tasks. Nor did treatment effects generalize to the classroom, as indicated by behavioral ratings of appropriateness and attentiveness.

Douglas, Parry, Marton, and Garson (1976) applied a training paradigm similar to that of Meichenbaum and Goodman's to a variety of tasks with 6-10 year-old hyperactive boys. Relative to the no-training control group, the trained group showed significantly greater improvement at posttesting and three-month follow-up on a variety of cognitive tasks such as listening, spelling, and oral comprehension tests. Children did not improve however, in terms of behavioral rating scales within the classroom, nor was much evidence found for generalized improvement in arithmetic and reading skills.

Camp, Blom, Herbert, and van Doorninck (1977) employed a "Think Aloud" program to improve self-control in six-to-eight year old aggressive boys. The procedures were reported to be very similar to those described by Meichenbaum and Goodman (1971) in emphasizing the modeling of cognitive strategies and developing answers to the following four basic questions: "What is my problem?; What is my plan?; Am I using my plan? and; How did I do?" (p. 160). Training tasks included cognitive tasks, interpersonal problem-solving games, and a complex version of "Simon says", among others. This training produced substantial improvement relative to controls on a variety of cognitive and psychomotor measures and in terms of classroom behavior as indicated by teacher ratings.

Palkes, Stewart, and Kahana (1968) taught hyperactive nine-year old boys a set of self-directed verbal commands, which essentially asked that they, "Stop! Listen, Look, and Think! Before! I answer." (p.821). While generalization of training effects to the classroom was not examined, the effects of training generalized from the training tasks (MFF), Embedded Figures Test (EFT) and Trail Making Test (TMT) to the posttest measure (Porteus Mazes).

Kendall and Finch (1978) trained impulsive children within a clinical population of emotionally disturbed children in verbal self-instructions via modeling with response-cost contingent upon errors during training. Relative to a control group which did not receive training in verbal self-instruction nor contingent response cost, intervention effects generalized from training tasks to the MFF and to teacher ratings of classroom behavior. Effects maintained at follow-up, two months after posttreatment evaluation.

Robin, Armel & O'Leary (1975) taught five- and six-year-old kindergartners with supposed writing deficiencies to print using either self-instructions or direct training (feedback and reinforcement only). A no-treatment control group was also included. Four upper case letters were used for training while the remaining letters of the alphabet were sampled to test for generalized effects of training. While SIT proved superior to direct training and the control condition in terms of acquisition on the previously trained letters, the effects of training did not generalize to any untrained target letters.

Bornstein and Quevillon (1976) used a self-instructional training package with three overactive four-year-old boys in a multiple baseline design across subjects. Training effects generalized from the training tasks, which included tests of simple sensory-motor skills and more complex tasks such as block design and conceptual grouping, to on-task behavior within the classroom. Treatment gains were maintained 22 1/2 weeks after baseline was initiated.

Friedling and O'Leary (1979) attempted to replicate the work of Bornstein and Quevillon with older children. They exposed seven and eight year-old hyperactive children to either a self-instructional training group or an attention-practice control group which omitted the self-instructional component.

The initial self-instruction, relative to the control condition, produced generalized effects to easy mathematics problems, but not to other academic measures nor to on-task behavior within the classroom. A subsequent program of contingent reinforcement for on-task behavior within the classroom, did not influence any of the academic measures but did yield substantial increases in on-task behavior.

To quote Meichenbaum and Asarnow (1979),

> As one surveys the CBM literature with children who manifest self-control problems, the evidence for treatment efficacy is promising, but the evidence for treatment generalization, especially across response modes and settings is less convincing and often equivocal (p. 11).

What accounted for these often disappointing results with respect to generalization? Several of the studies mentioned above (Meichenbaum & Goodman, 1971; Douglas et al., 1977) were particularly vague in their specification of training tasks, and the relationship between such tasks and the dependent measures on which generalization was sought.

What thus appeared to be at least an implicit assumption of several of these studies was that intervention would engender problem-solving strategies that exercised a functional autonomy independent of the task context. That problem-solving skills exist independent of the task has been debated (Skinner, 1966; Engelmann, 1971).

It is suggested here that if one is to train children under one set of circumstances or conditions and test in another (our defining characteristic of generalization) one must have some model or guideline by which he or she can reasonably expect generalization and justify probing for it, a priori. To illustrate, in looking at generalization across tasks, the hypothetical common stimulus components or common response demands might be presented (e.g., Solnick & Baer, 1984). In looking at generalization across responses, particularly responses whose topography is quite divergent, the hypothesized keystone response components might be explicated. Empirically deriving the limits and directionality of functional response classes may prove to be a very worthwhile endeavor (Dick & Roberts, 1982; Solnick & Baer, 1984). Natural contingencies operative in the test setting that might "trap" targeted behavior and thus foster generalization might be explicated (Friedling & O'Leary, 1979).

A theoretical model of the training and generalization conditions, the corresponding response requirements, or the interaction of these variables with the characteristics of the learner, might function to make generalization more predictable and controllable. The model might help avoid unrealistic efforts that, for example, trained children on modified intelligence tests and examined generalization on academic achievement tasks (e.g., Douglas et al., 1976) or trained analogue task performance in a one-to-one setting and examined generalization to on-task behavior in the classroom (Meichenbaum & Goodman, 1971).

A second question was how to interpret the inconsistent results of the SIT training, given that virtually all studies described above cited the original Meichenbaum and Goodman (1971) study as the model for their training procedures. Relatedly, SIT may be conceptualized as encompassing numerous treatment components, only a subset of which entail the self-instructions per se. How confident can one be, in the case where intervention was successful, that such success could be attributed to children's verbalization of instructions, and not the reinforcement contingencies, nor behavioral modeling, nor even the modeling of self-instructions? Other SIT components typically included overt and covert rehearsal, prompts to self-instruct, fading of prompts, feedback regarding self-instructions and outcome, and social and material reinforcement contingencies to shape and maintain self-instruction. In all fairness, it should be noted that some researchers whose immediate focus was on the outcome of SIT intervention, acknowledged their failure to address process questions and the need to do so in the future (Douglas, et al., 1976; Camp et al., 1976).

In some instances, to complicate matters further, elements of intervention not typically included within the SIT package and elsewhere implicated in fostering generalized treatment effects, were paired with SIT. For example, the Douglas et al. study included elements of training multiple stimulus exemplars in the form of trainers, teachers, and parents. To increase the likelihood that techniques generalized, Camp et al. encouraged children's development of alternative plans, solutions, and outcomes, a package itself shown to yield generalized effects (Shure & Spivack, 1980). In both of these studies, then, it was quite possible that training components other than the conventional self-instruction per se were responsible for generalization. In other studies, it was demonstrated with some certainty that generalized effects were due not to the rule-based intervention per se, but to other components. As we saw, Friedling and O'Leary failed to obtain generalization from the training of intelligence task performance to on-task behavior in the classroom, until such behavior was explicitly consequated, an example of sequential modification. Kendall and Finch (1978) found a direct correlation between response cost occurrences and behavioral improvement in the classroom, and suggested that the response cost enactments were causally responsible for the generalized effects in the classroom.

The conventional SIT package has been subjected to some process research in an effort to tease apart the degree of contribution of various components to training and generalization effects. For example, in a second experiment in the original Meichenbaum and Goodman study, SIT was compared to an attentional control group and a cognitive modeling group. The latter included all treatment components as SIT except for the training and instruction in self-instructions. This controlled for the influence of cognitive modeling per se. Results indicated that children trained in self-instruction showed greater generalization to the MFF in terms of response latency and errors than children in the remaining two conditions.

Complementing and further delineating these results, Palkes, Stewart, and Freedman (1972) trained hyperactive boys to vocalize aloud the self-directed commands or instructed them to read the commands silently, as they worked with the MFF, the Embedded Figure Test (EFT), and the Trail Making Test (TMT). Results indicated that only the children who vocalized aloud showed treatment effects which generalized to the Porteus Maze posttest, though this difference did not maintain at a two-week delayed retest.

While Robin, Armel, & O'Leary's (1975) study did not experimentally dismantle SIT components, process questions were nonetheless addressed. In contrast to the two studies cited above, these researchers cast some doubt on the utility of targeting children's self-verbalizations. In this study, children's rates of spontaneous, overt self-instruction were recorded as a means of providing an outcome-independent check on the effectiveness of self-instruction. While seven of ten children in the self-instruction group spontaneously self-instructed on the target letter posttest, none did so on the generalization posttest. Further, correlations between percent correct letter performance and the rate of self-instructions were nonsignificant. Moreover, it was observed that,

> While some subjects self-instructed correctly, they were simultaneously observed to make incorrect writing responses, suggesting that their verbal and motor response systems were often functionally independent (p. 185).

Roberts and Mullis (1980) found that impulsive first-grade children who received self-instructional training with arithmetic problems outperformed an instructions only and a control group on a test of arithmetic performance. However, they did not outperform children assigned to a behavioral modeling or a verbal modeling group, the latter differing from SIT only with respect to the self-verbalization component. Kendall and Braswell (1982) assigned non-self-controlled problem children, who ranged in age from eight to twelve years, to a cognitive-behavioral (SIT) treatment, a behavioral treatment which differed from SIT only in the absence of the cognitive modeling and self-instruction components, or an attention-control condition.

The results of posttreatment and ten-week follow-up indicated that the cognitive-behavioral treatment was superior to the other conditions on teachers' blind ratings of self-control and non-blind therapist ratings of improvement, while both the cognitive-behavioral and behavioral treatments were comparable to each other and superior to the controls on teacher ratings of hyperactivity and WRAT spelling performance. While claim was made that other academic measures and MFF performance showed cognitive-behavioral and behavioral conditions to be superior to controls, and that only the cognitive behavioral treatment showed improvements in children's self-report of self concept, these claims were made on the basis of post-hoc statistical analyses of trials effects, in the absence of conditions effects or trials x conditions interaction.

A summary of the results of the CBM literature reviewed thus far is tentatively offered at this point. Results are reviewed first with respect to outcome, distinguishing between generalization across tasks (analogue or academic) and settings, and subsequently with respect to process. In terms of treatment outcome, CBM-SIT intervention more often than not yielded behavior change (e.g., observer ratings of on-task behavior) across settings. A fair amount of generalization across tasks from training to testing conditions was observed when the tasks were analogue performance tasks (e.g., MFF, Porteus Mazes, or modified intelligence test tasks) but generalization was far less common, and perhaps somewhat surprising when training and generalization tasks were academically relevant ones. Several writers have indicated that the effectiveness of SIT in generating generalized effects over tasks may well hinge upon task analyses that take into account the degree to which the target skills already fall within the child's repertoire (Robin, Armel, and O'Leary, 1975; Feinberg and Roberts, 1980), the degree to which targeted behaviors are accompanied by stable ability factors (Bornstein & Quevillon, 1976), and the degree to which the tasks entail motoric response components (Roberts & Dick, 1982). These factors may partly account for the moderately greater success SIT has had in yielding generalized results across analogue versus academic measures. Subject factors have also been implicated in mediating successful SIT effects, most notably that of self-attribution, which will be addressed below.

In terms of process, the results of studies reviewed above favored an interpretation that self-instruction, most conservatively taken to include instructions to self-instruct, and modeling, prompting and chaining of self-instructions, was more effective in yielding generalization than treatments which did not include these elements. Less conservatively, there were occasional studies that implicated the self-instructing element per se, versus the modeling of self-instruction, and the overt versus covert rehearsal elements, as essential. However, given the inconsistency of the outcome data, more process research of this nature is needed. The SIT paradigm has yielded some promising results, but it has not been a panacea for failures to generalize, as originally anticipated.

Self-instructional content. It is somewhat surprising, given the strong conceptual impetus for SIT studies provided by the Soviet cognitive-developmental theorizing of Vygotsky (1962) and Luria (1961), that relatively little process research has attended to the nature or content of the self-instructions. Luria, for example, expounded a verbal-developmental self-regulatory progression, the first stage of which was characterized by the inability of the child's own speech to control behavior, and by the ability of adults' speech to exercise inhibitory but not excitatory control. The child progresses to the point where his or her own speech may initiate behavior but not inhibit it, and finally to the point where it is purely the semantic aspects of the child's own speech rather than the motoric, which come to control his or her behavior. Thus, the emphasis during the preschool years is on the child's newly developing and semantically based verbal self-regulatory skills.

The validity of the Lurian model of developmental self-regulation has been the subject of considerable debate and empirical scrutiny (e.g., Wozniak, 1973). In the present context, the model is widely acknowledged to have served as an important heuristic device in providing the original conceptual impetus for a host of CBM studies.

If the goal then, of self-instructional training programs is to teach children a set or repertoire of verbalizations that will serve as stimuli to control a large class of responses or multiple classes, what exactly is it that we wish to teach children to say? The original SIT paradigm of Meichenbaum and Goodman featured:

> (a) questions about the nature and demands of the task, to compensate for a child's failure to comprehend the nature of the problem; (b) answers to these questions in the form of cognitive rehearsal and planning, to overcome a possible deficiency in the spontaneous production of mediators; (c) self-instructions in the form of self-guidance while performing the task, in order to overcome a failure to mediate or regulate overt behavior verbally, and; (d) self-reinforcement (p. 117).

Very few direct attempts have been made to experimentally compare various types of self-instructions as conceptualized by Meichenbaum and Goodman or others, for that matter.

Three recent studies that did experimentally vary the nature of the self-instructions and looked at the effects upon generalized learning were those of Kendall and Wilcox

(1980), Feinberg and Roberts (1980), and Schleser, Meyers, Cohen, and Thackwray (1983). Kendall and Wilcox compared the effectiveness of self-instructional training that featured "conceptual" versus "concrete" statements in working with non-self-controlled problem children, 8 to 12 years of age. These writers reasoned that, "corresponding to notions of metacognitive development, where the focus is on awareness of the thinking process," (p. 81), the conceptual labeling/training procedures were thought to be more likely to affect behavior change and facilitate generalization of treatment effects.

Differential results were not evident on several performance measures (MFF, Porteus Mazes) or subject self-report. However, the conceptual SIT yielded greater generalization than remaining conditions as indicated by teachers' blind ratings of self-control and hyperactivity. This study was the first attempt to experimentally vary and examine the nature of the instructions typically employed within the SIT paradigm. It can, however, be criticized on a number of grounds. First, the authors failed to explicitly attend to the relationship, if any, between tasks used in training and tasks used as indices of generalization. The training tasks, from which generalization was sought, were described in little more detail than, "various psychoeducational exercises." (p. 83). These writers at worst, appeared to be presupposing that their intervention would be effective independent of the task context in which the verbalizations were trained. At best, they insufficiently explicated the stimulus or response requirement commonalities between training and transfer tasks. Doing so might have provided a logical basis from which generalization could have been posited to occur as a function of verbal training.

Second, Kendall & Wilcox wrote that,

> The 'concrete' directions were worded so as to apply specifically only to the task at hand, the 'conceptual' directions, by contrast, were worded more globally and abstractly, in such a way that they could apply to a wide range of situations (p. 82).

Examples of concrete and conceptual statements were given, but the above was as close to a definition of the statements as Kendall and Wilcox came. The difficulty with their definition was the ease with which it led toward circular explanations of differential treatment effects. The effect of intervention became its own explanation. Hence, if a "conceptual" or "general" approach yields greater generality of training effects, it is the result of the general training. And we know the training was general because it yielded stronger generalization effects. Clearly an account of this nature contributes little to an understanding of the generalization process. The results of studies of generalization which employ rule-based or verbally-based interventions can themselves be generalized only to the extent that a conceptual framework is explicated that identifies the commonalities within various sets of rules and verbalizations. This must not be done solely on the basis of the differential effects such sets of rules or verbalizations might have upon behavior across tasks and time.

One model which might prove useful in this respect is the problem-solving model of Grimm, Bijou, and Parsons (1973) which will be discussed below. Feinberg and Roberts (1980) experimentally varied the nature of the self-instructions within an SIT paradigm in an attempt to replicate Kendall and Wilcox. These writers trained first-through fifth-grade learning disabled children to employ either concrete or conceptual self-instructional statements as they worked with phonics, vocabulary, and reading-related tasks. Although a generalized effect to mathematics tasks was not found for either the concrete or conceptual training nor for the direct instruction control, differential effects were found on the Spache Reading Test.

Mildly deficient children profited more from the conceptual intervention than did children in the other two groups, while severely deficient children profited more from the direct instruction control than did children in the other two conditions. This finding was consistent with others that have implicated the role of available response repertoires as determinants of SIT outcome (Higa, Tharp, &

Calkins, 1978; Robin, Armel, & O'Leary, 1975; Bornstein & Quevillon, 1976).

With regard to the current discussion, Feinberg and Roberts objectively anchored their distinction between concrete and conceptual statements by adopting their concrete statements from the task directions in elementary textbooks and workbooks which provided the training tasks. The conceptual statements were adopted from those employed by Palkes et al., (1968; 1972) and Camp et al., (1977). While this attempt to objectively differentiate the conceptual and concrete statements was a step in the right direction, the general (conceptual) and specific (concrete) labels still encourage circular explanations. Also, as with the Kendall and Wilcox study, generalization was sought across widely divergent tasks without sufficient explanation of common stimulus elements or response requirements between tasks upon which the verbally based intervention might operate.

Schleser, Meyers, Cohen, & Thackwray (1983) included comparisons between treatments which featured,

> ...specific content instructions designed to provide an optimal strategy for successful performance on a math training task, [or a] broad problem-solving strategy applicable to a variety of tasks but not anchored to a particular task (p. 954).

The results pertaining to these two experimental conditions indicated that the specific content self-instructions proved superior to the general instructions on a task similar to training (PIAT Math) but the general instructions were superior to the specific on untrained generalization measures (PIAT Spelling and General Information). As with the previous two studies, no model was offered a priori to distinguish and independently validate the "specific" and "general" content self-instructions. Nor was an account offered which might have explicated the common task variables or task-approach variables upon which intervention might have successfully operated.

One model that might prove useful in this regard is the problem solving model proposed by Grimm, Bijou, and Parsons (1973). These writers suggested that problem-solving is not simply concerned with emitting a solution but with the techniques of finding the solution.

> A problem-solving sequence begins with an external stimulus having a discriminative function, goes through a series of observable and nonobservable stimulus-response-stimulus interactions, and ends with an external response and a reinforcement (p.27).

Failure to emit the complete problem-solving sequence including the solution response might occur because the response is unavailable, the conditioned reinforcers in the chain are not functional, or because stimuli in the chain do not exercise discriminative control. Grimm et al. applied their model to the learning of number-numeral correspondence problems in a single subject design with two boys enrolled in a class for retarded and emotionally disturbed children.

Focusing on the discriminative control element, training was initiated in which the covert part of the response chain was made overt so that mediating responses could be monitored and reinforced. This intervention led to high degrees of accuracy, whereas prior reinforcement of correct solution responses alone, did not.

Though not emphasized by these writers, making the covert part of the response chain overt entailed verbal training, specifically the verbal identification of written numerals and counting symbols aloud. That verbal training might be a particularly useful means of strengthening a problem-solving sequence was implied by Winokur (1976) who asked,

> Can the mander and the reinforcement mediator be one and the same person? It seems that they can. As a results of having been conditioned to (a) mand and (b) mediate, by someone else, we play both roles to others and ourselves. Adults as a accomplished manders-compliers, seem to approach everything as if it were a verbal problem, and they use the verbal responses as discriminative stimuli to cue nonverbal responses (p. 38).

Teaching children to verbalize stimulus elements, mand their own behavior, or state partial or full contingencies might prove to be an effective means of establishing problem-solving sequences of behavior. The problem-solving model of Grimm et al. importantly highlighted the need to attend to early overt or covert task requirements in the problem-solving sequence. Lloyd (1980), citing the Grimm et al. study among others, called for "attack strategy training," (p. 59), to promote generalization. This training featured the training of specific strategies as opposed to training in self-verbalization and self-instruction which Lloyd deemed unnecessary. He wrote that task-specific strategies appear to have a better chance of promoting generalization than general strategies at least in the realm of academics.

There is no logical reason, however, to dismiss self-verbalization training if one favors an approach which stresses the training of attack strategies. As already pointed out, the Grimm et al. study employed verbal training as a means of making implicit, covert, and specific problem-solving responses explicit and overt. Verbal training may well be an important means of fostering generalization by strengthening "specific" problem-solving responses common to a range of tasks. This model with its distinction between problem-solving and solution responses, in addition to guiding task analyses, may also help discourage nonconstructive conceptualizations of general and specific statements.

In sum, few studies directly examined the differential influence of the self-instructional content. The handful of studies that did yielded sharply diverging results for the effectiveness of the conceptual/general training relative to the concrete/specific. These ranged across studies from an absence of generalization effects for academic measures

(Feinberg and Roberts, 1980) to generalization on behavioral rating scales only (Kendall & Wilcox, 1980) to substantial generalization effects across widely different academic measures (Schleser et al., 1983).

All three studies can be faulted on two common grounds. First, they offered labels for their categories of verbalization training that may encourage circular accounts of the effects they tried to explain. Second, they failed to explicate adequately the relationship between training and transfer tasks. Both may hinder an understanding of the training and task circumstances that do or do not lead to generalization.

Grimm, Bijou, & Parsons (1973) offered a conceptual framework for problem-solving that may be useful in objectively anchoring verbal self-instructions and devising task analyses. It is also quite compatible with Stokes and Baer's (1977) concept of mediated generalization. A problem-solving response may form a common core for a range of tasks. As discussed above, verbal training has been proposed to be a theoretically important means of strengthening a problem-solving response and sequence. Training a verbal response which specifies this common problem-solving response would then be a potent means of securing generalization, and would fall within the realm of operations discussed earlier as mediated generalization. A response is established as part of the new learning that is likely to be utilized in other problems as well, because it specifies a shared response component that enhances the probability of solving the problems.

Metacognitive Development

Discussion will now, for the moment, turn away from cognitive-behavioral and behavioral conceptualizations and remediations of the generalization problem and turn toward applied investigation and conceptualization of metacognition and its development. Metacognitive development operates from a paradigm conceptually distinct from traditional behavioral and cognitive-behavioral approaches. It is concerned with the acquisition of knowledge and cognition about cognitive development.

Brown (1975) made a distinction between knowing how to know and knowing about knowing. The former referred to mnemonic strategies which were deliberately instigated for the purpose of remembering. In reference to the latter, which encompassed metamemory, Brown wrote that young children do not realize a need to memorize: they are oblivious to the limitations of their memory capacity. For Brown, it is the intention to use an appropriate strategy, "subordinated to the goal of remembering" (p. 113), that is deficient in the developmentally young, and not any specific memorial skill per se.

Flavell and Wellman (1977) suggested that sensitivity to instructions to remember and subsequently employ a mnemonic is an index of metamemory. They suggested that young children seldom deliberately try to retrieve or prepare for future retrieval, in response to situations that commonly elicit precisely those sorts of cognitive efforts in mature individuals.

Brown (1978) described the metacognitive processes as including: (a) prediction and planning which precede problem-solving attempts; (b) checking and monitoring which are subsequently performed to evaluate the outcome of these attempts; (c) checking outcomes for internal consistency and against common sense criteria. In short, such processes as checking, planning, self-questioning, self-testing, and monitoring ongoing attempts to solve problems are viewed as central components of metacognitive development.

For the present writer, the concept of metacognition became clearer as data that gave rise to notions of metastrategic failure were reviewed, and interventions which allegedly tapped metacognitive awareness were examined. For example, citing Meichenbaum and Asarnow (1979), while kindergartners know that a memory task is harder if it has a large number of items, only older children know that a recall task is harder if one has to learn two sets of words that are easily confused. According to Brown (1978) third-graders were said to demonstrate metacognition when 95% of those queried reported that they would prefer to phone a friend's number right after getting it rather than get a drink of water first. Kindergartners generally failed to demonstrate awareness of planful behavior, as only 40% preferred to phone immediately. Brown reported that if the task is sufficiently simple, evidence of planful (metastrategic) behavior can be seen in children as young as three. For example, Wellman, Ritter, and Flavell (1975) inferred metacognitive awareness on the part of three-year olds who, when asked to remember the location of a toy subsequently hidden by a cup, touched the cup or pointed to it, as a means of correctly responding.

Brown and Barclay (1976) claimed to implicitly address metamemorial processes in their intervention when they instructed children to look at the stimulus pictures as long as they wanted, and only when they knew them very well, to ring the bell and say them back, on a recall of common objects task. Kendall, Borkowski, and Cavanaugh (1980) were said to train metastrategies when with paired associated picture tasks, they taught children to verbalize a relationship (e.g., the <u>nurse</u> holds the <u>toaster</u>), and give a reason for the relationship. Metastrategy training also featured the provision of feedback as to the value of the strategy.

According to Brown and DeLoache (1978) these metastrategies of checking, planning, question-asking, self-testing, and self-monitoring are basic characteristics of efficient thought, and one of their most important properties is that they are transituational. If so, then what better way to engender generalizable training effects than to train these strategies? Meichenbaum and Asarnow pointed out the similarity of Brown's description of the elements of metacognitive processes and the content of the self-instructions as typically formulated by CBM interventions such as those of Palkes et al., (1968; 1972) and Camp et al. (1976). Brown, Campione, and Barclay (1979) also acknowledged the similarity and have in fact explicitly called for self-interrogatory training with the eventual aim of training the child to think dialectically, as fostered by a socratic teaching method. These authors endorsed providing the child with a routine set of questions to ask himself before proceeding, for example: (a) Stop and Think! (b) Do I know what to do? (c) Is there anything more I need to know before I can begin? and (d) Is there anything more I need to know that will help me (i.e., is this problem in any way like one I have done before)?

This routine is strikingly similar to that proposed originally by Meichenbaum and Goodman and many of the other SIT reformulations reviewed above. Thus, two distinct conceptual paradigms studying children's behavior or memory performance on quite dissimilar tasks and often in divergent settings with different population samples, converged in their implication of self-interrogatory strategies as a promising means of promoting generalization.

This section thus far has attempted to provide the reader with a working understanding of metacognitive training by offering some background in its conceptualization and instances of its application. Recent redirections of its application were also discussed.

This leads to a discussion of the empirical impetus for metastrategy training and a review of the results and implications of select studies which featured metastrategic interventions. The following section will attempt to integrate the empirical data and address the conceptual status and utility of the metamemory approach in comparison to a behavioral or cognitive-behavioral approach.

As with the CBM-SIT approach discussed above, the empirical impetus for studies of metastrategy training came from more conventional studies which often yielded training effects of surprisingly limited generality. These studies typically entailed training children in the use of various mnemonic strategies and testing for recall at another time, on another task, or in another context. Mentally retarded or learning-disabled children were often the population from which the sample was drawn, but parallel results were found with younger (typically preschool) children.

Brown (1978) wrote that slow learners and retardates do learn mnemonics, do improve their recall and that these effects can be maintained over time, but evidence for generalization is difficult to find. Brown and Barclay

(1976) found that explicit training in a suitable mnemonic was sufficient to improve and sustain the performance of older retardates on a serial recall of pictures task, but lasting effects were not found for younger retardates (mean MA, 6 years, 8 months). Brown interestingly and anecdotally reported that her lab never found reliable differences between experienced and naive subjects entering a new experiment. This informal observation offered an independent substantiation of the narrowness of children's learning within Brown's experiments. But results of this nature are not limited to samples of retarded children.

Studies of preschool, kindergarten, and second-grade children found that transfer on discrimination learning tasks did not occur when the task format changed from successive to simultaneous discrimination or the reverse (Campione & Beaton, 1972; Campione & Brown, 1973). Crisafi and Brown (1983) working with normal two, three, and four-year olds on a series of increasingly difficult inferential reasoning tasks, found no generalization across tasks for two-year olds and only limited generalization across tasks for three-year-olds. The extent of generalization for the latter group of children, limited as it was, was attributed to the stimulus similarity of two of the three tasks.

Brown and Barclay (1976) in a study referred to briefly above, "implicitly confront[ed] the executive control problem," (p. 73), in teaching young (mean MA, 6-8) and old (mean MA, 8-5) retardates to recall the names of pictures of common objects. The design compared the effectiveness of a label condition, an anticipation condition, and a rehearsal condition,the latter two of which were described to entail a critical self-test element.

Results indicated that on a prompted posttest the day after training all children in the anticipation and rehearsal conditions outperformed children in the label condition, but for the younger children these differences did not maintain two weeks later on an unprompted posttest. Some process (observational) data indicated that the younger children in the anticipation and rehearsal conditions still used their trained strategies during the unprompted posttest, prompting these writers to conclude that the younger subjects did not abandon the strategy, "but failed to monitor [its] efficient use." (p. 78). For Brown and Barclay the strategy was no longer, "subordinated in a meaningful way to the goal of the metamemory task." (p. 79).

Nonetheless, the results of the study were promising for older children at least, and invited Brown, Campione, and Barclay (1979) to continue and extend the investigation with the same children. With respect to maintenance of intervention effects on the training task, it was found that one year later and without prompts, the same reliable group differences maintained for the older children. For the younger children, no such differences were found and performance remained at very low levels. When children were reminded of their formerly trained strategies, differences emerged for the younger children but once again did not persist.

During a generalization phase three months later which entailed the older children only, children in the anticipation and rehearsal groups outperformed those in label and control conditions. The task required them to recall the gist of reading passages, as measured by ratings of the importance of the idea units recalled. While Brown et al. called this study their first successful attempt in teaching a generalized cognitive skill in EMR's on a quite dissimilar recall readiness task, there were some difficulties with the study. Claimed group differences during the generalization phase were apparently not substantiated with appropriate post-hoc analyses, and as the authors point out, observational data did not rule out the amount of training as a confound. Implicating contingencies, by telling children that it "helps some people to underline, mark the paper, take notes, etc., to check if they are ready for a test," (p. 507), did not improve recall nor did it lead to increased note-taking, underlining, etc., in this sample of EMR children.

That such intervention might prove helpful was indicated by the work of Kennedy and Miller (1976) with normal 6-7 year olds. Following training to verbally rehearse on a serial recall task, those who subsequently failed to spontaneously rehearse were divided into feedback and no feedback conditions. Only children in the feedback condition, who received information about the value of the verbal strategy, maintained its usage.

"Information," however, included verbal reinforcement, prompting, feedback, and had strong demand characteristics (e.g., "My goodness you did so much better when you whispered those names over and over. I guess whispering helped you remember the pictures better. Right?" p. 567). Kennedy and Miller properly highlighted the need to look at the effects of such feedback both across tasks and time.

Belmont, Butterfield, and Borkowski (1978) worked with 12-15 year olds on letter recall tasks with sequential letter presentation. All children were trained to use an appropriate mnemonic strategy on a first task, but on a second and highly similar task half the children were given "generalization training." The results two weeks later indicated that the generalization training was superior in terms of far generalization (to a letter position probe task) and long-term near generalization. A problematic aspect of this study was the vagueness with which the generalization training was described. The authors briefly described the strategies taught and their intent to "have the child understand the harmony of input and output processes and the similarities and differences between [the different strategies]," (p.421), but they were imprecise in their account of how such strategies were taught (e.g., via verbal instruction with rules, verbal modeling, verbal prompting, reinforcement, etc.). Another difficulty was that the generalization group received twice as much training as its control, and hence the amount of training was inextricably confounded with the nature of such training.

Kendall, Borkowski, and Cavanaugh (1980) trained EMR children (Mean MA = 6.9 years) on paired-associate picture tasks and tested them on the training items (retention), new paired associates lists (maintenance) and paired-associated triads (generalization). Results indicated that relative to two control groups one of which importantly received comparable task exposure, the interrogative training group was superior at retention and maintenance test for low MA children, and superior to controls at maintenance and generalization for high MA children. Interrogative training, as mentioned earlier, consisted of: training the children to verbalize the relationship (e.g., the <u>nurse</u> holds the <u>toaster</u>); training them to formulate a reason for the relationship; and verbal feedback regarding the value of the strategy.

Once again, a component analysis is called for to tease apart the relative contributions of the various elements of the interrogative training. Although interrogatory training yielded generalization effects over tasks and time, it is noteworthy that there were no differential effects of training on a metamemory questionnaire administered pre- and posttest, leading these authors to suggest that "a limited set of memory experiences will elevate only task-relevant metamemory." (p. 269).

A study by Lodico, Ghatala, Levin, Pressley, and Bell (1983) was impressive in the degree of generalization obtained across markedly divergent tasks with second graders. These writers proposed to experimentally manipulate the metacognitive processes only implicitly addressed by studies along the lines of Brown, Campione, and Barclay. During an initial phase, the experimental group was prompted to draw a circle both freehand, and with a cookie cutter, and was then asked which was better and why, and given feedback regarding their responses to those questions. Inferior and superior strategies were similarly trained on a letter jumble task again with questioning and feedback. A control group practiced the same strategies on the same tasks but received no instructions about the value of monitoring their performance or selecting the best strategy.

All children were subsequently trained to use both an inferior and a superior strategy on paired associate and free recall tasks, and were subsequently permitted to use the strategy of their choice. Results indicated that more experimental than control children utilized the more effective strategies when given a forced choice for both tasks. These writers concluded that instruction in general memory-monitoring principles was sufficient to effect a change in strategy usage, and that when first- through third-graders were given feedback regarding the strategy's usefulness, it maintained.

Although the generalization of training from circle-drawing to free recall was quite impressive, the results called for one caveat to be made. The experimental children exhibited recall performance superior to the controls on the paired associate tasks before the forced-choice (test) trial, and this, "motivational superiority" (p. 273), could have carried over to the forced-choice trial, and could not be ruled out as a confound. Lodico et al. wrote that their study provided direct experimental support for the presumed relationship between metacognitive knowledge and subsequent strategy use.

In the present writer's opinion, however, this statement was too strong and its implicit inferences unnecessary. Lodico et al.'s study effectively and, given the tasks employed creatively pointed to the utility in some instances of training skills in strategy-monitoring, or learning to discriminate the effectiveness of various strategies for various tasks. Not only was little gained by attributing the success of training to its impact upon metacognitive knowledge, but accounts of this nature may lead away from precise formulations of training elements or processes necessary and sufficient for highly generalizable training effects. Lodico et al.'s metacognitive study was not unique in drawing excessive inferences about cognitive processes from the data. This discussion of metacognitive conceptualization will resume following review of one other generalization study subsumed under the rubric of metacognitive training.

The last study to be reviewed in this section is that of Crisafi and Brown (1983) who tested the commonly held assumption that preschoolers are particularly poor at transferring information. In a cognitive-developmental study, normal two, three, and four-year olds were trained and tested on a series of increasingly difficult inferential reasoning tasks, on which for example, they were first trained to obtain either a penny or dime from a purse or piggy bank, and second, to use that coin to obtain a gumball from a gumball machine. More difficult versions of the task used novel coin containers, tokens, reinforcers, reinforcement containers, and required different topographical responses to obtain secondary and generalized reinforcers. Children were trained on the two phases of each task separately, and then tested with all stimuli present to determine if they could emit the entire motoric sequence. Four experiments were performed which examined the impact upon generalization of: the easy to hard sequence of task presentation (multiple task exemplars) alone; the sequence plus verbally emphasizing the similarity of the tasks, and; the sequence plus teaching children to verbalize task rules.

The results of this series of experiments indicated that the easy to hard sequence alone, in comparison to a sequence in which two irrelevant tasks preceded the third inferential reasoning task, produced: no generalization in two-year olds; some in three-year old when the problems were structurally very similar, and; transfer in four-year olds even when the problems were dissimilar. Providing hints that the tasks were similar and teaching children to state task rules, each superimposed over the easy-to-hard sequence, facilitated transfer relative to the sequence alone in children as young as two and three.

This study was important and unique in its demonstration of the power of a verbally based intervention in conjunction with multiple exemplars to promote generalization across tasks in comparisons with a more conventional and established means of doing so via the training of multiple task exemplars alone. As such, it constituted a process study, much needed within the literature. Crisafi and Brown also chose to examine generalization across a series of tasks bonded by common conditional discriminations and response requirements. This served as a structural basis from which Crisafi and Brown could reasonably and logically expect that their rule-based intervention might generalize.

As with much of the metastrategy training literature and the CBM training studies as well, the verbally based interventions in which task similarity was stressed and task rules were prompted, entailed many different training elements. These included instructing the children that the tasks were the same, the experimenter's verbal specification of initial response requirements, sometimes with strong demand characteristics and sometimes not, verbal prompting for the child to verbalize task rules, and so forth. Also, no mention was made of the experimenter's role in the case of poorly formulated rules. Were these rules ignored, corrected, differentially reinforced, shaped, etc.? Given the multiplicity of elements in the rule-based interventions, it was unwarranted of Crisafi and Brown to attribute generalization effects to the self-verbalization element per se. Nonetheless, the study remains one of the most important demonstrations of the influence of rule based interventions in addressing the learning of generalized skills.

Belmont, Butterfield, and Ferretti (1982) reviewed transfer of training studies, largely within the memory literature, and concluded that important transfer could be achieved only if general skills such as goal setting, strategy planning, and self-monitoring were trained in addition to the specific skills whose transfer was sought. Metaprocessing was conceptualized by these writers to be a superordinate function, qualitatively different in process than the specific skills to be controlled. These writers offered the observation that six of seven studies reviewed that produced substantial transfer not only instructed subordinate skills but led children to, "see the wisdom" (p. 150), of metastrategies such as defining goals, designing plans, and monitoring implementation and outcomes of those plans.

It is acknowledged that several of the metastrategy training studies are promising in their demonstration of generalization across tasks, and have involved procedurally inventive and important interventions. However, this literature is subject to criticism as well. It is argued that the invocation of the construct of metacognition and the demand for metastrategy training which follows involve unnecessary degrees of inference, which, rather than throwing light on the phenomena under study may, unfortunately, maintain vague explanations of generalization effects.

More specifically, it is argued here that deference to metacognition need not be made on occasions when operationally, children are taught to discriminate the products or process of their own behavior, and perhaps verbalize the discrimination as well. Relatedly, no special status is attributed to awareness, on whatever basis it is inferred, as a primary cause of behavior. Verbal report, as verbal behavior like other behavior, may have stimulus functions which set the occasion for subsequent behavior, overt or covert. There is no need, however to attribute special status, or assume a qualitatively different cognitive process has been engaged when one rule-based intervention is more successful than another, or when the child is trained to discriminate and report aspects of his own behavior. To do so entails unnecessary inference at the expense of parsimony.

This is not to say that the success of such training is not of both applied and theoretical interest. Verbally based interventions need to be subjected to process research to determine what components are critically implicated in acquisition, maintenance, and generalization processes. Explanatory accounts which take verbal and non-verbal behavior at face value offer greater parsimony, and in this writer's opinion are implicated by much of the data from metastrategy training studies themselves. For example, Salatas and Flavell (1976, cited in Brown, 1978) indicated that although children were aware of the need for strategic intervention, they did not apply such strategies successfully. The failure of Kendall et al.'s interrogative training condition to yield differential training effects on a metamemory questionnaire led them to speculate that a limited set of memory experiences will elevate only task-relevant metamemory. Recall that metamemory was discussed by its advocates as a superordinate process which transcends the particulars of the task at hand and is broadly deficient in the very young. The concept of task relevant metamemory is thus self-contradictory and illogical. That awareness, as indexed by verbal report, is not necessary for acquisition of a response was additionally supported by many studies of human operant conditioning (Hefferline & Keenan, 1963) though this finding was not undisputed (DeNike & Spielberger, 1963). Different sets of skills and behaviors may control different sized sets of responses. What is critical to learning and generalization may be the learning of relevant problem-solving responses that facilitate solution responses for a given problem or class or problems. Verbal training methods may serve as useful means to learn relevant problem-solving responses, but verbal report of conditional discriminations, or other indices of metacognition may only accompany learning and generalized learning and not cause it. In fact, as we have seen, such verbal report may not even accompany it. For many operating within the metacognitive training paradigm (e.g., Brown & Barclay, 1976) if B fails to affect C (given sufficient levels of other necessary factors such as attention, motivation, etc.), a deficit at A is presumed and targeted for intervention (see figure 1).

Figure 1

<u>Metacognitive Approach to Strategy Training</u>

Metastrategy training	A Metaknowledge	C Outcome performance
Strategy training	B Strategy knowledge	

If the strategy is trained and successfully used in one set of (training) conditions, but not used in another (maintenance or generalization) set of conditions, then a failure in strategy monitoring or awareness is postulated, and metastrategic intervention devised and implemented. An alternative and more precise conceptualization is offered by a behavioral account which ascribes no special status to awareness. In this account, reliance is placed upon direct observation and minimal inference in an attempt to remain objectively close to the data. Thus, the metacognitive psychologist's discussion and examination of strategic knowledge, for the behaviorist becomes a discussion and examination of strategic behavior and verbal report of such behavior or conditional behavior (see Figure 2).

Figure 2

Behavioral Approach to Strategy Training

	A	C
Train conditional strategies	Discriminate strategic behavior	Outcome behavior
Train to verbalize conditional strategies	Report of discriminated strategic behavior	
	B	
Train strategies	Strategic behavior	
Train to verbalize strategies	Report of strategic behavior	

In this account, verbal and nonverbal behavior are viewed as independent streams of behavior (Roberts, 1979) which may or may not, depending on the circumstances, affect each others probability of occurrence. Rather than appealing to contrary notions of "task-relevant metamemory" (Kendall et al., 1980, p. 269), we can discuss the relationship, if any, between children's report of strategy usage in one set of circumstances and their report of strategy usage in another. And we can examine the relationship between strategic behavior in one set of circumstances and in another.

If metamemory is not commonly demonstrated except on the particular task or class of tasks targeted for training (Kendall et al., 1980), if metamemory or at least its training is demonstrated sometimes to fail to affect strategic behavior (Salatas & Flavell, 1976), and if successful strategic behavior has no necessary accompanying metaawareness (Kendall et al., 1980), then its conceptual status as a higher-order, executive process superordinate to and distinct from strategic processes is dubious. This position would likely be endorsed by Engelmann (1971), who wrote,

> The notion of non-specific operations is rejected. An operation is applicable only to certain concrete problems. The subject must somehow be able to see that certain aspects of the problem imply a particular operation. Without this assumption, the operation would be used either universally or randomly. If it is used in a discriminated manner, there must be a basis for discrimination, which means that the operation is specific to a certain set of cues. The operation can be applied to a wide variety of situations, but the operation still remains quite specific (p. 463).

Loper (1980) suggested that asking children to monitor themselves with, "Do I understand?" may be effective if children have sufficient metacognitive awareness to ask if they are progressing. This implied that if the intervention proved unsuccessful, children had insufficient metacognitive awareness: its success would implicate sufficient metacognitive awareness. Clearly, in this instance, stopping at the construct of metacognitive awareness as an explanatory account of behavior is unsatisfactory and circular. It might prove more fruitful to ask, under what circumstance, with what kind of learning history with both verbal and non-verbal stimuli might such self-questioning or self-monitoring strategies prove effective? This leads to a brief summary and discussion of both the CBM-SIT studies and the metastrategy training studies. Before proceeding, however, it should be acknowledged that the preferences of metacognitive researchers to employ cognitive constructs such as metacognition as causal accounts of behavior, and for behavioral and cognitive-behavioral researchers to avoid or minimize the use of such constructs, reflect paradigmatic differences which entail different underlying theoretical assumptions about the nature of the human organism.

The present discussion of the various paradigmatic approaches to the issue of generalized learning does not presume to decide that one set of assumptions more truthfully addresses the nature of human learning than the other. Nor is the present study being touted as a critical test that will help decide the validity of one paradigm versus the other. It is suggested, however, that a behavioral or cognitive-behavioral approach more exactly deals with the data at hand and less often invokes explanatory accounts which stop at the level of inferential constructs. Such accounts may conceal more complex relationships between verbal and non-verbal behavior in a particular learning context, given the child's verbal and non-verbal learning history.

CBM-SIT and Metacognitive Training Studies

Acquisition and generalization studies operating out of the CBM-SIT and metacognitive training paradigms did what relatively few behavior analytic studies have attempted to do: examined the influence of rule-based and verbally-based interventions upon learning and generalization, primarily across tasks and time. As noted above, a distillation of the results of the CBM-SIT studies indicated that generalization was more commonly observed across tasks when the tasks were analogue rather than academic, and subject factors such as the child's relevant behavioral repertoire, were implicated. Generalization of on-task behavior across settings was often but not always observed.

With regard to the metacognitive training studies discussed above, generalization across tasks was often achieved, sometimes across tasks strongly divergent in terms of response requirements and stimulus features. Studies operating from both paradigms often failed to examine critical process questions. This was particularly true for the metacognitive training studies, though in fairness, this training represents a newer endeavor. It behooves the investigator to design his or her study around experimental questions that not only ask, "Does it work?", but "Why does it work?" The SIT and metacognitive training paradigms are clearly far too complex procedurally for a single process study to definitively tease apart unnecessary, necessary, and sufficient training elements, but each study should minimally attempt to isolate and scrutinize one facet of training. In this manner, various studies might build upon one another and converge to implicate critical treatment elements and processes, as was the case, for example, with the study of systematic desensitization (Marks, 1978).

One criticism common to the vast majority of both SIT and metacognitive training studies was that they typically assumed that the presentation of rules or training of self-verbalizations would control the behavior or strategies they specified. Relatedly, it was further assumed that whereas the trained verbal strategy or behavior might fail to maintain across conditions, the trained metastrategy or general, verbal problem-solving skill would itself maintain and subsequently control behavior.

That verbalization training alone may not be sufficient to control specified behavior was repeatedly demonstrated in the literature on verbal-nonverbal correspondence training, even when the responses specified by the verbal self-instructions were well within the child's repertoire (Risley & Hart, 1968; Israel & O'Leary, 1973; Feinberg & Roberts, 1982; 1984). For example, in the Risley and Hart study mentioned earlier, the enactment of reinforcement contingencies for preschoolers' reports of play, true or false, did not increase rates of play behavior, except when reinstated following a period in which only true reports of play were reinforced.

A study by Carnine, Kameenui, and Maggs (1982) more germane to the academic arena, is also illustrative. In a dismantling paradigm, these experimenters trained one group of first-graders on a task requiring a classification and discrimination sequence, to verbalize both the concept and the rule. A second group was additionally trained to classify in accord with the concept, while a third group received all training the above two groups received but additionally were trained to behave in accord with what the rule specified. Results indicated that only the last group's performance was significantly above chance. This suggested to these writers that for complicated rules, ensuring that primary students issued concept and rule statements did not guarantee that the students learned to apply the concepts.

For the training of children's verbalizations to have the generalized effects hoped for, it may prove helpful to explicitly assure that the verbalizations control the behavior and strategies specified. The role of this aspect of SIT and metacognitive studies in promoting generalization has typically been either neglected procedurally or only very casually acknowledged.

Targeting Verbal Antecedents: Process

Thus far, emphasis has been placed on the content, or stimulus control function of self-instructions as typically employed within the SIT and metacognitive training paradigms. Equally deserving of attention, however, may be the process by which children are led to verbally identify task stimuli, response requirements, or contingencies. Data generated from studies representing a variety of theoretical paradigms bear importantly on this issue. Studies to be discussed include those of: Tennyson, Youngers, and Suebsonthi (1983) and Richards and Siegler (1981) operating from a cognitive-developmental perspective; Schleser et al. (1983) operating from a cognitive-behavioral paradigm; Catania, Matthews, & Shimoff (1982) operating from a basic, operant approach, and Guthrie (1967) operating from an applied-educational perspective.

Tennyson, Youngers, and Suebsonthi (1983) compared four methods of teaching third grade children the concept of a regular polygon. In a 2 x 2 factorial design, presentations of polygons emphasized either best examples of polygons or a statement of the critical attributes. The polygon examples or statements of attributes were presented in either expository and interrogatory form or interrogatory form only. Results indicated that, in terms of children's ability to properly classify new regular polygon instances, the presentation of best examples along with a definition facilitated prototype formation more than did a statement of the critical attributes along with the definition.

Also, an interrogatory strategy that required children to respond (classify) and then provided feedback, was inferior to one that featured both an interrogatory and expository format. An earlier study by this lab (Tennyson et al., 1981, cited in Tennyson et al., 1983) similarly showed an additive effect of an instructional presentation that used both expository and interrogatory examples in comparison to either presentation alone, in classifying new concept instances. In neither study did the interrogatory training alone result in maximal gains in terms of children's generalization of their classification skills to new concept instances.

The results of this study also bear on the earlier discussion of the implementation of multiple exemplar training and rule-based intervention as means of facilitating generalized learning. It appeared from the results of the Tennyson et al. (1983) study, that children learned mathematical concepts more readily from an approach which emphasized the presentation of multiple exemplars of concept instances, than one which emphasized rules governing those concept instances. The presentation of rules specifying concept properties was not prerequisite to generalized concept classification.

Also operating from a cognitive-developmental perspective, Richards and Siegler (1981) compared the effectiveness of training children "to take an analytic attitude," (p. 1319) with a control group in teaching three-year olds to predict which side of a balance scale would descend. Balance scale problems varied in terms of the distances of weights from the fulcrum and the amount of weights on each side of the scale. Results indicated that children trained to take an analytic attitude produced more correct answers than control children who were only given right/wrong feedback after each training problem. After each answer during training, children in the analytic attitude condition were told to,

> ...look carefully at the balance scale, and see if you can tell why (this side went down) (it stayed balanced). See if you can figure out what made it (go down to this side) (stay balanced), (p. 1319).

It is unclear procedurally whether or not children's verbal responses to this prompt were prompted further or explicitly required, and if so, whether particular verbal responses were consequated or corrected.

A study by Schleser, et al. (1983) was discussed above in reference to the process studies of the content of the self-instructions employed within the SIT paradigm. In addition to the specific or general content self-instructions, a condition called specific content self-instruction via directed discovery was also compared to the control condition. In this discovery condition,

> Children were led, through a question-and-answer socratic dialogue with the experimenter, to 'discover' the [specific content] strategy statements... The experimenter rephrased the child's responses if necessary and had the child cumulatively rehearse all discovered statements (p.954).

Results indicated that the training which featured directed discovery was not different from the specific content training and superior to the two remaining conditions on arithmetic tasks similar to the training tasks. Directed discovery training was superior to both specific and general content training on the MFF, which served as an index of generalization. Further, only the directed discovery training yielded generalized effects to the PIAT Reading Recognition subtest.

The potency of this directed discovery procedure appeared very impressive in its demonstration of generalized training effects from arithmetical training tasks to generalization tasks which assessed spelling, reading, and even general information skills. Schleser et al. attributed the success of the directed discovery training to having taught these children not just what to think, but how to think (Meichenbaum & Asarnow, 1979). Such might not be the case, however, as children in the directed discovery condition differed from the other two conditions not only in the manner in which they were led to verbalize self-instructions, but also in terms of the fading of self-instructions. Only the directed discovery training required that children's discovered statements remain at an overt level throughout training, while the task-specific and general problem-solving training used the standard overt-to-covert fading procedure of Meichenbaum (1971). What was prerequisite to generalization then might not have been the process of discovering specific statements but simply arranging for such statements to be emitted overtly throughout the course of training. It was conceivable that children in the two training conditions that featured fading of self-instructions were no longer producing verbal mediators by the end of training. If so, mediated generalization could not be expected to occur.

A second question addresses the ease and the means by which children in the directed discovery condition were led to "discover" the specific statements which the investigators had a priori identified. Recall that Crisafi and Brown had no success in getting children to emit remotely appropriate rules for their training tasks, albeit their children were very young. Schleser et al. wrote that, "the experimenter rephrased the child's responses if necessary and had the child cumulatively rehearse all discovered statements," (p. 954). But how many children required their responses to be rephrased? For those who did, how different was this directed discovery procedure from the specific content self-instructional training?

Finally, it might be asked by what means the directed discovery method facilitated generalization across tasks which encompassed requisite responses as diverse as arithmetic and general information. Schleser et al. do not spell out their conceptualization of the means by which the directed discovery procedure yielded such a considerable degree of generalization.

In a human operant study, Catania, Matthews, and Shimoff (1982) reinforced college students' slow or fast key presses on one of two reinforcement schedules. Students were frequently required to write down their guesses as to what pressing behavior earned reinforcement. On some occasions, students' guesses were instructed (e.g., told to guess that slow presses earned reinforcement), on some occasions guesses were shaped, and on some occasions guesses were not differentially reinforced. The influence of this instructed, shaped, and non-reinforced verbal behavior upon the non-verbal key pressing behavior was examined, as was the reverse. Results indicated that consistent control of pressing rates by guesses occurred when guesses were shaped but not instructed. In general, shaped guesses controlled pressing rates regardless of the button-pressing contingencies, with rates conforming to guesses. Catania et al. also reported that control operated in the other direction as well, with pressing controlling guessing when guessing was nondifferentially reinforced or instructed. They concluded that verbal behavior was more likely to determine subsequent nonverbal behavior when it was shaped than when it was instructed, implying to these writers the clinical importance of changing private thought.

In the present context, these results suggest that shaped verbal behavior has greater potential for the maintenance of behavior it specifies than instructed or nondifferentially reinforced verbal behavior, in the face of competing contingencies that would otherwise dictate other behavior. These results further validate the need discussed earlier in the Schleser et al. study to objectify the degree to which verbalization of task rules is left for the child to discover, or is rephrased. Also of interest in the Catania et al. study was the difficulty the authors had in shaping verbalization of "press slow" and, "press fast." These guesses were successfully shaped in not quite half of the cases attempted, despite the simplicity of the targeted guesses and the sophistication of the subjects (college students).

In an earlier section, it was mentioned that subject variables have been implicated as mediators of the effectiveness of self-instructional training, one such variable being subjects' locus of control. Catania et al. suggested that their distinction between shaped and instructed verbal behavior related to the locus of control construct. They wrote that, in the terms of social psychology,

> Students whose guesses were shaped, unaware of the source of control, attributed them to their own behavior and thus responded in accordance with verbal behavior they believe they had generated themselves. [But] ...such an account begins by assuming what the present data show experimentally, that the students may say things to themselves that affect their subsequent nonverbal behavior (p. 246).

Guthrie (1967), from an applied educational perspective, compared various instruction sequences in teaching college seniors to decipher cryptograms. In the example-rule group, cryptogram examples were presented to criterion, following which subjects were taught to verbalize the rule upon request. The reverse sequence occurred in the Rule-Example (Expository) group. An Example (Discovery) condition consisted of presenting only examples of cryptograms to criterion. A control group which featured training on an irrelevant task was included as well. Following each training trial, all subjects received feedback which consisted of presenting the cryptogram with the correct word beside it.

Results indicated that the expository instruction facilitated retention relative to the remaining three conditions. The expository instruction brought subjects to criterion during training in roughly half the trials required of the other conditions, but actually impeded remote transfer, as the Discovery, Example-Rule, and even the control group had lower error scores on a remote transfer test. The discovery method did not facilitate retention but facilitated transfer, as subjects in this condition outperformed control and expository subjects on a near transfer test, and all other subjects on the remote transfer test.

Procedurally, Guthrie's discovery training was most analogous to multiple exemplar training, since subjects were presented with instances of cryptograms without accompanying rule training. In contrast to the other studies discussed in this section, Guthrie did not instruct, prompt, or shape subjects' emission of task rules. Thus, there was no empirical evidence, even indirect, that subjects "discovered" rules in this condition. Their behavior, as they learned the cryptograms, may have been consistent with task rules, but was not necessarily controlled by these rules at either a covert or overt level. This study then, interestingly suggested that, at least with respect to this set of tasks and subject sample, multiple exemplar training alone yielded a greater degree of generalization across tasks than such training which included training in rule verbalization--even when this rule verbalization training was shown to be facilitory on the directly trained task. When the structure of the task was changed, such that trained rules no longer specified appropriate problem-solving responses, this rule training component may have interfered with emission of newly-relevant problem-solving responses. On the other hand, when only multiple task examples were given, this may have evoked, as Guthrie speculated, searching, exploratory behavior that was reinforced and facilitated problem solution on the transfer tasks.

The studies discussed above were inspired by widely divergent conceptual paradigms, employed different methodologies with different population samples, and offered various rationales for teaching people to verbally identify what they were doing or needed to do. Yet their results for the most part converged to suggest that encouraging people to verbally identify task requirements was a promising means of promoting generalization, provided the responses specified were appropriate to the range of tasks. If we accept the implications of writers such as Vygotsky (1962) and Winokur (1976) that older children and adults typically and spontaneously engage a novel problem with a good amount of covert verbal problem-solving, then all the studies presented in this section are consistent with a position which calls for training children to verbalize task rules as they engage the tasks.

Specifying Contingencies

In an earlier section, it was suggested that the conceptualization of instructions and self-instructions as general or specific should be abandoned and reliance placed on a problem-solving model which comprised attending, problem-solving, and problem-solution phases, with emphasis upon establishing stimulus control within the problem-solving phase in particular (Grimm, Bijou, & Parsons, 1973).

Self-instructions could then be conceptualized in terms of whether they specified and targeted attending, problem-solving, or solution responses. Instructions and self-instructions, as verbal responses, may specify stimuli or reinforcers or contingencies as well, and still constitute functional problem-solving responses. Training the verbal identification of response-outcome contingencies has been subjected to limited experimental examination.

O'Leary (1980) wrote that the restatement of contingencies was a promising but relatively unevaluated procedure. To the present writer's knowledge, the only applied experiments that addressed this issue as either the focal point or sidelight of their studies were made by Kennedy and Miller (1976), Belmont, Butterfield, & Borkowski (1978), and Kendall, Borkowski, and Cavanaugh (1980). While children were not explicitly trained to verbalize the contingency between strategy use and outcome, the contingency was presented to children by the experimenter. Both the Kennedy and Miller study and the Belmont, Butterfield, and Borkowski study found that training which included feedback about the usefulness of the strategy led to maintenance of the strategy. In the latter study, such training yielded generalization across divergent tasks. Unfortunately, in neither of these studies was the role of feedback regarding the strategy's utility procedurally isolated from a variety of additional components.

The study by Kendall, Borkowski, and Cavanaugh did explicitly manipulate the strategy feedback element, within the interrogatory training condition which was the intervention of primary interest. In this case, strategy feedback versus its absence did not make a difference on children's paired associate recall. But, as the authors themselves acknowledged, all children had earlier received feedback about the value of the rehearsal strategy, perhaps negating the influence of the feedback element.

None of the studies that provided children with verbal feedback about the contingencies between responding and outcomes were successful in experimentally isolating this element of training. Training packages that included this element, however, yielded fair amounts of generalization across tasks. This training element deserves further experimental evaluation.

Matrices

The training and testing tasks used in the present study were, for the most part, borrowed directly from Bryant (1983) who acknowledged their similarity to the Raven Progressive Matrices and especially the Coloured Progressive Matrices (RCPM), the latter tailored for preschool children. It might be useful then, to briefly discuss the utility and validity of the RCPM. A pair of factor analytic studies (Carlson & Wiedl, 1976; Schmidtke & Schaller, 1980) of preschool and/or early elementary school children suggested that the RCPM is a measure of abstract reasoning, and the latter study described it further as testing both perceptual closure on the one hand, and the completion of homogenous patterns and recognition of given elements on the other. In Garrity and Donoghue's (1976) study of preschoolers, RCPM scores correlated with Peabody Picture Vocabulary Test

(PPVT) scores, at least with respect to the sample of preschool girls. Court (1983), in a review of the literature on RCPM and sex differences, described the RCPM as perhaps the best measure of "g", the general intelligence factor. Finally, Valencia (1984) citing his own work and that of others, called the RCPM a promising nonverbal intelligence measure when used with children of culturally and linguistically diverse backgrounds.

The present study called for training children to attend to and monitor up to two stimulus dimensions relevant to matrix completion. All children were tested on similar problems and some that even included a third relevant dimension. From a developmental perspective, it was important to examine the empirical evidence bearing on several procedural and task parameters, to help assure that tasks were not inordinately difficult and hence inappropriate. Relevant questions included: what evidence was there that some dimensions were easier for young children to attend to or label than others?; how many dimensions or attributes could young children successfully monitor and respond to? and; was it best for a problem-solving strategy within each matrix to target all dimensions of each stimulus before proceeding to the next stimulus, or to target one dimension of all stimuli before proceeding to the next dimension?

Much child developmental research has addressed the first question of the relative salience or preference exhibited towards various stimulus dimensions in preschool children. The importance of dimensional preference or salience paradigms lies in the predictive relationship between children's preferences and learning rate on problem-solving tasks in which the preferred dimension is relevant to task solution (Smiley, 1972; Odom & Guzman, 1972). Typically, young children's preferences have been assessed on only two dimensions, usually color and form. Early discussions of this literature (e.g., Suchman & Trabasso, 1966) were consistent in suggesting that after the age of six, children choose largely on the basis of form and before six, usually color. But more recent research has made this relationship between age and dimensional preference less clear. Offenbach (1983) found that most of their preschool (4-5 year old) children classified similarity on the basis of form and not color. Unlike earlier studies, Seitz and Weir (1971) found that a large number (just over 50%) of their subjects responded equivalently to both color and form dimensions. Roughly equal numbers of the remaining children exhibited color or form preferences. According to Smiley (1972) although very young (preschool) children make more color responses than older children do, form choices tend to predominate at all ages. The results of these studies were not sufficiently consistent to permit generalizations to be drawn at least with respect to young (preschool) children. These studies revealed a predominance of color, form, and mixed preference preschoolers. There was thus no a priori reason to suspect that the present study's presentation of matrices with both color and form (shape) relevant to solution would favor one dimension or the other for the children as a group.

With respect to the dimension of size, Suchman and Trabasso (1966) found that on subtests with the preferred stimulus removed, 3-6 year olds who preferred color, chose form over size and this preference increased with age. Form-preferring subjects, however, chose size as often as color at all age ranges. Given that nearly two-thirds of their children between 2-10 and 4-11 were form-preferring, size did indeed appear to be very nearly as salient a dimension as color in this age range in this study.

The fourth dimension manipulated in the present study, pattern within stimuli, has not to this writer's knowledge been experimentally investigated in these dimensional salience or preference paradigms. Thus there is no empirical base from which to estimate the relative salience of the pattern dimension. However, color and form have been simultaneously compared with other dimensions such as number (stimulus frequency) and position (Odom & Guzman, 1972), and results indicated that young children found the latter two dimensions to be substantially less salient than form and color. The degree to which the salience of color and form overshadowed the salience of number and position suggests that pattern might be relatively low salience, as well. Granted, in the absence of empirical data on the pattern dimension, this is little more than conjecture.

On the basis of the data discussed above, it was anticipated that preschoolers would find color and form dimensions the most salient and discriminable, closely followed by the size dimension, and lastly the pattern dimension, in the absence of any verbal instruction that might affect the probability of responding to dimensions.

A second question concerned the feasibility of training preschool children on two or more stimulus dimensions. Data that bear on this issue were reported by Odom and Guzman (1972), albeit the youngest children they worked with were kindergartners (mean age, five years, eight months). Following their salience test, children were administered an identity (problem-solving) task on which the correct choice matched all four dimensional values and an incorrect choice matched only three of four. The low incidence of errors to all items combined indicated that children were responding under the control of two or more dimensions, and that four dimensional identity problems were within these children's repertoires even without minimal right/wrong feedback.

Schuepfer and Gholson (1980, cited in Gholson, 1980) studied children's response hypotheses on no-feedback trials interspersed between feedback trials. The task required the child to identify the dimensional value (e.g., large) arbitrarily deemed correct by the experimenter. Results indicated that 65 of 71 preschoolers reached criterion on two-dimension problems and subsequently all of these children learned four-dimension problems in just one or two daily sessions. Schuepfer and Gholson in a second experiment suggested that in contrast to elementary school children, preschoolers as a group were far less systematic in eliminating dimensions, but they also acknowledged that their training did not focus on teaching strategy systems. They wrote that, "Whether preschoolers can be induced to exhibit sequences of hypotheses that correspond to the strategy categories remains to be investigated." (p. 76).

The tasks employed in the present study most likely fell in between the tasks of Odom and Guzman, and Schuepfer and Gholson in terms of difficulty. In comparison to the former's tasks the matrices demanded not only attention to relevant dimensions but stimulus patterns as well. In comparison to Schuepfer and Gholson, given that all stimuli on each matrix were presented simultaneously, the memory requirement in the present study's task was probably lesser.

The work of Schuepfer and Gholson (1980, cited in Gholson, 1980) addressed the last question: whether it was better strategically within each matrix to teach the monitoring of all dimensions of one stimulus before proceeding to the next stimulus, or to teach the monitoring of one dimension of all stimuli before proceeding to the next dimension. The superiority of the latter strategy was suggested by their data from elementary school children which showed that competent hypothesis testers tested one dimensional attribute before proceeding to another.

Before concluding this section, two studies which addressed children's problem-solving of RCPM or RCPM-like tasks will be briefly presented. While the studies were not conceptualized by their authors in either cognitive-behavioral or metacognitive terms, and did not address the generalization issue per se, their interventions were similar to many called cognitive-behavioral and had similar aspects to the intervention employed in the present study. Turner, Hall, and Grimmett (1973) subjected kindergartners to one of three interventions or a control (no feedback) condition as they worked on a derived RCPM-like task: Verbal (right/wrong) feedback; visual feedback (E placed correct choices in the empty space), and; explanation feedback (verbal feedback and an explanation, e.g., that's right, the lines match up; that's wrong, think about the shape). Testing on the RCPM revealed that all three interventions were superior to the control condition, and not different from each other. Providing strategies was no more effective than verbal (right/wrong) feedback. The authors concluded that the children failed to use the strategies because they were too young to benefit from such limited instruction, or the strategies themselves were deficient. These results offer another substantiation of the need to procedurally assure that verbal strategies control the specified nonverbal behavior.

Bethge, Carlson, & Wiedl (1982) administered the RCPM to third-graders via one of three methods. Children received either standard (no feedback) administration, elaborated feedback, or problem verbalization. Elaborated feedback entailed verbal right/wrong feedback, and informing the child why he was correct or not. The problem verbalization condition required the child to first describe the main stimulus pattern before making a response, and then to explain why the particular solution was chosen. Results on the RCPM indicated that the elaborated feedback and problem verbalization conditions were each superior to the control condition and not different from each other.

Process data indicated that the elaborated feedback and problem verbalization interventions versus control yielded more systematic and planful strategies. This was indicated by reduced numbers of omissions of comparing both rows of alternatives to the main pattern, and increased duration and frequency of eye fixations to main stimulus patterns and answer alternatives. The results thus showed the superiority of the two interventions in terms of process, as well as outcome data. Unfortunately, the validity of these results was called into question by a research design which failed to include any pretest measures which might have assured that group differences were a function of treatment and not extraneous factors. Further, as with many of the metastrategy training and "discovery" training studies, Bethge et al. were not very explicit about the degree to which children's verbalizations were prompted, shaped, consequated, or corrected in the problem verbalization condition. As such, it is difficult to generalize the results of their intervention beyond the immediate study. These methodological shortcomings aside, the results of the Bethge et al. and Turner et al. studies suggested that children's matrix performance was no more facilitated by interventions which entailed self-verbalization of stimulus patterns or verbal feedback specifying problem-solving responses, than it was by simple verbal (right/wrong) feedback or elaborated feedback.

Several caveats in addition to the methodological criticisms offered above suggest that it may be premature to abandon rule-based interventions as a means of promoting generalization on RCPM-like tasks. First, it should be recalled that neither of these studies explicitly addressed the generalization issue. It is conceivable that each study's interventions might have manifested differential effects had a series of logically related but diverse matrices been presented. Second, both studies worked with older children: kindergartners in the Turner et al. study and third graders in the Bethge et al. study. Crisafi and Brown (1983) showed that on an inferential reasoning task, only the younger (three but not four-year old) children transferred more efficiently when encouraged to describe task requirements than when presented simple multiple exemplars with feedback. For older children, it was sufficient merely to present them with the easy-to-hard exemplar sequence and verbal feedback. Perhaps, then, children in the Turner et al. and Bethge et al. studies had developmental learning histories sufficiently extensive to allow them to learn from feedback and elaborated feedback alone.

Finally, the work of Bryant (1983) addressed rule-based interventions for the learning of RCPM-like tasks. Although her study specifically addressed the relationship between learning and transfer abilities and general measures of intelligence, it suggested that four- and five-year old children profited considerably from adults' verbalization of rules or hints as they solved the matrices.

In sum, the tasks employed in the present study were derived from a task that has enjoyed theoretical and empirical support as a valid measure of intelligence. Developmental work in this age range suggested that: (1) at least three of the four stimulus dimensions were relatively high salience in the absence of instruction directed toward those dimensions; (2) children, with training, were likely to respond under the control of three stimulus dimensions, and; (3) a problem-solving strategy that targeted one dimension of all stimuli before proceeding to the next dimension was likely to be more effective than one which targeted all dimensions of each stimulus before proceeding to the next stimulus. Experiments featuring rule-based intervention with the RCPM did not examine generalization per se.

Summary

The frequent failure to find generalization effects across time, settings, responses, and tasks in many conventional behavior modification studies and mnemonics training studies led to reconceptualizations of generalization as an active process worthy of greater theoretical and empirical scrutiny. It provided impetus for the theoretical development of both the cognitive-behavioral self-instructional training paradigm and the metacognitive training paradigm, which despite different underlying theory and rationale, share an applied philosophy of training children to carry on verbal dialogue with themselves as a means of fostering generalization. Applied behavior analysis as well has implicated the importance of training verbal mediators as an aide to generalization, though its investigation within this paradigm has been relatively infrequent.

As a whole, SIT investigations of generalization across settings and tasks have yielded equivocal results. Metacognitive training studies have been somewhat more successful in effecting generalization across tasks, though process research in both paradigms is lacking, given the array of training components self-verbalization is typically embedded within. Both paradigms also subscribe to notions of general and specific self-statements which lead to circular explanations of generalization effects. The theoretical paradigm of metacognition may be faulted for attributing higher-order status to metastrategy training, at the expense of parsimony and perhaps in opposition to the data as well.

A behavioral approach places greater reliance on observables and assumes that different sets of verbal responses have stimulus properties that control variously sized classes of behavior. The maintenance of behavior and maintenance of generalized behavior is a function of the reinforcement schedule it was established under and the degree to which such behavior continues to be contingently reinforced by either arranged or natural consequences. It is also a function of the presence of controlling stimuli in the new conditions.

It has been established that the breadth of learning as assessed across tasks is a function of the range of tasks and stimuli used in training. Another theoretically important but scarcely tested means of promoting this type of generalization is to train children to verbally mediate generalization. If the interest is in generalization across tasks, the extent to which the trained verbal mediators specify shared problem-solving response components determines the success of generalization.

Rather than classify instructions and self-instructions as general or specific, it is more profitable, borrowing from an operant problem-solving model, to discuss them in terms of their specification of responses in attending, problem-solving, or problem-solution phases. Verbal training may focus on the specification of stimulus properties or response-reinforcer contingencies, the latter of which may be an important element in achieving generalization.

In addition, the process by which children's self-instructions come to be emitted may have implications for generalization effects. There is some evidence that prompting and shaping versus instructing task rules may lead to greater maintenance and generalization of behavior specified by those rules.

Hypotheses

The present study addressed this fundamental question: Is generalization facilitated by teaching children to verbally identify task requirements as they engage a series of tasks, relative to the training of multiple exemplars (differential feedback for solution responses across tasks)? The answers to two more focused questions were sought as well: (a) In teaching children to verbally identify task requirements as they engage a series of tasks, is generalization enhanced more by training (instructing, modeling, and prompting) predetermined statements or by encouraging children to induce their own rules?; (b) If teaching children to verbally identify task requirements is facilitative relative to multiple exemplars, can the differential effects be attributed to rule training (teaching a contingency-specifying rule via instructing, modeling, and prompting) per se, versus the training of problem-solving behavior specified by the rule?

The hypotheses which bear on these questions were generated largely on the basis of two sets of assumptions regarding conditions which contribute to training effects which generalize over tasks and time. It has been demonstrated that the programming of multiple stimulus and response exemplars is an effective means of promoting generalization across tasks. It is assumed however that training verbal mediators may further enhance generalization. Verbal mediators may be trained to act as problem-solving responses in a sequence to cue other problem-responses and facilitate acquisition on the directly trained task. To the extent that the mediators cue response components shared by various tasks, generalization is facilitated. Generalization is enhanced in this manner, given that the mediators continue to be emitted, overtly or covertly in the new conditions. Verbal responses are easily recalled, and make it easier to retain discriminations over time (Skinner, 1966). Verbal responses in the form of rules may thus serve as readily emitted responses with stimulus functions. In later stages of learning, the verbal mediators may no longer be facilitory and drop out of the sequence.

The second assumption is that the process by which rules come to be emitted is also important in determining the maintenance and generalization of behavior specified by the rules. When the contingencies change with a new task such that problem-solving responses are no longer effective, training which has encouraged children to examine the outcomes of their own behavior and induce effective strategies should yield the greatest degree of generalization to this new task.

The multiple exemplar condition (ME) exposed children to a series of increasingly difficult but related tasks, and examined maintenance on these tasks and generalization to other tasks. Beyond this, the rule training (RT) condition trained children to verbalize the contingency between problem-solving responses and task outcome, and then perform shared problem-solving responses. The problem-solving control (PSC) condition isolated the impact of the rule training per se by controlling for the modeling and performance of problem-solving responses specified by the rule. The rule discovery condition (RD) required the child, with some external guidance, to derive his or her own task rules.

Children were trained and tested on matrix completion tasks. Eleven sets of matrices (training variants 1, 1.5, 2, 3 and testing variants 1 through 7) differed on the basis of the number of dimensions that varied and the orderliness with which each dimension varied. All, however, shared required problem-solving responses. Training was conducted with training variants 1 through 3 on training days 1 through 3. Children were tested during posttests on the training variants (maintenance), testing variants 1 through 7 (generalization), and the Raven's Coloured Progressive Matrices (far generalization).

Given these questions, assumptions, and procedures, the following hypotheses were offered regarding acquisition, maintenance, generalization, and far generalization:

1. With respect to acquisition, it was hypothesized that children in the rule training (RT) and problem-solving control (PSC) conditions would require fewer trials to criterion than children in the remaining conditions on the training tasks. In these two conditions, the immediate training of problem-solving responses would be sufficient to facilitate task solution. Children in conditions ME and RD

did not receive explicit training in problem-solving. The rules RD children came to emit would likely depend on feedback over the course of trials. In contrast, the problem-solving responses of RT and PSC children were ensured from the very first trial. Thus, on trials to criterion during training the ordering of the groups was predicted to be: RT = PSC > RD = ME.

2. Regarding maintenance or generalization over time, on tasks earlier subjected directly to training, it was hypothesized that children in the rule training condition would outperform children in the rule discovery condition, who in turn would outperform children in the remaining two conditions. Group RT was predicted to select the correct choice more often than group RD, who in turn would select the correct choice more often than groups PSC and ME on training variants 1 through 3 during posttests. Training children to verbalize task rules in conditions RT and RD offered the possibility that these verbalizations, as readily emitted responses, would mediate generalization in untrained conditions. To the extent that RD children learned to identify useful rules their performance would approach that of RT children.

3. Regarding generalization to testing tasks (which lay beyond the range of training tasks in terms of the number of stimulus dimensions that systematically varied), the same hypothesis was offered as with maintenance: Group

RT would select the correct choice more often than group RD, who in turn would select the correct choice more often than groups PSC and ME on testing variants 1 through 7 during posttests. Training children to verbalize task rules in conditions RT and RD increased the possibility that on similar testing tasks, these verbalizations as easily emitted responses with stimulus functions, would mediate generalization. To the extent that RD children learned to identify useful rules their performance would approach that of RT children. Thus, during posttest, the ordering of the groups for measures of both maintenance (training variants 1 through 3) and generalization (testing variants 1 through 7) was predicted to be: RT > RD > PSC = ME.

4. With respect to generalization to a task (RCPM) which required some different (not just additional) problem-solving responses than the training tasks, it was hypothesized that children in the RD condition would outperform children in the ME condition, who in turn, would outperform children in the remaining two conditions. Training children in the RD condition to verbally identify task requirements would facilitate their performance on a new task, to the extent that children continued to verbally explore their behavior. Training children with rules and problem-solving responses that were no longer appropriate (instrumental) to the new task, as in RT and PSC conditions, was expected to interfere with performance to the extent

that problem-solving responses continued to be emitted at an overt or covert level. Thus, during the posttests the order of groups on the far measure of generalization (the RCPM) was predicted to be: RD > ME > RT = PSC.

CHAPTER II

METHOD

Subjects

The subjects were 62 children enrolled in three preschool classrooms at the Kamehameha Schools (Ulupono, Nanaikapono, and Nanakuli) and a fourth nearby preschool (Na Lei). All were private schools. Children ranged in age from 4 years, 6 months to 5 years, 4 months, with a mean age of 5 years, 0 months. At the request of the experimenter, teachers from each classroom identified preschoolers whom they thought lacked skills in labeling common colors, shapes, and sizes (small, medium, and big). These children were excluded from further participation.

The labeling skills were assessed with remaining children using an informal screening device. Those who met criterion (described below) at screening were assigned to one of the four training conditions and one of four trainers including this experimenter. Assignments were made randomly, with the following consideration: Peabody Picture Vocabulary Test (PPVT) scores were balanced across the training conditions.

The matrix completion tasks which comprised the pretest doubled as a second screening device, as those children performing at a rate of 50% correct or greater on particular sets of items were also dropped from the study. Table 1 shows the number and IQ scores of preschoolers withdrawn from the study prior to and following assignment to conditions, and the reasons for their withdrawal. Table 2 shows, for each training condition, the number of children withdrawn and a more detailed description of the reasons for withdrawal.

Of the five children withdrawn after assignment to conditions, only one child's withdrawal was clearly specific to his training demands, and hence it is concluded that the assumption of randomness in each training condition was not violated. The remaining 38 children who successfully participated were distributed among training groups, preschools, and trainers as displayed in table 3.

While an attempt was made to evenly distribute children among trainers and conditions, withdrawals and practical considerations rendered this distribution somewhat uneven. The Multiple Exemplar (ME) condition, Rule Discovery (RD) condition, Problem-Solving Control (PSC) condition, and Rule Training (RT) condition had 9,9,10, and 10 children in each group, respectively. The respective mean ages for the groups were: 4 years, 10 months; 5 years, 1 month; 4 years, 11 months, and; 5 years, 1 month.

The mean PPVT IQ scores for the groups were: 97.6; 98.3; 104.1, and; 103.3, respectively. A one-way analysis of variance on these scores revealed that they did not differ significantly among the four training conditions, $F(3, 34) = 0.995$, $p > .25$, table 4).

Materials

The children were trained and tested using matrices modeled after the Raven's Coloured Progressive Matrices and originally used by Bryant (1983). The tasks were well-suited to the purposes of the present study, as various stimulus properties could be manipulated along objective dimensions. This permitted the selection of task instances, and in particular variants, to be made on logical versus strictly intuitive grounds. It also guided the composition of verbal rules which were likely to facilitate learning.

The task domain employed to meet these criteria consisted of sets of 3 x 3 matrices of geometric stimuli printed on poster board. For each example of these stimulus cards, the lower right hand stimulus was missing and was to be chosen from a card of 4 to 6 stimuli situated in front of the child (see figure 3). All forms on this solution card were manipulable so that the child could select by hand the correct form and place it in the missing lower right-hand position of the stimulus card. The location of the correct stimulus was randomly determined for all stimulus cards.

Within each matrix, values of the dimensions of size, shape, color, or pattern (superimposed on each stimulus), or some combination varied. Additionally, one or more dimensions were scrambled across the rows or columns as opposed to being laid out logically. Each set of matrices that followed the same rules regarding the particular dimensions that varied and the degree to which they were scattered was referred to as a variant. Eleven variants were used in the present study and are described below (see figure 3 for examples of each variant).

Stimulus cards and solution cards were made of posterboard, and measured 10.63 inches by 8.38 inches. On the stimulus cards, a lower right hand area measuring 3.5 inches by 2.75 inches was cut away. Forms of various sizes and colors were professionally printed on the posterboard. The forms were circles, squares, isosceles right triangles, and five-point stars. At their greatest breadth, small, medium, and large instances of each were in an approximate ratio of 1:2:3, after Bryant, (1983).

The breadth of small, medium, and large forms were 0.63 inches, 1.31 inches, and 2.0 inches, respectively. The pattern attribute was applied to the stimuli of stimulus cards and choice cards via Letraset tape. Its width was 0.13 inches.

Variants used in training. Four variants were employed during training. In all instances of variant 1 (V-1), shape was the only dimension that varied, with each value (e.g. square) uniform within each row. In all instances of V-1.5, color was the only dimension that varied, with each attribute (e.g. blue) uniform within each column.

All instances of V-2 varied on both shape and color dimensions. Each value of shape was uniform within each row, and each color attribute was uniform within each column. Each V-2 matrix was thus a combination of V-1 and V-1.5. Instances of V-3 varied on the color dimension only. It differed from V-1.5 in that attributes of colors were uniform not within rows or columns, but on the diagonal axis giving a scattered appearance.

Variants used in testing. Examples of the four variants described above also appeared in pretest and posttests. Seven others appeared uniquely in the pretest and posttests. V-1 was identical to V-1 of training, except that values of size, not shape, were uniform within each row. V-2 of testing was identical to V-2 of training except that values of size, not shape, were uniform within each row, and attributes of pattern, not color, were uniform within each column. V-3 of testing was identical to V-3 of training, except that values of size, not color, were scattered.

Instances of V-4 and V-5 varied on both color and shape dimensions. Unlike V-2 of training however, color attributes were scattered for V-4, and both color attributes and shape values were scattered for V-5.

Instances of V-6 and V-7 each varied on 3 dimensions. Instances of V-6 varied on shape, color, and size dimensions. Values of shape were uniform within each row, values of size were uniform within each column, and attributes of color were scattered. Instances of V-7 varied on shape, color, and pattern dimensions. Values of shape were scattered, attributes of colors were scattered, and attributes of pattern were uniform within each column.

For all matrices, if the size of the stimuli varied, the correct choice brought the frequency of the different sizes to parity. If shape varied, the correct choice brought the frequency of the different shapes to parity. If both size and shape varied, the correct choice brought the frequency of both the different sizes and shapes to parity, and so forth. Thus for all matrices, the correct choice brought the stimulus frequency of a relevant dimension to parity, with those dimensions that systematically vary being defined as relevant.

An attempt was made to use progressively more difficult variants in each successive training day of the study. Proceeding across days, the critical stimulus features became more numerous or complex, building upon earlier variants. Thus, while the variant of the third day of training shared commonalities with those of the first day, they had different features as well.

On the basis of pilot and experimental data from Bryant (1983), variants were presented in the following manner to meet the requirements mentioned above: Training day one featured V-1 and V-1.5; day 2 featured V-2, and day 3 featured V-3.

The testing variants were presented in blocks of 4 examples, in a predetermined random order during pretest and each posttest. Two instances of each of the four training variants were presented during pretest and posttests, and served as an index of maintenance. There were four instances each of V-2, V-4, V-5, V-6, and V-7, and two each of V-1 and V-3. These served as indices of generalization, as children received no exposure to them during training. There were thus 8 maintenance items and 24 generalization items.

The Raven Coloured Progressive Matrices (RCPM), set B, was also administered as part of the pretest and posttests. Similar to the training and testing variants, the RCPM required the child to select the one stimulus from a solution set of 6 stimuli that best completed the matrix. Unlike the variants, many of the RCPM matrices have been described as, "component parts that need closure," (Turner, Hall, and Grimmett, 1973, p. 358). As such, it was presumed that many of the RCPM items required different problem-solving responses from those of the training and testing variants. Set B consisted of 12 items, bringing the total number of items for the pretest and each posttest to forty-four.

During pretest, training, and posttests, poker chips served as token reinforcers. At the end of the daily session, children could exchange tokens earned for backup reinforcers such as balloons, marbles, and stickers. Whatever the amount of tokens earned each day, they could be exchanged for one, and only one backup reinforcer. This controlled for the amount of reinforcement across all training and testing days, and between training conditions.

Design

Following the informal screening of shape, color, size, and pattern identification, children were assigned to one of four experimental conditions: Multiple Exemplar (ME); Rule Discovery (RD); Problem-Solving Control (PSC), and; Rule Training (RT). Children in all conditions were exposed to the pretest, at least 3 days of training, and two posttests, the first unprompted and the second prompted.

During training, all children received feedback. They were first told whether their choice was correct or incorrect. For all children, trainers then described the relevant characteristics of the correctly completed matrix. Children received no such feedback during the pretest and the posttests. Immediately prior to the presentation of pretest, training, and posttest materials, all children received minimal verbal instructions to perform the tasks, and verbal prompts, as necessary, to continue working.

Children in the ME condition were exposed to the progressively more complex sequence of materials of training days one through three. As no attempt was made to instruct, model, or prompt rules, or train problem-solving responses, this condition served as a control for the exposure to multiple exemplars in comparison to the remaining three conditions.

In the RT condition, children were trained to verbalize a rule which specified a relationship between problem-solving responses and outcome. They were also trained to do those problem-solving responses and related nonverbal problem-solving responses before choosing their answers.

In the PSC condition, children were trained to emit the verbal and nonverbal problem-solving responses, but were not instructed or required to verbalize the rule. Nor was the rule ever modeled.

Immediately prior to solving each matrix, trainers prompted children in the RD condition to say what they would do to find the right answer.

Children worked on the matrices for approximately 20-40 minutes each school day. For each child, the pretest occurred on two consecutive school days. The day after pretest, training began. It lasted three consecutive school days, for all but three children. Two children in the RD condition required an extra training day for V-2. One child in the ME condition required one additional day for V-3. Children spent the next two school days working on the first posttest. They spent the following two days working on the second posttest. Barring absences, all children spent 9 or 10 consecutive school days from the first to last day of the study (see table 5 for a design overview).

Procedure

Screening. Prior to the pretest, all children were informally screened with stimulus cards to determine and assure their competence in discriminating verbally all attributes of color and pattern, and all values of size and shape that were to appear on the matrices. Trainers corrected children's errors. To participate in the study, children had to identify correctly all values of a dimension twice consecutively, without prompting. For example, they had to identify the sequence, red-blue-green-yellow-red-blue-green-yellow, without error or prompt.

Pretest. On the pretest days, all children received exposure to a set of 32 matrices, composed of: two instances each of training variants 1, 1.5, 2, and 3; two instances each of testing variants 1 and 3, and; four instances each of testing variants 2, 4, 5, 6, and 7. Variants were presented in a predetermined random order of blocks of four to minimize acquisition and generalization via the logical programming of matrices during this phase. The Raven

Coloured Progressive Matrices (RCPM), set B, was also administered as a block within the matrices. It consists of twelve items, bringing the total number of pretest items to forty-four (see table 6).

As the trainer presented each item, he or she simply instructed the child to, "Find the picture that belongs here," (pointing to the lower right hand position of the stimulus card) "from the ones here," (sweeping his or her hand across the solution card). No feedback was given regarding the accuracy of each child's answer.

Trainers put a token in a plastic cup in front of the child on alternate matrices. After the child was given a few seconds to study the matrix but before he or she selected an answer, the token was delivered. At the same time the trainer said, "Here's a token for working so hard."

The purpose of the pretesting was to assure that children could not reliably perform the matrices and to provide a baseline against which to compare the differential effectiveness of the various training conditions.

Training

Multiple Exemplar condition. In this condition, children were exposed to the sequence of progressively more complex and difficult tasks in training without the benefit of instruction or prompting of rules or labels: Nor were any problem-solving responses modeled or trained.

The child was given minimal instructions to, "Find the one that belongs here," (trainer pointed to the lower right hand position) "from the ones here," (trainer swept his or her hand across the solution set). This instruction and other verbal prompts were repeated by the trainers as often as necessary to sustain the child's on-task behavior. They were faded out as the child responded readily.

After the child had a few seconds to study the matrix but before he or she could select an answer, the trainer placed a token in the child's cup. At the same time the trainer said, "Here's a token for working so hard." Material reinforcement was given every trial of training, contingent upon on-task behavior only.

Immediately after the child responded, he or she was given feedback regarding the accuracy of the choice. Selection of the correct choice was praised verbally: i.e., "Good job! That's the right one." In the case of a wrong choice the trainer said, "Good try, but that's not the right answer. This is the one that belongs," (the trainer removed the child's incorrect choice and placed the correct choice in position).

Correct or incorrect, at this point the trainer described the relevant characteristics of the correctly completed matrix: e.g., "This is the right one because you have three squares, three circles, and three triangles." The trainer permitted the child to observe the correct choice in place for a few seconds, and then presented the next matrix.

Rule Training condition. Children in this condition were exposed to the same task sequence, minimal instructions, praise, and feedback as were children in the ME condition. Additionally, a rule which specified the relationship between verbal problem-solving responses and task outcome was taught. The verbal and nonverbal problem-solving responses were also taught.

Immediately after the matrix was presented, the trainer modeled the appropriate rule. For a V-1 matrix, where shape varied, the trainer said, "If I count the shapes, it may help me get the right answer."

The trainer then modeled the verbal and nonverbal problem-solving. This entailed counting the frequency of each stimulus value for each relevant dimension, labeling the value, and calling out each needed (missing) value. With a V-2 matrix for example, the trainer would say, "One, two, three circles; one, two, three, squares; one, two triangles... I need a triangle. One, two, three, smalls; one, two, three mediums; one, two larges... I need a large." At the same time, trainers pointed with their fingers, tracking each stimulus in a corresponding manner. This was the nonverbal component of the problem-solving. The trainer then prompted the child to verbalize the rule, modeling as needed, and praising longer chains of behavior across training trials. The trainer similarly prompted the child to emit appropriate problem-solving. The problem-solving was further modeled as needed, and longer chains of behavior were praised across training trials.

After the child verbalized the rule and emitted all the required problem-solving, but before he or she selected an answer, the trainer placed a token in the child's cup. At the same time the trainer said, "Here's a token for saying and doing all those things." Thus, material reinforcement was contingent upon emission of the entire rule and problem-solving sequence. If necessary, the child was then encouraged to select an answer. Trainers modeled the rule and problem-solving responses on the first trial with each new variant. On successive trials, trainers faded out the modeling to wean children from reliance upon the trainer. If, on a given trial, the child was unsuccessful in emitting the rule or problem-solving sequence, he was prompted with the modeling of its initial portion. If still not sufficient, more was modeled until the child complied. Note that while modeled rules and problem-solving were faded, the child was required to verbalize the rule overtly and perform corresponding problem-solving responses throughout training.

Problem-Solving Control condition. Children in this condition were exposed to the same task sequence, minimal instructions, praise, and feedback as were children in the

ME and RT conditions. As in the RT condition, children were trained to emit the appropriate problem-solving responses. However, they were not trained to verbalize the formal rule which specified the relationship between problem-solving responses and outcome. The trainer never modeled the rule, nor did the trainer instruct the child to verbalize the rule.

Relative to the RT condition, the present condition thus controlled for the effects of the modeling of the problem-solving and the effects of the problem-solving behavior itself.

After each matrix was presented, the trainer modeled counting the frequency of each stimulus value for each relevant dimension, labeling the value, and calling out each needed value. With a V-2 matrix for example, the trainer said, "One, two, three circles; one, two, three squares; one, two, triangles... I need a triangle. One, two, three, smalls; one, two, three, mediums; one, two, larges... I need a large." At the same time, the trainer pointed with his or her finger to each stimulus in a corresponding manner.

After the child emitted all the required problem-solving, but before he or she selected an answer, the trainer placed a token in the child's cup. At the same time the trainer said, "Here's a token for saying and doing all those things." Thus, material reinforcement was contingent upon emission of the entire problem-solving sequence. If necessary, the child was then encouraged to select an answer.

As in the RT condition, trainers modeled the problem-solving responses on the first trial with each new variant, and subsequently faded the modeling of problem-solving responses while praising longer problem-solving chains. As in the RT condition, the problem-solving sequence was prompted and modeled as necessary to assure that problem-solving continued throughout training.

Rule Discovery condition. Children in this condition were exposed to the same task sequence, minimal instructions, praise, and feedback as were children in the previous conditions. However, they were additionally prompted to identify rules regarding matrix solution. On every trial, immediately before the matrix was presented, the trainer prompted the child to name rules, specify problem-solving responses, or problem-solve by asking the child to, "Tell me out loud how to do this," or, "Tell me what you have to do to get the right answer."

Children were enthusiastically praised for stating appropriate rules, specifying appropriate problem-solving responses, or for doing appropriate problem-solving. If for example, the child responded, "three circles, three squares, three triangles," on an appropriate V-1 matrix, the trainer said, "Terrific! You did a really good job!" Before the child selected his or her answer, the trainer would put a token in the child's cup saying, "Here's a token for doing such a good job telling me how to do it."

If the child's verbal response did not meet these criteria, the trainer said, "Good try." Before the child selected his or her answer, the trainer would put a token in the child's cup saying, "Here's a token for trying to tell me how to do it."

If the child failed to verbalize an appropriate rule, specify an appropriate problem-solving response, or do appropriate problem-solving, in response to the trainer's first prompt, the trainer prompted the child a second time on that trial. No matter what the child's verbalization in response to this second prompt, no third prompts were permitted on a given trial. Thus, children had either one or two opportunities to verbalize rules on each trial, depending on the appropriateness of their first attempt. Material reinforcement was contingent upon verbalization regardless of content, and only one token could be earned on each trial.

As in the RT and PSC conditions, children were prompted if necessary to assure that the rules they came to emit remained overt throughout training. See appendix A for the protocols for all four training conditions.

Meeting criterion. To advance from one training variant to another, or from variant 3 to posttest, all children had to meet a criterion of three consecutive correct problems within each variant.

Children in the RT condition had to meet an additional criterion: On two consecutive trials within each variant, they had to verbalize the appropriate rule and do appropriate problem-solving without modeling and prompting from the trainer.

Similarly, children in the PSC condition had to meet the same criterion for problem-solving. Children in the RD condition had to verbalize rules on two consecutive trials without prompts from the trainer, regardless of the appropriateness of the rules.

If on a given variant, the child met criterion on outcome before meeting the rule criterion, subsequent matrices were presented without the solution card. This enabled children to continue working toward the two-trial criterion for rules, without overlearning correct solutions to the matrices.

Unprompted posttest. The unprompted posttest and its administration was identical to the pretest. In the two days of the unprompted posttest, children were exposed to the same set of 44 matrices as in the pretest. As in pretest, blocks of variants were presented in the same predetermined random order to avoid presentation of a logical sequence of matrices that might foster acquisition or generalization. Trainers' instructions and reinforcement contingencies remained the same as in pretest, and again no feedback regarding accuracy was given.

The purpose of this posttest was to provide an assessment of maintenance via training variants 1 through 3, and an assessment of generalization across tasks via testing variants 1 through 7. The inclusion of RCPM items provided an index of far generalization.

Prompted posttest. The prompted posttest was always administered on the two school days immediately following the unprompted posttest. The prompted posttest and its administration was identical to the pretest and unprompted posttest with one exception: On each trial, immediately following the minimal instructions to do each problem, the trainers prompted children to use their formerly trained strategies.

For children in groups RT, PSC, and RD the prompt was, "Remember to say and do the things you learned before you answer." For children in group ME, who had not learned to say anything during training, the prompt was, "Remember to do the things you learned before you answer."

Consistent with Brown and Campione's (1978) notion of dynamic measures of transfer, the purpose of the prompted posttest was to provide a more sensitive measure of generalization effects in the event that differential training effects failed to materialize on the unprompted posttest.

Acquisition. Trials to criterion on variants 1 through 3 of training served as measures of acquisition on materials immediately subjected to training.

Maintenance. Eight examples of the four training variants were included in each posttest. The number correct thus served as an index of maintenance.

Generalization across tasks. Thirty-two examples of the seven testing variants were included in each posttest. Not present during training, they served as an index of generalization.

Far generalization. The number of items correct on set B of the Raven Coloured Progressive Matrices (RCPM) provided an index of far generalization on each posttest. Trained rules and problem-solving strategies were not anticipated to be instrumental to RCPM solution.

Process and Verbalization Measures

In conditions RT and RD it was thought that the verbal behavior children came to emit during training would mediate generalization in altered settings by increasing the probability of behavior it specified. In contrast, group PSC was expected to exercise little influence upon children's verbalization of rules. Instead, generalization effects which occurred were likely to be the result of continued application of trained problem-solving responses.

Given that different interventions were expected to have their impact on different types of responses (rules versus problem-solving), assessing the types of responses that occurred during training and testing became an important means of validating treatment elements.

For all training conditions, children's pretest, training, and posttest verbalizations were recorded via tape recorder and trainer transcription to: assure that the various training regimens successfully determined children's rules and problem-solving during training, and; determine whether children's trained or prompted rules and problem-solving maintained during nontraining conditions, at least on an overt level.

The recording of children's rules in the RT and RD conditions during training permitted a comparison of rule quality or face validity. In the event of group differences on measures of matrix solution, this comparison might have permitted those differences to be discussed primarily in terms of the process by which rules came to be emitted, and not the content.

Coding. After the study's completion, children's verbal and nonverbal behavior was coded on every trial of pretest, training, and posttest.

The coding scheme for children's verbal and nonverbal behavior reflected the problem-solving model which inspired the Rule Training and Problem-Solving Control training conditions. In the RT training, children were taught a rule which specified the relationship between a problem-solving response and an outcome: e.g., "If I count the colors and shapes, it may help me get it right." Such responses were coded as rules (RULE-).

Verbal responses were coded (SpPROB-) if they only specified the problem-solving response (e.g.,"If I count the colors and shapes."). They were coded (SOL) if they only specified the solution response or outcome (e.g., "It will help me get it right.").

Both RT and PSC conditions taught children to engage in verbal problem-solving: e.g., "1,2,3 circles; 1,2,3 squares; 1,2 triangles, I need a triangle. 1,2,3 blue; 1,2,3 reds; 1,2 greens, I need a green." Such responses were scored (PROB-). The simple labeling of a relevant dimension or attribute (e.g., "circle,") was scored (LAB).

Verbalizations with little face validity were scored as irrelevant (IRR).

Both the RT and PSC conditions taught children to do nonverbal problem-solving as well. These children were taught to touch, in sequence, identical attributes of each relevant dimension. They would, for example, track with their fingers each of the three circles, then the three stars, then the three triangles followed by the three red forms, three blue, and three green. This was scored as tracking (T-). Complete failure to track was scored as no pointing (NP).

In addition to these categories, verbal and nonverbal responses were coded according to the accuracy with which they conformed to relevant dimensions. Responses that conformed to all relevant dimensions, a subset of relevant dimensions, or no relevant dimensions were coded as appropriate (ap), partially appropriate (ptap), or inappropriate (inap), respectively.

To illustrate, consider a matrix with color and shape relevant. If the child said, "I have to count the colors and shapes," it was scored as (spPROB-ap), as all relevant dimensions were specified. "I have to count the colors to get the right answer," would be coded (RULE-ptap), because the shape dimension was not specified.

Should the child problem-solve by counting, " 1,2,3 big; 1,2,3 medium; 1,2 small," it would be scored (PROB-inap), as neither shape nor color was referenced. Similarly, if the child tracked color but not shape with his or her fingers, it would be scored (T-ptap).

Definitions and more examples of each coded category are presented in Appendix B.

Trainers. The four trainers were: the experimenter; an M.A. level psychologist (trainer C); an M.S.W. (trainer A), and ; a graduate holding a B.S. degree (trainer N). Trainers receive approximately twelve hours of practice sessions in the administration of the experimental conditions. Trainers initially practiced with each other, and then with several preschoolers. These sessions featured instruction, modeling, rehearsal, and feedback. All trainers demonstrated mastery of the training procedures prior to commencement of the study. During the course of training, trainers' performance was monitored occasionally by one another to assure that training was conducted properly. With the exception of the experimenter, all trainers were naive as to experimental questions and hypotheses.

Reliabilities

Reliability between pairs of trainers was obtained prior to commencement of the study with the help of pilot preschoolers from the Ulupono classroom. After the study's completion, audiotapes and videotapes of all training and testing sessions were reliably transcribed and then reliably coded. Also, transcribed nonverbal behavior was reliably coded. All reliabilities were calculated via the following formula: Reliability = no. of agreements/ (no. of agreements + no. of disagreements) x 100.

Pre-study reliabilities. The prestudy reliabilities included trainers' monitoring on each trial: (1) the number corresponding to the chosen answer; (2) the answer's correctness; (3) the child's verbalization of the required rule, as trained in the RT condition; (4) the child's counting and naming aspects of all relevant dimensions, as trained in the PSC and RT conditions; (5) the child's tracking values of each relevant dimension, as trained in the PSC and RT conditions, and; (6) the child's verbalization of self-determined rules as encouraged in the RD condition. Reliability was also obtained on the recording of: (7) children's tracking that did not conform to trained sequences.

Trainers were randomly paired to work with each pilot child. A particular training variant or pretest and a particular training condition were selected for one trainer to implement and monitor while the other trainer sat nearby and simultaneously monitored children's verbal and nonverbal behavior.

Trainers were seated so they could not observe each other's monitoring sheets. The pre-study reliabilities corresponding to each of the seven categories described above were: (1) 91.3% (42/46); (2) 95.7% (44/46); (3) 100% (52/52); (4) 90.3% (65/72); (5) 98.6% (71/72); (6) undetermined--no self-rules observed by either trainer, and; (7) 66.7% (2/3).

Once the study had begun, reliability checks for these categories were again determined by pairs of trainers. The particular pair of trainers, child, training condition and training or testing phase for each check was not determined randomly but by the pragmatics of the situation: When a child was absent or otherwise unavailable, the free trainer would assess reliability with the trainer of the next child to begin a session.

Again corresponding to the seven categories above, the checks yielded reliabilities of: (1) 95.7% (154/161); (2) 97.5% (157/161); (3) 100% (51/51); (4) 98.0% (50/51); (5) 94.1% (48/51); (6) 87.5% (42/48), and; (7) 89.3% (25/28).

It might be noted that recordings of items 3 through 6 above, rule verbalization, verbal problem-solving, tracking, and discovered rules, served less as dependent measures per se than they did as self-check measures or markers which signaled to the trainer that training on a particular training variant was complete. For example when the trainer checklisted items (4) and (5) on two consecutive training trials in the PSC condition, training on that particular variant was complete (provided the child had, as well, produced three consecutive correct solutions on that variant).

<u>Post-study reliabilities</u>. After training and testing were completed for all children, reliability between the experimenter and trainers C. and N. was determined for the <u>transcription</u> of verbal behavior from videotape (prompted posttest, first day only) and audiotape (all remaining testing and training sessions). Reliability was trained to a criterion of 85.0%. When this criterion was met, the experimenter and trainers C. and N. independently

transcribed audio and videotapes with occasional reliability checks.

After all transcription was complete, the reliability between the experimenter and trainer C. was determined for the coding of the transcribed verbalizations and the coding of the transcribed nonverbal behavior. Reliability was trained to a criterion of 85.0%, after which the experimenter and trainer coded transcriptions independently with occasional reliability checks.

Transcription. The reliability of transcription was assessed on 9.8% of the total number of trials available for all children. Over the final 4.7%, reliability was achieved. Reliability for overall transcription between the experimenter and trainers C. and N. was 93.0% (436/469). Adjusting for inflation by eliminating trials on which both trainers agreed no speech occurred yielded agreement of 92.3% (398/431). Agreement that the same number of speech units occurred on each trial was 92.0% (276/300).

Independent reliability checks revealed sufficiently high levels: overall transcription, 89.1% (172/193); adjusted for no-speech trials, 84.5% (93/110), and; number of speech units per trial, 86.4% (102/118).

Coding of verbal behavior. The reliability for coding across all categories was 89.9% (286/318). Of the thirteen total speech categories, six categories each accounted for 1.8% or fewer of all coded instances (table 7).

Reliabilities for these low occurrence categories were generally low: (LAB), 50.0% (2/4); (SOL), 25.0% (1/4); (spPROB-inap), 20.0% (1/5); and; (PROB-ap), 100% (6/6). Instances of these categories were arbitrated by the experimenter. The experimenter and trainer C. agreed that no instances of categories (INT) or (SpPROB-ap) occurred.

Reliabilities for the remaining categories, which each accounted for 4.7% or more of all coded instances ranged from 80.0% to 100.0%: (PROB-ptap), 100% (22/22); (PROB-inap), 100% (74/74); (SpPROB-ptap), 80.0% (28/35); (RULE-ap), 100% (15/15); (RULE-ptap), 94.4% (17/18); (RULE-inap), 95.7% (45/47), and; (IRR), 85.2% (75/88).

Independent reliability checks revealed sufficiently high levels summed across all categories: 87.2% (171/196). Three categories showed very low reliability, 0% (0/4), but these were very low occurrence categories as all three combined accounted for only 2.0% (4/196) of all coded instances (see table 7). The coding of these instances was arbitrated by the experimenter.

<u>Coding of nonverbal behavior</u>. Reliability was assessed on 11.6% of the total number of trials available for all children. Over the final 9.0%, reliability was achieved. Reliability for overall coding of nonverbal behavior between the experimenter and trainer C. was 94.4% (510/540). Adjusting for inflation by eliminating trials on which

experimenter and trainer agreed no pointing occurred yielded agreement of 91.0% (294/333).

The reliabilities for each category were: (T-ap), 84.5% (49/58); (T-ptap), 94.8% (128/135); (T-inap), 90.0% (117/130), and; (NP), 99.5% (216/217, see table 7). Independent reliability checks revealed sufficiently high levels across all categories: 95.0% (115/121). Over 90% of the trials during the checks were coded NP (no pointing).

Summary

Thirty-eight preschoolers were assigned to one of four conditions and were trained with various sets of matrix completion tasks in a pre-post design.

The Multiple Exemplar (ME) group received minimal instructions and feedback across training tasks, as did all groups. The Rule Training (RT) condition additionally required children to verbalize a rule and then perform specified problem-solving responses. The Problem-Solving Control (PSC) group isolated the impact of the rule training per se, by requiring problem-solving responses alone. Rule Discovery (RD) training encouraged children to verbalize task requirements.

Following training, children were tested on the training matrices (maintenance), logical extrapolations (generalization), and matrices that required alternate strategies (far generalization). A second posttest was administered on which children were prompted to use their formerly trained strategies.

CHAPTER III

RESULTS

Overview

First, the differential effects of training on matrix solution, or outcome, was examined. There were five outcome measures: trials to criterion during training; number correct on maintenance items; number correct on generalization items; partial generalization scores, and; number correct on far generalization items.

The Rule Discovery (RD) and Multiple Exemplar (ME) training, unlike the Rule Training (RT) and Problem-Solving Control (PSC) conditions, did not ensure that children would issue predetermined rules and problem-solving. For that reason, a brief description of the verbalizations of RD and ME children during training was next presented.

Finally, rules, verbal problem-solving, and nonverbal problem-solving were examined during posttests to validate training effects and to explore their relationship with outcome.

Effects of Training on Trials to Criterion

The number of trials to criterion during training was summed across the four training variants and subjected to a one-way analysis of variance. This analysis yielded a significant groups effect, $F(3, 34) = 3.076$, $p<.05$. Duncan multiple range tests revealed that children receiving the Rule Training (RT) required significantly fewer trials to reach criterion than did children in the Rule Discovery (RD) condition ($p<.05$) and Multiple Exemplar (ME) condition ($p<.05$).

Children in the Problem-Solving Control (PSC) condition required fewer trials to criterion than children in the RD and ME conditions, though these differences were marginally significant ($p<.10$, table 8). Including these marginal differences then, the RT and PSC conditions were comparable and each superior to the RD and ME conditions (figure 4). These results are identical to those predicted by the first hypothesis. The first hypothesis predicted that the immediate training of problem-solving skills during this phase would be most facilitative.

Method of Analyses

To assess the differential effects of training, all maintenance and generalization measures were subjected to analyses of covariance on posttest scores with the respective pretest scores serving as the covariate. First, the maintenance items of the unprompted posttest were subjected to the ancova with the pretest score serving as covariate. Then the maintenance items of the prompted posttest were subjected to the ancova, again with the pretest score serving as covariate.

The same procedure was repeated for the generalization items, the partial generalization scores, and the far generalization items. These analyses of adjusted posttest scores served to eliminate the variance associated with chance differences among the treatment groups with respect to pretraining dependent measures.

These analyses of maintenance, generalization, and far generalization measures were supplemented with two-way (groups X trials) repeated measures analyses of variance to determine absolute improvement over trials, for each training condition. Significant effects were examined via Duncan Multiple Range tests, applied to adjusted cell means in the case of ANCOVA. Significant interaction effects were followed by F-tests for simple effects and Duncan Multiple Range tests in the case of repeated measures ANOVA.

Maintenance Items

For maintenance items at the unprompted posttest, an analysis of covariance on the adjusted scores revealed a significant groups effect, $\underline{F}(3, 33) = 2.886$, $\underline{p} < .05$. Duncan Multiple Range tests revealed that children in the PSC and RT conditions outperformed children in the ME condition ($\underline{p} < .05$). No other comparisons were significant (see figure 5 and table 9).

For maintenance items at the prompted posttest, the analysis of covariance on the adjusted scores revealed a significant groups effect, $\underline{F}(3, 33) = 3.360$, $\underline{p} < .03$. Duncan

Multiple Range tests revealed, as in the unprompted posttest, that children in the PSC and RT conditions outperformed children in the ME condition ($\underline{p}<.05$), with no other comparisons significant (see figure 6 and table 10).

These results were not anticipated on the basis of the second hypothesis, which predicted that RT children would outperform RD children, who would in turn outperform PSC and ME children. The rule component of the RT and RD conditions was presumed essential to maintenance. This hypothesis was not confirmed by the data.

A two-way repeated measures analysis of variance on the maintenance items revealed a significant main effect for groups, $\underline{F}(3, 34) = 3.491$, $\underline{p}<.026$, a significant main effect for trials, $\underline{F}(2, 68) = 99.895$, $\underline{p}<.001$, and a nonsignificant groups x trials interaction, $\underline{F}(6, 68) = 1.795$, $\underline{p}>.113$, (table 11). Of primary interest here was the trials effect.

Duncan Multiple Range tests indicated that children's performance on maintenance items improved significantly from pretest to the unprompted posttest ($\underline{p}<.001$), from the unprompted posttest to the prompted posttest ($\underline{p}<.05$), and from the pretest to the prompted posttest ($\underline{p}<.001$), summed across groups (figure 7, bottom).

Regarding the main effect for groups, Duncan tests indicated that children in the PSC and RT conditions outperformed children in the ME condition ($\underline{p}<.05$), summed over trials. No other comparisons were significant (figure

7, top). These group differences were thus the same as those revealed by the covariance analyses of the unprompted and prompted posttests. On maintenance items then, RT and PSC training yielded comparable improvements, superior to that of ME training. RD training was not distinguished from any other training by its effects on maintenance items.

Generalization Items

For generalization items at the unprompted posttest, an analysis of covariance on the adjusted scores yielded a nonsignificant groups effect, $F(3, 33) = 1.052$, $p > .383$, (table 12).

For generalization items at the prompted posttest, the ancova on the adjusted scores revealed a significant groups effect, $F(3, 33) = 4.725$, $p < .008$. According to Duncan Multiple Range tests, children in the RT condition outperformed children in both the RD ($p < .05$) and ME ($p < .005$) conditions. Children in the PSC condition outperformed those in the ME condition ($p < .05$) and were marginally superior to those in the RD condition ($p < .10$, table 13, figure 8). Including the marginal differences, RT and PSC children were comparable and superior to RD and ME children, who also were comparable. Thus, on the generalization items of the prompted posttest, RT and PSC children displayed the greatest gains.

A two-way repeated measures analysis of variance on the generalization items revealed a significant main effect for groups, F(3, 34) = 4.024, p<.015, a significant effect for trials, F(2, 68) = 83.791, p<.001, and a significant groups x trials interaction, F(6, 68) = 2.542, p<.028, (table 14). F-tests for simple effects, conducted to determine which particular groups improved over trials, revealed that all improved: ME, F(2, 68) = 7.625, p<.01; RD, F(2, 68) = 14.063, p<.001; PSC, F(2, 68) = 32.304, p<.001; RT, F(2, 68) = 40.084, p<.001, (figure 9). To determine the exact locus of these improvements, Duncan Multiple Range tests were applied. Children's scores in the ME condition increased from pretest to the unprompted posttest (p<.005), and from pretest to the prompted posttest (p<.001), but not from the unprompted to the prompted posttest (p>.10).

Similarly, the only significant increases for children's scores in the RD condition occurred from pretest to unprompted posttest (p<.001) and from pretest to the prompted posttest (p<.001).

Children's scores on generalization items in the PSC condition increased from pretest to unprompted posttest (p<.001), increased marginally from unprompted to prompted posttest (p<.10), and increased from pretest to prompted posttest (p<.001).

Similarly, children's scores in the RT condition increased from pretest to unprompted posttest (p<.001), from unprompted to prompted posttest (p<.005), and from pretest to prompted posttest (p<.001).

To summarize, all groups improved from pretest to the unprompted posttest, but only the RT and PSC conditions improved from the unprompted to prompted posttests. This likely reflects the significance of the role played by the prompts to the RT and PSC training.

The repeated measures anova provided an index of absolute improvement on generalization items from pretest to posttests. An alternate and independent means of doing so was via a large sample approximation of the binomial test. This test revealed that children in each of the four groups performed at or below chance levels on the 24 generalization items during pretest (alpha = .05, one-tailed test). Further, children in each of the four groups, including ME and RD, performed significantly above chance levels on the generalization items of both the unprompted and prompted posttests (alpha = .05, one-tailed test, table 15). Thus, these results were in agreement with the repeated measures anova for generalization items in suggesting that all groups improved significantly from pretest to posttests.

Partial generalization scores. Given that most of the generalization items varied on two or three dimensions, it was possible for a child's incorrect answer to be correct on a subset of relevant dimensions. Partial generalization scores were derived from a pooling of each child's incorrect responses to yield a more sensitive index of generalization. Children were given credit for each appropriate or correct attribute of their incorrect answers, and a percentage score was calculated. Given that correct choices did not directly contribute to this error analysis, the partial generalization measure was orthogonal to the generalization measure discussed above.

An analysis of covariance on the adjusted partial generalization scores of the unprompted posttest yielded a nonsignificant groups effect, $F(3, 33) = 1.146$, $p>.345$, (table 16).

An analysis of covariance on the adjusted scores of the prompted posttest indicated a significant groups effect, $F(3, 33) = 4.439$, $p<.01$. Duncan Multiple Range tests showed that children in group RT were superior to children in groups ME ($p<.01$) and RD ($p<.05$). Children in group PSC were superior to those in group ME ($p<.05$) and marginally superior to RD children ($p<.10$, table 17, figure 10).

Including the marginal differences, RT and PSC children were comparable and superior to RD and ME children. This pattern of group differences replicated that of the generalization measure. A two-way repeated measures analysis of variance on the partial generalization scores revealed a significant groups effect, $F(3, 34) = 3.258$, $p<.033$, a significant trials effect, $F(2, 68) = 34.363$, $p<.001$, and a nonsignificant groups x trials interaction, $F(6, 68) = 1.739$, $p>.125$, (table 18).

Duncan tests indicated that children's partial generalization scores significantly improved from pretest to the unprompted posttest ($\underline{p}<.001$), from the unprompted to the prompted posttest ($\underline{p}<.05$), and from the pretest to the prompted posttest ($\underline{p}<.001$), summed across groups (figure 11, bottom).

Unlike the generalization items then, the partial generalization scores revealed that all groups improved from the unprompted to the prompted posttest.

Duncan tests showed that children's scores in the PSC condition were higher than those in the RD and ME conditions ($\underline{p}<.05$). Scores in the RT condition were marginally higher than those of the RD and ME conditions ($\underline{p}<.10$). Thus, including marginal differences, scores in the PSC and RT conditions were comparable and superior to those of the RD and ME conditions, which were also comparable (figure 11, top). This pattern of group differences is identical to that which resulted from the ANCOVAS performed on generalization and partial generalization measures of the prompted posttest.

Results of these generalization and partial generalization measures were not consistent with the third hypothesis. It predicted that RT children would outperform RD children who would, in turn, outperform PSC and ME children. The rule component of the former two conditions was presumed to be essential to broad generalization across tasks and time, as stated above.

Far Generalization: (RCPM Set B)

For RCPM items at the unprompted posttest, an analysis of covariance on the adjusted scores revealed a nonsignificant groups effect, *F*(3, 33) = 0.132, *p*>.940, (table 19). For RCPM items at the prompted posttest, the analysis of the adjusted scores also revealed a nonsignificant groups effect, *F*(3, 33) = 0.701, *p*>.558, (table 20).

Carlson and Wiedl (1976) subjected the RCPM to factor analysis and found that items 8 through 12 of set B loaded clearly on a dimension they labeled, "concrete and abstract reasoning (p. 176)." Perhaps a separate examination of these items would yield a measure more sensitive to training effects. An analysis of covariance on the adjusted scores for these items at unprompted posttest yielded a nonsignificant groups effect, *F*(3, 33) = 0.510, *p*>.678, (table 21). At prompted posttest, the analysis also yielded a nonsignificant groups effect, *F*(3, 33) = 0.285, *p*>.836, (table 22).

The failure to find group differences on the far generalization measure was not consistent with the fourth hypothesis, which predicted that children's scores in the RD condition would exceed those of the ME condition, which would in turn exceed those of the RT and PSC conditions.

It was anticipated that the problem-solving training of groups PSC and RT would hinder performance on the novel RCPM items. At face value, counting appeared to be incompatible with the "abstract reasoning" requirement of the RCPM. It was also anticipated that RD training would facilitate children's identification of helpful rules as they encountered the novel problems. Apparently, this did not occur.

Absolute improvements were explored via a two-way repeated measures analysis of variance on the RCPM items. This showed a nonsignificant main effect for groups, $F(3, 34) = 1.576$, $p>.213$, a significant main effect for trials, $F(2, 68) = 11.618$, $p<.001$, and a nonsignificant group x trials interaction, $F(6, 68) = 0.496$, $p>.809$, (table 23).

Duncan Multiple Range tests suggested that RCPM scores increased from pretest to the unprompted posttest ($p<.005$), and from pretest to the prompted posttest ($p<.001$), but not from unprompted to prompted posttest ($p>.10$, figure 12).

Process Measures: Did Training Do What it was Expected to?

RD and ME children. During training, children in the RD condition were prompted once or twice on each trial with, "How are you going to do this to get it right?". ME children received no such prompting. Children's verbal and nonverbal responses on each trial were coded in the same manner as were their responses during pretest and posttests (see method section and appendix B). Three categories, (SpPROB-ap), (SpPROB-inap), and (INT) were eliminated because they accounted for only two coded instances in the entire study.

Results indicated that 61.7% (317/514) of RD children's prompted verbalizations during training were scored as irrelevant(IRR), while 95.7% (222/232) of ME children's responses were so scored (table 24). Recall that irrelevant(IRR) verbalizations included trials on which no verbalizations occurred at all. The bulk of the ME children's IRR responses consisted of these no verbalization trials (86.4% of total trials).

RD children engaged in verbal problem-solving, partially appropriate, on 17.2% (PROB-ptap:49/285) of all trials, and inappropriately on 1.4% (PROB-inap:4/285). ME children engaged in verbal problem-solving on 1.8% (PROB-ptap:4/221) of all trials.

With respect to specification of problem-solving responses, RD children did so on 24.9% of all trials (SpPROB-ptap:71/285). ME children did so on 0% (0/221). RD children verbalized a solution response on 4.9% of all trials (SOL: 14/285), while ME children did so on 0.5% of all trials (1/221).

Simple labeling of relevant stimulus aspects (LAB) occurred on 15.4% (44/285) of RD children's trials and 2.3% of ME trials (5/221).

Finally, RD children engaged in nonverbal tracking (T-ap or T-ptap) on 14.7% of all trials (42/285). ME children did so on 0.9% of all trials (2/221).

It is clear from these comparisons that RD training prompted more verbal problem-solving and specification of responses and outcomes than ME training. During the latter, children were preponderantly silent. See table 25 for examples of RD children's most frequent verbalizations.

How successful, in absolute terms was the RD training in prompting viable rules and verbal problem-solving? It should be noted that only 5 of the 9 RD children verbally problem-solved (PROB-ap, -ptap, or inap) on one or more trials during training. These five did so on less than 1/3 (30.8%, 49/159) of their training trials.

With respect to verbal rules, 5 of 9 RD children had verbal responses scored (SpPROB-ptap) or (SOL) on more than one trial during training. These occurred on 54.6% (83/152) and 9.2% (14/152) of their training trials, respectively. Only one child, failed to emit any verbal problem-solving or any verbal rules.

Most importantly, no child in the RD condition on any trial specified the appropriate contingency, e.g., "I have to count the colors to get it right," (RULE-ap), nor did they ever manage simply to say the appropriate problem-solving response, e.g., "I have to count the colors," (SpPROB-ap). Thus, although RD children offered more relevant verbalizations than ME children, in absolute terms they were only moderately successful in identifying rules.

RT and PSC children. In these conditions, varying degrees of modeling, prompting, and shaping were used to assure that children engaged in specified rules, and verbal and nonverbal strategies on every trial of training. On every trial then, children in the RT group verbalized appropriate rules (RULE-ap), appropriate verbal problem-solving (PROB-ap), and appropriate nonverbal tracking (T-ap). PSC children engaged in appropriate verbal problem-solving and appropriate nonverbal tracking (but no rules) on every trial. For an illustration of RT training, see appendix C.

Training validation. Group differences emerged for generalization items at the second, prompted posttest but not at the initial, unprompted posttest. Given that the order of unprompted and prompted posttests was not counterbalanced, it cannot be concluded that prompting was a necessary condition for group differences to emerge, despite the intuitive appeal of such an argument.

The expected effect of prompting was to increase the problem-solving of RT and PSC children from the unprompted to the prompted posttest. If this increase occurred, it could be argued more persuasively that prompts per se were responsible for group differences at prompted posttest. Conversely, if problem-solving did not increase from the unprompted to the prompted posttest, the prompts would thus be shown unsuccessful in producing their desired effect, and group differences at prompted posttest could not readily be attributed to prompting.

Dependent t-tests were employed to determine if problem-solving increased between posttest administrations. Because of the number of t-tests employed to address this and subsequent questions, an alpha level of .02 was adopted to provide some containment of experimentwise error rates.

A t-test performed on the number of instances of verbal problem-solving (coded as PROB-ap, -ptap, or -inap) combined with the number of instances of nonverbal problem-solving (coded as T-ap,-ptap, or -inap), for PSC and RT children combined suggested significant increases between posttests, $t(19) = -3.407$, $p<.01$, (table 26). The same measures with respect to RT children alone yielded similar results, $t(9) = -3.465$, $p<.01$. Problem-solving for PSC children alone did not increase between posttests, $t(9) = -1.502$, $p>.10$.

It was noted however, that one child in the PSC condition problem-solved frequently at unprompted posttest (more than any other PSC child) but problem-solved very little at prompted posttest. Also, pertinent, Ken showed the greatest improvement in generalization from pretest to each posttest of any PSC child. Excluding his data, children in the PSC condition display a trend toward increased problem-solving from unprompted to prompted posttest, $\underline{t}(8) = -2.824$, $.02<\underline{p}<.05$.

For RT children, emission of the formal rule, which included instances coded as RULE-ap, -ptap, -inap, SpPROB-ptap, and SOL, increased from unprompted to prompted posttest, $\underline{t}(9) = -3.743$, $\underline{p}<.01$, (table 27). The increase in problem-solving and rule recitation from unprompted to prompted posttest is consistent with the argument that group differences emerged at the prompted posttest as a function of prompting.

Rule training distinguished the RT training from the PSC training condition. It was anticipated to increase the probability of problem-solving, particularly at the unprompted posttest. The failure of RT children to outperform PSC children on any of the outcome measures might suggest that the rule did not function effectively in this manner. However, it is also plausible that RT children did problem-solve more than PSC children at posttests but, for whatever reason or reasons, these differences were not reflected by the outcome measures.

Independent t-tests were conducted on the combined number of instances of verbal and nonverbal problem-solving to examine the influence of the formal rule training. During the unprompted posttest, there was a marginal difference between PSC and RT children's problem-solving, $\underline{t}(18) = 2.351$, $.02<\underline{p}<.05$. This difference however, was in favor of PSC children, who tended to do more problem-solving during the unprompted posttest than RT children (see again table 26). At prompted posttest, RT children's problem-solving approached that of PSC children, as no differences were apparent, $\underline{t}(18) = -0.180$, $\underline{p} > .80$. It is clear from these results that RT children were not more likely to problem-solve at unprompted or prompted posttests than PSC children. In fact, at unprompted posttest they were somewhat less likely than PSC children.

During training, RT and PSC children were taught verbal problem-solving that named values of shape, attributes of color, or both. Additionally, RT children were taught rules which specified the counting of shape and color dimensions. Sixteen of the twenty-four generalization items of the pretest and each posttest featured size, pattern, or both size and pattern as relevant and novel dimensions.

Thus it was of interest to examine whether children's verbalizations changed to reflect the newly relevant dimensions. This was addressed for children in all conditions and groups RT and PSC in particular, on the generalization items of posttests where size, pattern, or both were relevant.

The measure of this form of generalization was the number of children who generalized appropriately on one or more trials of the posttests.

Six of the ten children in the RT group emitted verbal problem-solving that was appropriate to a novel dimension during the posttests. This was exemplified by an RT child who said, "1,2,3 big; 1,2,3 medium; 1,2 small," on item 9 of the prompted posttest.

Five of the ten PSC children problem-solved in similar fashion. Three of nine RD children did so, while only one of nine ME children issued verbal problem-solving that reflected a novel dimension during posttests.

Only one child in the RT condition verbalized rules which reflected a novel dimension. On several items of the prompted posttest, the child said, "If I count the big ones and small ones it will help me."

Thus, while the majority of children in the RT condition issued verbal problem-solving that was appropriate to a novel dimension, only one child's rules were appropriate. The generalization of the formal properties of the rule appeared less probable than the generalization of the formal properties of the verbal problem-solving which the rule specified.

One child in the PSC condition specified a problem-solving response that was appropriate to a novel dimension: "Look at the lines."

No child in either the RD or ME condition emitted a rule or specified a problem-solving response that was appropriate to a novel dimension.

The limited effectiveness of the RD intervention in prompting verbal problem-solving and rules during training was discussed above. It would follow that such rules and problem-solving would be lacking at posttests as well. This was examined via independent t-tests using the multiple exemplar group for comparison, with an alpha level of .02. At unprompted posttest, results indicated that RD children were not more likely to say rules (RULE-ap, -ptap, or -inap, SpPROB-ptap, or SOL) than ME children, $\underline{t}(16) = -1.000$, $\underline{p} > .20$, (table 28). Nor were RD children more likely to do so at the prompted posttest, $\underline{t}(16) = -1.356$, $\underline{p} > .10$).

RD children were not more likely to do verbal problem-solving (PROB-ap, -ptap, or -inap) than ME children at unprompted posttest, $\underline{t}(16) = -0.899$, $\underline{p} > .20$, (table 29), nor were they more likely at prompted posttest, $\underline{t}(16) = -0.816$, $\underline{p} > .20$. Prompted rules and verbal problem-solving did not appear to generalize from training to posttests, when compared to the data of ME children. This is not surprising given the modest success in prompting rules from RD children during training.

The Relationship Between Problem-Solving and Outcome

To facilitate interpretation, coded verbalizations were retabulated to form three new categories: "Appropriate" verbalizations included appropriate verbal problem-solving and contingency specification (PROB-ap, RULE-ap); "Partially appropriate" verbalizations included partially appropriate problem-solving (PROB-ptap) and partially appropriate rules (RULE-ptap, SpPROB-ptap, and SOL), the latter two categories entailing only a portion of the rule, and; "Inappropriate/Irrelevant" verbalizations which included simple labeling (LAB), inappropriate problem-solving (PROB-inap), inappropriate rules (RULE-inap), and irrelevant verbalizations (IRR).

Because it was the case that, for a given trial, several or all of the three categories could be represented, the following decision rule was imposed: Any trial that included an appropriate verbalization was scored as such, while any trial that included both partially appropriate and inappropriate/ irrelevant verbalizations was scored as partially appropriate. The relationship between problem-solving and outcome was not examined on the 12 RCPM items of each posttest, because there was no reason to expect problem-solving, as employed in this study, to facilitate RCPM performance. The relationship was examined on the remaining items of the unprompted and prompted posttests combined, for PSC and RT children, combined.

A chi-square test of homogeneity revealed that these three categories were not homogeneous in their relationship to correct outcomes, x2(2, $\underline{N}$ = 1240) = 66.254, $\underline{p}$<.005, (table 30). Subsequent partitionings of chi-square revealed that appropriate verbalizations were more predictive of correct outcomes than partially appropriate verbalizations, x2(1, *N* = 1240) = 25.669, *p*<.005), while partially appropriate verbalizations were, in turn, more predictive of correct answers than inappropriate/ irrelevant, x2(1, *N* = 1240) = 7.546, *p*<.01.

The four nonverbal categories were: tracking-appropriate (T-ap); partially appropriate (T-ptap); inappropriate (T-inap), and; no pointing (NP). A chi-square test of homogeneity revealed that these categories were not homogeneous in their relationship to correct outcomes, x2(3, *N* = 1238) = 100.234, *p*<.005, (table 31). A partitioning of chi-square indicated that the "no pointing" and inappropriate nonverbal categories were not different from one another in their relationship to outcome, x2(1, *N* = 1238) = 1.588, *p*>.10. Nor were "no pointing" and inappropriate combined different from partially appropriate x2(1, *N* = 1238) = 1.367, *p*>.10). However, a comparison of these three combined categories with appropriate nonverbal problem-solving revealed they were not homogeneous in their relationship to correct outcomes, x2(1, *N* = 1238) = 97.275, *p*<.005): Appropriate tracking was far more predictive of correct solutions than the 3 remaining categories, which did not differ from each other.

The chi-square tests revealed the relative predictiveness of several verbal and nonverbal categories to correct outcomes. An alternative means of addressing the relationship between problem-solving and outcome in absolute terms is to determine, for a given problem-solving category, if children were performing above levels expected by chance.

A large sample approximation of the binomial test was applied to the data of PSC and RT children combined, at unprompted and prompted posttests combined, for all except RCPM problems. Results indicated that for all 3 of the verbal categories, appropriate, partially appropriate, and inappropriate/irrelevant, children were performing significantly above chance levels, one-tailed test, alpha =.05 (table 32). The same held true for the nonverbal problem-solving categories, with performance significantly above that expected by chance for no pointing, inappropriate, partially appropriate, and appropriate categories, one-tailed test, alpha=.05 (table 33).

Thus, regardless of the degree of appropriateness of the verbal or nonverbal problem-solving on each trial for PSC and RT children during posttests, they answered correctly more often than would be expected by chance.

Children in groups RD and ME issued relatively few verbal rules and virtually no problem-solving at posttests, and hence their data have not been subjected to the same analyses of problem-solving and outcome as data from PSC and RT children.

Summary of Results

Outcome. Groups RT and PSC reached criterion during training in a comparable number of trials. Groups ME and RD also reached criterion in a comparable number of trials, but required more trials than did the RT and PSC groups (figure 4).

The pretest and posttests yielded measures of maintenance, generalization, partial generalization, and far generalization. The maintenance items were the only items to show differential training effects at the unprompted posttest (figure 5). At both the unprompted and prompted posttests, an analysis of covariance revealed that children in the RT and PSC conditions outperformed children in the ME condition, with no other comparisons significant. Summed across groups, children's scores improved from the pretest to the unprompted posttest, and from the unprompted posttest to the prompted posttest (figure 7, bottom).

Generalization items revealed differential training effects at the prompted posttest. RT and PSC children were comparable and superior to RD and ME children, who also were comparable (figure 8). While all groups improved from the pretest to the unprompted posttest, only the RT and PSC children improved from the unprompted to the prompted posttest (figure 9).

Partial generalization scores also revealed differential training effects at the prompted posttest. As with the generalization items, groups RT and PSC were comparable and superior to groups RD and ME, which also were comparable (figure 10). Summed across groups, partial generalization scores improved from pretest to the unprompted posttest, and from the unprompted posttest to the prompted posttest (figure 11, bottom).

Unlike the other measures, far generalization items failed to show differential training effects at either posttest. Summed across groups, children's RCPM scores increased from the pretest to the unprompted posttest but not from the unprompted posttest to the prompted posttest (figure 12).

Process. During training, ME children were silent on 86.4% of their trials. Although RD children specified problem-solving responses on 24.9% of trials, and verbally problem solved on 17.2% of trials, no RD child ever specified an appropriate rule.

Problem-solving, as trained in the RT and PSC conditions, tended to increase between posttests for children in these conditions. This was consistent with an interpretation that the prompt of the prompted posttest was essential to differential training effects for the generalization items.

Children who received training in the formal rule did not do more problem-solving at either posttest than PSC children. Thus, the rule training component did not appear to exercise special functional properties in increasing the probability of problem-solving. In fact the rule itself, much as the problem-solving behavior it specified, increased as the result of the prompt of the second posttest.

RD children did not verbalize more rules, nor did they do more problem-solving than ME children at posttests.

For RT and PSC children at posttests, rules, verbal problem-solving, and nonverbal problem-solving were coded for their degree of appropriateness, on the basis of the portion of relevant dimensions captured. On a trial to trial basis, the appropriateness of rules and problem-solving was predictive of correct outcomes.

It was also true however, that regardless of the degree of appropriateness of rules or problem-solving, children performed above the level of responding predicted by chance. Failure to problem-solve appropriately, or failure to issue an appropriate rule by no means guaranteed an incorrect solution.

CHAPTER IV

DISCUSSION

The present study evaluated the functional utility of rules and problem-solving in fostering generalization across tasks in a sample of normal preschool children. Trained rules, prompted rules, and problem-solving were examined in a learning context that featured the training of multiple exemplars. A discussion of the results as they pertain to the hypotheses will begin this chapter.

Hypothesis I: Rule Training (RT) and Problem-Solving Control (PSC) Conditions Would Require Fewer Trials to Criterion than Rule Discovery (RD) and Multiple Exemplar (ME) Conditions

The results were consistent with the first hypothesis. Children in the RT and PSC conditions required fewer total trials to criterion during training on variants 1 through 3 than did children in conditions RD and ME.

An operant account of problem-solving (Skinner, 1966) and an applied operant model (Grimm, Bijou, & Parsons, 1973) were presented in the introductory chapter. Initially, task and learning contexts set the occasion for problem-solving responses which alter the problem-situation and are reinforced for doing so. Stimuli associated with the altered problem situation set the occasion for solution responses which produce reinforcement that maintains the entire chain.

RT and PSC children were taught a series of problem-solving steps designed to increase the probability of correct solutions. Specifically, children were taught to count, label, and track values of relevant dimensions. Once emitted, these problem-solving responses altered the task situation such that stimuli from the altered situation set the occasion for problem-solution. Counting, labeling, and tracking made relevant stimuli more discriminable, and increased the probability of emission of a solution response.

During the RT and PSC training, trainers ensured that appropriate problem-solving occurred on every trial. Hence the superior performance of these groups validated the utility of the problem-solving responses that were targeted for the training tasks.

The formal rules taught to RT children specified a relationship between problem-solving responses and outcome. On trials to criterion, RT training was not expected to yield greater effects than PSC training, as the rules simply specified the problem-solving that was already guaranteed to occur during training for both the RT and PSC children.

For children in groups RD and ME, no explicit problem-solving techniques were trained. RD children were prompted to verbally identify task demands on each training trial with each variant. Their success in doing so was expected to increase over the course of trials and variants, with feedback. In contrast, the problem-solving common to groups RT and PSC was ensured from the very first training trial. Hence, children in groups RD and ME required significantly more trials to reach criterion than did children in groups RT and PSC.

It should be noted that RD and ME children reached criterion on all four variants of training. These children may have learned covert problem-solving techniques of an undisclosed nature. Or they may simply have learned to discriminate correct choices on the basis of the feedback component of training. In either case, their performance was inferior to RT and PSC children as measured by trials to criterion.

Hypothesis II: On the Maintenance Items of the Unprompted Posttest, Children in the RT Group Would Outperform Children in the RD Group, Who Would in Turn Outperform Children in the PSC and ME Groups

The results of the present study failed to confirm the second hypothesis. The maintenance items of both the unprompted and the prompted posttests revealed that the RT and the PSC training did not differ from one another. Both were superior to ME training, and the RD training was not different from any other training condition.

One assumption that generated the second hypothesis was that trained rules would act as verbal mediators. Emitted rules are responses with stimulus functions. As salient common stimuli, they could be carried hypothetically from any training context to any testing context, and hence mediate generalization over time. Verbal responses are easily recalled, and make it easier to retain discriminations over time (Skinner, 1966). In the extratraining conditions of the posttests, the rules taught to RT children would be more readily emitted than the problem-solving responses that the rules specified. If this were so, RT children might outperform PSC children in solving the matrices. However, RT training effects failed to surpass the effects of the PSC training on maintenance items.

What was the locus of this failure within the problem-solving sequence that was taught to RT children? Three possibilities presented themselves: (1) The testing context failed to control rule emission at posttests; (2) rule emission occurred at posttests, but it failed to control the problem-solving it specified; (3) rules and problem-solving responses occurred at posttests, but problem-solving did not control solution responses. These issues are also germane to the third hypothesis, and will be explored there.

As in the RT condition, the RD condition was expected to exert its effects beyond those of the PSC and ME groups as a function of the rules emitted over the course of training. To the extent that RD children learned to identify viable rules, these group differences would be evident. The RD training was not superior to PSC and ME training on maintenance items. Its effects did not differ from any other training condition at either the unprompted or prompted posttest. Correspondingly, RD children did not emit more rules or verbal problem-solving than did ME children at either posttest. These results might be expected, however, if the RD intervention was less than successful in prompting viable rules from children during training.

Apparently, this was the case. As reported in the previous chapter, process data indicated that RD children were only moderately effective in producing appropriate verbalizations during training: During training, the majority of RD children's verbalizations were coded as irrelevant; On no trial did any RD child specify that counting was an appropriate strategy.

Conceptual impetus for the RD intervention came from the assumption that the process by which rules were emitted was important in determining the maintenance and generalization of behavior the rules specified. Rules that were prompted and differentially reinforced might control behavior more strongly than a rule of identical topography that was modeled and instructed. During training however, the RD intervention did not generate verbalizations comparable to the rules that were taught to RT children. As such, the RD intervention could not address the validity of this assumption. Conceptual and pragmatic implications of the RD intervention will be discussed below.

RT and PSC training effects were comparable and superior to the Multiple Exemplar training. Common to the former training regimens was the inclusion of problem-solving that required children to count and track values of relevant stimulus dimensions. The problem-solving responses altered the task situation and set the occasion for problem solution. The ME training did not feature the training of problem-solving strategies.

Thus, the problem-solving component may have mediated generalization on maintenance items relative to the ME condition. However, the effects of the RT and PSC training were not different from that of the RD training, which, like the ME training, did not include training in problem-solving. The fact that RD training effects did not differ from RT and PSC training effects, nor from ME training effects is not readily explained. The role of the problem-solving component will be addressed further in the discussion of the third hypothesis.

Hypothesis III: On the Generalization Items of the Unprompted Posttest, Children in the RT Group Would

Outperform Children in the RD Group, Who Would in Turn Outperform Children in the PSC and ME Groups

Analyses of the generalization items on both the unprompted and the prompted posttests failed to confirm the third hypothesis. There were no group differences on the unprompted posttest. On the prompted posttest, the RT and PSC conditions were comparable and superior to the RD and ME conditions, which also did not differ. The analysis of the partial generalization scores revealed an identical pattern of group differences: No groups differed at the unprompted posttest, and; groups RT and PSC outperformed groups RD and ME at the prompted posttest.

The same assumption that generated the second hypothesis generated the third: In extratraining conditions, appropriate rules, as readily emitted behavior, would act as verbal mediators to facilitate the problem-solving behavior specified. In turn, the problem-solving behavior would increase the probability of correct solutions to the matrices. But the generalization items, like the maintenance items, revealed no greater effects for RT training than PSC training.

The failure of RT effects to surpass PSC effects on the generalization and maintenance items raised questions that differentially implicated the source of the failure in the problem-solving model: Did the rule training component of the RT condition fail to facilitate the occurrence of problem-solving at posttests? or; did rule training effectively facilitate problem-solving at posttests, but this increase, for whatever reason, was not reflected by correct outcomes?

Before proceeding, a caveat is called for regarding the inferences drawn from the process data. Measures of children's overt emission of rules and problem-solving were dependent variables. The relationships established between rules and problem-solving and each with outcome were thus correlational in nature. Causality could not be determined as it could from a true experimental manipulation.

Moreover a measurement problem, common to rule training paradigms, complicated matters further. It was conceivable that rules and problem-solving were emitted covertly at posttests. These covert responses might have had important stimulus properties which led to correct solutions. If rules and problem-solving occurred covertly and facilitated matrix solution, observed relationships between these process and outcome variables would necessarily be lowered.

Nonetheless, the relationships that emerged between process measures permitted plausible causal inferences given the training, testing, and task contexts of the present study. The requirement for overt verbalization and problem-solving was never relaxed during training. Thus training did not encourage children to emit any problem-solving or rules at a covert level. The Soviet developmental literature has suggested that increases in the level of task difficulty will increase the probability of overt self-regulatory speech in young children (Vygotsky, 1962; Roberts, 1979). The bulk of posttest items were difficult and complex in comparison to training items. Thus this literature suggests that task difficulty would serve to maintain children's overt verbalizations at posttests. Covert problem-solving might only minimally mitigate observed relationships between process measures. The relationships between process measures are explored tentatively, in the hope of elaborating the results generated by the true experimental manipulations.

Returning to the question posed above, the process data presented in the preceding chapter implicated the former event: rule training failed to facilitate problem-solving at posttests. Relative to the PSC children, the RT children did no more verbal and nonverbal problem-solving at each posttest. In fact, at the unprompted posttest, children in the PSC condition tended towards greater amounts of problem-solving than RT children.

The fact that RT children did no more problem-solving at posttests than did PSC children raised another critical question: Did RT children fail to emit rules at the unprompted posttest? Or was it the case that RT children successfully emitted rules at the unprompted posttest, but these rules failed to control corresponding problem-solving?

The former situation appeared to be accurate. Rather than inadequate control of problem-solving by rules, it was simply the case that RT children verbalized a rule on only 9.5 percent (42/440) of trials at the unprompted posttest (see table 34).

The functional control exercised by RT children's rules appeared strong. When a rule was verbalized by RT children at the unprompted posttest, problem-solving followed immediately 92.9 percent (39/42) of the time. In the absence of rules, problem-solving occurred on only 2.3 percent (9/398) of trials. This strong relationship between rules and problem-solving may be a function of the reinforcement contingency for both during training.

Given these results, the assumption above will again be addressed. The rules that RT children were trained to emit were effective in controlling specified problem-solving at posttest. However, they emitted these rules only infrequently at the unprompted posttest. Rules did not appear to function at the unprompted posttest as readily emitted behavior. They were emitted with no greater frequency than the problem-solving responses (counting and tracking) they specified.

Why did RT children emit rules infrequently at posttest? One possibility was that in the absence of prompts, the stimuli of the task and learning context did not control rule emission. To reduce this possibility, prompts for rules and problem-solving were faded during training. On two consecutive trials of each training variant, RT children were required to emit the entire rule and problem-solving sequence without any prompts or aid from the trainers.

Alternatively, it was possible that rules were not emitted because reinforcement was no longer contingent upon rules or problem-solving at either posttest. And it may be the case that for these preschoolers, correct matrix solutions did not serve as reinforcers, maintaining earlier problem-solving including rule emission. Brown and DeLoache (1978) suggested that preschoolers' failure to regulate and monitor necessary steps followed from a "lack of familiarity with the game at hand," (p. 128). If the game is to do well in school, their statement points to the possibility that answering questions and solving problems have not yet acquired reinforcing properties for the typical preschooler. As such, it may be necessary to arrange contingent reinforcement to ensure continued success in problem solution and problem-solving.

As discussed above, the RD intervention was only moderately effective in prompting appropriate verbalizations during training and was ineffective in prompting rules. It was not surprising then, that it yielded no greater training effects than the ME condition at either posttest.

On the generalization items of the prompted posttest, RT and PSC training proved comparable and superior to the ME training, as was the case with maintenance items. Unlike the maintenance items, the generalization items of the prompted posttest also revealed that RT and PSC training were each superior to the RD training. RD and ME training did not reveal differential training effects on the generalization items.

Given that both RT and PSC training featured the training of problem-solving strategies, and RD and ME training did not include problem-solving training, it was probable that the problem-solving component mediated generalization to novel tasks on the prompted posttest.

Convergent support for this supposition, albeit correlational in nature, came from a variety of sources. First, recall that group differences were manifest at prompted posttest only. It was observed that problem-solving occurred on a high percentage of generalization items for RT and PSC children at prompted posttest, but only a small percentage of items on unprompted posttest (see again table 34). This increase in problem-solving between posttests was significant.

Evidence that the problem-solving of RT and PSC children mediated correct outcomes came from data that showed that the degree of appropriateness of rules and problem-solving was directly related to correct outcomes on the posttests. Rules and verbal problem-solving were more predictive of correct matrix solutions when they reflected all relevant dimensions than when they represented only a subset.

In turn, rules and verbal problem-solving that reflected a subset of relevant dimensions were more predictive of correct solutions than were inappropriate or irrelevant verbalizations. A similar relationship held for nonverbal problem-solving or tracking, as the tracking of all relevant dimensions predicted correct solutions more than tracking a subset of dimensions or tracking no dimensions at all.

Thus, the degree of appropriateness of verbal and nonverbal problem-solving predicted correct solutions at both posttests. However, significantly more problem-solving occurred at the prompted posttest, and this was where group differences emerged for both generalization items and partial generalization scores. Group differences emerged at prompted posttest as the result of the problem-solving training common to groups RT and PSC. But problem-solving exercised its differential effects at the prompted posttest as the result of the prompting. The fact that the frequency of occurrence of rules and problem-solving increased significantly from unprompted to prompted posttests is consistent with the conclusion that prompting was a necessary condition for the problem-solving training to exercise differential training effects.

In sum, this was the picture of the generalization data: The prompts of the prompted posttest frequently cued problem-solving responses for RT and PSC children. This problem-solving may have been instrumental in facilitating correct matrix solution. At the first posttest, without prompts to use previously trained strategies, problem-solving occurred infrequently.

Contrary to expectation, RT children's rules also occurred infrequently. The few occasions on which RT children emitted rules at the unprompted posttest, the rules appeared to effectively control corresponding problem-solving. Thus, emitted rules did not appear deficient in controlling problem-solving at the unprompted posttest. Unprompted, they simply were not emitted with any regularity.

Hypothesis IV: On the Far Generalization (RCPM Set B) Items of the Unprompted Posttest, Children in the RD Group Would Outperform Children in the ME Group, Who Would in Turn, Outperform Children in the RT and PSC Groups

Analyses of the RCPM items at both the unprompted and prompted posttest failed to confirm the fourth hypothesis: There were no group differences at either the unprompted or prompted posttest. It was hoped that RD children would have verbally identified appropriate strategies on the RCPM items, having been taught to do so on the training items. As discussed above, however, RD children gave few relevant verbalizations during training and no appropriate rules.

On the posttests, including the prompted posttest, RD children failed to emit any more rules or verbal problem-solving than ME children. Children in the RD condition had not reliably learned to verbally identify task requirements during training. Therefore, they did not outperform ME children on the RCPM items at posttests.

The rules and problem-solving taught to RT and PSC children during training were not amenable to the bulk of the RCPM items. To the extent that RT and PSC children continued to emit these strategies, performance was expected to be depressed relative to ME children, who received no rules or problem-solving training. RT and PSC children problem-solved on RCPM items approximately half as frequently as they did on remaining maintenance and generalization items, even on the prompted posttest (see table 34). Thus, RT and PSC children discriminated to some degree that trained problem-solving strategies were inappropriate on RCPM items.

The failure of group differences to emerge on RCPM items was explained by the fact that: RD children did not learn to verbally identify appropriate rules during training, and RT and PSC children discriminated that formerly trained problem-solving was inappropriate to RCPM completion.

As a sidelight, the failure to find group differences on RCPM items in favor of groups RT and PSC helped rule out trainer enthusiasm or other nonspecific factors as determinants of the group differences on the generalization items.

The Rule Discovery (RD) Intervention

Process data suggested that children in this condition were only moderately successful at verbalizing relevant problem-solving steps during training. They were clearly unsuccessful in identifying viable rules that specified a relationship between problem-solving responses and outcomes. At subsequent posttests, RD children emitted rules and verbal problem-solving with no greater frequency than ME children.

Children in the RD condition did not emit rules that were comparable to those taught to RT children. Thus, the RD training failed to permit an evaluation of the process by which rules were emitted and the impact of that process upon generalization. However this failure was not without conceptual and practical implications. RD children were given only one or two prompts on each trial to identify appropriate rules during training. This limit was imposed to minimize the aversiveness of repeated questioning. It was also true that trainers never modeled appropriate rules to ensure that this procedure remained distinct from other training conditions.

But it was also the case that RD children failed to verbally identify relevant rules despite the fact that they all learned each of the four training variants to criterion. Additionally, trainers provided elaborated feedback which entailed the verbal identification of relevant stimulus characteristics of correctly completed matrices (e.g., "This is the right answer because there are 3 blue, 3 green, and 3 red").

Difficulty in prompting young children's verbalizations of task requirements has been noted elsewhere (Crisafi & Brown, 1983). Further, discovery methods that have led children to identify viable rules (and yielded superior generalization) may not have constituted discovery methods at all. Schleser, Meyers, Cohen, & Thackwray (1983) employed discovery training with third and fourth graders. The experimenter "rephrased" the child's responses if necessary and had the child cumulatively rehearse all "discovered" statements. Clearly this intervention entailed much more than the prompting and shaping of rules, and may well have provided more practice with "discovered" statements than with instructed statements of the comparison conditions.

Thus, data from various studies including the present one have suggested that a relatively great degree of environmental support is necessary for children to emit ostensibly valid rules. Perhaps then, developmental factors hindered children's identification of viable rules in the RD condition. Perhaps these factors limited the effectiveness of the rule component of the RT training as well.

There are data that suggest that developmental factors should not be overemphasized as strict determinants of the present results. As discussed in the first chapter, difficulty in shaping rules has been observed even with sophisticated subjects and seemingly simple response requirements. Catania, Matthews, and Shimoff (1982) successfully shaped guesses of "press fast," or "press slow," in not quite half of their sample of college students.

What may be critical to successful rule emission is a relevant learning history with similar rules and circumstances. Just as generalized task performance is facilitated by training multiple exemplars, generalized rules may result from training multiple rule instances.

The present results offered some support for this conceptualization of rule discovery. Verbal problem-solving that was appropriate to <u>novel</u> dimensions appeared most frequently among children in groups RT and PSC. Eleven of the fifteen children who named and counted values of novel dimensions on generalization items were in the RT and PSC groups. Having been trained to count values of colors and shapes during training, they counted values of size or pattern at posttest. This occurred despite its never having been prompted or reinforced at any time.

Training which is successful in producing the verbal identification of task contingencies likely entails substantial amounts of prompting and environmental supports, or a history of appropriate rule identification with similar task demands.

Were developmental factors responsible for the failure of the rule component of RT training to produce training effects beyond those of PSC training?

The five-year olds of the present study were at or near the age at which the child's speech is said to begin to be transformed into thought and goes underground, as a function of the child's social-psychological history (Luria, 1961). It is in this five to seven year age range that language comes to exercise a mediational function.

Metastrategic behavior has been observed in children as young as three years (Flavell, Ritter, & Wellman, 1975, cited in Flavell & Wellman, 1977), while metastrategic failure has been observed in adults. As an example of the latter, Gick and Holyoke (1983, cited in Crisafi & Brown, 1983) found that explicit instructions to generalize had been required for generalization to occur in their adult sample.

Higa, Tharp, and Calkins (1978) found that requiring young children to verbalize as they performed a nonverbal task interfered with its execution. For kindergartners versus second graders, self-instructions retarded performance. In the present study, rules and verbal problem-solving, as well as problem solution, were all independently trained to criterion to minimize interference generated by dual task requirements. Reinforcement during training was contingent on the joint occurrence of the rule and specified problem-solving for RT children. This is precisely the means by which correspondence has been produced in preschool children (e.g., Risley & Hart, 1968).

Thus, the ability to verbally self-regulate behavior does not appear to be a strict function of age. More important perhaps in predicting the effectiveness of rule training is the child's learning history with specified tasks and rules. Task and rule complexity has elsewhere been implicated as a critical variable in the learning of rules and their effective application (Carnine, Kameenui, & Maggs, 1982).

Summary Implications

There is a paucity of research investigating the impact of verbal antecedents upon the generalization of training effects to extratraining conditions. The results of applied studies generated by diverse theoretical paradigms have typically failed to isolate the effects of rule-based interventions from those of alternative training components, including the training of multiple exemplars.

The present study procedurally isolated a rule-training component from the multiple task exemplars that the rule was designed to address. Additionally, it uniquely controlled for the influence of the problem-solving behaviors specified by the rules. An implicit assumption of both the self-instructional training studies and the metastrategy training studies was that instructed rules or trained self-verbalizations would control the behaviors or strategies specified. It was further assumed implicitly in these paradigms that whereas the trained strategies or behavior might fail to maintain in extratraining conditions, the metastrategy or rule would itself maintain and subsequently control behavior in superordinate fashion.

For children in groups RT and PSC, a problem-solving sequence entailing verbal and nonverbal components was made overt and reinforced during training. The problem-solving was appropriate to the range of training tasks, but was appropriate beyond that range as well. This problem-solving training yielded greater generalization effects than the multiple exemplar training alone, to tasks which lay beyond the training range at prompted posttest.

Although prompting was necessary for differential training effects to occur, this constituted generalization as defined by Stokes and Baer (1977):

> Generalization may be claimed when no extratraining manipulations are needed for extratraining changes; or may be claimed when some extra manipulations are necessary but their cost or extent is clearly less than that of the direct intervention. (p. 350).

According to the same writers, the teaching of problem-solving during training constituted mediated generalization: A response was established,

> as part of the new learning that [was] likely to be utilized in other problems as well, and [constituted] sufficient commonality between the original learning and the new problem to result in generalization. (p. 361).

In addition to multiple exemplar training and problem-solving, RT children were trained to verbalize formal rules which specified the contingency between problem-solving and outcome. This training yielded no greater training effects than the training of problem-solving in conjunction with multiple exemplars, despite the rule-correspondence feature of the RT training.

The rules of RT children appeared to strongly control problem-solving at the unprompted posttest, most likely as the result of the reinforcement contingency for correspondence between rules and problem-solving during training. Rules simply occurred at low frequencies at the unprompted posttest. The prompts of the prompted posttest were necessary to do what was hoped the rule training would do: promote high frequencies of problem-solving.

The findings of Guevremont, Osnes, and Stokes (1986) have some noteworthy parallels to the present results.

These experimenters employed a correspondence training procedure to develop consistency between children's verbalizations and their subsequent behavior across settings and time. For each of three preschool children, once correspondence was established in one or two settings, with one or two target behaviors, behaviors became modifiable merely by prompting relevant antecedent verbalizations. When verbalizations were no longer prompted, however, corresponding targeted behaviors did not usually maintain.

Similarly, in the present study, correspondence between rules and problem-solving was established for RT children during training. At the unprompted posttest, neither rules nor corresponding problem-solving were prompted, and differential treatment effects were not observed. At prompted posttest, rules were prompted. These likely controlled corresponding problem-solving which yielded differential treatment effects.

If the goal of intervention is to maximize correct outcomes in the absence of any form of intervention, it may be necessary to arrange for solution responses that generate their own reinforcement, or fade arbitrary reinforcers. Verbal antecedents and problem-solving might then maintain as long as they were instrumental to correct solutions.

Should rules be trained? In the present paradigm, their training appeared to have no special generalization properties over and above the training of behaviors they specified. For a given child and target behavior, if rules are functional, prompting them might prove more expedient, if not more effective, than prompting specified behavior.

BIBLIOGRAPHY

Baer, D. M., Peterson, R. F., & Sherman, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal of the Experimental Analysis of Behavior, 10,* 405-416.

Becker, W. C. (1971). Teaching concepts and operations, or how to make kids smart. In W. C. Becker (Ed.), *An empirical basis for change in education*. Chicago Science Research Associates.

Belmont, J. M., Butterfield, E. C. & Borkowski, J. G. (1978). Training retarded people to generalize memorization methods across memory tasks. In M. M. Gruneberg, P. E. Morris, & R. N. Synkes (Eds.), *Practical aspects of memory*. London, LTD: Academic Press.

Belmont, J. M. Butterfield, E. C., & Ferretti, R. P. (1982). To secure transfer of training, instruct self-management skills. In D. K. Detterman & R. J. Sternberg (Eds.), *How and how much can intelligence be increased*. Norwood, J. J.: Ablex Pub. Corp.

Bethge, H., Carlson, J. S. & Wiedl, K. H. (1982). The effects of dynamic assessment procedures on Raven matrices performance, visual search behavior, test anxiety, and test orientation. *Intelligence, 6,* 89-97.

Bornstein, P. H., & Hamilton, S. B. (1975). Token rewards and straw men. *American Psychologist, 30,* 780-781.

Bornstein, P. H. & Quevillon, R. P. (1976). The effects of a self-instructional package on overactive preschool boys. *Journal of Applied Behavior Analysis, 9,* 179-188.

Brown, A. L. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. In H. Reese (Ed.), *Advances in child development and behavior: Vol 10*. New York Academic Press.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology: Vol. 1*. Hillsdale, N.J.: Erlbaum.

Brown, A. L., & Barclay, C. R. (1976). The effects of training specific mnemonics on the metamnemonic efficiency of retarded children. Child Development, 47, 71-80.

Brown, A. L., Campione, J. C., & Barclay, C. R. (1979). Training self-checking routines for estimating test readiness: Generalization from list learning to prose recall. Child Development, 50, 501-512.

Brown, A. L. & DeLoache, J. S. (1978). Skills, plans, and self-regulation. In R. S. Siegler (Ed.), Children's thinking: What develops? Hillsdale, N.J.: Erlbaum.

Bryant, W. R. (1983). Preschool children's learning and transfer of matrices problems: A study of proximal development. Unpublished manuscript, University of Illinois.

Camp, B. W., Blom, G. E., Herbert, F., & Van Doorninck, W. J. (1977). "Think aloud": A program for developing self-control in young aggressive boys. Journal of Abnormal Child Psychology, 5, 157-169.

Campione, J. C., & Beaton, V. L. (1972). Transfer of training: Some boundary conditions and initial theory. Journal of Experimental Child Psychology, 13, 94-114.

Campione, J. C., & Brown, A. L. (1973). The role of contextual cues in mediating transfer. Journal of Experimental Child Psychology, 16, 217-224.

Carlson, J. S., & Wiedl, K. H. (1976). The factorial analysis of perceptual and abstract reasoning abilities in tests of concrete opeational thought. Educational and Psychological Measurement, 36, 1015-1019.

Carnine, D. W. (1980). Three procedures for presenting minimally different positive and negative instances. Journal of Educational Psychology, 72, 452-456.

Carnine, D. W., & Becker, W. C. (1982). Theory of instruction: Generalization issues. Educational Psychology, 2, 249-262.

Carnine, D. W., Kameenui, E., & Maggs, A. (1982). Components of analytic assistance: Statement saying, concept training, and strategy training. Journal of Educational Research, 75, 374-377.

Catania, A. C., Matthews, B. A., & Shimoff, E. (1982). Instructed versus shaped human verbal behavior: Interactions with nonverbal responding. Journal of the Experimental Analysis of Behavior, 38, 233-248.

Court, J. H. (1983). Sex differences in performance of Raven's Progressive Matrices: A review. Alberta Journal of Educational Research, 29, 54-74.

Crisafi, M. & Brown, A. L. (1983). Flexible use of an inferential reasoning rule by very young children. Paper presented at the meeting of the Society for Research in Child Development, Detroit.

De Nike, L. D., & Spielberger, C. D. (1963). Induced mediating states in verbal conditioning. Journal of Verbal Learning and Verbal Behavior, 1, 339-345.

Dick, M. L., & Roberts, R. N. (1982). Manipulating antecedent self-instructions to increase response generalization in an academic task. Paper presented at the meeting of the American Psychological Association, Washington, D. C.

Douglas, V. S., Parry, P., Marton, P., & Garson, C. (1976). Assessment of a cognitive training program for hyperactive children. Journal of Abnormal Child Psychology, 4, 389-410.

Engelmann, S. (1971). The effectiveness of direct verbal instruction on IQ performance and achievement in reading arithmetic. In W. C. Becker (Ed.) An empirical basis for change in education. Chicago: Science Research Associates.

Feinberg, H., & Roberts, R. N. (1980). A comparison of self-instructional training procedures in the remediation of academic performance in the LD classroom: Generality of treatment effects. Paper presented at the meeting of the Association for the Advancement Behavior Therapy, New York.

Feinberg, H., & Roberts, R. N. (1982). Teaching academic skills through correspondence-training. Unpublished manuscript, University of North Carolina at Greensboro, Greensboro, North Carolina.

Feinberg, H., & Roberts, R. N. (1984). Teaching academic skills through correspondence training. (A replication). Paper presented at the meeting of the American Psychological Association, Toronto.

Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail, Jr., & J. W. Hagen (Eds.), Perspectives on the development of memory and cognition. New Jersey: Erlbaum.

Gagne, R. M. (1966). Human problem-solving: Internal and external events. In B. Kleinmuntz (Ed.), Problem-solving. New York: Wiley.

Garrity, L. I., & Donoghue, J. T. (1976). Preschool children's performance on the Raven's coloured Progressive Matrices and the Peabody Picture Vocabulary Test. Educational and Psychological Measurement, 36, 1043-1047.

Gholson, B. (1980). The cognitive-developmental basis of human learning studies in hypothesis testing. New York: Academic Press.

Grimm, J. A., Bijou, S. W., & Parsons, J. A. (1973). A problem-solving model for teaching remedial arithmetic to handicapped young children. Journal of Abnormal Child Psychology, 1, 26-39.

Guevremont, D.C., Osnes, P.G., & Stokes, T.K. (1986). Preparation for effective self-regulation: The development of generalized verbal control. Journal of Applied Behavior Analysis, 19, 99-104.

Guthrie, J. T. (1967). Expository instruction versus a discovery method. Journal of Educational Psychology, 58, 45-49.

Hefferline, R. F., & Keenan, B. (1963). Amplitude-induction gradient of a small-scale (covert) operant. Journal of the Experimental Analysis of Behavior, 6, 307-3165.

Higa, W. R., Tharp, R. G., & Calkins, R. P. (1978). Developmental verbal control of behavior: Implications for self-instructional training. Journal of Experimental Child Psychology, 26, 489-497.

Israel, A. C., & O'Leary, K. D. (1973). Developing correspondence between children's words and deeds. Child Development, 44, 575-581.

Kazdin, A. E. (1973). The failure of some patients to respond to token programs. Journal of Behavior Therapy and Experimental Psychiatry, 4, 7-14.

Kendall, C. R., Borkowski, J. G., & Cavanaugh, J. C. (1980). Metamemory and the transfer of an interrogative strategy by EMR children. Intelligence, 4, 255-270.

Kendall, P. C., & Braswell, L. (1982). Cognitive-behavioral self-control therapy for children: A components analysis. Journal of Consulting and Clinical Psychology, 50, 672-689.

Kendall, P. C., & Finch, A. J. (1978). Cognitive-behavioral treatment for impulsivity: A group comparison study. Journal of Consulting and Clinical Psychology, 46, 110-118.

Kendall, P. C., & Wilcox, L. E. (1980). Cognitive-behavioral treatment for impulsivity: Concrete versus conceptual training in nonself-controlled problem children. Journal of Consulting and Clinical Psychology, 48, 80-91.

Kennedy, B. A., & Miller, O. J. (1976). Persistent use of verbal rehearsal as a function of information about its value. Child Development, 47, 566-569.

Kirschenbaum, D. S., & Tomarken, A. JH. (1982). On facing the generalization problem: The study of self-regulatory failure. In P. C. Kendall (Ed.), Advances in cognitive-behavioral research and therapy: Vol. 1. Orlando, Fla: Academic Press.

Lloyd, J. (1980). Academic instruction and cognitive behavior modification: The need for attack strategy training. Exceptional Education Quarterly, 1, 53-63.

Lodico, M. G., Ghatala, E. S., Levin, J. R., Pressley, M., & Bell, J. A. (1983). The effects of strategy-monitoring training on children's selection of effective memory strategies. Journal of Experimental Child Psychology, 35, 263-277.

Loper, A. B. (1980). Metacognitive development: Implications for cognitive training. Exceptional Education Quarterly, 1, 1-8.

Lovitt, T. C., & Curtiss, K. A. (1968). Effects of manipulating an antecedent event on mathematics response rate. Journal of Applied Behavior Analysis, 1, 329-333.

Marks, I. (1978). Behavioral psychotherapy of adult neurosis. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change: An empirical analysis*. New York: Wiley.

Meichenbaum, D. H., & Asarnow, J. (1979). Cognitive-behavior modification and metacognitive development: Implications for the classroom. In P. Kendall & S. Hollon, (Eds.), *Cognitive-behavioral interventions: Theory, research, and procedures,* New York: Academic Press.

Meichenbaum, D. H., & Goodman, J. (1971). Training impulsive children to talk to themselves: A means of developing self-control. *Journal of Abnormal Psychology, 77,* 115-126.

Odom, R. D. & Guzman, R. D. (1972). Development of hierarchies of dimensional salience. *Developmental Psychology, 6,* 271-287.

Offenbach, S. I. (1983). The concept of dimension in research on children's learning. *Monographs of the Society for Research in Child Development, 48,* (6, Serial No. 204).

O'Leary, K. D., & Kent, R. (1973). Behavior modification for social action: research tactics and problems. In L. Hamerlynck, L. Handy, and E. Mash (Eds.), *Behavior change: methodology, concepts and practice*. Champagne, Illinois: Research Press.

O'Leary, S. G. (1980). A response to cognitive training. *Exceptional Education Quarterly, 1,* 89-94.

Palkes, H., Stewart, M., & Freedman, J. (1972). Improvement in maze performance of hyperactive boys as a function of verbal-training procedures. *Journal of Special Education, 5,* 337-342.

Palkes, H., Stewart, M., & Kahana, B. (1968). Porteus maze performance of hyperactive boys after training in self-directed verbal commands. *Child Development, 39,* 817-826.

Richards, D. D. & Siegler, R. S. (1981). Very young children's acquisition of systematic strategies. *Child Development, 52,* 1318-1321.

Risley, T. R., & Hart, B. (1968). Developing correspondence between the non-verbal and verbal behavior of preschool children. Journal of Applied Behavior Analysis, 1, 267-281.

Roberts, R. N. (1979). Private speech in academic problem-solving: A naturalistic perspective. In G. Zivin (Ed.), The development of self-regulation through private speech. New York: Wiley.

Roberts, R. N. & Dick, M. L. (1982). Self-control in the classroom: Theoretical issues and practical applications. In T. R. Kratochwill (Ed.), Advances in school psychology: Vol. 2. Hillsdale, New Jersey: Lawrence Erlbaum.

Roberts, R. N. & Mullis, M. (1980). A component analysis of self-instructional training. Paper presented at the meeting of the Western Psychological Association, Honolulu.

Robin, A. L., Armel, S., & O'Leary, K. D. (1975). The effects of self-instruction on writing deficiencies. Behavior Therapy, 6, 315-323.

Schleser, R., Meyers, A. W., Cohen, R., & Thackwray, D. (1983). Self-instructions with non-self controlled children: Effects of discovery versus faded rehearsal. Journal of Consulting and Clinical Psychology, 51, 954-955.

Schmidtke, A. & Schaller, S. (1980). Comparative study of factor structure of Raven's Coloured Progressive Matrices. Perpetual and Motor Skills, 51, 1244-1246.

Seitz, V., & Weir, M. W. (1971). Strength of dimensional preferences as a predictor of nursery-school children's performance on a concept-shift task. Journal of Experimental Child Psychology, 12, 370-386.

Shure, M. B., & Spivack, G. (1980). Interpersonal problem solving as a mediator of behavioral adjustment in preschool and kindergarten children. Journal of Applied Developmental Psychology, 1, 29-44.

Skinner, B. F. (1966). An operant analysis of problem solving. In B. Kleinmuntz (Ed.), Problem solving. New York: Wiley.

Smiley, S. S. (1972). Instability of dimensional preference following changes in relative cue similarity. Journal of Experimental Child Psychology, 13, 394-403.

Solnick, J. V. & Baer, D. M. (1984). Using multiple exemplars for teaching number-numeral correspondence: Some structural aspects. Analysis and Intervention in Developmental Disabilities, 4, 47-63.

Stokes, T. F. & Baer, D. M. (1977). An implicit technology of generalization. Journal of Applied Behavior Analysis, 10, 349-367.

Stokes, T. F., Baer, D. M., & Jackson, R. L. (1974). Programming the generalization of a operant response in four retarded children. Journal of Applied Behavior Analysis, 7, 599-610.

Suchman, R. G., & Trabasso, T. (1966). Color and form preference in young children. Journal of Experimental Child Psychology, 3, 177-187.

Tennyson, R. D., Youngers, J., & Suebsonthi, P. (1983). Concept learning by children using instructional presentation forms for prototype formation and classification-skill development. Journal of Educational Psychology, 75, 280-291.

Turner, R. R., Hall, V. C., & Grimmett, S. (1973). Effects of familiarization feedback on the performance of lower-class and middle-class kindergartners on the Raven Coloured Progressive Matrices. Journal of Educational Psychology, 65, 356-363.

Valencia, R. R. (1984). Reliability of the Raven Coloured Progressive Matrices for Anglo and for Mexican-American children. Psychology in the Schools, 21, 49-52.

Vygotsky, L. S. (1962). Thought and language. Cambridge, Mass: MIT Press.

Wellman, H. M., Ritter, K., & Flavell, J. H. (1975). Deliberate memory behavior in the delayed reactions of very young children. Developmental Psychology, 11, 780-787.

Winokur, S. (1976). A primer of verbal behavior: An operant view. Englewood Cliffs, New Jersey: Prentice-Hall.

Wozniak, R. (1973). Verbal regulation of motor behavior: Soviet research and non-Soviet replications. Human Development, 15, 13-57.

Table 1

Withdrawn and Participating Children (Mean PPVT Scores in Parentheses).

Children withdrawn because:	Children withdrawn prior to group assignment and first training day.	Children withdrawn after group assignment and first training day.	Total
Inadequate requisite skills (naming colors & shapes, counting).	6 (87.8)	-	6
Failed training criterions (for correct answer or problem-solving).	-	2 (82.5)	2
Pretest ceiling (> 50% correct on maint. or gen. items).	12 (108.5)	-	12
Parents withdrew children from preschool.	1 (99.0)	1 (70.0)	2
Children refused.	0	2 (95.0)	2
Total withdrawn.	19 (101.5)	5 (85.0)	24 (98.1)
Total participating.	-	-	38(101.0)
Percent participating.	-	-	61.3

Table 2

<u>Number of Children Withdrawn from Each Training Group Following Assignment.</u>

Training group:	Multiple Exemplar	Rule Discovery	Problem-Solving Control	Rule Training	Total
Children withdrawn after group assignment and 1st training day because:					
Failed criterion for unprompted rule verbalization.	-	-	-	1	1
Failed criterion for correct answers.	0	1	0	0	1
Parents withdrew children.	0	0	1	0	1
Children refused.	0	0	1	1	2
Total	0	1	2	2	5

Table 3

Assignment of Children to: Groups and Trainers; Groups and Schools, and; Trainers and Schools.

Group x trainer

Trainer: / Group:	ME	RD	PSC	RT	Total
Experimenter	3	4	3	4	14
Trainer C.	3	2	5	2	12
Trainer N.	2	2	1	3	8
Trainer A.	1	1	1	1	4
Total	9	9	10	10	38

Group x school

School: / Group:	ME	RD	PSC	RT	Total
Ulupono	2	1	1	1	5
Nanaikapono	2	2	2	2	8
Nanakuli	2	3	2	2	9
Na Lei	3	3	5	5	16
Total	9	9	10	10	38

Trainer x school

School: / Trainer:	E.	C.	N.	A.	Total
Ulupono	5	0	0	0	5
Nanaikapono	0	0	4	4	8
Nanakuli	5	4	0	0	9
Na Lei	4	8	4	0	16
Total	14	12	8	4	38

Table 4

<u>One-way Analysis of Variance on Children's Peabody Picture Vocabulary Test (PPVT) Scores.</u>

Source	Sum of Squares	df	Mean Squares	F	Significance
Group	319.726	3	106.575	0.995	p>.25
Error	3641.222	34	107.095		

Table 5

Design and Measures.

Group:	Multiple Exemplar	Rule Training	Problem-Solving Control	Rule Discovery
Training Highlights:	Minimal instructions.	Minimal instr.	Minimal instr.	Minimal instr.
	Multiple exemplars.	Multiple exemplars.	Multiple exemplars.	Multiple exemplars.
	Feedback.	Feedback.	Feedback.	Feedback.
	--	Train rules.	--	--
	--	Train problem solving.	Train problem solving.	--
	--	--	--	Prompt rules.

Pretest. 44 items total: 8 maintenance items identical to those of training; 24 generalization items (logical extrapolations of maintenance items); 12 far generalization items (novel matrices of RCPM).

Training day 1: Variants 1 and 1.5 to criterion.

Training day 2: Variant 2 to criterion.

Training day 3: Variant 3 to criterion.

Unprompted posttest (2 days). 44 items total: identical to pretest.

Prompted posttest (2 days). 44 items total: identical to pretest *except* all children were reminded to, "Say and do the things you learned, before you answer," as needed.

Dependent measures.

Outcome: Trials to criterion during training.
Number correct during posttests, for maintenance, generalization, and far generalization items.
Process: Prompted rules during training.
Rules and problem-solving at posttests.
Relationship between rules, problem-solving, and outcome.

Table 6

Description and Position of Variants Comprising Pretest and Posttests.

Item nos.	Name of variant	Properties: ? dimensions uniform within ? axes.
1-4	Testing V-5	Shape-diagonal; Color-diagonal
5-8	Testing V-4	Shape-horizontal; Color-diagonal
9-10	Testing V-1	Size-horizontal
11-12	Testing V-3	Size-diagonal
13-16	Testing V-7	Shape-diagonal; Color-diagonal; Pattern-vertical
17-20	Testing V-6	Shape-horizontal; Color-diagonal; Size-vertical
21-22	Training V-1	Shape-horizontal
23-24	Training V-1.5	Color-vertical
25-36	RCPM, set B	Closure, abstract reasoning.
37-38	Training V-2	Shape-horizontal; Color-vertical
39-40	Training V-3	Color diagonal
41-44	Testing V-2	Size-horizontal; Pattern-vertical

Table 7

Reliability of Coding and Reliability Checks.

Verbal coding categories	Training to criterion (%)		Reliability checks (%)	
(LAB)	2/4	50.0	6/8	75.0
(SOL)	1/4	25.0	(no occurrence)	
(PROB-ap)	6/6	100.0	1/1	100.0
(PROB-ptap)	22/22	100.0	4/5	80.0
(PROB-inap)	74/74	100.0	0/1	0
(SpPROB-ap)	(no occurrence)		(no occurrence)	
(SpPROB-ptap)	28/35	80.0	0/2	0
(SpPROB-inap)	1/5	20.0	0/1	0
(RULE-ap)	15/15	100.0	(no occurrence)	
(RULE-ptap)	17/18	94.4	(no occurrence)	
(RULE-inap)	45/47	95.7	(no occurrence)	
(IRR)	75/88	85.2	160/178	89.9
(INT)	(no occurrence)		(no occurrence)	
Overall:	286/318	89.9	171/196	87.2
Nonverbal coding categories				
(T-ap)	49/58	84.5	2/3	66.7
(T-ptap)	128/135	94.8	3/6	50.0
(T-inap)	117/130	90.0	1/3	33.0
(NP)	216/217	99.5	109/109	100.0
Overall:	510/540	94.4	115/121	95.0

Table 8

One-way (Group) Analysis of Variance on Trials to Criterion During Training and Duncan Multiple Range Tests.

Source	Sum of Squares	df	Mean Squares	F	Significance
Group	1041.710	3	347.237	3.076	p<.05
Error	3838.511	34	112.897		

		RT	PSC	ME	RD			
Group means:		13.50	15.10	24.56	24.89	r	CV.05	CV.10
RT	13.50	--	1.60	11.06**	11.39**	4	10.77	9.01
PSC	15.10		--	9.46*	9.79*	3	10.44	8.72
ME	24.56			--	0.33	2	9.93	8.26
RD	24.89				--			

** p<.05
* p<.10

Table 9

Analysis of Covariance on Maintenance Items of Unprompted Posttest, Adjusted for Pretest, and Duncan Multiple Range Test.

Source	Sum of Squares	df	Mean Squares	F	Significance
Covariate	4.064	1	4.064	0.906	0.348
Group	38.827	3	12.942	2.886	p<.050
Error	147.978	33	4.484		
Total	190.868	37	5.159		

		ME	RD	RT	PSC			
Adjusted means:		4.15	5.46	6.55	6.70	r	CV.10	CV.05
ME	4.15	--	1.31	2.40*	2.56*	4	1.80	2.15
RD	5.46		--	1.09	1.24	3	1.74	2.08
RT	6.55			--	0.15	2	1.65	1.98
PSC	6.70				--			

* $p<.05$

Table 10

Analysis of Covariance on Maintenance Items of Prompted Posttest, Adjusted for Pretest, and Duncan Multiple Range Test.

Source	Sum of Squares	df	Mean Squares	F	Significance
Covariate	6.845	1	6.845	2.197	0.148
Group	31.412	3	10.471	3.360	p<.030
Error	102.822	33	3.116		
Total	141.079	37	3.813		

		ME	RD	RT	PSC			
Adjusted means:		5.06	5.91	7.12	7.30	r	CV.10	CV.05
ME	5.06	--	0.86	2.07*	2.25*	4	1.50	1.79
RD	5.91		--	1.21	1.39	3	1.45	1.74
RT	7.12			--	0.18	2	1.37	1.65
PSC	7.30				--			

* p<.05

Table 11

Two-way (Groups x Trials) Repeated Measures Analysis of Variance on Maintenance Items, and Duncan Multiple Range Tests for Main Effects.

Source	Sum of Squares	df	Mean Squares	F	Significance
Groups	57.732	3	19.244	3.491	p<.026
Error	187.426	34	5.513		
Trials	357.619	2	178.810	99.895	p<.001
Group x trials	19.281	6	3.214	1.795	.113
Error	121.718	68	1.790		

		ME	RD	PSC	RT			
Group means:		3.74	4.52	5.43	5.47	r	CV.10	CV.05
ME	3.74	--	0.78	1.69*	1.73*	4	1.15	1.37
RD	4.52		--	0.91	0.95	3	1.11	1.33
PSC	5.43			--	0.04	2	1.05	1.27
RT	5.47				--			

* p<.05

		Pretest	Unprompted posttest	Prompted posttest			
Trial means:		2.32	5.76	6.39	r	CV.05	CV.001
Pre.	2.32	--	3.44**	4.07**	3	0.65	1.10
Unpr. post.	5.76		--	0.63*	2	0.61	1.06
Prpt. post.	6.39			--			

** p<.001
* p<.05

Table 12

Analysis of Covariance on Generalization Items of Unprompted Posttest, Adjusted for Pretest.

Source	Sum of Squares	df	Mean Squares	F	Significance
Covariate	206.583	1	206.583	13.819	p<.001
Group	47.162	3	15.721	1.052	0.383
Error	493.308	33	14.949		
Total	747.053	37	20.191		

Table 13

Analysis of Covariance on Generalization Items of Prompted Posttest, Adjusted for Pretest, and Duncan Multiple Range Test.

Source	Sum of Squares	df	Mean Squares	F	Significance
Covariate	160.460	1	160.460	12.693	p<.001
Group	179.218	3	59.739	4.725	p<.008
Error	417.190	33	12.642		
Total	756.868	37	20.456		

		ME	RD	PSC	RT				
Adjusted means:		7.15	8.84	11.63	12.88	r	CV.10	CV.05	CV.005
ME	7.15	--	1.69	4.49**	5.74***	4	3.02	3.61	5.25
RD	8.84		--	2.79*	4.04**	3	2.92	3.50	5.12
PSC	11.63			--	1.25	2	2.77	3.33	4.92
RT	12.88				--				

*** p<.005
** p<.05
* p<.10

Table 14

Two-way (Groups x Trials) Repeated Measures Analysis of Variance on Generalization Items, F-tests for Simple Effects, and Duncan Multiple Range Tests for Simple Effects.

Source	Sum of Squares	df	Mean Squares	F	Significance
Groups	300.928	3	100.309	4.024	p<.015
Error	847.563	34	24.928		
Trials	998.791	2	499.395	83.791	p<.001
Groups x trials	90.912	6	15.152	2.542	p<.028
Error	405.281	68	5.960		

Source	Mean Squares	F	Significance
ME	45.444	7.625	p<.01
RD	83.815	14.063	p<.001
PSC	192.533	32.304	p<.001
RT	238.900	40.084	p<.001
Error	5.960		

		Pre.	Unpr. post.	Prpt. post.				
Trial means at ME:		2.44	5.89	6.67	r	CV.10	CV.005	CV.001
Pre.	2.44	--	3.45*	4.23**	3	1.98	3.40	4.01
Unpr. post.	5.89		--	0.78	2	1.87	3.27	3.88
Prpt. post.	6.67			--				

** p<.001

* p<.005

Table 14 (continued)

Two-way (Groups x Trials) Repeated Measures Analysis of Variance on Generalization Items, F-tests for Simple Effects, and Duncan Multiple Range Tests for Simple Effects.

		Pre.	Unpr. post.	Prpt. post.			
Trial means at RD:		2.56	6.89	8.44	r	CV.10	CV.001
Pre.	2.56	--	4.33*	5.88*	3	1.98	4.01
Unpr. post.	6.89		--	1.55	2	1.87	3.88
Prpt. post.	8.44			--			

* p<.001

		Pre.	Unpr. post.	Prpt. post.			
Trial means at PSC:		3.70	10.10	12.10	r	CV.10	CV.001
Pre.	3.70	--	6.40**	8.40**	3	1.98	4.01
Unpr. post.	10.10		--	2.00*	2	1.87	3.88
Prpt. post.	12.10			--			

** p<.001
* p<.10

		Pre.	Unpr. post.	Prpt. post.			
Trial means at RT:		3.50	9.40	13.20	r	CV.005	CV.001
Pre.	3.50	--	5.90**	9.70**	3	3.40	4.01
Unpr. post.	9.40		--	3.80*	2	3.27	3.88
Prpt. post.	13.20			--			

** p<.001
* p<.005

Table 15

Total Number of Generalization Items Correct, Groups x Trials (Critical Values for Chance Responding Yielded by a Large Sample Approximation of the Binomial Test, Appear in Parentheses).

Group:	ME	RD	PSC	RT
Trials				
Pretest	22	23	37	35
Unprompted post.	53*	62*	101*	94*
Prompted post.	60*	76*	121*	132*
CV	(46.1)	(46.1)	(50.7)	(50.7)

* $p<.05$, 1-tailed.

Table 16

Analysis of Covariance on Partial Generalization Scores of Unprompted Posttest, Adjusted for Pretest.

Source	Sum of Squares	df	Mean Squares	F	Significance
Regression	12,979.488	1	12,979.488	3.207	0.083
Constant	103,085.343	1	103,085.343	25.468	p<.001
Group	13,917.720	3	4,639.240	1.146	0.345
Error	133,573.401	33	4047.679		

Table 17

Analysis of Covariance on Partial Generalization Scores of Prompted Posttest, Adjusted for Pretest, and Duncan Multiple Range Test.

Source	Sum of Squares	df	Mean Squares	F	Significance
Regression	1444.071	1	1444.071	0.255	0.617
Constant	255,804.356	1	255,804.356	45.156	p<.001
Group	75,431.444	3	25,143.815	4.439	p<.010
Error	186,941.018	33	5,664.879		

		ME	RD	PSC	RT				
Adjusted means:		419.1	428.3	497.0	524.6	r	CV.10	CV.05	CV.01
ME	419.1	--	9.2	77.9**	105.5***	4	63.85	76.34	101.41
RD	428.3		--	68.7*	96.3**	3	61.79	74.00	98.69
PSC	497.0			--	27.6	2	58.54	70.40	94.63
RT	524.6				--				

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

Table 18

Two-way (Groups x Trials) Repeated Measures Analysis of Variance on Partial Generalization Scores, and Duncan Multiple Range Tests for Main Effects.

Source	Sum of Squares	df	Mean Squares	F	Significance
Groups	64,475.051	3	21,491.684	3.258	p<.033
Error	224,255.204	34	6,595.741		
Trials	244,232.944	2	122,116.472	34.363	p<.001
Group x trials	37,077.858	6	6,179.643	1.739	.125
Error	241,653.230	68	3,553.724		

Group means:		ME 395.04	RD 395.48	RT 436.73	PSC 447.70	r	CV.10	CV.05
ME	395.04	--	0.44	41.69*	52.66**	4	39.75	47.51
RD	395.48		--	41.25*	52.22**	3	38.47	46.05
RT	436.73			--	10.97	2	36.44	43.81
PSC	447.70				--			

** p<.05
* p<.10

Trial means:		Pretest 356.76	Unpr. post. 433.61	Prpt. post. 469.55	r	CV.05	CV.001
Pre.	356.76	--	76.85**	112.79**	3	28.82	48.95
Unpr. post.	433.61		--	35.94*	2	27.40	47.39
Prpt. post.	469.55			--			

** p<.001
* p<.05

Table 19

Analysis of Covariance on RCPM (Set B Items) of Unprompted Posttest, Adjusted for Pretest.

Source	Sum of Squares	df	Mean Squares	F	Significance
Covariate	8.057	1	8.057	1.858	0.182
Group	1.714	3	0.571	0.132	0.940
Error	143.097	33	4.336		
Total	152.868	37	4.132		

Table 20

Analysis of Covariance on RCPM (Set B Items) of Prompted Posttest, Adjusted for Pretest.

Source	Sum of Squares	df	Mean Squares	F	Significance
Covariate	14.411	1	14.411	2.366	0.134
Group	12.803	3	4.268	0.701	0.558
Error	200.997	33	6.091		
Total	228.211	37	6.168		

Table 21

Analysis of Covariance on RCPM (Set B Items 8-12) of Unprompted Posttest, Adjusted for Pretest.

Source	Sum of Squares	df	Mean Squares	F	Significance
Covariate	0.703	1	0.703	1.321	0.259
Group	0.814	3	0.271	0.510	0.678
Error	17.562	33	0.532		
Total	19.079	37	0.516		

Table 22

Analysis of Covariance on RCPM (Set B Items 8-12) of Prompted Posttest, Adjusted for Pretest.

Source	Sum of Squares	df	Mean Squares	F	Significance
Covariate	3.654	1	3.654	4.442	p<.043
Group	0.703	3	0.234	0.285	0.836
Error	27.143	33	0.823		
Total	31.500	37	0.851		

Table 23

Two-way (Groups x Trials) Repeated Measures Analysis of Variance on RCPM Items, and Duncan Multiple Range Test.

Source	Sum of Squares	df	Mean Squares	F	Significance
Groups	34.777	3	11.592	1.576	.213
Error	250.074	34	7.355		
Trials	59.416	2	29.708	11.618	p<.001
Groups x trials	7.610	6	1.268	0.496	.809
Error	173.881	68	2.557		

Trial means:		Pre. 2.58	Unpr. post. 3.76	Prpt. post. 4.32	r	CV.10	CV.005	CV.001
Pre.	2.58	--	1.18*	1.74**	3	0.648	1.113	1.313
Unpr. post.	3.76		--	0.56	2	0.613	1.071	1.271
Prpt. post.	4.32			--				

** p<.001
 * p<.005

Table 24

Frequency of Verbal and Nonverbal Codes for Groups RD and ME During Training.

Group:	RD	ME
Coded verbal		
(LAB)	44	5
(SOL)	14	1
(PROB-ap)	0	0
(PROB-ptap)	49	4
(PROB-inap)	4	0
(SpPROB-ptap)	84	0
(RULE-ap)	0	0
(RULE-ptap)	0	0
(RULE-inap)	0	0
(IRR)	317	222
Trials to criterion:	285	221
Trials silent:	0	191
Coded nonverbal		
(T-ap) or (T-ptap)	42	2
(T-inap) or (NP)	243	219

Table 25

Most Frequent Verbalizations Prompted from each Child in the RD Condition During Training (Codes in Parentheses).

Child		Verbalization
Brent	(S-10)	This one (IRR); Listen to my teacher (IRR).
Jan	(S-11)	This one match this one (SpPROB-ptap).
Kimo	(S-12)	This one (IRR); Three blue, three red, three yellow (PROB-ptap).
Karen	(S-13)	You have to put it down (IRR).
Lani	(S-14)	Match 'em with this one (SpPROB-ptap).
Dan	(S-15)	Find the right one (SOL).
Kenni	(S-16)	Three circles, three squares, three stars (PROB-ptap).
Joan	(S-17)	From saying it (IRR).
Della	(S-18)	Three blues, three greens, three reds (PROB-ptap).

Table 26

Frequency of Verbal Problem-solving (PROB-ap, -ptap, or -inap) and Nonverbal Problem-solving (T-ap, -ptap, or -inap) for RT and PSC Children at Unprompted and Prompted Posttests.

Group		Unprompted posttest Problem-solving Verbal	Unprompted posttest Problem-solving Nonverbal	Prompted posttest Problem-solving Verbal	Prompted posttest Problem-solving Nonverbal
PSC	(S-19)	12	12	33	26
	(S-20)	6	6	15	15
	(S-21)	3	3	43	26
	(S-22)	30	29	3	4
	(S-23)	20	11	13	11
	(S-24)	14	14	42	42
	(S-25)	8	10	8	8
	(S-26)	0	0	0	0
	(S-27)	19	11	33	26
	(S-28)	22	23	33	31
	$\bar{X}$ =	13.4	11.9	22.3	18.9
RT	(S-29)	0	0	8	8
	(S-30)	0	0	0	0
	(S-31)	0	2	43	30
	(S-32)	0	3	44	31
	(S-33)	0	0	43	43
	(S-34)	0	1	1	1
	(S-35)	12	12	25	25
	(S-36)	10	9	17	16
	(S-37)	7	8	26	23
	(S-38)	16	15	27	25
	$\bar{X}$ =	4.5	5.0	23.4	20.2

Table 27

Frequency of: Rules; Specification of Problem-solving; and Specification of Outcomes, Combined (RULE-ap, -ptap, -inap; SpPROB-ptap; & SOL), for RT and PSC Children at Unprompted and Prompted Posttests.

Group		Unprompted posttest	Prompted posttest
PSC	(S-19)	6	2
	(S-20)	0	0
	(S-21)	0	3
	(S-22)	4	1
	(S-23)	41	61
	(S-24)	0	0
	(S-25)	0	0
	(S-26)	0	0
	(S-27)	3	0
	(S-28)	0	0
		$\bar{X}$= 5.4	6.7
RT	(S-29)	0	8
	(S-30)	0	0
	(S-31)	12	42
	(S-32)	0	42
	(S-33)	0	43
	(S-34)	0	2
	(S-35)	9	25
	(S-36)	10	22
	(S-37)	6	38
	(S-38)	16	23
		$\bar{X}$= 5.3	24.5

Table 28

Frequency of: Rules; Specification of Problem-solving; and Specification of Outcomes, Combined (RULE-ap, -ptap, -inap; SpPROB-ptap; & SOL), for ME and RD Children at Unprompted and Prompted Posttests.

Group			Unprompted posttest	Prompted posttest
ME	(S-1)		0	0
	(S-2)		0	0
	(S-3)		0	0
	(S-4)		0	0
	(S-5)		0	0
	(S-6)		0	0
	(S-7)		0	0
	(S-8)		0	0
	(S-9)		0	0
		$\bar{X}=$	0	0
RD	(S-10)		1	123
	(S-11)		0	0
	(S-12)		0	0
	(S-13)		0	0
	(S-14)		0	43
	(S-15)		0	0
	(S-16)		0	3
	(S-17)		0	0
	(S-18)		0	0
		$\bar{X}=$	0.11	18.78

Table 29

Frequency of Verbal Problem-solving (PROB-ap, -ptap, & -inap) for ME and RD Children at Unprompted and Prompted Posttests.

Group			Unprompted posttest	Prompted posttest
ME	(S-1)		0	0
	(S-2)		0	1
	(S-3)		1	4
	(S-4)		0	0
	(S-5)		2	1
	(S-6)		0	0
	(S-7)		0	0
	(S-8)		0	0
	(S-9)		0	0
		$\bar{X}$=	0.33	0.67
RD	(S-10)		1	1
	(S-11)		0	0
	(S-12)		0	1
	(S-13)		0	1
	(S-14)		0	0
	(S-15)		0	0
	(S-16)		20	17
	(S-17)		0	0
	(S-18)		0	0
		$\bar{X}$=	2.33	2.22

Table 30

Overall Chi-square and Partitionings of Chi-square on Appropriateness of Verbalizations x Outcome. Data for RT and PSC Children Combined, and each Posttest Combined, excluding RCPM Items (Expected Means in Parentheses).

	Verbalizations			
Outcome	Appropriate	Partially appropriate	Inapprop./ Irrelevant	Total
Example correct:	154 (106.2)	175 (168.6)	375 (429.2)	704
Example incorrect:	33 (80.8)	122 (128.4)	381 (326.8)	536
Total:	187	297	756	1240

	Source	df	x2	Significance
Overall:	Ap. vs. Ptap. vs. Inap.	2	66.254	p<.005
Partition I:	Ptap. vs. Inap.	1	7.546	p<.01
	(Ptap.--Inap.) vs. Ap.	1	58.708	p<.005
Partition II:	Ap. vs. Ptap.	1	25.669	p<.005
	(Ap.--Ptap.) vs. Inap.	1	40.585	p<.005

Table 31

Overall Chi-square and Partitioning of Chi-square on Appropriateness of Tracking x Outcome. Data for RT and PSC Children Combined, and each Posttest Combined, excluding RCPM Items (Expected Means in Parentheses).

	Tracking				
	Appropriate	Partially Approp.	Inapprop.	No Pointing	Total
Outcome					
Example correct:	161(101.1)	157(164.8)	31(40.8)	362(404.3)	711
Example incorrect:	15 (74.9)	130(122.2)	40(30.2)	342(299.7)	527
Total:	176	287	71	704	1238

Source	df	x2	Significance
Overall: Ap vs. Ptap vs. Inap vs. NP	3	100.234	p<.005
Partition: Inap vs. NP	1	1.588	p>.10
Ptap vs. (Inap--NP)	1	1.367	p>.10
Ap vs. (Ptap--Inap--NP)	1	97.275	p<.005

Table 32

Correct Examples x Appropriateness of Verbalizations, with Data for RT and PSC Children Combined, and both Posttests Combined, Excluding RCPM Items. Critical Values for Chance Responding, Yielded by a Large Sample Approximation of the Binomial Test, are in Parentheses.

	Verbalizations		
Outcome	Appropriate	Partially Appropriate	Inappropriate/ Irrelevant
Example correct:	154*	175*	375*
Example incorrect:	33	122	381
Total	187	297	756
CV	(41.9)	(63.8)	(152.0)

* p<.05, 1-tailed

Table 33

<u>Correct Examples x Appropriateness of Tracking, with Data for RT and PSC Children Combined, and Both Posttests Combined, Excluding RCPM Items. Critical Values for Chance Responding, Yielded by a Large Sample Approximation of the Binomial Test, are in Parentheses.</u>

	Tracking			
Outcome	Appropriate	Partially Appropriate	Inappropriate	No Pointing
Example correct:	161*	157*	31*	362*
Example incorrect:	15	130	40	342
Total	176	287	71	704
CV	(39.7)	(61.8)	(18.0)	(142.1)

* p<.05, 1-tailed

Table 34

Proportion of Maintenance Items, Generalization Items, and RCPM Items on which Children Problem-solved Verbally (PROB-ap,-ptap, or -inap) and Emitted Rules (RULE-ap,-ptap, or -inap). Data Examined for Groups RT and PSC at Both Posttests.

	Items	Unprompted Posttest	Prompted Posttest
Rules: RT group	Maint.	.050 (4/80)	.613 (49/80)
	Gen.	.158 (38/240)	.613 (147/240)
	RCPM	0 (0/120)	.417 (50/120)
	All	.095 (42/440)	.559 (246/440)
Verbal Problem-Solving: RT & PSC combined.	Maint.	.281 (45/160)	.569 (91/160)
	Gen.	.267 (128/480)	.608 (292/480)
	RCPM	.008 (2/240)	.288 (69/240)
	All	.199 (175/880)	.514 (452/880)
Verbal Problem-Solving: Group RT only.	Maint.	.100 (8/80)	.575 (46/80)
	Gen.	.167 (40/240)	.625 (150/240)
	RCPM	0 (0/120)	.325 (39/120)
	All	.109 (48/440)	.534 (235/440)
Verbal Problem-Solving: Group PSC only.	Maint.	.463 (37/80)	.563 (45/80)
	Gen.	.367 (88/240)	.592 (142/240)
	RCPM	.017 (2/120)	.250 (30/120)
	All	.289 (127/440)	.493 (217/440)

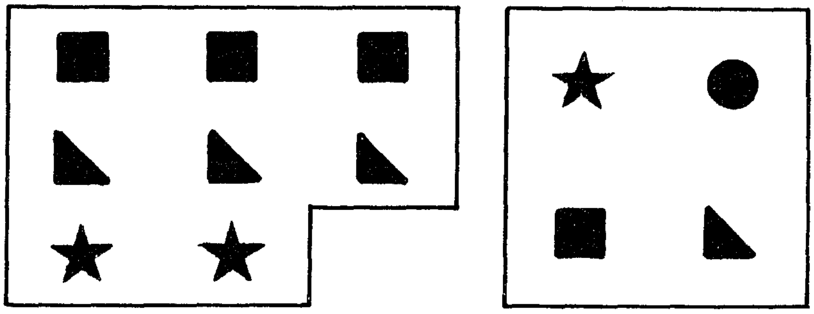
Figure 3

Examples of Variants Used in Training and Testing.

Training

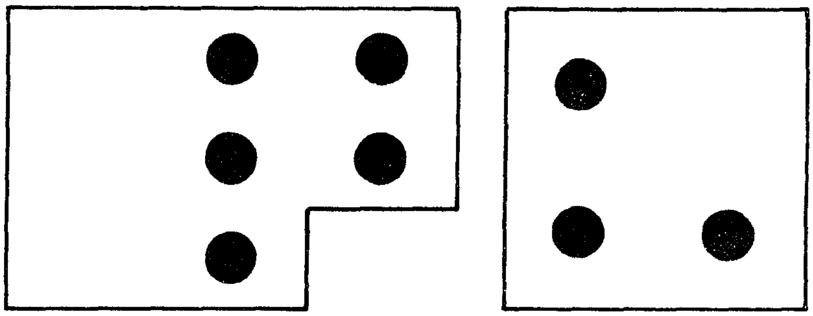
V-1

Shape-horizontal



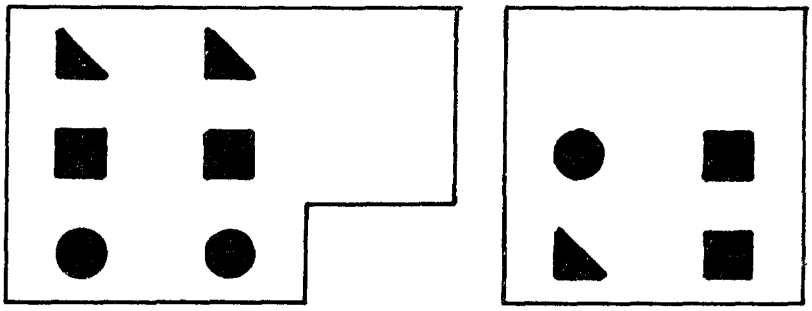
V-1.5

Color-vertical



V-2

Shape-horizontal
Color-vertical



V-3

Color-diagonal

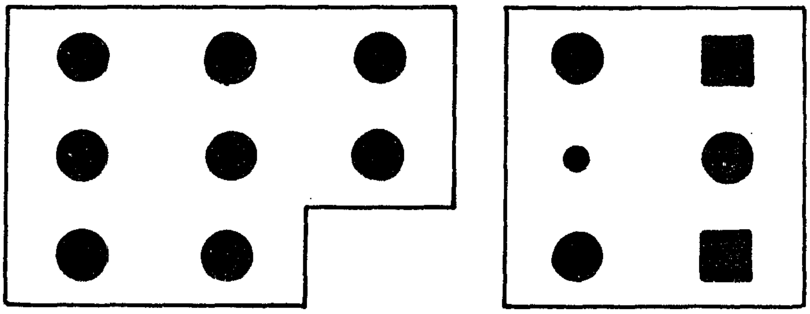


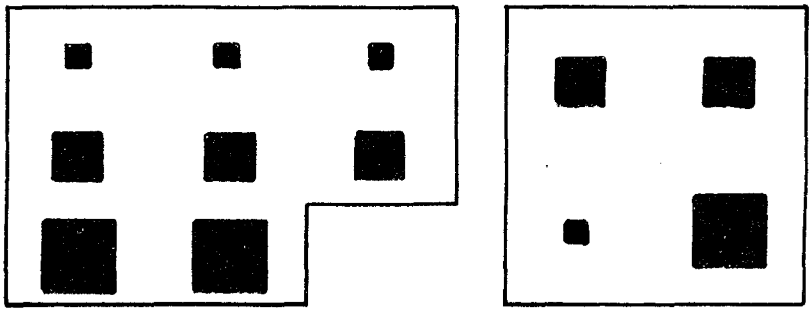
Figure 3 (continued)

Examples of Variants Used in Training and Testing.

Testing

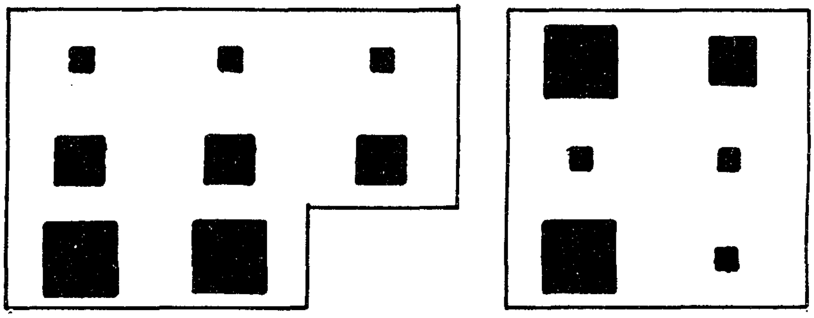
V-1

Size-horizontal



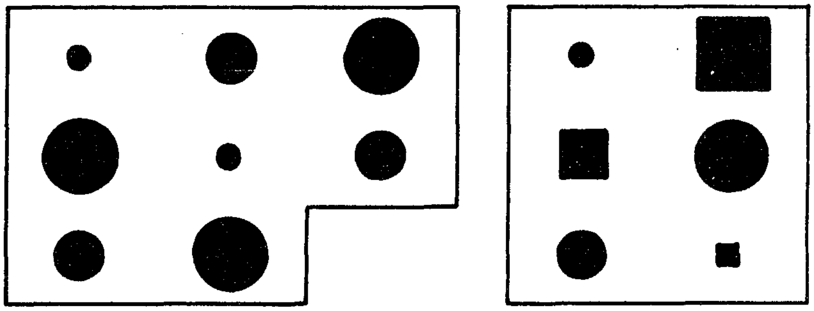
V-2

Size-horizontal
Pattern-vertical



V-3

Size-diagonal



V-4

Shape-horizontal
Color-diagonal

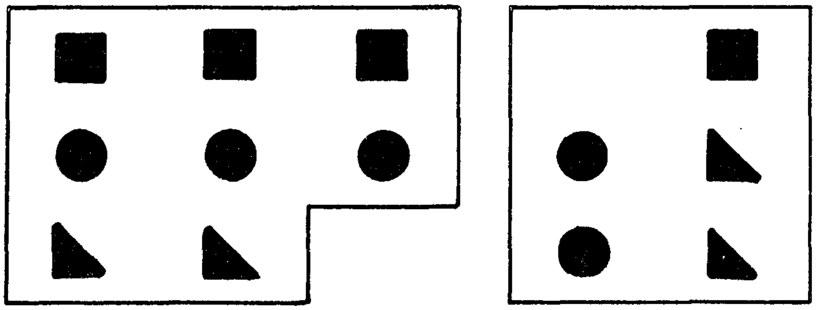


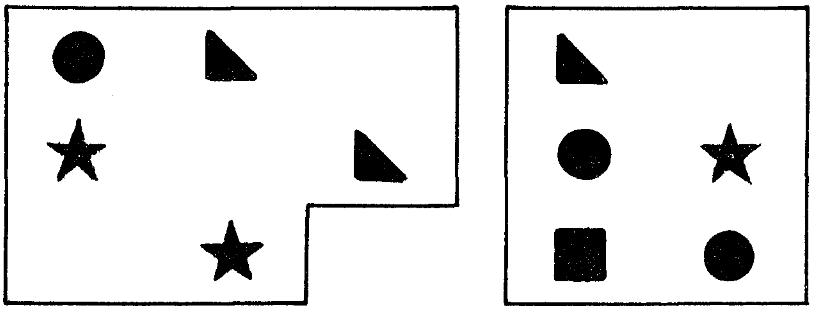
Figure 3 (continued)

Examples of Variants Used in Training and Testing.

Testing

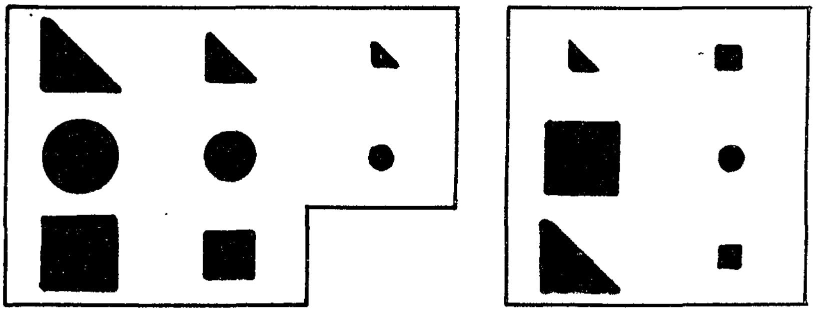
V-5

Color-diagonal
Shape-diagonal



V-6

Shape-horizontal
Size-vertical
Color-diagonal



V-7

Shape-diagonal
Color-diagonal
Pattern-vertical

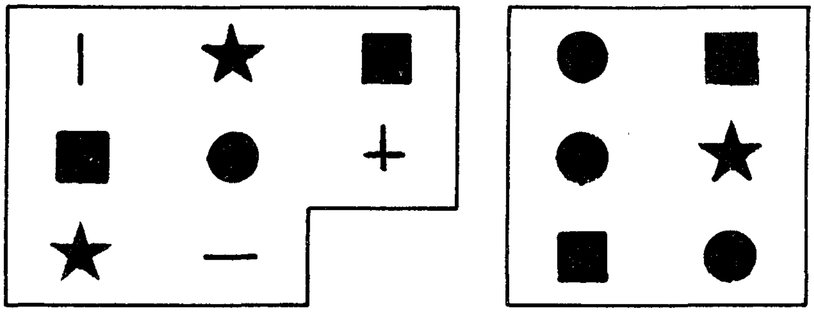


Figure 4

Mean Trials to Criterion During Training, as a Function of Training Group (V-1 through V-3 Combined).

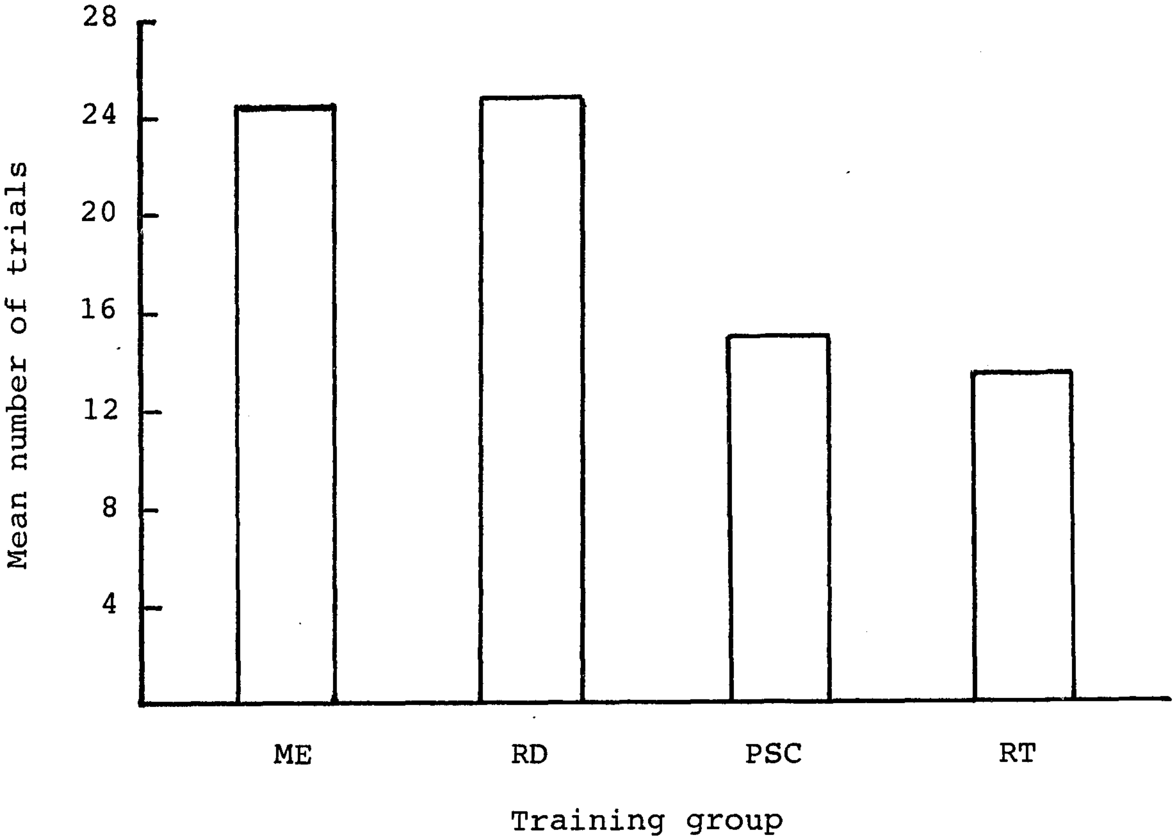


Figure 5

Mean Number Correct on Maintenance Items of Unprompted Posttest, as a Function of Training Group, Adjusted for Pretest.

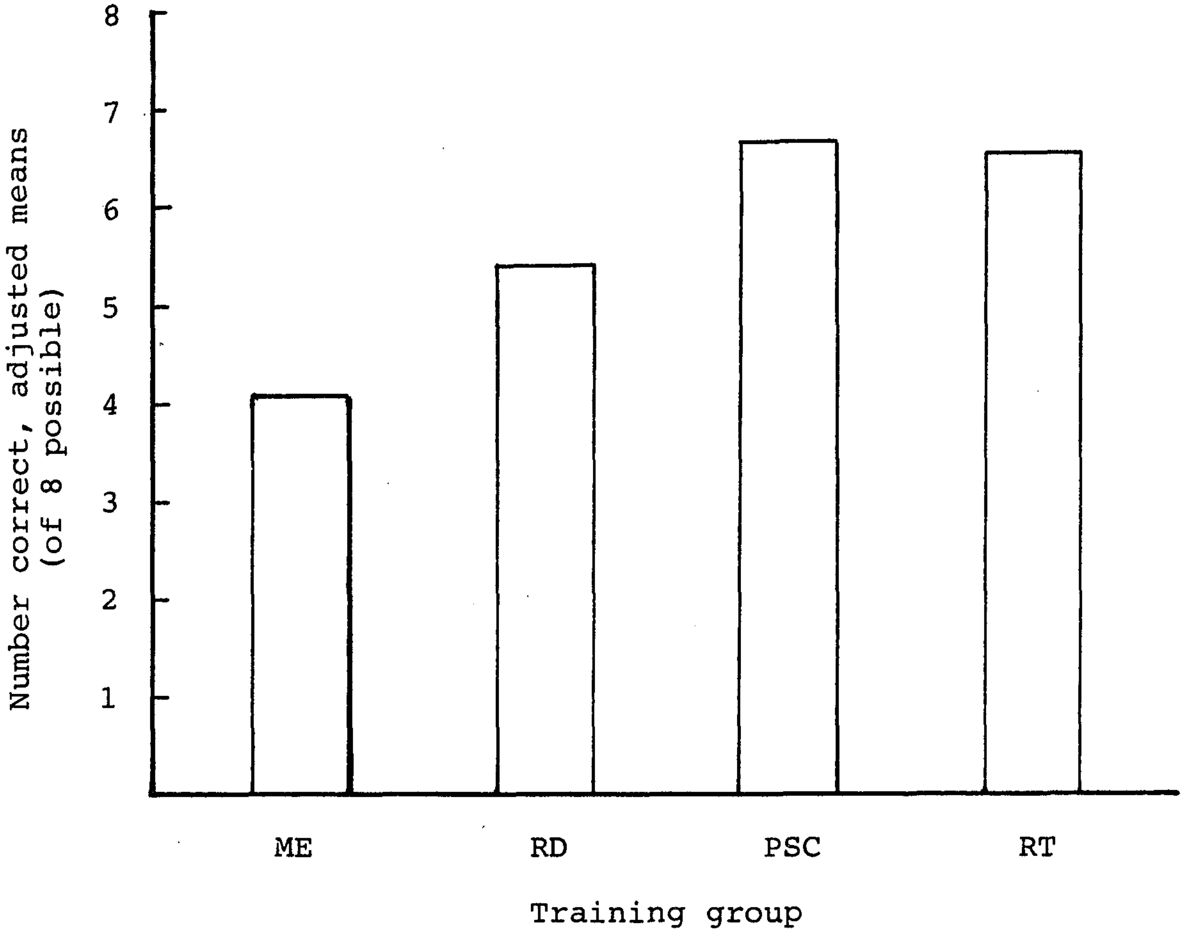


Figure 6

Mean Number Correct on Maintenance Items of Prompted Posttest, as a Function of Training Group, Adjusted for Pretest.

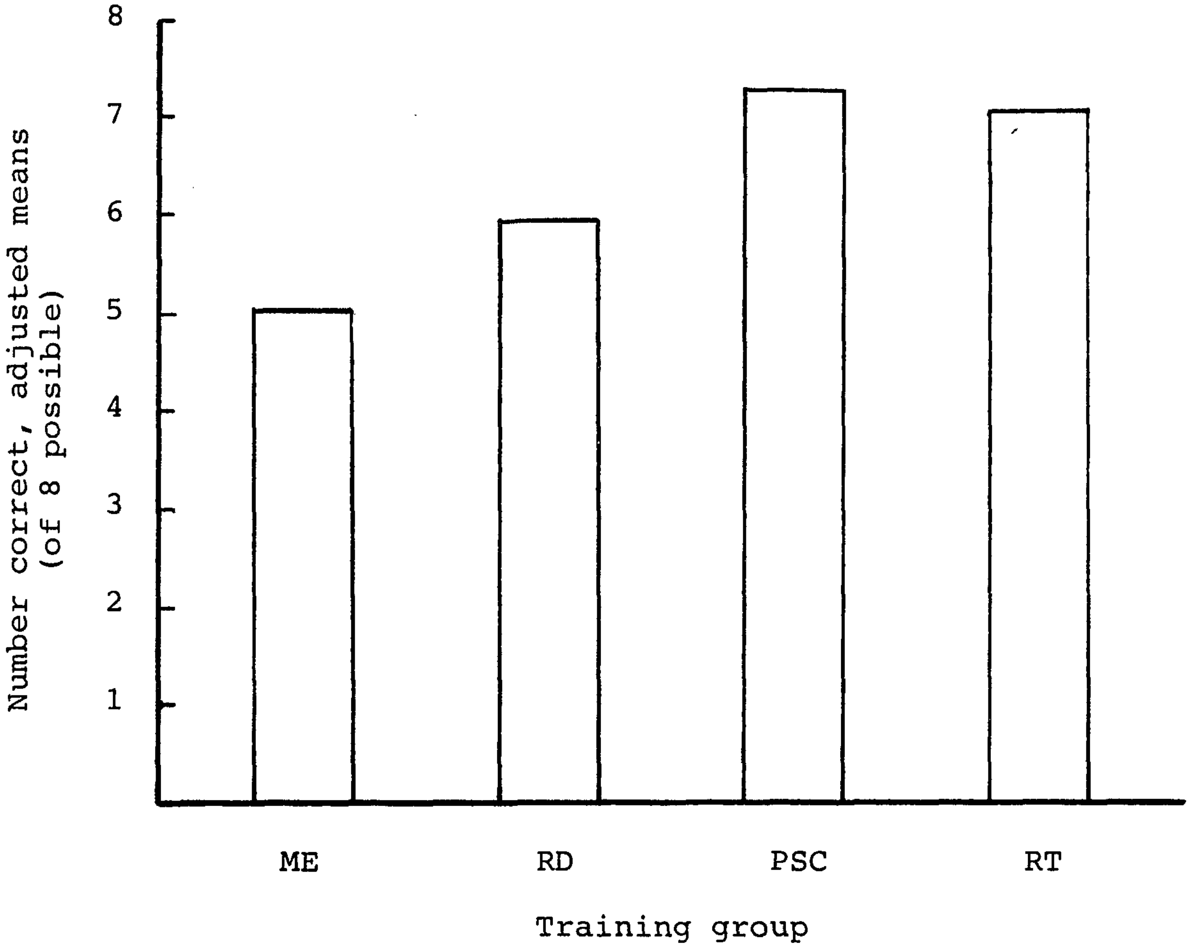


Figure 7

Mean Number Correct on Maintenance Items as a Function of Training Group, and as a Function of Trials.

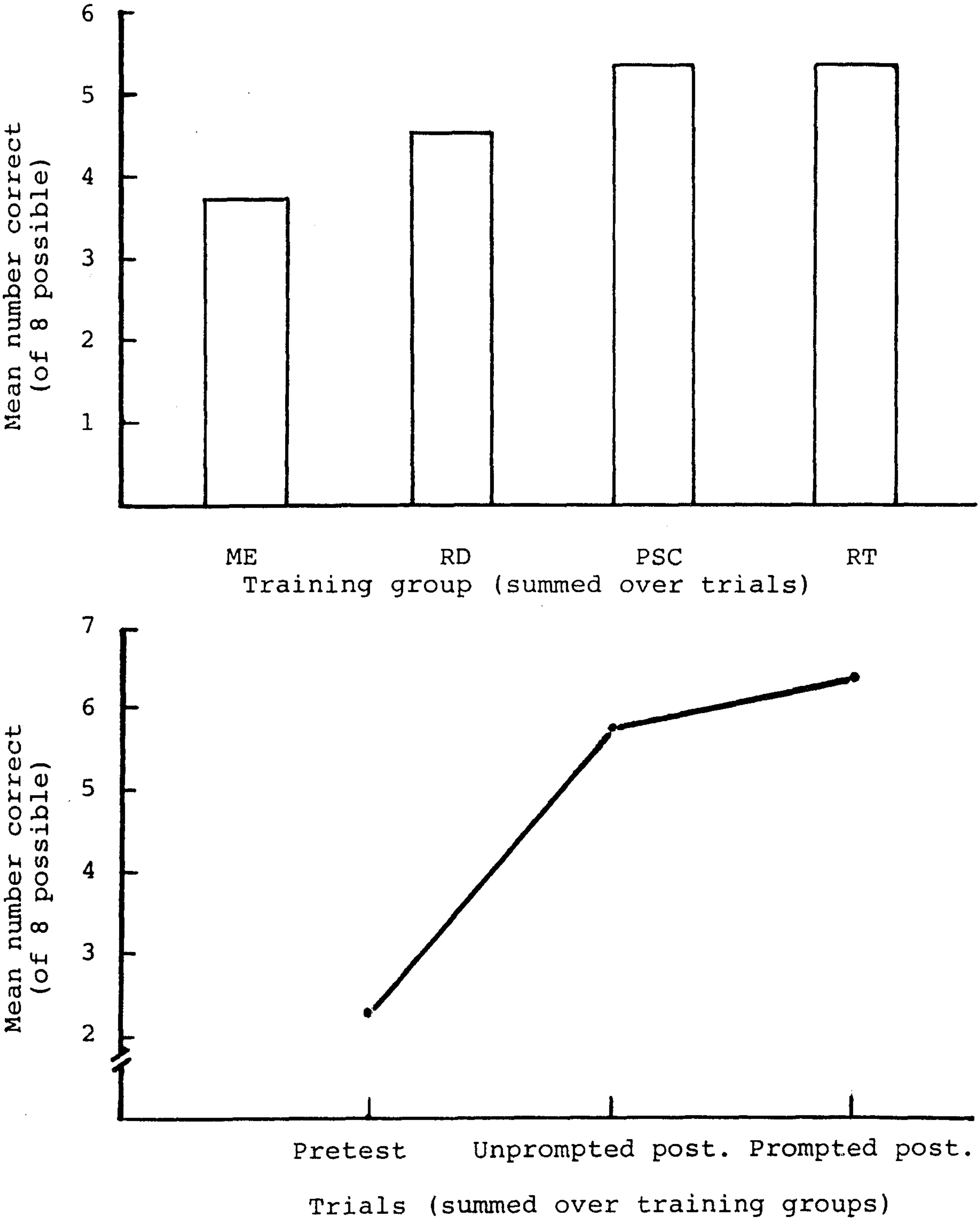


Figure 8

Mean Number Correct on Generalization Items of Prompted Posttest, as a Function of Training Group, Adjusted for Pretest.

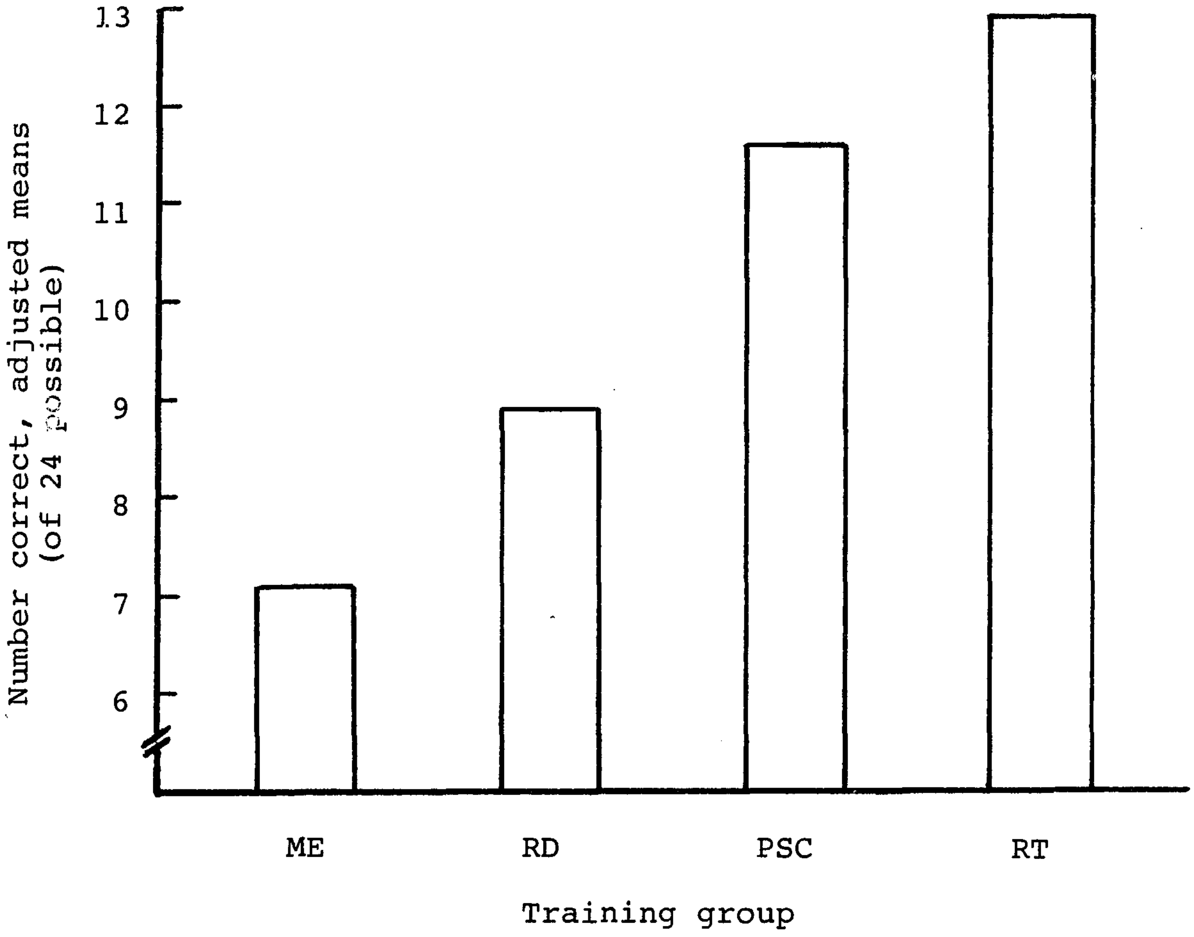


Figure 9

Mean Number Correct on Generalization Items as a Function of Training Group and Trials: Interaction Effect.

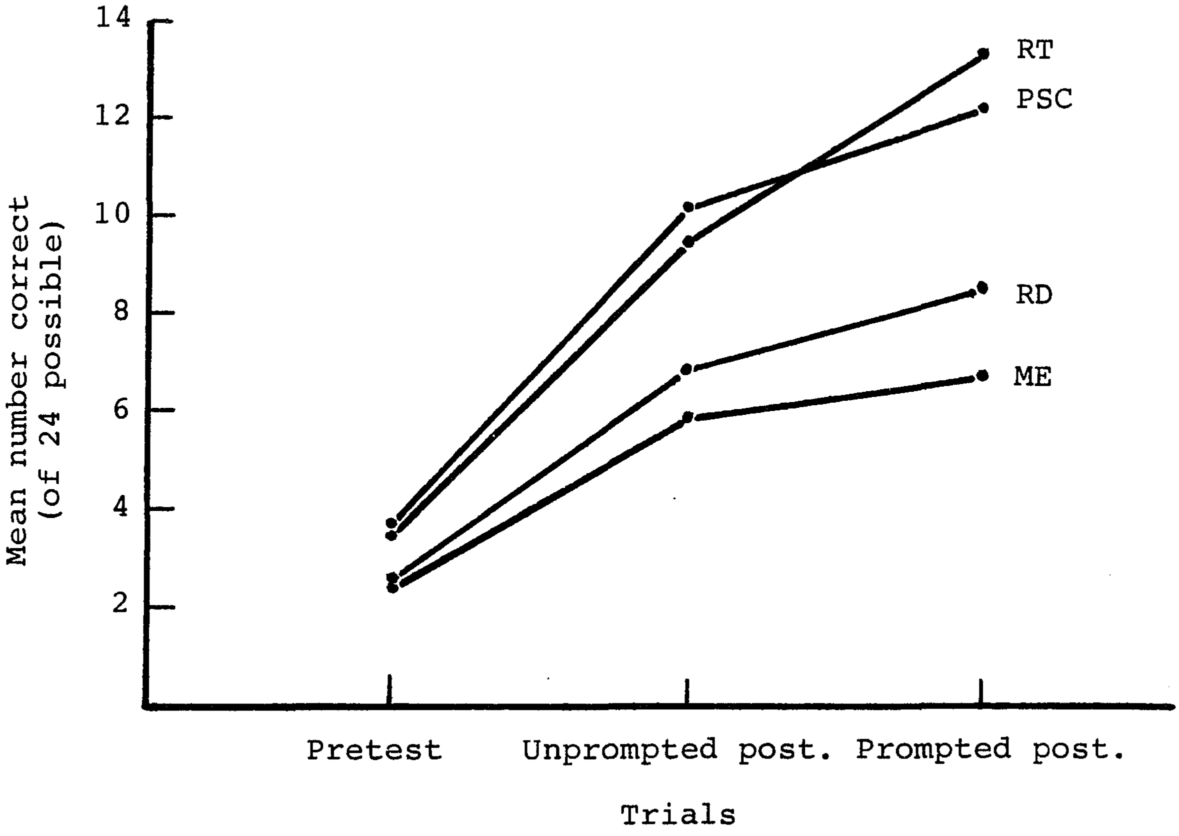


Figure 10

Mean Percent Correct on Partial Gen. Scores of Prompted Posttest, as a Function of Training Group, Adjusted for Pretest.

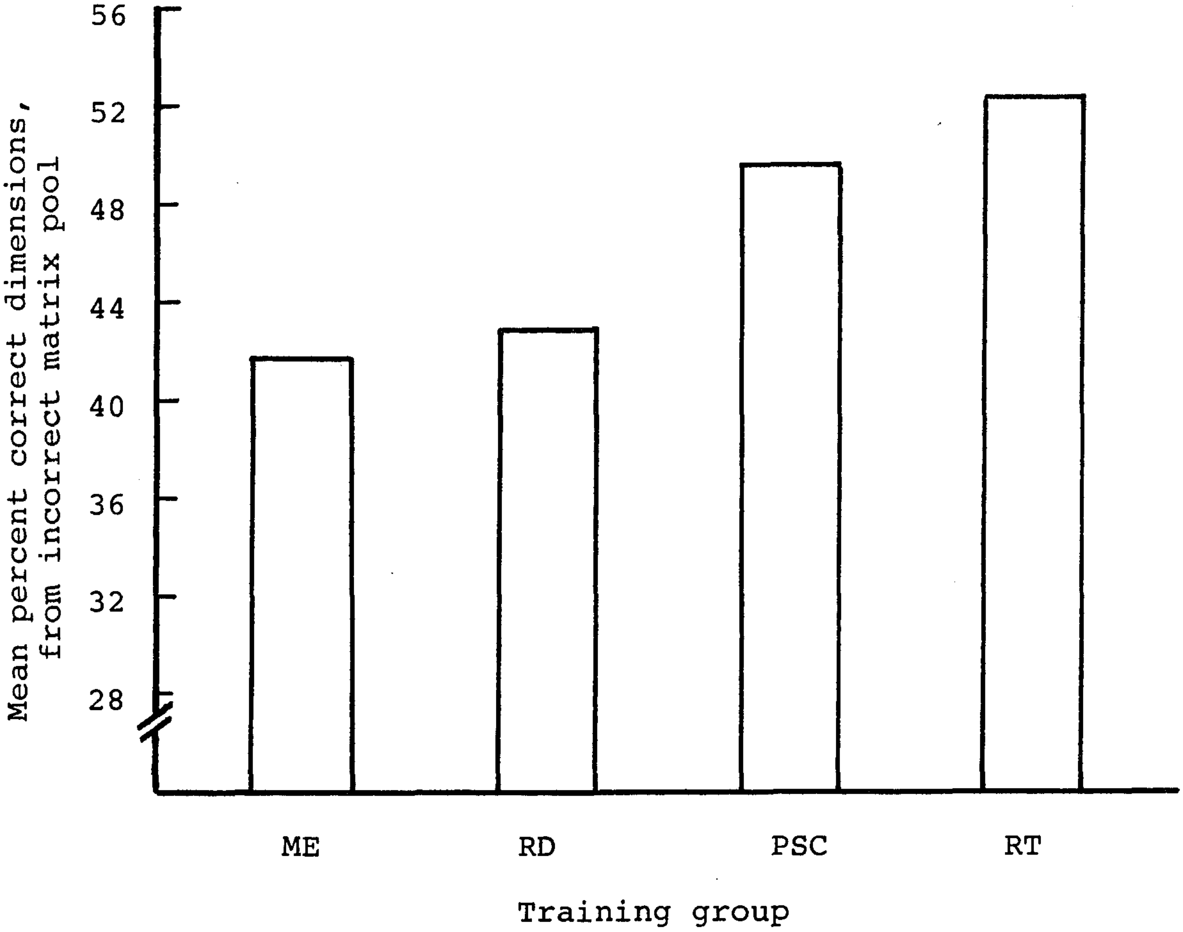


Figure 11

Mean Percent Correct on Partial Gen. Scores, as a Function of Training Group, and as a Function of Trials.

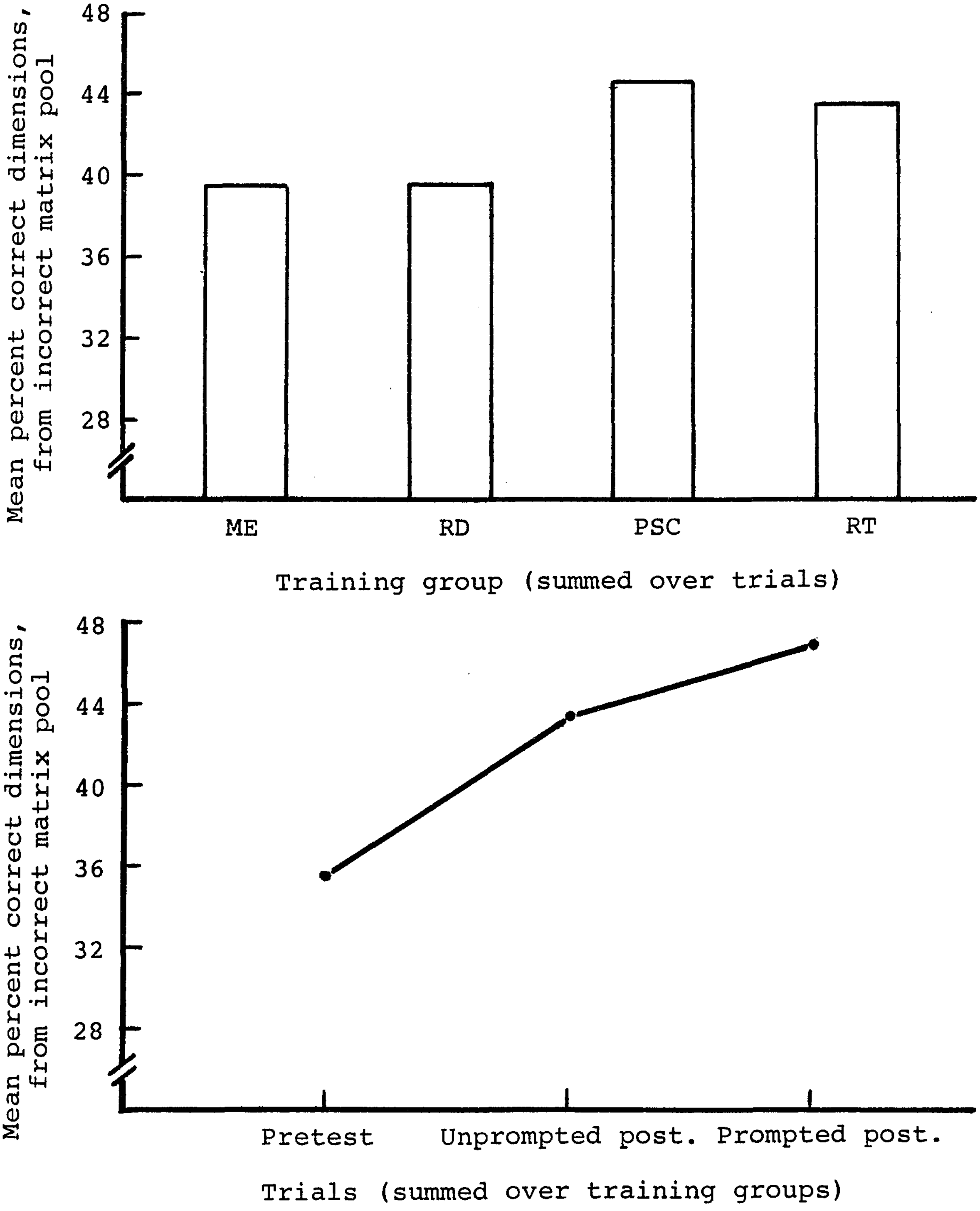
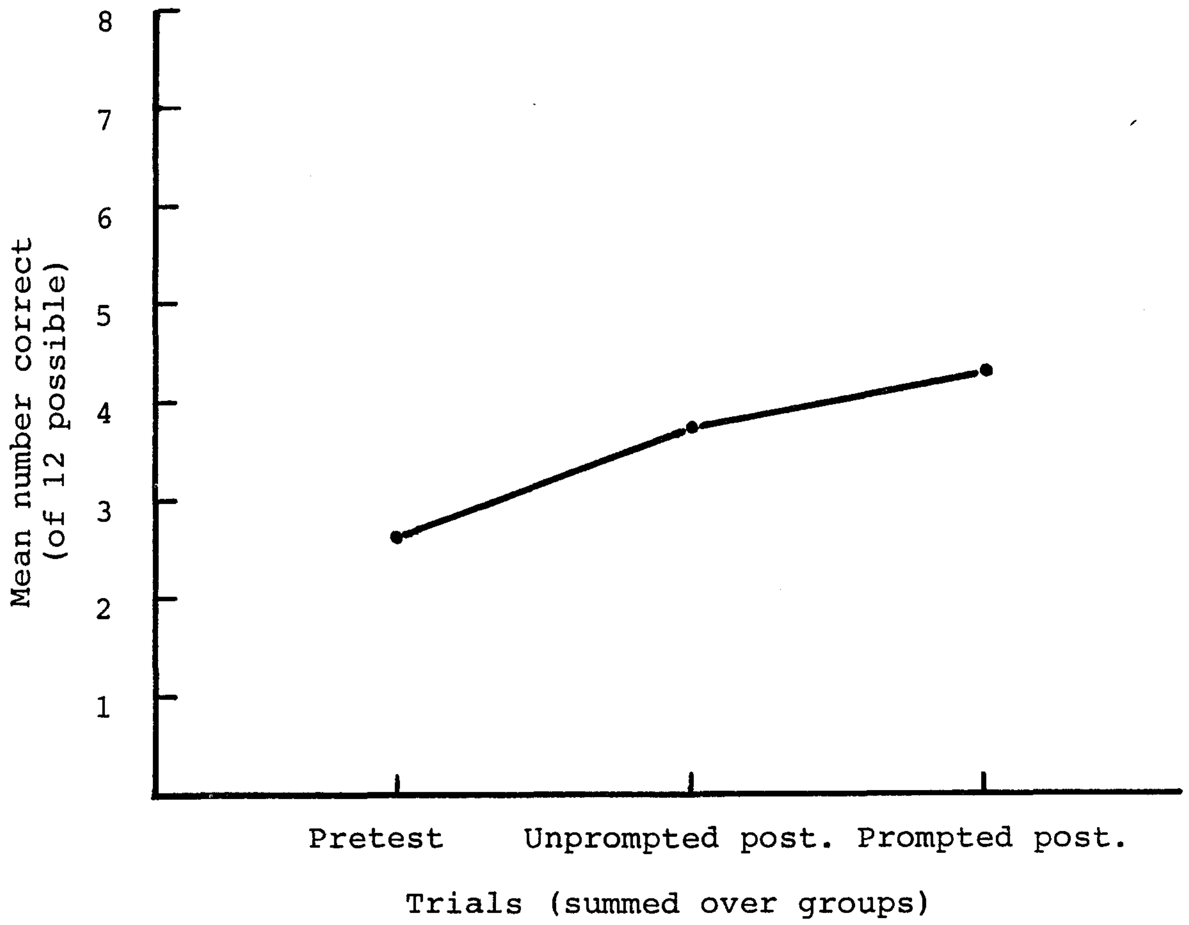


Figure 12

Mean Number Correct on RCPM Set B Items, as a Function of Trials



Appendix A

Pretest/Posttest and Training Protocols

Pretest

Today, we're going to see how you do with some puzzles. Some of the puzzles are easy. Many of the puzzles are very hard, so it's O.K. if you don't get them all right. All I want you to do is try your best, O.K.? Later on, we're going to give you some hints to do them. I'm not going to tell you today if you're right or wrong--I just want to see how well you do them.

O.K., let's start. (T. takes first matrix and solution set and lays it out before the child). I'm going to show you puzzles with a piece missing like this one. Here are some pieces and you need to figure out which one finishes the puzzle--which one makes the most sense. Only one of the pieces here (T. gestures) is the right one.

Pretest and Posttest

(Before each trial, T. says) Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here that goes here (T. points).

Do not offer praise after the child selects his answer, right or wrong. Occasional "O.K."'s noncontingently are permissible. "Let's try the next one," is preferred.

Deliver material reinforcement immediately prior to child response, contingent upon on-task behavior only. Material reinforcement may be earned on alternate trials only.

If child appears discouraged, you may repeat, "These are real hard, just try your best," as needed.

ME Training Protocols

PRODUCT: The essence of this condition is to teach children by providing feedback only on each trial. The child is given no training in verbal or nonverbal problem-solving, or rule verbalization.

The trainer accomplishes this by doing the following on all training trials:

PROCESS: (1) Trainer asks child to look at all stimuli in matrix and solution set.

(2) Trainer delivers social and material reward noncontingently, and immediately before the child selects an answer.

(3) Trainer praises correct answer, corrects incorrect, and "explains" correct choice.

For example:

First trial only:

T. (Child's name), now we're going to begin helping you do these problems. When you do them, you'll get a chip. At the end of the day, you can trade in your chips for the prizes we showed you before.

All trials:

T. O.K., (child's name). Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that goes here (T. points). Before the child responds, "Good! You're working really hard and here's a token," (T. puts a token in the child's cup). Now what's your answer? (if necessary).

C. Child selects answer.

T. (If correct) That's right! That's the right answer! Very good! That's right because 3 yellows, 3 greens, and 3 blues (T. gestures).

T. (If incorrect) Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are 3 yellows, 3 greens, and 3 blues (T. gestures).

RD Training Protocols

PRODUCT: The essence of this condition is to teach children on each trial to verbally identify what they do as they attempt each task. The idea is to strike some balance between:

(1) Repeatedly prompting the child on a given trial to verbalize an appropriate rule, and;

(2) Accepting on a given trial any verbal response the child offers.

The trainer accomplishes this by doing the following on all training trials:

PROCESS: (1) Trainer asks child to look at all stimuli in matrix and solution set.

(2) Trainer prompts child to verbalize what he needs to do to solve the matrix:

T. But first, tell me, how are you going to do this?

or, T. But first, tell me what you have to do to get the right answer?

(3) Trainer delivers praise enthusiastically contingent upon child's verbalizing the:

(a) contingency (e.g., count the shapes to get it right);

(b) specification of a problem-solving response (e.g., count the shapes);

(c) problem-solving (e.g., 'cause 3 green, 3 red & 2 blue).

T. delivers material reinforcement, prompts child to answer, and gives feedback.

(4) If the child verbalizes:

(d) the relevant stimulus only (e.g., the square one)

(e) anything else (e.g., just do it; I don't know),

then T. offers a follow-up prompt:

T. What makes you think that's the right answer?

(5) Again, if the child verbalizes the contingency, specifies a p.s. response, or verbally problem-solves, praise enthusiastically and reinforce materially.

(6) If the child verbalizes anything else, offer no more follow-up prompts and materially reinforce. Say, "Here's a token for trying to tell me how to do it."

(7) Trainer praises correct answer, corrects incorrect, and explains correct choice.

Keep in mind:
Material reinforcement for verbalizing, without regard for content.

Strong social reinforcement contingent upon verbalizing the contingency, specifying a problem-solving response, or verbal problem-solving.

Maximally one introductory prompt to verbalize and one follow-up prompt per trial.

This is illustrated by the following examples:

First trial:

T. (Child's name), now we're going to begin helping you do these problems. I want to see if you can tell me how you're doing them. When you tell me how you're doing them, you'll get a chip. At the end of the day, you can trade in your chips for the prize we showed you before.

(Child's name), look at all of these (T. gestures) and all of these (T. gestures) and find the one from here that goes here (T. points). But first tell me what you have to do to get the right answer?

C. It's this one.

T. What makes you think that's the right answer?

C. I know it.

T. Here's a token for trying to tell me how to do it (puts token in cup). Now what's your answer?

C. (Child responds).

T. (If correct), That's right! That's the right answer! Very good! That's right because 3 reds, 3 yellows, and 3 blues (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are 3 reds, 3 yellows, and 3 blues (T. gestures).

Next trial:

T. (Child's name), look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that belongs here (T. points). But first tell me, what do you have to do to get this right?

C. This one.

T. What makes you think that's the right answer?

C. Because 2 yellows and you need one more.

T. Very good! That's terrific! You did a great job telling me how to do it, and here's your token (puts token in cup). Now what's your answer?

C. (Child responds).

T. (If correct) That's right! That's the right answer! Very good! That's right because 3 greens, 3 reds, and 3 yellows (T. gestures).

(If incorrect) Good try, but that's not the right answer. This is the right one. (T. places correct answer in its spot) because there are 3 greens, 3 reds, and 3 yellows.

PSC Training Protocols

PRODUCT: The essence of this condition is to teach children on each trial to count, label, and touch values of each relevant condition, prior to selecting an answer.

Over the course of training, children should learn to count values of relevant dimensions (identifying those relevant), independent of aid from the trainer.

The trainer accomplishes this by doing the following on all training trials:

PROCESS:

(1) Trainer asks child to look at all stimuli in matrix and solution set.

(2) Trainer models, as needed, problem-solving: counting and labeling values of relevant stimulus dimensions while touching stimuli in a corresponding manner.

(3) Trainer prompts child, as needed, to problem-solve.

(4) Trainer delivers social and material reward contingent upon counting/ labeling/ tracking regardless of aid required, and then encourages child to select answer.

(5) Trainer praises correct answer, corrects incorrect, and explains correct choice.

Over the course of training trials, models and prompts are faded. Instead of simply providing the child with what he needs to do and say on each new trial, the trainer tests to see if the child can generate required responses without aid, and if he cannot, he leads the child toward required responses.

The procedure, including the fading element, is illustrated by the following examples:

First trial:

T. (Child's name), now we're going to begin teaching you some things to do to help you with these problems. When you do these things, you'll get a token. At the end of the day, you can trade in your chips for the prizes we showed you before.

O.K., (child's name). Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here that goes here (T. points). But first I want you to do this:

(T. gestures) One, two, three yellows. One, two, three greens. One, two blues, I need a blue. Now you try.

C. N.R.

T. (tracking) 1, 2, 3 yellows.

C. 1, 2, 3 yellows (tracking).

T. Good! (C. pauses). 1, 2, 3 greens (tracking).

C. 1, 2, 3 greens (tracking).

T. Good! (C. pauses). 1, 2 blue. I need a blue (tracking).

C. 1, 2 blue. I need a blue (tracking)

T. Good! Here's a token for doing and saying all those things (puts token in cup). Now what's your answer?

C. (Child responds).

T. (If correct), That's right! That's the right answer! Very good! That's right because 3 yellows, 3 greens, and 3 blues (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are 3 yellows, 3 greens, and three blues (T. gestures).

A few trials later:

T. Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that goes here (T. points) but first... (T. pauses).

C. (Child starts to count along horizontal axis although colors are uniform on vertical).

T. Wait. Count this way. (T. sweeps finger down first column).

C. 1, 2, 3 (tracking; then pauses).

T. Three what?

C. Triangles.

T. 1, 2, 3 blues (tracking). Say that.

C. 1, 2, 3 blues (tracking).

T. Good! (C. pauses). Keep counting.

C. 1, 2, 3 yellow. 1, 2, 3 green (tracking).

T. How many green? Count them again.

C. 1, 2 green (tracking). (C. pauses).

T. 1, 2 greens. I need a green (tracking).

C. 1, 2 greens. I need a green (tracking).

T. Very good! Here's a token for doing and saying all those things (puts token in cup). What's your answer?

C. (C. selects answer).

T. (If correct), That's right! That's the right answer! Very good! That's right because 3 blues, 3 yellows, and 3 greens (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are 3 blues, 3 yellows, and three greens (T. gestures).

Later:

T. Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that belongs here (T. points). But first...

C. Triangle, triangle, triangle (tracking).

T. Wait. What do you need to count?

C. Colors (C. pauses).

T. Go ahead.

C. Red, red, red (tracking).

T. 1, 2, 3 reds

C. 1, 2, 3 reds (tracking; C. pauses).

T. Keep counting.

C. 1, 2, 3 yellows. 1, 2 greens (tracking; C. pauses).

T. I need a...

C. I need a green.

T. Terrific! Here's a chip for doing and saying all those things (puts token in cup). What's your answer?

C. (C. selects).

T. (If correct), That's right! That's the right answer! Very good! That's right because 3 reds, 3 yellows, and 3 greens (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are 3 reds, 3 yellows, and 3 greens (T. gestures).

For V-2, shape and color relevant:

T. Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that belongs here (T. points). But first...

C. 1, 2, 3 reds. 1, 2, 3 yellows. 1, 2 greens, I need a green (tracking).

T. Good!

C. (Pauses).

T. Count the shapes.

C. N.R.

T. 1, 2, 3 circles...

C. 1, 2, 3 circles. 1, 2, 3 squares. 1, 2 triangles (tracking). (C. pauses).

T. I need a...

C. I need a triangle.

T. Good! Here's a chip for doing and saying all those things (puts token in cup).

C. (C. selects answer).

T. (If correct), That's right! That's the right answer! Very good! That's right because 3 reds, 3 yellows, 3 greens (T. gestures), and 3 circles, 3 squares, 3 triangles (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are 3 reds, 3 yellows, 3 greens (T. gestures), and 3 circles, 3 squares, 3 triangles (T. gestures).

Later training trial, V-1.5 problem:

T. Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that belongs here (T. points).

C. 1, 2, 3 blues. 1, 2, 3 yellows. 1,2 red (tracking; C. begins to pick answer).

T. Wait. 1, 2 red. I need a ...

C. Red.

T. Say it.

C. I need a red.

T. Good! Here's a chip for doing and saying all those things (puts chip in cup).

C. (C. selects answer).

T. (If correct), That's right! That's the right answer! Very good! That's right because 3 blue, 3 yellow, and 3 reds (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its place) because there are 3 blue, 3 yellow, and 3 red (T. gestures).

RT Training Protocols

PRODUCT: The essence of this condition is to teach children to: say the rule; and count, label, and track values of each relevant dimension, prior to selecting their answers.

Over the course of training, children should learn to say the rule, and count values of relevant dimensions (identifying those relevant) independent of aid from the trainer.

PROCESS:

(1) Trainer asks child to look at all stimuli in matrix and solution set.

(2) Trainer models, as needed, verbalization of the rule.

(3) Trainer models, as needed, rule following: counting and labeling values of relevant stimulus dimensions while tracking stimuli in a corresponding manner.

(4) Trainer prompts child, as needed, to verbalize the rule.

(5) Trainer prompts child, as needed, to problem-solve.

(6) Trainer delivers social and material reward contingent upon rule verbalization and counting/ labeling/ tracking regardless of aid required. He then encourages the child to select an answer.

(7) Trainer praises correct answer, corrects incorrect, and explains correct choice.

Over the course of training trials, models and prompts are faded. Instead of simply providing the child with what he needs to do and say on each new trial, the trainer tests to see if the child can generate required responses without aid, and if he cannot, he leads the child to required responses.

The procedure, including the fading element, is illustrated by the following examples:

First trial:

T. (Child's name), now we're going to begin teaching you some things to say and do to help you with these problems. When you say and do these things, you'll get a chip. At the end of the day, you can trade in your chips for the prizes we showed you before.

O.K., (child's name). Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that goes here (T. points). But first I want you to say:

If I count the different colors, it may help me get it right. 1, 2, 3 yellows. 1, 2, 3 greens. 1, 2, blues. I need a blue (tracking). Now you try.

C. N.R.

T. If I count the different colors...

C. If I count the different colors...(C. pauses).

T. it may help me get it right.

C. it may help me get it right.

T. Good! O.K., now count.

C. N.R.

T. 1, 2, 3 yellows (tracking).

C. 1, 2, 3 yellows (tracking).

T. Good! (C. pauses). 1, 2, 3 greens (tracking)

C. 1, 2, 3 greens (tracking).

T. Good! (C. pauses). 1, 2 blue. I need a blue (tracking).

C. 1, 2 blue. I need a blue (tracking).

T. Very good! Here's a chip for doing and saying all those things (puts token in cup). Now what's your answer?

C. (Child selects answer).

T. (If correct), That's right! That's the right answer! Very Good! That's right because 3 yellows, 3 greens, 3 blues (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are three yellows, 3 greens, and 3 blues (T. gestures).

A few trials later:

T. Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that goes here (T. points). But first, what do you say?

C. (C. pauses).

T. If I count the colors. Say that.

C. If I count the colors...(C. pauses).

T. It may help me get it right.

C. It may help me get it right.

T. Good! Now count the colors.

C. (C. starts to count along horizontal axis when colors are uniform on vertical).

T. Wait. Count this way. (T. sweeps finger down first column).

C. 1, 2, 3 (tracking; pauses).

T. 3 what?

C. Triangles.

T. 1, 2, 3 blues (tracking).

C. 1, 2, 3 blues (tracking).

T. Good! Keep counting.

C. 1, 2, 3 yellows. 1, 2, 3 greens (tracking).

T. How many greens? Count them again.

C. 1, 2 green (tracking; C. pauses).

T. 1, 2 greens. I need a green.

C. 1, 2 greens. I need a green (tracking).

T. Very good! Here's a chip for doing and saying all those things (puts token in cup). What's your answer?

C. (Selects answer).

T. (If correct), That's right! That's the right answer! Very Good! That's right because 3 blues, 3 yellows, and 3 greens (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are 3 blues, 3 yellows, and 3 greens (T. gestures).

Later:

T. Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that goes here (T. points). But first...

C. If I count... (C. pauses).

T. What are you going to count on this one?

C. the colors it will help me get it right.

T. Very good!

C. N.R.

T. Go ahead, count.

C. Triangle, triangle, triangle (tracking).

T. Wait, you said you'd count what?

C. Colors (C. pauses).

T. Go ahead.

C. Red, red, red (tracking).

T. 1, 2, 3 reds.

C. 1, 2, 3 reds (tracking; C. pauses).

T. Keep counting.

C. 1, 2, 3 yellows. 1, 2, greens (tracking; C. pauses).

T. I need a...

C. I need a green.

T. Terrific! Here's a token for doing and saying all those things (puts token in cup). What's your answer?

C. (C. selects answer).

T. (If correct), That's right! That's the right answer! Very Good! That's right because 3 reds, 3 yellows, and 3 greens (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are 3 reds, 3 yellows, and 3 greens (T. gestures).

For V-2, shape and color relevant:

T. Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that goes here (T. points). But first, what do you say?

C. If I count... (C. pauses).

T. Count what?

C. the colors it may help me get it right.

T. the colors and...

C. the colors and shapes (C. pauses).

T. it may...

C. it may help me get it right.

T. Very good!

C. N.R.

T. Go ahead, count the colors and shapes.

C. 1, 2, 3 reds. 1, 2, 3 yellows. 1, 2 greens (tracking). I need a green.

T. Very good!

C. N.R.

T. What else did you tell me you'd count?

C. N.R.

T. 1, 2, 3 circles...

C. 1, 2, 3 circles. 1, 2, 3 squares. 1, 2 triangles (tracking; C. pauses).

T. I need a...

C. I need a triangle.

T. Very good! Here's a token for doing and saying all those things (puts token in cup).

C. (C. selects answer).

T. (If correct), That's right! That's the right answer! Very Good! That's right because 3 reds, 3 yellows, 3 greens (T. gestures) and 3 circles, 3 squares, 3 triangles (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are 3 reds, 3 yellows, 3 greens (T. gestures) and 3 circles, 3 squares, 3 triangles (T. gestures).

Later, V-1.5 problem:

T. Look at all of these (T. gestures) and all of these (T. gestures) and find the one from here (T. gestures) that goes here (T. points).

C. If I count the colors, it may help me get it right. 1, 2, 3 blue. 1, 2, 3 yellow. 1, 2 red (tracking).

T. 1, 2 red. I need a...

C. red.

T. Say 1, 2, red. I need a red.

C. 1, 2 red. I need a red.

T. Terrific! Here's a token for doing and saying all those things (puts token in cup).

C. (Child selects answer).

T. (If correct), That's right! That's the right answer! Very Good! That's right because 3 blues, 3 yellows, and 3 reds (T. gestures).

(If incorrect), Good try, but that's not the right answer. This is the right one (T. places correct answer in its spot) because there are 3 blues, 3 yellows, and 3 reds (T. gestures).

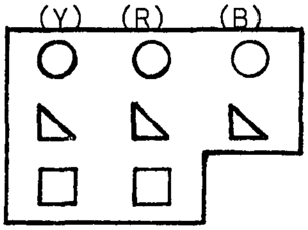
Appendix B

Verbal and Nonverbal Coding: Definitions and Examples (Defining Elements of Examples are Underlined).

Verbal Coding

I. (LAB). Labeling of attributes of relevant dimensions. If the child labels stimulus aspects relevant to matrix solution, or if he or she labels a relevant dimension, without meeting the criteria for any verbal categories below.

Examples:



Triangles.
Square.
Red, red, red.
Color.
Shapes.
I need a square.
Get the right shape.

II. (SOL). Solution response. Scored if the child's verbal response includes specification of the solution response.

Examples:
Find the right one.
Put the answer here.
I have to finish the puzzle.
Find the one from here that goes here.
Put down the right piece.

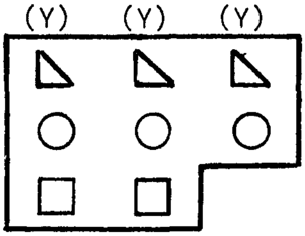
III. (PROB-) Problem-solving response. To be scored if the verbal response, at face value, constitutes a problem-solving response.

(inap). Inappropriate. Child incorrectly labels all attributes, or labels attributes of irrelevant dimensions only; child simply counts.

(ptap). Partially appropriate. Counts and labels or just labels all attributes of at least one relevant dimension (but not all relevant dimensions); or, counts or labels attributes of the underrepresented color, shape, size and/or pattern of at least one relevant dimension (but not all relevant dimensions).

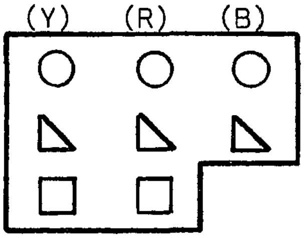
(ap). Appropriate. Child counts and labels, or just labels all attributes of all relevant dimensions and says, "I need a [relevant attribute]", for one or more dimension, or: child counts or labels attributes of the underrepresented color, shape size, and/or pattern in all relevant dimensions, and says, "I need a [relevant attribute]" for one or more relevant dimensions.

Examples:



(PROB-inap).

1,2,3 blue, 1,2,3 blue, 1,2 blue.
1,2,3,4,5,6.
1,2,3,4,5, red.
1,2, triangles.
1,2 orange.



(PROB-ptap).

1,2,3 triangles, 1,2,3 circles, 1,2 squares.
Yellow, yellow, yellow; red, red, red; blue, blue.
1,2 blue.
1,2 squares.

(PROB-ap).

1,2,3 t., 1,2,3 c., 1,2 sq., I need a square.
    1,2,3 y., 1,2,3, r., 1,2, b.
t.,t.,t., c.,c.,c., sq..sq., y.,y.,y., r.,r.,r.,
    b.,b., I need a blue.
1,2, sq, I need a square. 1,2, b.
2 sq., 2 b., I need a blue.

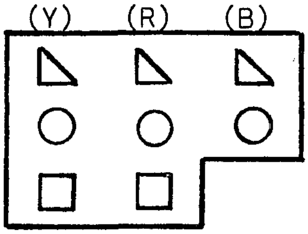
IV. (*SpPROB*-). *Specification of a problem-solving response*. Scored if the vebal response specifies a problem-solving response and referent (relevant dimensions or attributes).

(*ap*). *Appropriate*. Child verbalizes a problem-solving response and all relevant dimensions.

(*ptap*). *Partially appropriate*. Child names some but not all relevant dimensions.

(*inap*). *Inappropriate*. Child names dimensions, none of which are relevant.

Examples:



(*SpPROB-ap*).

Look at all the *colors and shapes*.
Count the *colors and shapes*.
Match the *blue*, *match the square*.

(*SpPROB-ptap*).

Match this with this.
Count 'em.
Look at all of these.
Try my best.
Count the colors.

(*SpPROB-inap*).

Count the *sizes*.
Match the *patterns*.
Look at all the *sizes*.

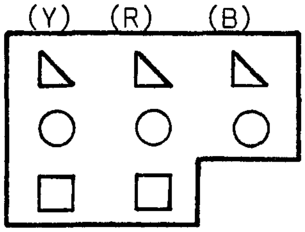
V. (*RULE*-)Scored if the child's verbal response includes specification of both a problem-solving response and outcome.

(*RULE-ap*). Scored if the child names the problem-solving response, all relevant dimensions, and outcome.

(*RULE-ptap*). Child names problem-solving response, one or more relevant dimensions (but not all), and outcome.

(*RULE-inap*). Child names problem-solving response, irrelevant dimensions only, and outcome.

Examples:



(RULE-ap).

If I count the shapes and colors it will help me.
Look at all the shapes and colors to get it right.
Count the shapes and colors to get a chip/prize/toy.
Match the blue and the square to get it right.

(RULE-ptap).

Count the shapes to get a chip.
Look at all to get it right.
Match 'em to get it right.

(RULE-inap).

Look to get it right.
Look at all the sizes to get it right.
Count the patterns to get a token.

VI. (IRR). Irrelevant. If child's verbal response fails to specify problem-solving responses, rules, or outcomes, and fails to identify relevant stimulus dimensions or attributes. Also scored if no verbal response.

Examples:

Just do it.
I don't know.
Look at 'em.
(No response).
My brother has a new bike.

VII. (INT). Interfering. Specification of an irrelevant plan of action. A verbal response that, at face value, might lead the child toward behavior incompatible with problem solution.

Examples:

Say the alphabet.
Sing my songs.

Nonverbal Coding

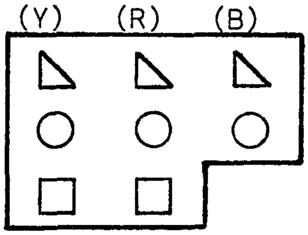
I. (T-). Tracking. Scored if the child uses his finger(s) to "mark" sequences of stimuli on the matrix completion cards.

(ap). Scored if the child systematically points to or touches all stimuli of all relevant dimensions, or: Scored if the child points to or touches all stimuli of the underrepresented color, shape, size or pattern, for all relevant dimensions.

(ptap). Scored if the child systematically points to or touches all stimuli of at least one dimension (but not all), or: Scored if the child systematically points to or touches all stimuli of the underrepresented color, shape, size, or pattern for at least one relevant dimension (but not all).

(inap). Any pointing or touching of stimuli that does not meet the criterion for above two categories.

Examples:



(T-ap).

(To triangles), (to circles), (to squares), (to yellows), (to reds), (to blues).

(To squares), (to blues).

(T-ptap).

(To triangles), (to circles), (to squares).

(To squares).

(T-inap).

(To triangles), (to circles).

II. (NP). No pointing. Scorable if child fails to point to the matrix completion card.

Appendix C

Transcript of Rule Verbalization Training with Don (S-31).

[V-1, trial #1]

E Watch what I do. If I count the shapes it may help me get it right. Let me hear you say that. If I count the shapes...
C (no response)
E Say that out loud. If I count the shapes...
C If I count the shapes...
E It may help me get it right.
C It may help me get it right.
E Good! Terrific! Now watch what I do. 1,2,3 squares (tracking with fingers) 1,2,3 triangles, 1,2 stars, I need a star. Let me see you do that. Start up here (pointing to top left stimulus).
C 1,2,3-3 squares.
E good!
C 1,2,3...
E 3 triangles.
C triangles. 1,2 stars...
E I need a...
C I need a star.
E Terrific! Here's a chip for doing and saying all those things. Let's see your answer.
C (child selects answer).
E O.K. terrific. That's the right answer because you have 3 squares, 3 triangles, and 3 stars(E gestures to each form). Let's do another one.

[V-1, trial #2]

E Look at all of these and all of these (E gestures) and figure out which one goes here (pointing to empty space). But first say, if I count the shapes...
C If I count the shapes...
E It may help me get it right.
C It may help me get it right.
E Good! Go ahead, count.
C 1,2,3-3 stars.
E Good.
C 1,2,3-3 circles.
E Good.
C 1,2-2 squares...
E I need...
C I need one square.
E Terrific! Here's a chip for doing and saying all those things. Let's see your answer.
C (child selects answer).

E Okay, that's exactly right. That's the right answer because you have 3 stars, 3 circles, and 3 squares (E gestures to each form). Very good.

[2 trials later, V-1.5, trial #1]

E Look at all of these and all of these (E gestures), and figure out which one goes here (E points to empty space) but first say, If I count the colors...
C If I count the colors...
E It may help me get it right.
C It may help me...
E get it right.
C get it right...
E Go ahead.
C It's mixed up colors!
E Go ahead, count the colors.
C 1,2,3-3...
E 3 yellow.
C 3 yellow. 1,2,3-3 blue.
E Good.
C 1,2 greens.
E I...
C (no response)
E I need...
C I need a green circle.
E Good job! Here's a chip for doing and saying all those things.
C (Child selects answer).
E O.K., and that's the right answer because you have 3 yellow, 3 blue, and 3 green (E gestures to each color).

[4 trials later, V-1.5, trial #5 (Don has reached criterion on outcome, and is presented the matrix without the solution set present)].

E Go ahead, do this one.
C If I count these colors it will help me.
E Good job!
C 1,2,3-3 blue square.
E O.K.
C 1,2,3-3 green square.
E Good!
C 1,2-2 yellow square. I need one... 1,2-2 yellows. I need one more yellow square.

E Good job, terrific! Here's a chip for doing and saying all those things. And that's the right answer 'cause 3 blues, 3 greens, 3 yellows (E gestures to colors).

[Next day, V-2, trial #1].

E Look at all of these, and all of these, but I want you to say, If I count the shapes <u>and colors,</u> it may help me get it right.

C If I count the shapes and colors.

E It may help me...

C it may help me.

E Good! Now watch. 1,2,3 triangles, 1,2,3 squares, 1,2 circles, I need a circle. 1,2,3 blue, 1,2,3 red, 1,2 yellow, I need a yellow. Let me see if you can say all that.

C 1,2,3-3 triangles.

E Good!

C 1,2,3-3 squares.

E Good.

C 1,2 circles...

E I need a...

C I need a circle.

E Now count <u>this</u> way (gestures across vertical axis).

C 1,2,3...

E Blue.

C 3 blue, 1,2,3 wait, (you) no tell me. Red.

E Good!.

C 1,2-2 yellows, I need a yellow.

E Terrific! Here's a chip for doing and saying all those things. Let's see your answer.

C (Child selects).

E That's a good try. That's not quite the right answer. Let me show you which one is right (E puts correct answer in place). This is the right answer because you have 3 triangles, 3 squares, 3 circles, and you have 3 blue, 3 red, and 3 yellow. Let's try another.

[5 trials later, V-2, trial # 6].

E Look at all of these and these (E gestures), and figure out which one goes there (E points).

C If I count the colors and shapes, it will help me.

E Good!

C 1,2,3-3 stars.

E Good.

C 1,2,3-3 rectangles.

E Triangles.

C Triangles. 1,2-2 circles. I need a... I need a circle.

E O.K., good.

C 1,2,3-3 blues.

E Good.

C 1,2,3-3 greens.

E Um-hmm.

C 1,2-2 reds.
E Um-hmm.
C I need a red.
E Terrific job doing and saying all those things. Let's see your answer.
C (Child selects).
E O.K., and that's the right answer, because you have 3 stars, 3 triangles, 3 circles; 3 blue, 3 green, and 3 red (E gestures).