DOCUMENT RESUME

ED 243 781 SO 015 628

AUTHOR	Pellegrino, James W.; Glaser, Robert
TITLE	Analyzing Aptitudes for Learning: Inductive Reasoning.
INSTITUTION	Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	84
NOTE	79p.; For related documents, see SO 015 626-627. In: Glaser, R., Ed. Advances in Instructional Psychology. Hillsdale, N.J., Lawrence Erlbaum Associates, 1982. v2, p269-345.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Aptitude; *Aptitude Tests; *Cognitive Ability; *Cognitive Processes; Educational Research; *Induction; Instruction; Skill Analysis; Skill Development

ABSTRACT

A major focus of the psychology of instruction is understanding and facilitating the changes in cognition and performance that occur as an individual moves from low to higher competence in a domain of knowledge and skill. A new program of research which examines the initial state of the learner as a component of this transition in competence is described in this 8-part document. Part 1 introduces two previous attempts to relate aptitude to instruction: differential aptitude tests and aptitude treatment interaction. Part 2 provides an overview of the two general research approaches to aptitude analysis, the cognitive correlates approach and the cognitive components approach. The approach taken in the present research effort is described as a task analytic approach that considers basic processes, executive strategies, and content knowledge in aptitude test performances of skilled and less skilled individuals. Part 3 considers the relationship between inductive reasoning and general ability. Common generic properties of inductive tasks are discussed as are different types of inductive items on aptitude tests. Part 4 outlines earlier theories on analogical reasoning performance. Problem features, processing models, item processing data and theory, item errors, and performance are analyzed for figural, numerical, and verbal analogy solutions in parts 5 through 7. A final part discusses results of research, concluding that a number of interrelated factors differentiate high and low skill individuals. These factors include management of memory load, organization of an appropriate knowledge base, and procedural knowledge of task constraints. (LP)

ANALYZING APTITUDES FOR LEARNING: INDUCTIVE REASONING

James W. Pellegrino

University of California, Santa Barbara

Robert Glaser

University of Pittsburgh

Learning Research and Development Center

University of Pittsburgh

1984

Reprinted by permission from *Advances in Instructional Psychology* (Vol. 2), Hillsdale, N.J.: Lawrence Erlbaum Associates, 1982.

The research reported herein was supported by the Learning Research and Development Center, supported in part as a research and development center by funds from the National Institute of Education (NIE), Department of Education. The opinions expressed do not necessarily reflect the position or policy of NIE and no official endorsement should be inferred.

5 Analyzing Aptitudes for Learning: Inductive Reasoning

James W. Pellegrino
University of California, Santa Barbara
Robert Glaser
University of Pittsburgh

I. INTRODUCTION

A major focus of the psychology of instruction is understanding and facilitating the changes in cognition and performance that occur as an individual moves from low to higher competence in a domain of knowledge and skill. A framework for research devoted to analyzing this transition in competence can be specified in terms of several integral components. These are: (1) the nature of competent performance and of intermediate states; (2) the initial performance state of the learner; (3) the transition processes between this initial state and a state of competence; and (4) the monitoring and assessment of performance changes (Atkinson & Paulson, 1972; Glaser, 1976). We are concerned here with the second component—the initial state of the learner. Instruction begins with the learner's initial knowledge and skill, and proceeds forward from this base. These initial state characteristics facilitate or retard the learning of subject-matter competence. They are comprised of subject-matter skills that can assist in learning and that are transformed into more advanced states of competence. Initial state also consists of learning skills that contribute to the acquisition of new knowledge. This chapter describes the beginning of an attempt to gain theoretical and practical understanding of these skills for learning.

There are a number of ways in which the initial competence with which an individual begins a course of learning has been considered in educational practice. One is assessment through the use of aptitude and intelligence test scores that are predictive of scholastic achievement. A second is the diagnostic assessment of a student's strengths and weaknesses in a subject matter that might be attended to in remedial programs in the course of specific instruction. A third,

used primarily with young children, is the assessment and training of readiness skills; i.e., certain perceptual and language competencies required for instruction in reading and elementary arithmetic. While these three approaches are interrelated, they emphasize different aspects of performance; namely, general and specific aptitudes, subject matter prerequisities, and developmental level. Our concern in this chapter is with the first of these.

As is evidenced by current debate, aptitude and intelligence tests have become a well established aspect of our educational system. Hardly an individual has emerged from our educational system in the recent past without having his or her ability to learn measured by an aptitude test. These tests of general intelligence and of verbal and quantitative aptitude measure the kind of intellectual performance that is most accurately called "general scholastic ability." Correlational evidence has shown that the abilities tested are predictive of success in school learning. This point is to be emphasized because no test is simply valid in general, but is intended for a specific purpose and situation. Aware of this operational fact, textbooks and articles on the subject (e.g., Cronbach, 1970; Scarr, 1978; Tyler, 1963) carefully point out that these tests are not tests of intelligence in some abstract way. Rather, if we base our conclusions about what these tests measure on their most effective use, that is, their predictive validity, then the verdict is that they are primarily tests of abilities that are helpful in present day school situations. However, our understanding of these abilities for learning is very incomplete. We know that the abilities measured by such tests account for 35 to 45% of the variation in school performance over all school levels. We also know that this correlational knowledge does not provide the kind of understanding that enables us to enhance or remediate these abilities for learning.

There is current scientific and social dissatisfaction with such tests for three main reasons: (1) The present operational definition (validity coefficients) of these tests seems to have reached a plateau of efficiency with our present technology. Efforts to improve predictive and diagnostic validity have run into diminishing returns. (2) The tests reflect a restrictive assessment of abilities that limits their utility in the guidance and improvement of student achievement. This is so because the tests provide information useful primarily for go/no-go selective decisions about program entrance; they do not provide information that could change the course of instruction. To be considered adequate, diagnostic measures should assess differences in learning abilities and acquired knowledge, thereby indicating how schools can adapt their learning environments to diverse individual needs. (3) Finally, scientists are now recognizing that current test theory and technique have not made contact with modern psychological theories of learning and cognition. New efforts should be influenced by the developments in these areas; modern theory now appears capable of bringing us closer to understanding the components of cognition that underlie the abilities for learning, which lead to success in school.

With the above needs in mind, our program of research uses the concepts and methods of cognitive psychology to analyze the intellectual functions assessed by measures of scholastic aptitude. The eventual goal is to understand the constituent processes and content involved, and to show how they can be influenced through instruction. Before we describe our analyses however, it is appropriate to review briefly two previous attempts to relate measured aptitudes to different learning environments.

A. Previous Attempts to Relate Aptitudes to Instruction

1. *Differential Aptitudes.* Psychometricians have not been remiss in attempting to probe deeper into the different facets of human cognition so that tests might be more sensitive to individual differences. Some years ago, dissatisfaction with research on the IQ and on multiple factors led to a de-emphasis of the concept of general intelligence and increasing popularity of differential aptitude tests. In addition to an overall measure of general aptitude, schools began to employ tests that provided measures of a variety of factors such as spatial, mechanical, and abstract reasoning. These test batteries attempted to predict success in vocational programs that appeared to require different aptitude patterns.

In 1964, a careful analysis was done by McNemar of the validity coefficients of certain widely used differential aptitude batteries. He argued from his analysis that, "Aside from tests of numerical ability having differential value for predicting school grades in math, it seems safe to conclude that the worth of multitest batteries as differential predictors of achievement in school has not been demonstrated . . . It is far from clear that tests of general intelligence have been outmoded by the multitest batteries as the more useful predictors of school achievement" (p. 875). More recent work reaffirms McNemar's conclusion (Carroll, 1978).

Thus, the attempt to further differentiate specific ability patterns and relate them to specific educational programs was, at best, no more successful than the use of general ability measures. Differential aptitude tests followed the accepted practice of attempting to predict the final outcomes of learning, and did no better than general ability tests in attempting to identify and measure abilities that could be related to models of learning; and to prerequisite skills required for learning various tasks. McNemar (1964) reflected on a possible reason for this lack of success and stated the following:

> Abilities, or capacities, or aptitudes, or intellectual skills, or whatever you choose to call them, are measured in terms of response products to standardized stimulus situations. The stimulus is presented to an organism which by some process comes up with a response; thus any attempt to theorize and/or study intellect in terms of a simple stimulus-response (S-R) paradigm seems doomed to failure unless drastically modified and complicated by the insertion of O for organism and P for

> process. . . . Studies of individual differences never come to grips with the process, or operation by which a given organism achieves an intellectual response. Indeed, it is difficult to see how the available individual difference data can be used even as a starting point for generating a theory as to the process nature of general intelligence or of any other specified ability. (p. 881)

2. *Aptitude-Treatment Interaction*. Psychologists and educational researchers have been concerned about the relationship between measures of individual differences and learning variables. To a large extent, this work was heralded by the 1957 book by Cronbach and Gleser entitled *Psychological Tests and Personnel Decisions* and its second edition in 1965. This book developed a decision-theory model for the selection and placement of individuals into various "treatments." The word treatment was given a broad meaning, referring to what is done with an individual in an institutional setting, e.g., in education it refers to the particular programs or instructional methods a student is assigned to or has the opportunity to select. This theoretical analysis pointed out that aptitude information is useful in modifying and selecting treatments only when aptitude and treatment can be shown to interact. This research is different from that of the previously mentioned work on differential aptitude testing in which emphasis was placed on determining the relationship between measured aptitudes and learning outcomes resulting from relatively fixed curricula. In the ATI work, the emphasis is on determining whether aptitudes can predict which of several different learning methods might help different individuals attain similar educational outcomes.

Comprehensive reviews report detailed analyses of ATI studies (Bracht, 1969; Bracht & Glass, 1968). Cronbach and Snow (1977) have carried out a very extensive review and analysis of many of the ramifications of the ATI problem. They conclude that, with a few notable exceptions, ATI effects have not been solidly demonstrated. The frequency of studies in which the appropriate interactions have been found is low and the empirical evidence found in favor of such interactions is often not very convincing. In those occasional instances when positive results have been obtained, no general principles have emerged because of the lack of consistent findings in replicated studies and in transfer to new subject matter tasks.

While one is struck by the absence of any prescriptive assistance to instruction, certain sections of the Cronbach and Snow book suggest a trend that bears further study. A pattern of more promising results appeared in situations where investigators were forced to construct an aptitude measure because no ready made and labeled aptitude tests were already available. There appears to be a tradeoff between the reliability offered by established tests of aptitude and the information about acquisition processes afforded by tests specially constructed for experimental work. In research using specific aptitudes it seems as though

researchers presumed that the label of a particular aptitude measure had direct implications for instructional practice. For example, a spatial aptitude test was paired with procedures that de-emphasized verbal content in instruction. But the mere absence of words (diagrams, for example) by no means implies the presence of abilities required in these tests.

These results certainly do not recommend that standardized tests be abandoned as inappropriate measures in ATI research—the fault in these efforts appears to be in the absence of adequate theories of test performance rather than in the tests themselves. The clear need is that the use of traditional psychometric instruments will have to be accompanied by careful analyses of processes that relate aptitude, treatment, and the knowledge or skills being learned. Testable theories are required that describe competencies measured in the pretest, competencies required for task performance, and treatment procedures that connect the two (Snow, 1980a). At the present time, generally used aptitude constructs are not productive dimensions for measuring those individual differences that interact with different ways of learning. These measures, derived from a psychometric selection-oriented tradition, do not appear to relate to the processes of learning and performance that have been under investigation in experimental and developmental psychology. The treatments investigated in ATI studies have not been generated by any systematic analysis of the kinds of psychological processes called upon in particular instructional methods, and individual differences have not been assessed in terms of related performance processes.

B. Cognitive Performance and Individual Differences

The lesson implied in the results obtained from such endeavors as differential aptitude testing and ATI research has been learned slowly. In 1957, Cronbach suggested that "Constructs originating in differential psychology are now being tied to experimental variables. As a result, the whole theoretical picture in such an area as human abilities is changing . . . It now becomes possible . . . ultimately to unite the psychology of intelligence with the psychology of learning" (p. 682). The point was reiterated in 1972 by Glaser who called for research on the "new aptitudes" that would be interpreted in terms of process constructs. However, the lines of research reported in this chapter have only recently begun to conceptualize individual differences in aptitudes in terms of structure and process constructs of contemporary theories of human cognition and cognitive development.

The particular orientation and the specific problem that we address can now be succinctly stated. The global objective is to contribute to an understanding of the ways in which individuals differ in abilities for learning. In the long run, our goal will be achieved if we can couch abilities to learn in terms of the concepts of modern cognitive theory, and then develop procedures for identifying school-

related capabilities based upon these interpretations. In this chapter, our initial step is to accept the robust correlational fact of a relationship between certain abilities measured by test tasks and school achievement. We then identify classes of test tasks that have consistently appeared on scholastic aptitude tests and use current techniques of task analysis to understand the nature of the performance elicited by these tasks. A logical next step would be to relate the aptitude processes to similar task analytic work being pursued in school subject matter areas, e.g., beginning reading, text comprehension, elementary arithmetic, science problem solving, etc. Such an approach should begin to explain the predictive validity of the skills of learning measured by scholastic aptitude tests, and the reasons for limitations in validity, and may suggest how instruction could improve the intellectual performances involved. As Carroll (1978) has written:

> The performances required on many types of mental ability tests—tests of language competence, of ability to manipulate abstract concepts and relationships, of ability to apply knowledge to the solution of problems, and even of the ability to make simple and rapid comparisons of stimuli (as in a test of perceptual speed)—have great and obvious resemblances to performances required in school learning, and indeed in many other fields of human activity. If these performances are seen as based on learned, developed abilities of a rather generalized character, it would frequently be useful to assess the extent to which an individual has acquired these abilities. This could be for the purpose of determining the extent to which these abilities would need to be improved to prepare the individual for further experiences or learning activities, or of determining what kinds and amounts of intervention might be required to effect such improvements. These determinations, however, would have to be based on more exact information than we now have concerning the effects of different types of learning experiences . . . on the improvement of these abilities. (p. 93-94)

II. APPROACHES TO APTITUDE ANALYSIS

A. Overview

Within this developing area of aptitude research, there appear to be two general research approaches (Pellegrino & Glaser, 1979). The first of these, which we have termed the "cognitive correlates" approach, seeks to specify the elementary information processes that correlate with high and low levels of aptitude. Scores on tests of aptitude and intelligence are used to define subgroups to be compared on laboratory tasks that have relatively well-defined processing characteristics. The particular tasks chosen and their hypothesized underlying processes can be interpreted in the broader context of general models of the human cognitive system. Examples of this type of research can be found in the work of Hunt and his colleagues (e.g., Hunt, 1976, 1978; Hunt, Frost, & Lunneborg, 1973;

Hunt & Lansman, 1975). The second research approach, which we have termed the "cognitive components" or "task analytical" approach, attempts to directly identify the information processing components of performance on tasks used to assess aptitude. In this approach, performance on psychometric test tasks becomes the object of theoretical and empirical analyses. The objective of this work is to develop models of task performance and to use these process models as a basis for individual difference analysis. Examples of this type of task analytic research include the discussion by Estes (1974) and the recent work of Pellegrino & Glaser (1980), Sternberg (1977), and Carroll (1976).

As a result of our review of recent aptitude research (Pellegrino & Glaser, 1979), we have argued that the cognitive-components approach incorporates the cognitive-correlates approach, avoids the explanatory inadequacy of indirect correlational methods, and has the theoretical power to consider differences on various dimensions of cognitive functioning. Cognitive correlates research has typically focused on relatively simple processing tasks that measure the speed of accessing codes in long term memory or manipulating information in short-term memory. However, processing speed is only one of several cognitive components of complex tasks. Many memory and problem-solving tasks require higher level executive strategies that facilitate the appropriate sequencing of basic processing activities such as the activation and manipulation of memory codes. The efficiency, speed, and coordination of these higher-level operations may contribute substantially to individual differences in performance, particularly since increases in task complexity involve increased demands upon limited memory space and processing resources. Thus it becomes important to raise questions about individual differences in "executive routines," which have become central in most cognitive theories. In addition, semantic or declarative knowledge structures further determine how task content is encoded and represented and how it interacts with processing capabilities. These structural properties influence storage capacity, knowledge organization, and the availability of different types of semantic and procedural knowledge.

Our program of research employs a task analytic approach that attempts to consider these mutally dependent aspects—"basic" processes (automatic and controlled attention-demanding processes), executive strategies, and content knowledge—in the analysis of the performance characteristics of skilled and less-skilled performers on aptitude tasks. Work to date indicates that the analysis of high and low aptitude groups shows differences in the speed of accessing stored memory representations. Task analytic studies of performance on psychometric tasks implicate the efficiency of executive control and the knowledge structure properties. These are potential sources of individual differences that are related to developmental level, educational history, and general experience. An adequate explanation of individual differences in aptitudes for learning must come to grips with these interactive aspects of information processing.

B. A General Analytic Scheme

Before proceeding to a discussion of our research, we must briefly describe the general framework and analytic scheme that has guided our efforts. This plan, which has evolved from our own work and that of others (see Glaser & Pellegrino, 1978), prescribes the following stages and sets of issues.

1. The first step in a systematic analysis of individual differences in aptitudes for learning is to identify the domain of tasks associated with an aptitude factor. By this we mean identifying a core set of tasks that frequently occur across many widely used tests and that have been shown, in factor-analytic studies, to have consistent relationships to certain basic aptitude constructs. Thus, the tasks chosen for analysis should have: (1) reliable association with a reasonably general aptitude construct, and (2) consistent predictive validity with respect to a criterion performance of significance (for example, academic achievement).

An adequate understanding of individual differences in a particular aptitude cannot be based upon an intensive analysis of only a single task with a high loading on that aptitude construct. Rather, it is necessary to conduct analyses that consider the various intercorrelated tasks that define more completely a substantial set of performance, comprising a particular first-order or higher-order aptitude construct. A successful process analysis of multiple tasks provides the basis for understanding the patterns of intercorrelations among tasks. More importantly, the analysis of multiple tasks should permit the differentiation of general and specific cognitive processes and help focus on a level of analysis where research can identify the extent of process trainability and transfer effects.

2. Once the domain of tasks associated with an aptitude construct of interest is defined, it is then necessary to develop and validate information processing models for the different tasks. The theories and models can be derived from computer simulation programs and/or empirical studies of the effects of task properties on latency, solution protocols, and error patterns. These models must differentiate between basic cognitive processes and higher-level strategies that control process integration and sequencing. This flexibility is necessary because individual differences may be manifested at different levels as a function of the range and distribution of ability being considered. The analysis of a particular task must also explicate the sources of difficulty that differentiate test items, thus providing the basis for individual variation in test performance. Test tasks are composed of heterogeneous item sets where the individual items vary considerably in difficulty as a function of ability or developmental level. Thus, an understanding of individual differences in task performance must include a process theory of item difficulty. For this purpose, the processes specified as the components of performance must involve a level of analysis that is sufficient to explain individual item characteristics, individual subject performance, and the interaction of the two.

3. The third major step in the analysis is to use the models of task performance as the basis for individual differences analyses in each task. In this way, the utility of a model for explaining the source(s) of individual differences can be further tested and validated. Individual differences can be investigated in terms of the parameters of a model, or in terms of the applicability of different models for the performance of different individuals. Such an analysis must also investigate the sources of developmental differences. It is necessary to map out the relationship between overall developmental change in mental ability and sources of individual differences within separate age groups. There is no reason to assume that the sources of individual differences within one age group are necessarily applicable to individuals at a higher or lower maturational level.

4. The next step in the analysis of individual differences is the examination of cross task consistency in sources of individual differences. Based upon the results of the preceding steps, one can then attempt to specify and test the cognitive components that are general across all task forms representative of the aptitude construct (for example, all induction tasks), and those that are specific to a given task form or content area. In this work, the attempt should be made to account for the consistent patterns of relationships found in the psychometric literature.

In the later stages of this research, as individual and developmental differences in cognitive processes are identified, work can proceed on the analysis of criterion tasks similar to those used to establish aptitude test validities. The goal of this effort is to identify process and knowledge structure characteristics of cognitive performance that account for the correlations of aptitude measures with criterion performance. Concurrently, research can be carried out on the modifiability and instructability of these characteristics of intellectual functioning as they serve to support or enhance an individual's abilities for learning.

III. INDUCTIVE REASONING AND GENERAL ABILITY

A. Overview

Our work covers primarily the early stages of the general analytic scheme described above. In this chapter, we will focus on the analysis of performance on tasks representing a central aptitude or ability factor—the induction factor. We have chosen this factor because the tasks that load highest on the inductive reasoning factor also load the highest on any general factor. Furthermore, the task forms cut across all the major content dimensions. Thus, many forms of symbolic input serve to assess inductive reasoning skill. Also, this general factor, which can be extracted from most aptitude and intelligence tests, is the single best predictor of academic performance and achievement test scores (Snow, 1980b).

All inductive reasoning tasks have the same basic form or generic property: The individual must induce a rule governing a set of elements. Thus, a set of elements is presented and the task is to infer a pattern or rule structure that will allow generation or selection of an appropriate continuation, or verification that the pattern as shown is legitimate. This generic structure is manifest in a variety of task forms ranging from simple classification problems to highly complex matrix items. Figure 5.1 provides an illustration of the various task forms associated with inductive reasoning. Also shown in Fig. 5.1 are the different content dimensions that are typically utilized for a given task form. For the verbal and figural classification problems, the task is to determine the relationship (semantic, logical, geometric, etc.) governing the base set and select the alternative that is consistent with the inferred rule. In the case of the letter and number series problems, the task is to determine the relational and periodic structure of the element string and extrapolate it to complete the blank spaces. The figural, numerical, and verbal analogy items require the individual to choose the alternative that is related to the third term in the same way that the second term is related to the first. The numerical analogy is a similar form of this task that uses two initial item pairs to reduce ambiguity in specifying the appropriate type of relationship governing the problem. Finally, the figural matrix problem requires the individual to select the alternative that completes the matrix and that is consistent with the relationships governing the column and row structure.

Induction problems have been a part of aptitude tests almost from the inception of the testing movement. One or more of these task forms can be found on virtually any standardized aptitude or intelligence test at any development level. As an example, the Cognitive Abilities Test (Thorndike & Hagen, 1971) includes the following item types in its multilevel battery intended for grades 3-12: verbal classification and analogy, number series, and figural classification and analogy. The Raven Progressive Matrix Test consists entirely of figural matrix problems of the type shown in Fig. 5.1. Multiple aptitude batteries such as the Differential Aptitude Test and the Primary Mental Abilities Test provide separate inductive reasoning scores. Evidence is available for the claims that inductive reasoning tasks are: (1) highly correlated with academic achievement; (2) load the highest on a general factor extracted from performance on multitest batteries; and (3) differ in their intercorrelations as a function of task form and content domain. To illustrate these points, we focus on data available for the Cognitive Abilities Test.

In assessing validity, the CAT provides three separate scores representing performance on verbal, quantitative, and figural problem types, each of which is related to academic achievement. When CAT verbal, quantitative, and figural performance scores are correlated with achievement test performance as measured by the Iowa Tests of Basic Skills, the average correlations over grades 3-8 are uniformly high, as shown in Table 5.1. A similar pattern emerges when CAT performance is correlated with actual school grades. The specific subtests that contribute to these three separate CAT scores are shown in Table 5.2. This table

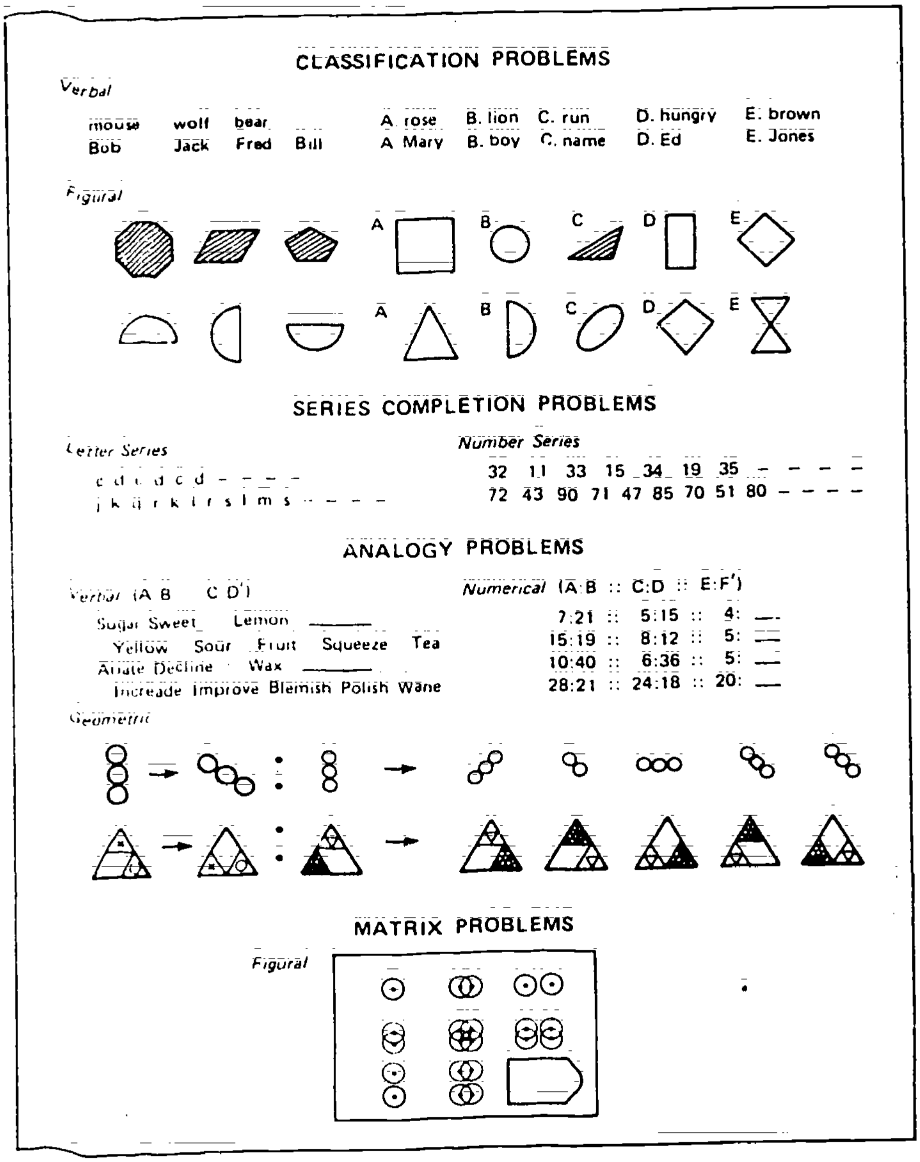


FIG. 5.1. Task forms associated with inductive reasoning.

Table 5.1
Achievement Test Correlations for the Cognitive Abilities Test

Iowa Subtest	CAT Form Verbal	Quantitative	Non-verbal
Vocabulary	.81	.66	.55
Reading	.80	.66	.56
Language	.78	.69	.57
Work Study	.77	.74	.65
Arithmetic	.73	.77	.62

also provides the loadings of each subtest on the general factor extracted from the pattern of subtest intercorrelations. In each grade, the induction tasks have high loading on the general factor.

Interrelations among verbal and figural induction tasks are shown in Table 5.3. The pattern of correlations is indicative of relatively strong relationships among all the induction tasks. However, the pattern also shows that intercorrelations are stronger within content domains (double-underlined coefficients) rather than for common task forms (single underlining); i.e., the verbal-verbal and figural-figural correlations are higher than the analogy-analogy and

Table 5.2
Subtest Loadings on a General Factor

Subtest	Grade 3	5	7	9	11
Verbal					
Vocabulary	.68	.67	.67	.71	.67
Sentence Completion	.77	.72	.73	.72	.73
Verbal Classification[a]	.82	.78	.70	.74	.70
Verbal Analogies[a]	.79	.79	.80	.83	.81
Quantitative					
Quantitative Relations	.72	.76	.74	.85	.85
Number Series[a]	.80	.83	.82	.82	.82
Equation Building	.74	.75	.79	.73	.68
Non-Verbal					
Figure Classification[a]	.66	.64	.67	.69	.76
Figure Analogies[a]	.74	.73	.77	.76	.78
Figure Synthesis	.60	.64	.62	.58	.64

[a]Denotes induction tasks

Table 5.3
Correlations Among Verbal and Figural Induction Tasks

Type of Induction Task	Grade 3 VC	Grade 3 FA	Grade 3 FC	Grade 7 VC	Grade 7 FA	Grade 7 FC	Grade 11 VC	Grade 11 FA	Grade 11 FC
Verbal Analogy	.74	.62	.55	.67	.63	.54	.71	.63	.62
Verbal Classification		.63	.57		.56	.51		.56	.53
Figural Analogy			.67			.68			.74

Note. VC - Verbal Classification
FA - Figural Analogy
FC - Figural ClassificationType of

classification-classification correlations. Finally, it should be noted that in both Tables 5.2 and 5.3, the analogy task seems to be one of the most stable in its patterning of factor loadings and intercorrelations. This is consistent with the fact that the analogy task is a most prominent induction task.

The extensive use of analogy items in intelligence and aptitude tests was documented by Dawis and Siojo (Note 1), and more recently, Sternberg (1977) has provided a detailed review and discussion of the importance of analogical reasoning within the field of differential psychology. Spearman (1923) and Raven (1938) both argued that inductive reasoning was central to the concept and measurement of intelligence. The only debate about the different task forms and content dimensions is whether they represent a single aptitude construct or can be subdivided into separate aptitude factors representing different relational types such as the induction of semantic as opposed to figural relations.

Rule induction tasks are not simply a psychometric curiosity, but are relevant in the broader domain of cognitive research and theory. Greeno (1978) for example, has characterized rule induction problems as instances of a major type of problem solving task within a general problem typology. He has also suggested that comprehension can, in many respects, be viewed as a special instance of rule induction. Simon and Lea (1974) have discussed the similarities between the processes utilized in rule induction, as incorporated in their General Rule Inducer program, and the components of concept formation as used in a program developed by Gregg and Simon (1967). Simon and Lea liken the instances presented in a concept attainment task to the elements of a series completion problem. Egan and Greeno (1974) consider rule induction in the framework of semantic memory research, and have pointed out that analogical reasoning, series completion, problem solving, and concept formation all require a search for relations among elements resulting in new interconnections between the nodes of a network structure. In this same semantic memory framework, Nor-

man, Gentner, and Stevens (1976) have argued for the essentially inductive nature of effective instruction. Structures of interrelated concepts are communicated to students by a teacher or through some other instructional medium. In order to comprehend and remember the material, the student must induce the structure of the presented information by detecting the relational pattern of the presented concepts and discovering the connections between the newly communicated material and the knowledge structures already in permanent memory. The importance of inductive thought processes and reasoning by analogy has been emphasized in science (e.g., Oppenheimer, 1956), mathematics (Polya, 1965), and in the acquisition of information in the classroom (e.g., Bruner, 1957; Forehand, 1974).

B. General Components of Induction Tasks

The different inductive reasoning tasks described above have been the subject of various empirical and theoretical studies. Serial pattern acquisition has been studied extensively in the psychological literature (e.g., Kotovsky & Simon, 1973; Restle, 1970; Simon & Kotovsky, 1963; Vitz & Todd, 1969). The extensive theoretical analysis of letter series problems in the form of a computer simulation program (Simon & Kotovsky, 1963) has been shown to be applicable to process training with children (Holzman, Glaser, & Pellegrino, 1976). Empirical and theoretical work has been done with verbal and figural analogy problems (e.g., Evans, 1968; Mulholland, Pellegrino, & Glaser, 1980; Pellegrino & Glaser, 1980; Reitman, 1965; Sternberg, 1977; Whitely, 1976). Figural matrix problems of the type found on the Raven's test have been discussed by Hunt (1974) and Jacobs and Vanderventer (1976). Rather than review each of the various studies in detail (see Sternberg, 1977; or Holzman, 1979, for detailed reviews), we would like to present a synthesis of these theoretical and empirical efforts in the form of a general model of the components representative of inductive reasoning tasks. This general model will serve as the background for our subsequent discussions of detailed models of analogical reasoning tasks.

All inductive reasoning tasks can be said to require the following processes: (1) encoding or representational processes that depend on the information stored in permanent memory; (2) inference processes that identify and/or generate relational features shared by two or more encoded elements; (3) rule assembly or monitoring processes that organize individual relational features into simple or complex relational structures; (4) comparison or match processes that can evaluate the similarities among relational structures; (5) discrimination processes capable of selecting among competing relational structures; and (6) decision and response or output processes.

In inductive reasoning tasks, these processes are called upon one or more times during the course of solution. Differences between task forms and content domains, as well as differences in item difficulty within a given task, are a

function of one or more of these components of solution. For example, letter series problems minimize encoding or representational difficulties by restricting the nature of the content (the alphabet) and by using relatively simple relational features (identity next (A–B), and backwards next (B–A) relations). Problem difficulty arises in the rule assembly process, where discovery of the periodic structure of the pattern, (i.e., period size) must precede discovery of the complex rule governing the entire sequence. Difficult items are those having larger period lengths and more complex relational structures (see Holzman et al. 1976). In the case of figural analogies or figural matrices, there is a significant encoding or representational component since one must often decompose complex and embedded patterns to identify the individual elements that are relationally transformed. Encoding and representational processes are also of obvious importance in verbal classification and analogy items because of the fuzzy nature of many verbal concepts and their multi-attribute internal representations.

Another major difference among the various induction tasks is the number of constraints on the solution to the problem. In a classification item, the inferred rule must govern the entire set of elements. In a series item, the inferred rule must be applicable to each successive period and be capable of continued extrapolation. In an analogy problem, the inferred rule for the first pair of terms must be consistent in relation and direction with the inferred rule for the second pair of terms. Finally, in a matrix problem, the inferred rule for rows must be applicable to all rows, the inferred rule for columns must show similar consistency and the two sets of rules must not be in conflict. The number of different constraints on final solution affects item and task difficulty by either requiring that more information be held in working memory, or that encoding, inference, rule assembly, comparison, and/or discrimination processes be executed repeatedly to achieve an unambiguous answer.

As an illustration, solution of a classification item typically requires fewer executions of inference, rule assembly, and comparison processes than the matrix task where both row and column rules must be inferred, assembled, and compared, and both sets of information must be held in memory. At a general level, successful performance in any rule induction task requires that the individual understand the constraints of the task and that those constraints serve as a significant component of the individual's problem space for the task. Differences among tasks, and among items within a task, can arise in part from the number and type of constraints that have to be met for problem solution.

The preceding general discussion of inductive reasoning tasks is intended as an overview of the psychological processes demanded by this set of tasks and the analytic issues that need to be addressed. What follows is a discussion of theoretical models of analogical reasoning. Because analogical reasoning items appear in different content forms and involve a sufficiently large set of constraints, understanding this task should contribute to a general understanding of processes involved in this pervasive form of aptitude test task.

IV. THEORETICAL ANALYSES OF ANALOGICAL REASONING PERFORMANCE

A. Overview of Previous Theories

One of the earliest "process" theories of analogical reasoning was formulated by Spearman (1923). According to Spearman, analogical reasoning involves three processes. First, one must "apprehend" or encode and understand the elements of the item. The second process involves the "eduction" of the relation between the first two terms of the analogy. The third process is the "eduction of correlates," in which one uses the relationship inferred between the first two terms together with the third term to find the solution to the item. Unfortunately, given the psychology of the time, Spearman's description of these processes was not sufficiently well specified to lead to direct experimental tests. In addition, there was implicit in his description an automaticity of function that fails to capture some of the apparent difficulty associated with solving many analogy items.

The process theory sketched out by Spearman has been greatly expanded and refined in the work of Sternberg (1977). Like Spearman, Sternberg has proposed a theory that is intended to apply across all analogical reasoning tasks. The component processes in Sternberg's theory include: (1) encoding the individual terms of the analogy; (2) inferring the relationship between the first two terms; (3) mapping the relationship between the first and third terms; (4) applying the results of the inference and mapping processes to the third term to generate an ideal fourth term that is then used as the basis for evaluating the alternative answers; (5) an optional justification process that is used to select among alternative answers when none precisely matches the ideal answer; and (6) a response process that indicates the choice of an answer. The processes specified in this theory are consistent with the general list presented earlier for all induction tasks.

Pellegrino and Lyon (1979) have provided a detailed discussion of the method and theory of Sternberg (1977). An important point in their commentary was addressed to the issue of understanding and modeling errors as well as correct performance (see also Pellegrino & Glaser, 1980). The theory and models that Sternberg postulates reflect algorithmic solution methods for items that are, for the most part, relatively easy and unambiguous. A different and more general theory is needed to represent performance for all levels of analogy difficulty and individual solution skill. Consequently, the study of analogy must be extended in two directions. One direction is to unpack the individual component processes by specifying operations performed on various types of information such as verbal, numerical, and figural stimuli. We will have considerably more to say about this subsequently. The second direction is to elaborate a performance theory that explicitly considers a much wider range of both item difficulty and individual ability than the relatively narrow range used in previous research.

The directions for theory development suggested above represent the program of research that we have undertaken over the past few years. Within this program

we have conducted studies that systematically examined performance in analogy tasks as a function of (1) content (figural, numerical, and verbal), (2) item difficulty, and (3) age and ability level. Latency, error, and protocol data, representing the performance of different age and skill groups in different content areas, has led to the development of a theory of performance in multiple-choice and forced-choice analogical reasoning tasks that relates directly to general theories of problem solving (e.g., Newell & Simon, 1972) and more specific theories of analogical reasoning. (e.g., Sternberg, 1977).

B. Elements of a Problem-solving Theory of Performance

Effective analogy solution behavior can be described as a series of steps toward satisfaction of a very specific and highly constrained goal: that of selecting a completion term from a set of options such that the C–D′ relation is matched with the A–B relation and is more closely matched than any of the other alternative C–D relations. Satisfaction of this goal requires satisfaction of implicit subgoals through use of three major sets of processes—relational inference (induction of structure), relational comparison (feature matching), and relative match comparison (discrimination among complex stimuli) (Heller, 1979). Execution of these processes satisfies necessary subgoals of the general or top goal stated earlier. It is necessary to identify two relations, to establish the degree of their correspondence or match, and to determine that the correspondence is greater than alternative matches. A complete process model consists of various sequences of processes by which these goals can be satisfied. The operation of these processes can be described as outlined in the following.

1. *Relational Inference.* Problem elements, i.e., the A and B terms of an item and their subparts, must first be encoded before a search for relations between problem elements is conducted. Then, in the case of verbal stimuli, information in semantic memory about features that link the A and B concepts is examined. In the case of numerical stimuli, an operation and a value are applied to one element (A) to produce the other element (B). Such information may be directly retrieved or computed. In the case of figural stimuli, spatial, logical, or numerical transformations are sought that change elements in A into their corresponding form in B. As relational links among elements are discovered or verified, the solver can be said to construct a cognitive representation of the elements and relations. The latency and accuracy of this representation will be a function of the thoroughness of the search and the extent of an individual's declarative knowledge. Representations may be changed during the course of solution either by adding or deleting item features or by discarding the entire structure from consideration and constructing a new representation. Representations of relational features must be constructed for the A–B pair and for each available C–D pair.

2. *Relational Comparison.* This can be described as a feature matching process. The outcome of relational comparison is a decision that relations either match or do not match. Given that representations of two relations have been constructed, features in both relations must be compared systematically to determine correspondence across representations. The goal of executing relational comparison is to determine whether relations correspond and the degree of correspondence. To conclude that two relations are analogous, the solver needs to identify a sufficient number of matching features; the criterion number or degree of correspondence may vary across individuals, or across items for each individual. (This process leads to final decision in true–false verification type tasks.) The latency and accuracy of decision will be a function of the completeness of the relationship representations and the thoroughness of feature matching processes, i.e., whether a sufficient number of features are present in the representations and whether they are compared as exhaustively as necessary. In the case of complex rules or relational structures, the feature matching process may produce an incorrect evaluation due to loss or degradation of information in working memory.

3. *Relative Match Comparison.* This can be described as an optional discrimination process. Given that two or more relations have been judged to be aligned with a referent relationship, a comparison of their relative match is performed to determine the one closest to the referent. Conversely, if matches have not been found, the alternatives must be re-examined to determine which relation comes closest to matching the referent relation. Essentially, discriminations must be made among alternative relations in order to select a best choice.

The outcome of match comparison processes is a decision to accept one option on the basis of the greatest degree of alignment with the referent relationship. The accuracy of this decision will be a function of the accuracy and completeness of representational and relational comparison processes, in addition to the thoroughness with which feature matches are compared. When processes do not yield definitive outcomes, i.e., when subgoals are not satisfied, additional sequences of processes must be invoked. These activities include re-execution of relation identification processes when relational comparison or match comparison is indefinite. In some cases where the initial relation identification process fails, the subgoal of identifying the A–B relation must be satisfied in some way for the analogical top goal to be reached. As a result, the search for an A–B relation continues throughout option examination. Relationship identification processes are executed for each C–D pair. Each C–D relation is then considered in conjunction with the A and B terms to determine whether it is applicable. If a representation of the A–B relation is constructed as a result of this process, the usual processes continue through to solution.

If no A–B relation is identified, then the analogical top goal must be suspended and relationship identification processes are relied upon to identify any

B-D features that could aid in selecting the most likely completion term. However, this does not actually satisfy the goals of either identifying two relations, establishing that the relations are aligned, or selecting a completion term that results in the two most closely aligned relations. Similarly, if match comparison processes do not yield a clear determination of the most closely matched relation, and no further relationship modifications can be determined, then final option selection must be arbitrary. In these cases, the subgoal of identifying the best possible match has not been satisfied.

Our discussion of the processes required for analogy solution has implicitly represented performance in terms of an "infer-infer-compare" solution proce-

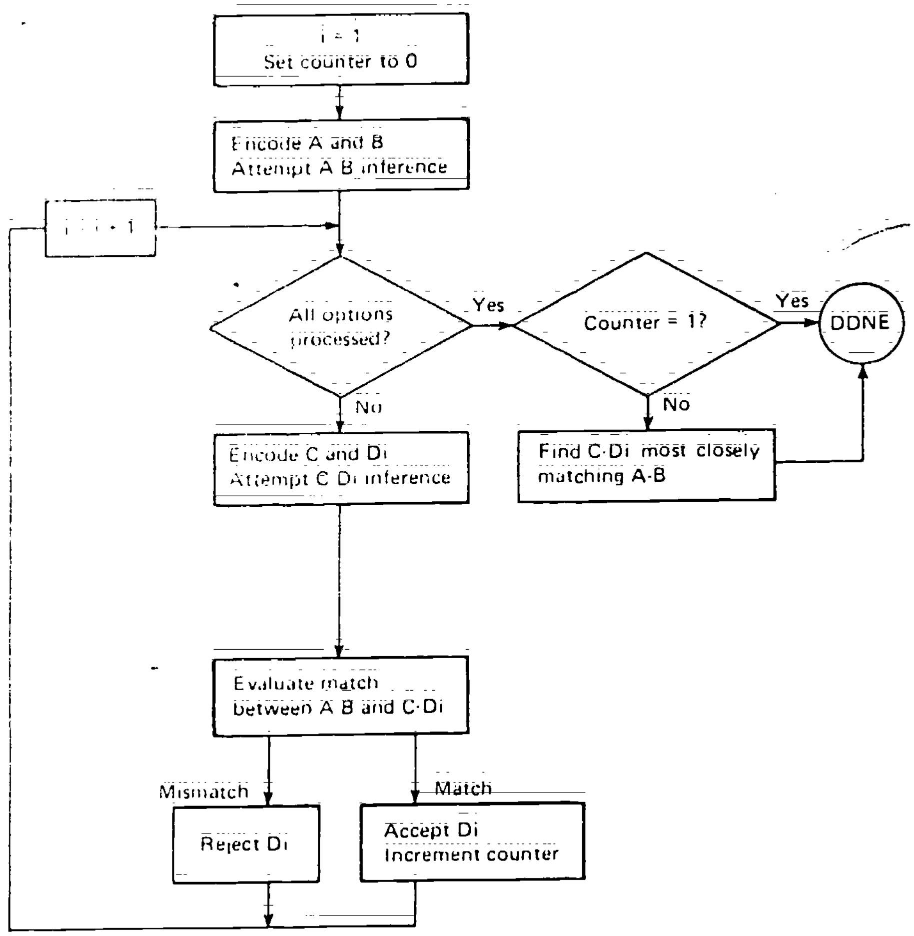


FIG. 5.2. Flow diagram of conceptually driven solution sequence.

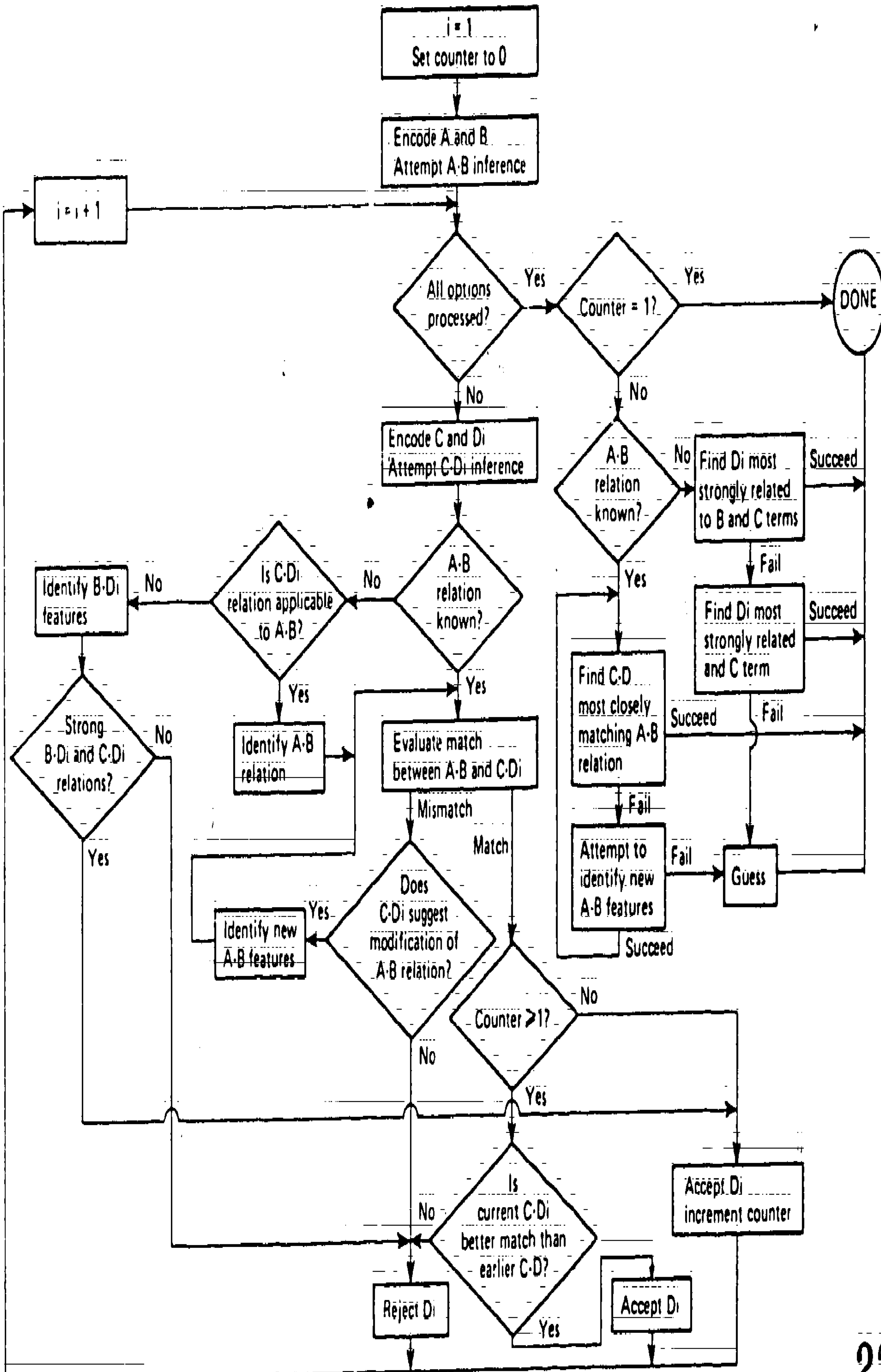


FIG. 5.3. Flow diagram of possible interactive solution sequences.

dure. That is, an A-B relation is identified, and each C-D relation is identified and compared against the A-B relation. Additional processing to achieve solution includes relative match comparison, modifications of A-B or C-D relations by execution of repeated relationship identification processes, and processing to identify the A-B relation after an initial failure to do so. In our data, it was found that these processes are often executed repeatedly within a single solution episode. When a goal is not satisfied, it can be pursued through iterative or alternative sequences of processes.

Two general types of analogical solution sequences have been identified. In "conceptually driven" solutions, the A-B relation is initially identified, each option is evaluated by comparison of relations, and one completion term is selected without revision or modification of the initially identified rule. A general flow diagram for conceptually driven solutions is shown in Fig. 5.2. This particular solution sequence is representative of performance on easy and unambiguous items where there is little problem in identifying an A-B rule, and the information in the original rule specification is sufficient to discriminate among the available options. Thus, at no point during the course of solution is there a problem in satisfying any of the necessary subgoals.

The second general type of solution sequence is "interactive," where identification or modification of the analogical rule or A-B relation is a result of information obtained in the set of completion terms. That is, whether or not a rule was initially identified, identification of C-D relations cues recognition of alternative A-B relations. In Fig. 5.3, the additional processes required for interactive solutions have been added to the flow diagram for conceptually driven solutions, and thus the figure represents the many possible analogy solution sequences that may be observed. The additional components of the model are a function of the interactive and recursive nature of solution in cases where there is a failure to initially satisfy the goal of identifying an A-B relation, or a failure to achieve matching relations because of an inappropriate or incomplete A-B relationship. The representation in Fig. 5.3 is intended to reflect performance on a broad range of item difficulty. In fact, easy items tend to evoke conceptually-driven solutions, while difficult items often require complex and interactive solution sequences. We will discuss this in more detail in the section on verbal analogy performance.

V. COMPONENTS OF FIGURAL ANALOGY SOLUTION

A. Overview

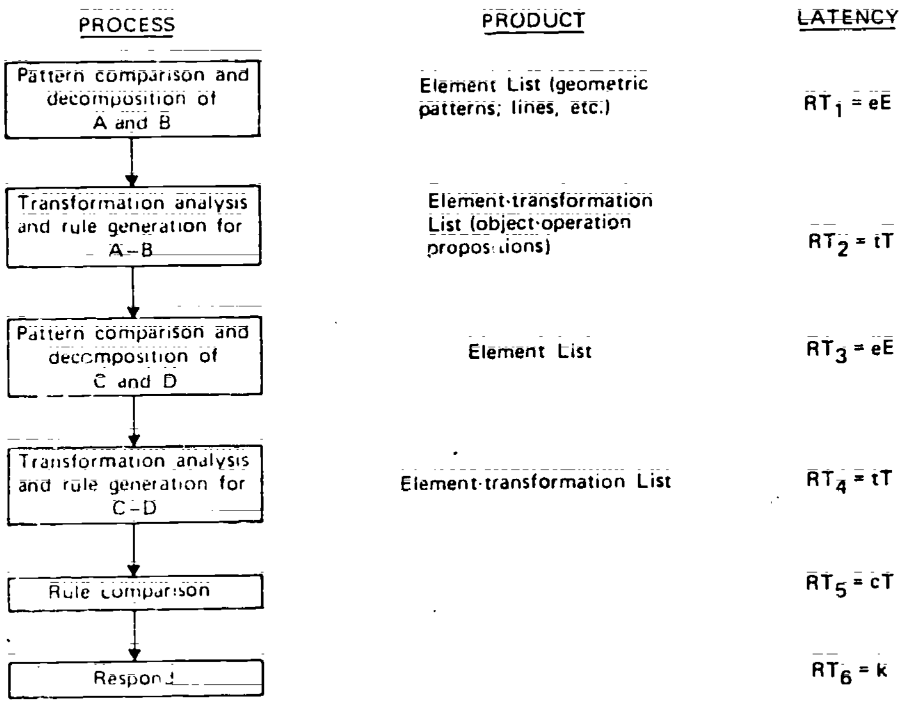
This is the first of three sections that discuss data and theory on the processing of specific types of analogies. We discuss figural analogies first because it is easier to describe and analyze the type of item features in these problems; in contrast to

the symbolic aspects of numerical and verbal analogies, the information necessary for item solution is externally represented in the physical problem array. The elements of a figural analogy are directly represented in the figures or patterns shown and the relational rules are based upon spatial and logical transformations of these physical features.

In the foregoing section on theories of analogy, we provided an overview of the components of a generalized solution model for all analogy tasks. The purpose of this section is to relate the specific task of solving figural analogies to that more general theory. Our goal is to show how the general theory can be instantiated in the form of process models for the solution of a specific item type. We will first review relevant models of figural analogy solution that attempt to relate item features to cognitive processing activities and that provide a skeletal model for item feature processing. We then discuss data relevant to such a model and attempt to link together data and theory on item features, response latency, and errors in a theory of performance in figural analogy tasks.

B. Problem Features and Processing Models

As noted earlier, Sternberg (1977) has provided a general theory of analogy solution that is applicable to the solution of figural analogies. A detailed specification of the representational and processing assumptions necessary for the solution of figural analogies has also been provided by Evans (1968). His theory is embodied in a computer program that was intended as an artificial intelligence analysis of task requirements. It was not, however, intended as a theory of human performance. Evans' model represents an "infer-infer-compare" solution procedure, consistent with our earlier theory discussion. In Evans' theory, the major processes can be described as (1) pattern comparison and decomposition (encoding or representation), (2) rule generation and matching (relational inference and relational matching), and (3) rule discrimination (relative match comparison), which is optional. The processes in Evans' theory are tied directly to basic aspects of item content. The specific aspects of item content are: (a) the individual elements used to construct the separate analogy terms, and (b) the individual spatial and logical transformations applied to the elements to construct overall rules. The elements are easily perceived plane geometric figures such as lines, circles, triangles, and quadrilaterals. The basic transformations include: removing or adding elements; rotating, reflecting, and displacing elements; size changes; and variations in element shading. The relationship between item content and processing operation is illustrated in Fig. 5.4, which shows a simplified process model for the verification of the truth or falsity of a completed analogy. As shown in the model, we assume that there is an initial pattern comparison and decomposition process that yields units of information that represent the individual elements involved in a pair of analogy terms. The time to execute such a process should be a function of the number of elements or dimensions that must



$$RT_{total} = xE + yT + k$$

FIG. 5.4. Simplified process model for figural analogy verification task. From "Components of Geometric Analogy Solution" by T. M. Mulholland, J. W. Pellegrino, and R. Glaser, *Cognitive Psychology*, 1980, *12*, 252–284. Copyright 1980 by Academic Press. Reprinted by permission.

be isolated. Thus, as shown in Fig. 5.4, the total time for this initial stage of processing (RT_1) is the product of the average time to represent a single element (e) and the total number of elements (E) to be represented.

The second stage of processing involves transformation analysis and rule generation. This phase of processing attempts to determine the transformations (operations) that specify the rule for changing the A stimulus into the B stimulus. The outcome of such a process is assumed to be a propositional list in working memory that represents element-transformation pairs, i.e., an object-operation list. The time to identify and order a set of transformations should be a direct function of the number of transformations involved in an item. The total time for this second stage of processing (RT_2) is the product of the average time for a single transformation (t) and the total number of transformations (T).

The two components of processing; (1) pattern comparison-decomposition, and (2) transformation analysis-rule generation, occur more than once, since analogies contain two separate pairs of terms. They are followed by a final process involving rule comparison or matching in which the propositional lists are compared to determine if there is equality or correspondence of rules. This process should also be a function of the number of transformations that describe the rule for a specific item. Thus, the total time for solution of the analogy verification task should be the combination of the separate times for the individual processing stages. This can be reduced to the following simple expression: $RT = xE + yT + k$. This assumption about item representation and processing was tested in an experiment that systematically varied item content (Mulholland et al., 1980). The outcomes of that experiment and the implications for more specific processing issues are summarized in the next section.

C. Data and Theory on Item Processing

In the study conducted by Mulholland et al. (1980), a set of analogies was constructed in which the number and type of elements and transformations were systematically varied across items. The items were presented in a true-false verification format, and latency data were used to evaluate the hypothesized processing model shown in Fig. 5.4. The analogies used were generated from six types of elements and six types of transformations that frequently occur in items found on aptitude tests. Figure 5.5 gives some examples of the types of true and false items. The true items had from one to three elements in each analogy term, and the rules were based on zero to three transformations of the elements in the analogy terms. Items were made false by adding incorrect element information or by replacing correct transformations with incorrect ones in the C–D pair. The items were presented to 28 undergraduates who had been previously administered a standardized test consisting of 25 multiple-choice items (see Mulholland et al., 1980, for additional procedural details).

1. *Processing True Analogies.* Verification of the truth of an analogy requires exhaustive search of all the element and transformation information in the A–B and C–D pair of terms. Thus, as shown in Fig. 5.4, the time to solution should be a monotonic function of increases in the structural complexity of items. The important issue in this regard is the absolute and relative amount of time that is associated with element versus transformation processing and whether the effects of these two factors are independent and additive, as represented in the model, or interactive. The basic reaction time data are shown in Fig. 5.6. Increasing the number of elements systematically increased the time to solution as predicted. This suggests that the patterns comprising the terms of the analogies were decomposed element by element as hypothesized. The rate of processing elements was nearly constant (additive) within each transformation condition.

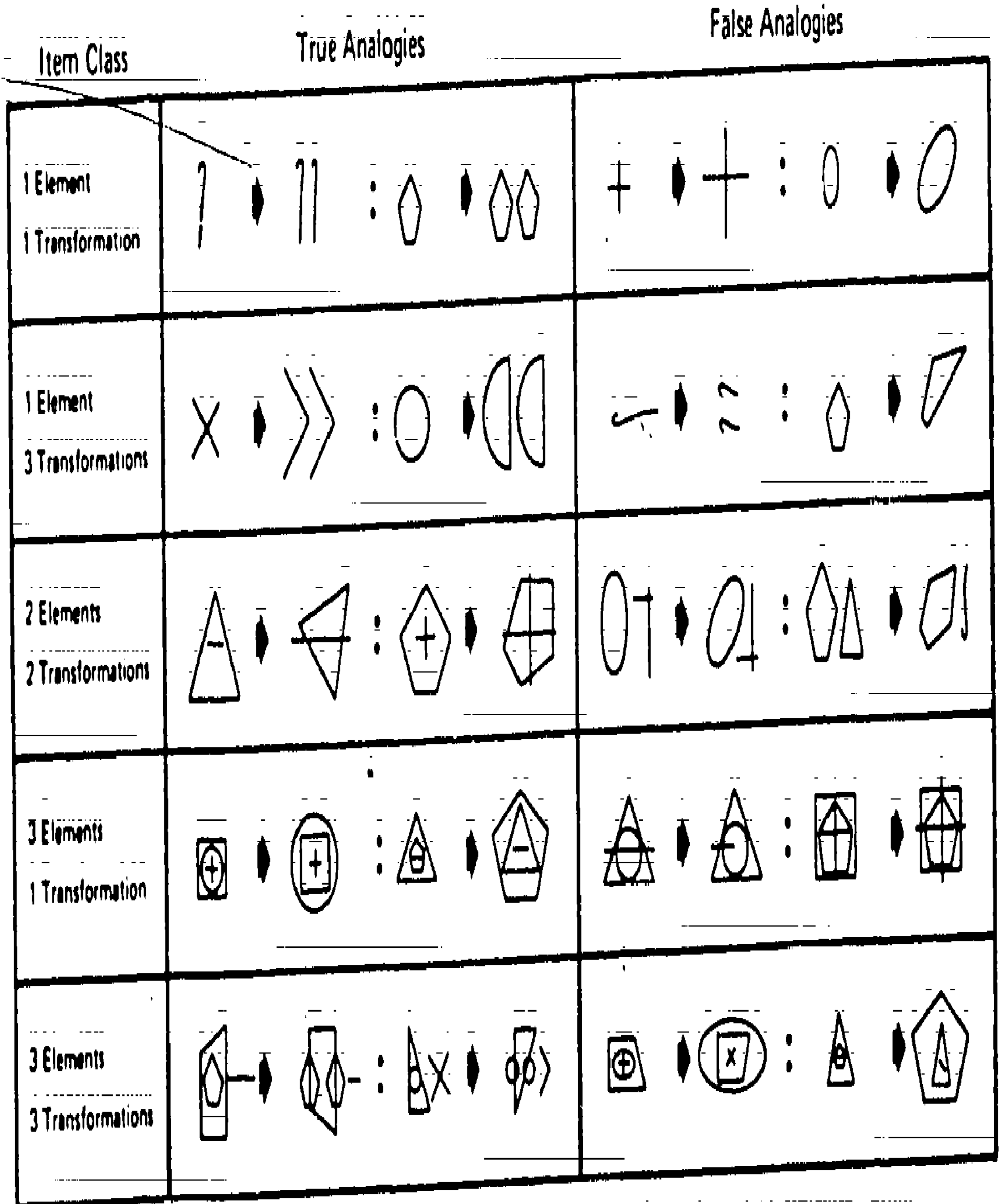


Fig. 5.5. Examples of figural analogies that show element and transformation structure. From "Components of Geometric Analogy Solution" by T. M. Mulholland, J. W. Pellegrino, and R. Glaser, *Cognitive Psychology*, 1980, *12*, 252-284. Copyright 1980 by Academic Press. Reprinted by permission.

but it varied as a function of the total number of transformations. Likewise, transformations appear to have been processed in a serial fashion as solution proceeded additively within each number of elements condition, but increasing in time per transformation with increases in the number of elements.

The data shown in Fig. 5.6 are generally consistent with the model shown earlier, but the interaction of element and transformation processing was not predicted. The predicted points shown in Fig. 5.6 are based upon the following function, which accounted for 97% of the variance in the group mean data: $RT(\text{msec}) = 425T + 358E + 75TE + 797$. Our explanation for the violation of simple additivity is in terms of capacity or resource limits in item processing. As item complexity increases, there is a problem of mental bookkeeping that will begin to draw upon a limited capacity processing system. Each operation per-

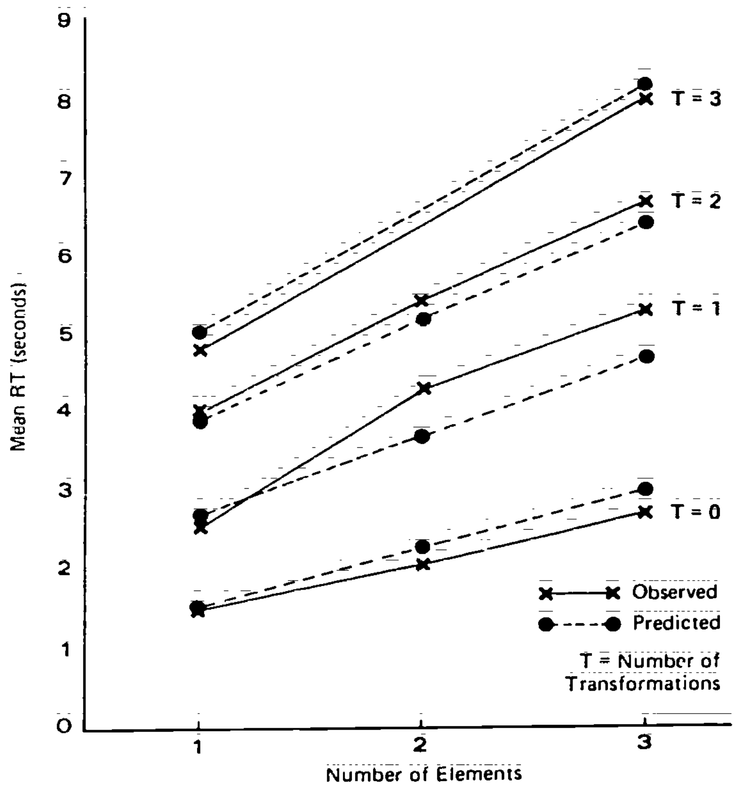


FIG. 5.6. Mean reaction time for true analogies as a function of elements and transformations. From "Components of Geometric Analogy Solution" by T. M. Mulholland, J. W. Pellegrino, and R. Glaser, *Cognitive Psychology*, 1980, *12*, 252–284. Copyright 1980 by Academic Press. Reprinted by permission.

formed in decomposing patterns and identifying transformations in the analogy terms yields units of information that require space in a working memory system. As more partial information is accumulated and entered into working memory, one may begin to approach the limits of this system. When this occurs, processing time and effort may have to be deployed in the service of updating and maintaining the contents of working memory. Note that the overall solution times vary over a range of seconds, requiring that solution information be available in memory and not become degraded for more than just a brief period.

Given these assumptions and data, we argue that the processes involved in pattern comparison-decomposition and transformation analysis-rule generation can be considered to be essentially additive, but the factor of memory load, which increases at a nonlinear rate, produces the increasingly longer times for complex items. When extrapolated to even more complex items than those used in this study, the mental bookkeeping operations become more apparent and may well take up the largest share of time. As both elements and transformations increase, solution may require substantial external memory that is not available, thereby creating a need for alternative processing strategies that are time consuming and require a considerable amount of conscious control and monitoring. This would represent a shift in the proportion of the total solution time that goes to the actual processing operations as compared to that required for information management.

2. *Processing False Analogies.* In order to discuss performance on false analogies, it is necessary to elaborate the model shown in Fig. 5.4 to incorporate representational and processing assumptions about element and transformation features. Figure 5.7 shows a more detailed model that is intended to represent assumptions such as the sequential and self-terminating processing of element and transformation features. The leftmost section of Fig. 5.7 represents the processing of elements and transformations in the A-B pair of terms. The middle section of the figure represents the components of C-D processing and overall rule comparison. This section of the model incorporates a self-terminating processing strategy, since in contrast to true items, it is possible to terminate processing of false items at the point where incorrect information is first detected.

The data on false item processing clearly supported a model incorporating a self-terminating solution procedure. The left panel of Fig. 5.8 shows the predictions based on this self-terminating processing model in terms of the number of elements that would be processed before each item type could be declared false. Items were made false by the replacement of correct elements in the C-D pair with incorrect ones. An exhaustive processing strategy would yield flat functions for the different conditions shown in Fig. 5.8. The right panel of Fig. 5.8 shows the mean reaction time for each type of incorrect element item. The data strongly support a model that assumes a strategy of process termination followed by response whenever incorrect element information is detected in the problem.

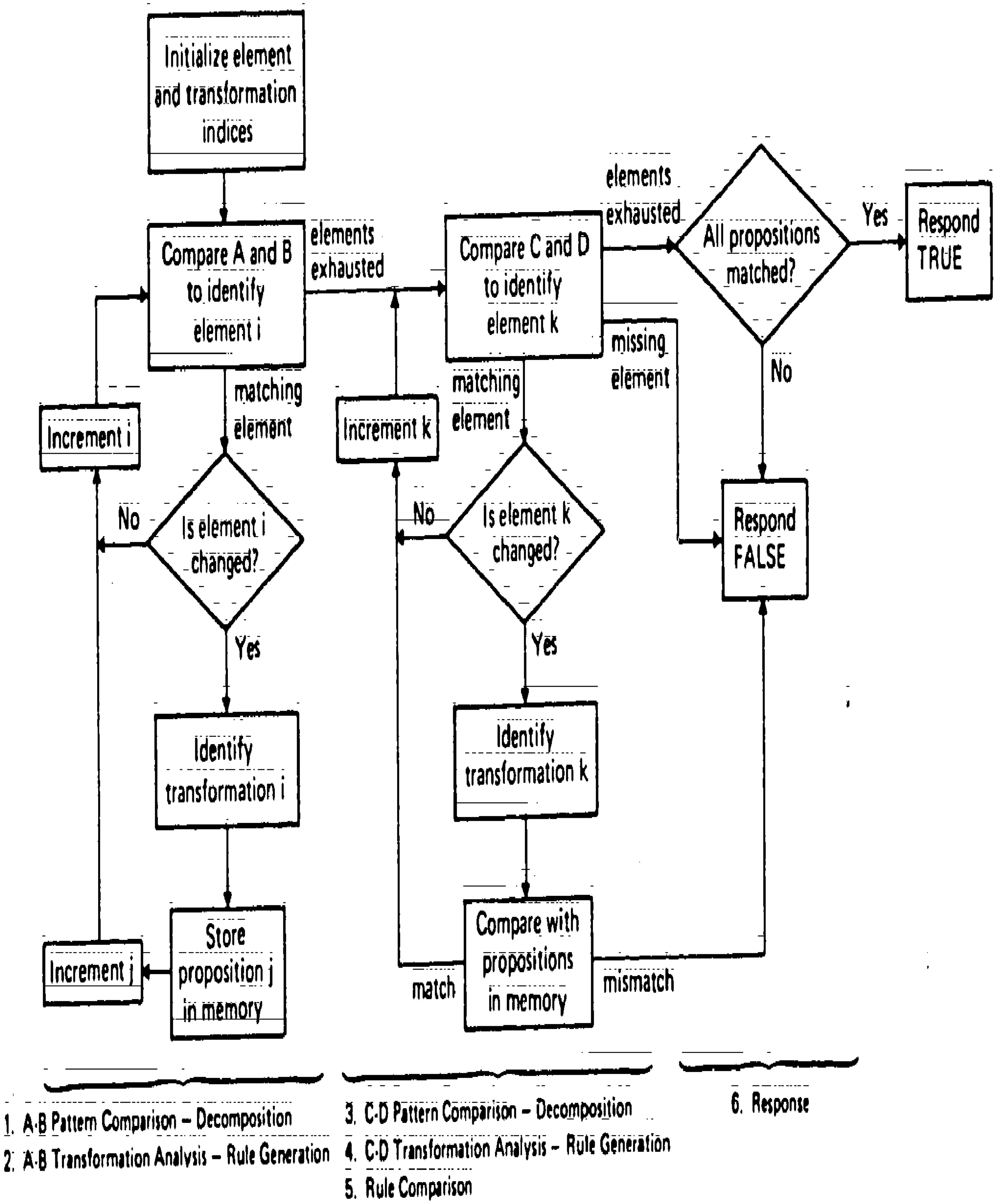


Fig. 5.7. Expanded process model for figural analogy verification task. Adapted from "Components of Geometric Analogy Solution" by T. M. Mulholland, J. W. Pellegrino, and R. Glaser, *Cognitive Psychology*, 1980, *12*, 252-284. Copyright 1980 by Academic Press. Reprinted by permission.

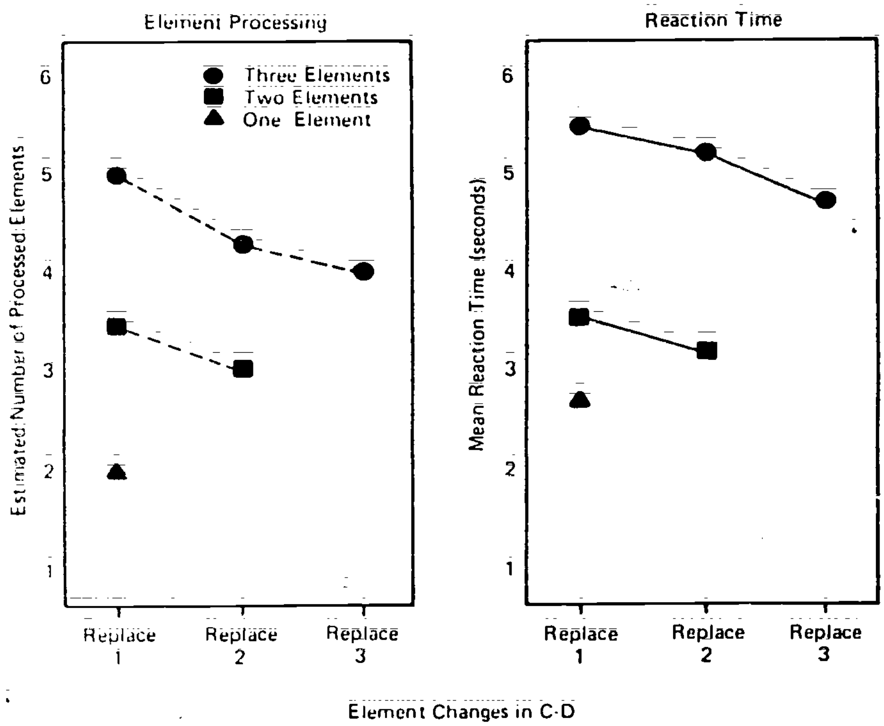


FIG. 5.8. Predicted and observed performance on analogies containing incorrect elements. Adapted from "Components of Geometric Analogy Solution" by T. M. Mulholland, J. W. Pellegrino, and R. Glaser, *Cognitive Psychology*, 1980, *12*, 252–284. Copyright 1980 by Academic Press. Reprinted by permission.

A similar analysis of processing was possible for the items that were made false by introducing incorrect transformation information. The upper panels of Fig. 5.9 show the predictions for both self-terminating and exhaustive processing strategies in terms of the number of transformations that would have to be processed to declare an item false given its transformational structure. The bottom panels of Fig. 5.9 show the mean rejection time for each type of item. Again, the data support a serial self-terminating processing strategy over an exhaustive processing strategy.

Thus, to summarize the data on item feature processing, it is clear that the major content components of figural analogy problems are systematically processed in a manner consistent with the detailed model shown in Fig. 5.7. The overall strategy represents an efficient, self-terminating solution mode that allows for the rapid rejection of incorrect items. Although not discussed here, the data also support a model in which element and transformation information is

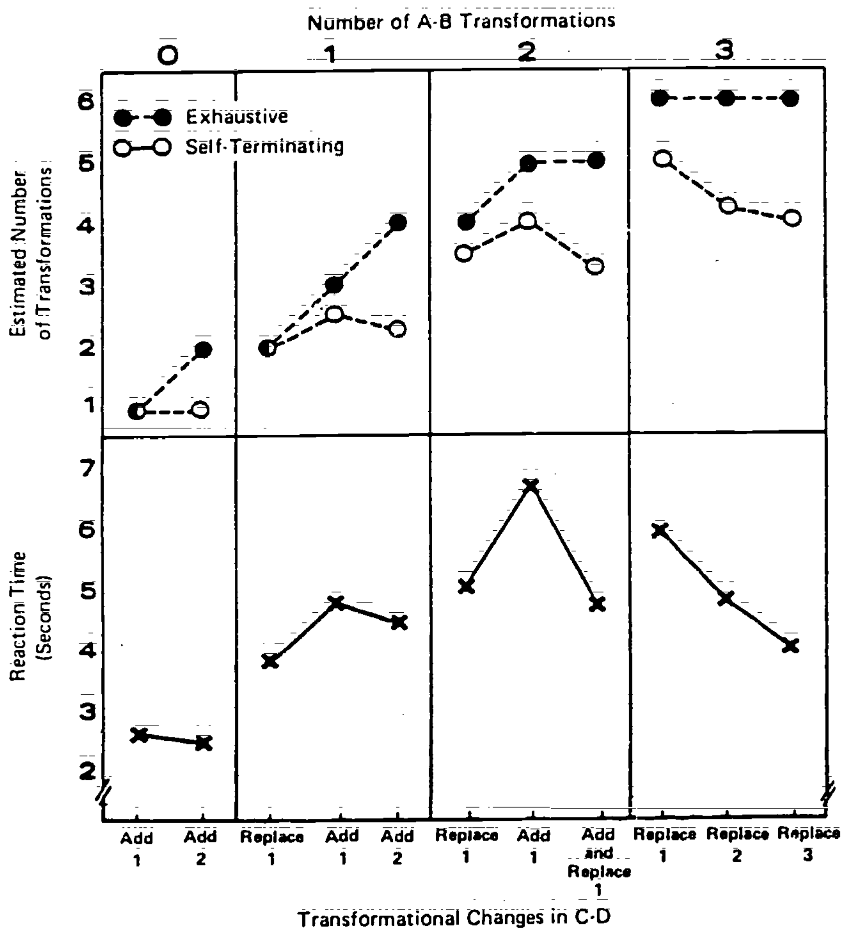


FIG. 5.9. Predicted and observed performance on analogies containing incorrect transformations. Adapted from "Components of Geometric Analogy Solution" by T. M. Mulholland, J. W. Pellegrino, and R. Glaser, *Cognitive Psychology*, 1980, *12*, 252-284. Copyright 1980 by Academic Press. Reprinted by permission.

processed in a pairwise sequential manner, with element information being processed before transformation information.

D. Item Features, Errors, and Performance

As in the case of latency data, theoretical assumptions about feature processing can be related to the error data to contribute to an overall theory of performance. The critical error data are shown in Fig. 5.10, which represents performance on

the true items. The error data clearly show that the major factor leading to verification errors involves the transformations that serve to define the rule for an item. There appears to be two ways in which increases in rule complexity (transformational complexity) lead to increased errors. In the two and three element items, where transformations are mapped one-to-one onto individual elements, the data show an independence and additivity of error probabilities associated with the separately transformed elements. In such cases, the increased error rate on more complex problems can be understood in terms of a simple accumulation of independent, incorrect representations of transformed elements, any one of which could lead to an overall incorrect response. For most adults this error rate should be low, particularly given the types of transformation used by Mulholland, et al.

The other way that rule complexity affects error rate is when multiple transformations of a single element are required. The probability of this type of error is much larger, and it may be due to the amount of information that must be retained in working memory. Assuming that each transformation applied to an

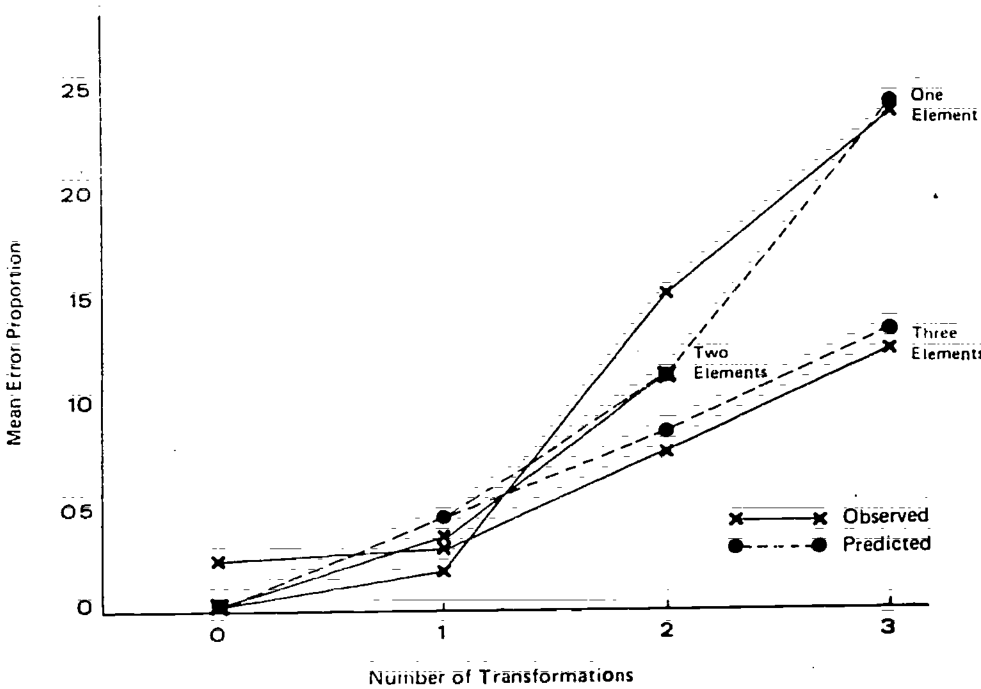


FIG. 5.10. Predicted and observed error rates for true analogies. Adapted from "Components of Geometric Analogy Solution" by T. M. Mulholland, J. W. Pellegrino, and R. Glaser, *Cognitive Psychology*, 1980, *12*, 252–284. Copyright 1980 by Academic Press. Reprinted by permission.

element requires at least one placekeeper in working memory (see Kotovsky & Simon, 1973), an item with two transformations of the same element will require at least two and possibly three memory placekeepers, since the order or sequence of applying transformations may constitute a third component of rule information. Additionally, in the one element multiple-transformation case, testing and memory must be internal for all but the final product of the application of several transformations. Thus, an additional memory placekeeper would be required for storing the intermediate products of solution processes during the evaluation of the C–D pair. These intermediate products that have to be stored, in addition to the other record keeping required, tend to further tax memory capacity, thereby increasing error.

Memory load problems and loss of information can also occur in items with several separately transformed elements. This assumption, along with those specified in the preceding, was used to derive the following general function for error probability: $P(\text{error}) = 1 - (1-\alpha)^T(1-\lambda^{M-\lambda})$ where:

- α = the probability of incorrectly representing or applying a given transformation
- T = the total number of transformations that determine the rule for an item
- λ = the limit or maximum amount of information that can be held in working memory
- M = the number of memory placekeepers required during the solution

This function was used to fit the data in Fig. 5.10 with best fitting values of α equal to .044 and λ equal to 5.8. The predicted values from the best fit are shown in Fig. 5.10 and the R^2 for the fit between observed and predicted was .93. The value of α is low as would be expected given that it represents the probability of misrepresenting or misapplying a simple spatial or logical transformation. The value of λ is quite close to the magical number 7 that is often cited as the capacity limit of short-term or working memory. Although the specific value of λ may vary for a group or an individual, the function described above captures the fact that the probability of information loss, and thus problem error, increases very rapidly as one approaches the limits of working memory. Thus, the true item latency and error data suggested that working memory factors associated with the representation and management of item features appear to provide the basis for nonadditive increases in solution latency as well as significant increases in the probability of error for certain levels of transformational complexity.

The analysis of the relationships among item features, solution latency, error rates, and processing models supports a general theory of figural analogy solution that can be summarized in terms of assumptions about processing, representation, and memory storage. In the first phase of processing, the A and B terms of the analogy are globally encoded, and the two internal representations are compared to determine specific elements or subpatterns common to both items. This process may be viewed as parsing the figure and determining the

most appropriate level of element representation. Such a pattern comparison-decomposition process is necessary since there are often multiple representations of patterns, and the relevant attributes of a figure can only be ascertained in the presence of a comparison stimulus. The relevance of an attribute is determined by a transformation as represented by the B term, which creates an encoding context for the problem. The comparison-decomposition process provides the basic information necessary for the execution of the subsequent inference and rule generation process during which the precise transformation(s) applied to each element or subpattern is determined. This process attempts to define an operation that is capable of changing an element in A into its corresponding form in B. The inference and rule generation process is assumed to be exhaustive for all the elements in the A-B pair. The outcome of the inference and rule generation process can be conceived of as an object-operation list, i.e., a propositional list, stored in working memory. This list provides the basis for evaluating the C-D half of the item.

The next major phase of processing also involves encoding, pattern comparison, decomposition, and inference processes, and, in addition, a rule comparison process. The C and D terms must be encoded and then decomposed into basic elements that correspond to those specified for A and B. If a basic element of C is not present in D, then processing terminates and the item is declared false. If an element is present, but the inferred transformation does not match the value stored in memory, then processing terminates and the item is declared false. In the event of a successful match, a new element-transformation pairing is tested and this continues until no further elements remain to be considered. At this point, the item is declared true. This description of processing is consistent with all the latency data for the true and false items obtained for adults (Mulholland et al., 1980) and children (Bisanz, 1979).

The general processing assumptions described in the preceding are also consistent with the error rates associated with certain item structure manipulations. The largest single source of error was multiple transformations of single elements. If the preceding model is applied in this case, then the intermediate results of the C-D inference process must be retained in memory and the entire transformation sequence inferred before the truth value of the D term can be judged. Thus, additional demands on working memory may cause some of the original element-transformation information to be lost or degraded.

Within the theory, it is assumed that each operation associated with pattern decomposition and transformation analysis of an A-B pair yields a unit of information that needs to be stored in working memory. The information can be conceived of as a list of element-transformation or object-operation propositions, as in the Simon and Kotovsky (1963) notational system for series problems (see also Rumelhart, 1977). As the number of transformations in a figural analogy problem increases, the load on working memory can become substantial and give rise to errors. Increases in memory load may also require the individual to

allocate substantial processing resources to activities designed to avoid or reduce information loss. Young children (Bisanz, 1979) and less proficient solvers may be particularly inefficient in these aspects of performance. It is important to note that the memory load explanation has been verified by empirical studies of performance on series extrapolation problems (e.g., Holzman, et al., 1976; Kotovsky & Simon, 1973). A similar set of representational, processing, and working memory assumptions may also be applicable to performance differences across items on figural matrix problems from the Raven's Progressive Matrices Test (e.g., see Hunt's, 1974, analysis of this task).

VI. COMPONENTS OF NUMERICAL ANALOGY SOLUTION

A. Overview

In our preceding discussion of figural analogy solution, we largely ignored the issue of the declarative knowledge base necessary for solving such problems—the elements of figural problems are readily identified patterns that can be presumed to have some universality over individuals and, perhaps, cultures. The solution of numerical analogy problems, however, requires a consideration of the knowledge base necessary to represent both the individual numerical stimuli and the relations between pairs of numbers. Such knowledge is variable across individuals of the same age, depending on their background and experiences, and it is highly variable across age groups. In addition, the organization of this knowledge may differentially affect performance across individuals and age groups.

In numerical analogy problems, there are also a variety of simple and complex relations between pairs of numbers that introduce a problem of representational variability or ambiguity in the analogical rule. Consider for example the pair 2:16, which can be represented as having several relationships, e.g., +14, ×8, or 2^4. Contrast this with the pair 2:15, which can be simply represented as +13. It is because of such possibilities, which relate directly to both the information represented in semantic memory and the way in which that information is accessed, that the multiple relational features of such problems require a careful analysis and must be accommodated in any process model and performance theory.

In this section, we will first present an analysis of problem features as related to a general process model for task performance. This serves as the context for our discussion of data and theory on item feature processing. The results are then related to assumptions about the storage and retrieval of quantitative information in permanent memory. Finally, we will consider the relationships among item features, errors, and performance theories. One aspect of our analysis of these relationships will be aspects of developmental changes in performance and their apparent relationship to changes in knowledge structures as a function of age and schooling.

B. Problem Features and Processing Models

Numerical analogy solution has not been the subject of any intensive empirical analyses or artificial intelligence theories. It is possible, however, to use our previously stated theory of analogy solution as the basis for describing a processing model of the infer-infer-compare type and apply it to the case of numerical analogies with the form A:B :: C:D :: E:_____. As we illustrated earlier, numerical analogies require two number pairs to specify and disambiguate the rule for the item. In the prototype format shown here, this rule is then applied to another number to complete the third pair. Figure 5.11 (ignoring the loop, which will be discussed later) is a simplified process model for solution of problems of this type. As shown in the model, the A and B terms must be encoded and an inference must be made about the rule that relates the two members of the pair. This rule is an operation or set of operations and a specific value for each operation, e.g., for the pair 2:5 it would be +3.

The next major phase of processing involves similar encoding and relational inference processes for the C:D pair. This stage of processing can be contextually insensitive: The type and value of the A–B relationship will not influence the inference process for the C–D pair. Or, the process can be a contextually sensitive or guided one that attempts to evaluate the applicability of the A–B relational information for the C–D pair. In either case, the outcome of this stage of processing is a match or mismatch of relational information or rules. In the case of a match, the rule can be applied to the E term to generate a response (i.e., a conceptually driven solution). In the case of a mismatching relationship, the inference process has to be re-executed for either or both the A–B and C–D pairs and another attempt has to be made to match information to find an acceptable rule (i.e., an interactive solution). The various possibilities for the case of mismatching relationships will be developed later in more detail.

The processing model provides a basis for considering the problem features that influence solution and the locus of their effect. Unlike figural analogies, the terms of numerical analogies are individual whole numbers and they are not composites of separate elements. Thus, the elements of the problem are straightforward and their encoding involves activation of conceptual knowledge about the properties of the particuair element: odd-even, prime, perfect square, etc.

The most critical feature of numerical analogy problems involves the relationships governing the problem as a whole and the possible relational information associated with the A–B and C–D pairs. It is possible to have problems where the overall rule is addition, but that differ in terms of other relational information that may affect solution. Consider two problems. In the problem 4:9 :: 7:12 :: 2:_____, the rule is +5 and no other simple relational information is present to influence solution. Thus, the solution of such an item should involve a straightforward, conceptually guided solution as shown in Fig. 5.11. Contrast the solution of the following problem with the one above: 6:30 :: 3:27 :: 2:_____.

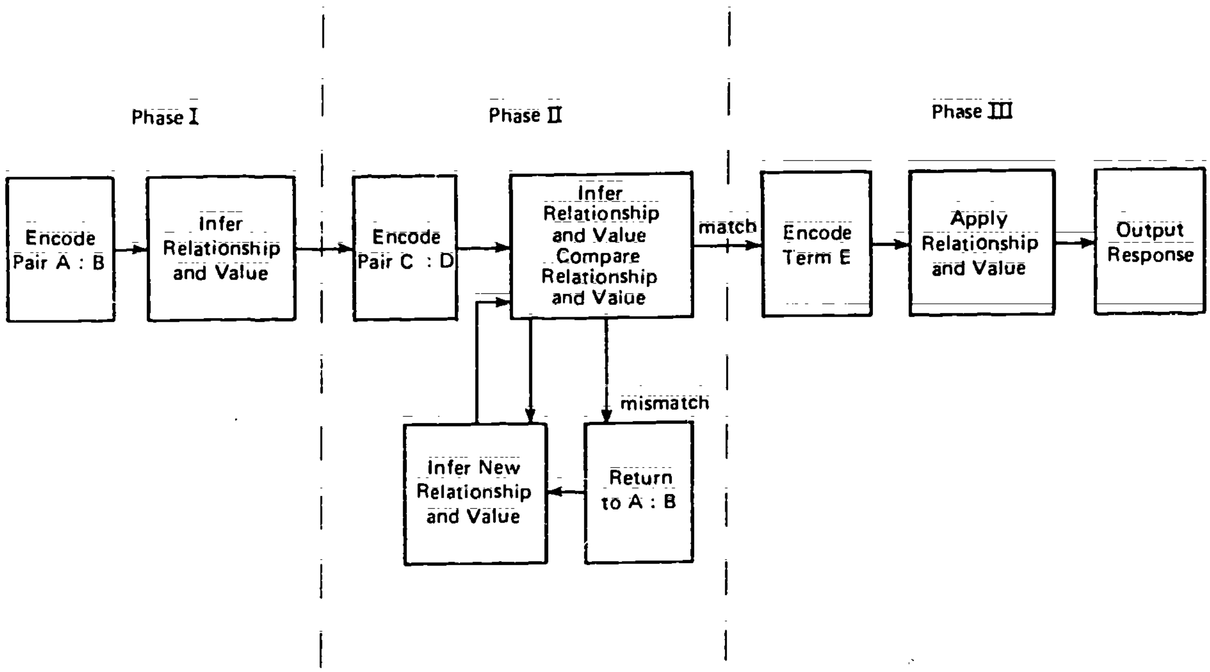


FIG. 5.11. Simplified process model for numerical analogy task.

The rule is + 24 but there is other simple relational information in both the A–B and C–D pairs. The influence of such information on solution can be minimal or substantial depending on the organization and automaticity of retrieval of certain types of relational knowledge. Thus, it is very likely that the information about multiplicative relationships ($6 \times 5 = 30$ and $3 \times 9 = 27$) will influence either or both the A–B and C–D relational inference process and lead to more difficult and complex solutions of the interactive type.

When there are several simple relations in both the A–B and C–D pairs, the organization and automaticity of retrieval of certain types of simple relational knowledge will affect item difficulty. Similarly, problems with either a low salience (e.g., powers, roots, proportions) or compound rule (e.g., $\times 2 + 3$) will also be more difficult to solve since there are a variety of simple types of relationships that are likely to be inferred, evaluated, and rejected prior to considering relationships such as cubes, cube roots, noninteger multiplication, or compound rules. These types of problems are not curiosities but represent actual manipulations of problem features on tests. Thus an important feature of numerical analogy problems is the type of relational knowledge required for solution and the number, sequence, and locus of processing operations involved in identifying or accessing that information.

The preceding issues about item features, information retrieval, and processing components were addressed in a study that presented simple analogy problems to college students of high and low aptitude. In this study, the testing format made it possible to derive separate latency estimates for each phase of processing (Pellegrino, Chi, & Majetic, Note 2). Table 5.4 shows the procedure for successively adding problem information and obtaining separate latencies during each phase of processing. The problems represented a variety of possible relationship

Table 5.4
Incremental Procedure for Presenting Numerical Analogy Problems

	Stimuli	Response	Latency Components
Phase I	4:7	Button press indicating inference of a relation	Encoding + Relational Inference
Phase II	4:7 8:11	Button press indicating determination of problem rule	Encoding + Relational Inference + Relational Comparison (+ additional inference)
Phase III	4:7 :: 8:11 :: 15:__	Keyboard responce indicating first digit of response	Encoding + Relational Application

types and values and allowed us to test assumptions about the order of accessing information in memory and the role of such information in subsequent phases of processing.

C. Data and Theory on Item Feature Processing

Problems were classified in terms of the global processing sequence governing solution, i.e., conceptually driven versus interactive solutions. For problems involving a conceptually driven solution, the initial relational inference for the A–B pair was applicable to the C–D pair, allowing the individual to move rapidly to the final application stage of solution. For problems with interactive solutions, the overall relation governing the problem was not immediately accessible due to the presence of more salient or competing relational information in the A–B or C–D pairs of terms. We will consider performance on these two general classes of problems separately to illustrate different points about item feature processing and knowledge retrieval. We will also focus on differences in the college sample that was tested. The 16 subjects were divided into a high and low aptitude group based upon SAT scores, the high group having an average score of 675 and the low group an average of 524.

1. *Conceptually Driven Solutions.* Table 5.5 shows the type of inferred relationships in the problems with conceptually driven solutions. Successful performance on problems of this type requires an available quantitative knowledge base that contains either direct declarative knowledge of specific interrelationships between pairs of numbers, or procedures for determining values and types of relationships that may exist. We have assumed that the necessary declarative and procedural knowledge is available in the college student sample and that simple problems of this type will not pose any major difficulty. However, there may be systematic individual differences in the speed of performing the processing activities necessary for solution of even simple analogy problems.

Problems requiring conceptually driven solutions had high levels of accuracy in both skill groups. In Fig. 5.12, the similarities and differences between skill

Table 5.5
Examples of Problems Generally Resulting in Conceptually Driven Solutions

Problem Type	Example
Simple Addition (single integer)	4:9 :: 7:12 :: 2:__
Complex Addition (multiple integer)	3:19 :: 6:22 :: 5:__
Simple Multiplication (single integer)	8:16 :: 5:10 :: 4:__
Complex Multiplication (multiple integer)	3:51 :: 4:68 :: 5:__
Squares (single integer)	4:16 :: 8:64 :: 3:__

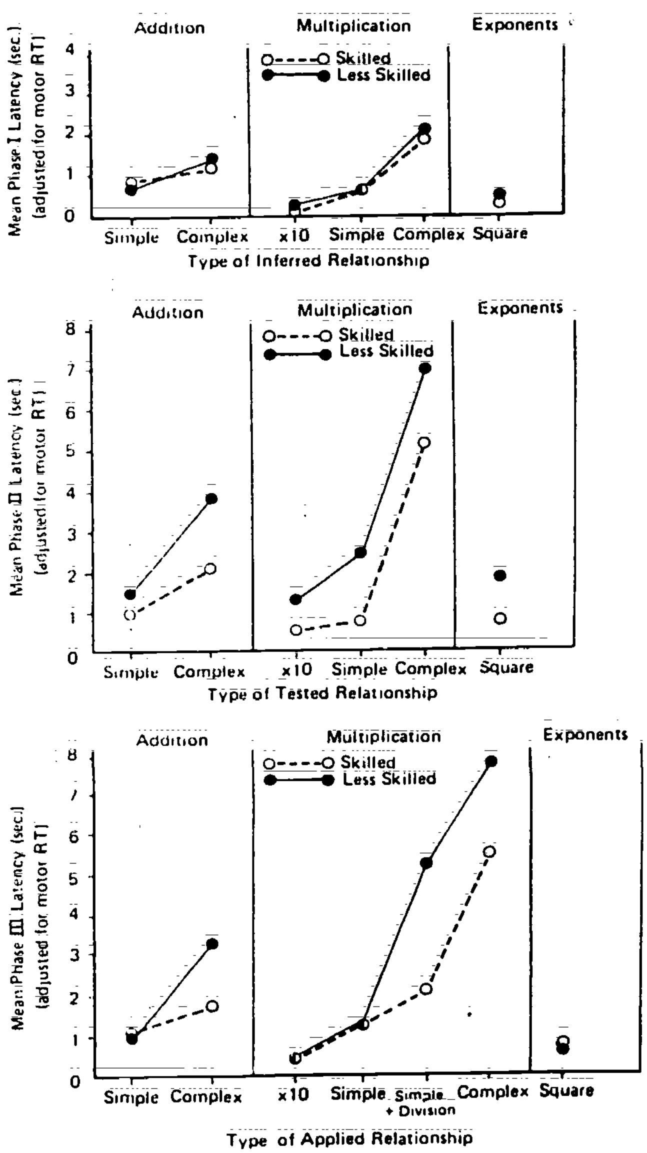


FIG. 5.12. Mean solution latencies for skilled and less-skilled individuals as a function of phase of processing and type of relationship.

groups over all three phases of processing can be examined. The top panel shows initial inference latency for the different relational types. As can be seen, there were highly systematic and significant differences in initial inference latency both within and between types of relationships. Furthermore, the skill groups did not differ in either the pattern of problem differences or overall latency. The middle panel of Fig. 5.12 shows the data for the second phase of processing. Again, there were highly systematic and significant differences in the second stage evaluation latency both within and across types of relationships. While the skill groups did not differ in the pattern of problem differences, there was an overall latency difference. For all problem conditions, the skilled group was faster during the second phase of processing. Coupled with the lack of a latency difference in the initial inference processing, this latter difference suggests the possibility that the skill groups may have differed in the precision of the inference carried over into the second phase of processing. The bottom panel of Fig. 5.12 shows the data for the third phase of processing. There were systematic and significant differences in the final application latency both within and across relationship types. In addition, skill groups differed significantly on problems requiring application of a multi-digit operation and successive application of two single digit operations as occurs in multiplication by fractions.

Two general conclusions can be drawn from these data. First, there are systematic problem differences in the speed with which certain types of quantitative relationships are inferred, evaluated, and applied in relatively simple analogy problems. Second, while skill differences do not appear in the patterning of problem differences during any of the three phases of processing, they are present in the speed of evaluating the relational overlap between two pairs of numbers and the application of relational information to generate final solutions.

2. *Interactive Solutions*. The systematic latency differences both between and within relationship types have important implications about the types of relational information that may be directly and automatically retrieved versus those that are actually computed. It seems reasonable to assume that the countless hours of practice on the multiplication tables plus the numerous times that these facts are used in everyday problem solving lead to a direct access and retrieval for multiplication relationships and that simple multiplication relationships can take precedence over addition; i.e., for the pair 4:28. It is possible to test this hypothesis by creating potentially garden path analogies of the type shown in Table 5.6. Such garden path problems should lead to interactive solutions if multiplication is the preferred inference for such number pairs. The three different problem types labeled A–I to A–III represent cases where a simple multiplication inference is possible sometime during the course of solution. The problems differ in whether such a relationship can be inferred in the second, first, or both pairs.

The data of interest are latencies during the second stage evaluation for problems where the correct rule is multi-digit addition, but where one or both pairs

Table 5.6
Examples of Problems Generally Resulting in Interactive Solutions

Problem Types	Relational Inferences	Examples
	Rule = Complex Addition	
A I	A:B is Complex Addition	8:35 :: 3:30 :: 4:___
	C:D is Simple Multiplication or Complex Addition	5:41 :: 9:45 :: 3:___
	Rule = Complex Addition	
A II	A:B is Simple Multiplication or Complex Addition	3:17 :: 7:31 :: 2:___
	C:D is Complex Addition	7:42 :: 3:38 :: 4:___
	Rule = Complex Addition	
A III	A:B is Simple Multiplication or Complex Addition	6:30 :: 3:27 :: 2:___
	C:D is Simple Multiplication or Complex Addition	10:40 :: 6:36 :: 5:___
	Rule = Simple Multiplication	
M I	A:B is Simple Multiplication	3:24 :: 8:64 :: 5:___
	C:D is Square or Simple Multiplication	4:12 :: 3:9 :: 2:___
	Rule = Simple Multiplication	
M II	A:B is Square or Simple Multiplication	5:25 :: 2:10 :: 4:___
	C:D is Simple Multiplication	8:64 :: 4:32 :: 3:___

have a possible simple multiplication inference. The left panel of Fig. 5.13 shows that, relative to the baseline addition problem, there is no interference when only the second pair involves a possible multiplication inference, i.e., problem type A-I. This suggests that second stage relational evaluation involves a contextually directed process based upon the specific outcome of the first stage inference process. Both skill groups showed interference effects when the first pair involved a possible multiplication inference, i.e., problem type A-II. These interference effects are increased when there is a second pair that involves a different possible multiplication inference, i.e., problem type A-III. The less

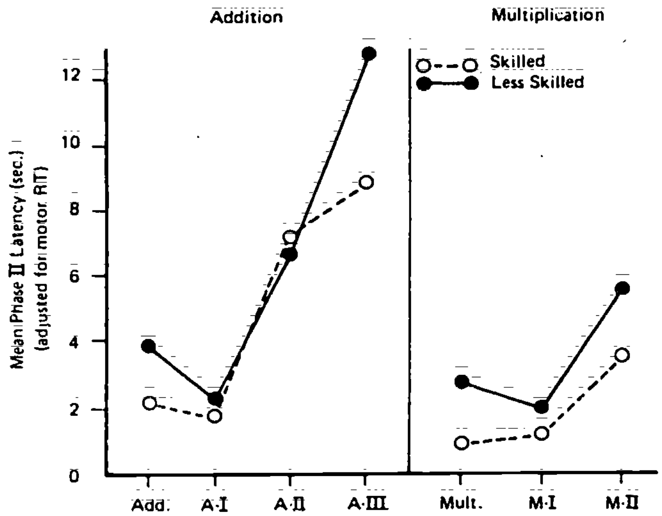


FIG. 5.13. Mean Phase II Latency as a function of skill and problem type.

skilled group shows a substantial increment in time for these problem types. The data for both groups, when coupled with the previous data on initial inference times, suggest that multiplication has precedence over addition, at least within the range of simple multiplication values associated with the highly overlearned multiplication tables. The precedence is most likely a function of the direct or spontaneous retrieval of the stored relational information rather than any computation process per se.

It is also possible to examine the possibility that simple exponent relationships such as squares are directly retrievable and that such retrieval may take precedence over or coincide with the retrieval of simple multiplication relationships. Potential garden path or interactive solution problems of this type were also created and these are shown in Table 5.6, i.e., problem type M-II. The right panel of Fig. 5.13 shows performance relative to a baseline multiplication problem where an inference of an exponent relationship is not possible, resulting in a simple conceptually driven solution. As can be seen in the figure, both skill groups showed that there were no interference effects when the second number pair involved a possible conflicting relationship, i.e., problem type M-I relative to the baseline condition. This again supports a processing model where the second stage relational evaluation is directed by specific inferences resulting from the first stage of processing. Both skill groups show interference effects when the first pair may be represented as a square relationship. Additionally, the skilled individuals are faster overall in the evaluation of all multiplication problems.

Thus, the data for the second set of potentially garden path or interactive solution items suggest that simple exponents take precedence over simple multiplication. Again this may result from the direct representation and spontaneous retrieval of this type of information in the knowledge structure. Clearly, the knowledge base sustains multiple representations of relations that affect problem solution in the context of a model such as the one outlined earlier. Furthermore, it should be noted that overall performance on the garden path or interactive solution problems shows that skill differences appear in the speed of evaluating certain types of relational information but not in the apparent order or precedence of encoding or inferring quantitative relationships. Such order differences would not be expected given that college students have available all the multiplication facts as well as the simple squares. However, skill differences appear in the speed and likelihood of successfully finding the addition rule in problems that capitalize on the encoding of multiplication relationships.

A final class of problems for which skill differences might be expected are those with relationships that are much less salient than squares and integer addition and multiplication, e.g., multiplication by fractions and the use of cubes and fourth powers. Examples of these problems are provided in Table 5.7. In these problems, it is expected that multiplication or addition will be the initial inference. Thus, the second stage of processing will require a search for alternative relationships, and skill differences may appear in both the time and success of such processing. Table 5.7 shows that there were substantial skill differences for both evaluation latency and application latency for the noninteger multiplication problems. Even larger differences were obtained when the problems involved cubes or fourth powers, and there were substantial accuracy differences on these problems as well.

3. *General Processing Models.* The data that we have presented on the processing of different item types and skill differences in processing can be used to evaluate a model of the type shown earlier in Fig. 5.11. The major

Table 5.7
Skill Differences in Performance for Problems Containing Low Salience Rules

Problem Type	Performance Measure	Skilled	Less-Skilled
Noninteger Multiplication	Phase II latency	5.06 sec.	10.25 sec.
14.21 : 8.12 :: 2:__	Phase III latency	2.38 sec.	5.28 sec.
12:16 : 9.12 :: 3:__	Error rate	4%	10%
Cubes and Fourth Powers	Phase II latency	4.17 sec.	14.96 sec.
4.64 : 2:8 :: 3:__	Phase III latency	10.01 sec.	24.62 sec.
2:16 : 3 81 :: 4:__	Error rate	12%	31%

modifications pertain to second stage processing activities given specific assumptions about the type of relational inferences made during the first phase of processing. In particular, we have assumed that exponentiation is a relational inference that has precedence over simple integer multiplication, which in turn has precedence over addition. These assumptions were more formally tested in the context of the full model shown in Fig. 5.11. The data previously discussed also favor a model in which the second stage evaluation involves directed relational processing. The legitimacy of such an assumption was further evaluated by comparing model fits for both directed and undirected relational evaluation. Finally, this processing model has two possible variants in the course of processing when second stage testing detects a mismatch or relational incongruity. In one mode, a relational inference process is applied to the second pair and the result is evaluated for the initial pair (a flip-flop procedure). In the other, given the presence of a mismatch, a new inference is attempted for the first pair and again evaluated for the second pair (a return-to-start procedure).

The data for both conceptually driven and interactive solution conditions was used to evaluate a variety of model variants. In addition, separate model fitting was done for the skilled and less skilled subjects as a group. Both skill groups were well fit by models that assume directed processing rather than a context independent processing procedure. Furthermore, the best fit in both skill groups was the flip-flop model, although the two directed processing models are virtually indistinguishable in the high skill group. The skill groups differed, however, in the parameters estimated for the various types of inferences. When individual subjects were fit, the directed processing models also did well, and the flip-flop model was either consistently superior to the return-to-start model or indistinguishable from it.

The model fitting simply confirms the previously stated conclusions about the types of processing activities involved in these simple types of analogy problems, the relationship of such processing to problem features, and the assumptions about knowledge base influences on performance. Skill differences over the range sampled do not appear in the sequence of processing activities, but are primarily restricted to the duration of processing events. Differences also appear in the likelihood of discovering some less salient numerical relationships and in overcoming certain set or einstellung effects.

D. Item Features, Errors, and Performance

A study conducted by Holzman (1979), attempted to explore the contribution of different problem features to both the latency and accuracy of solution on numerical analogies of the type we have been discussing. Rather than focusing on individual phases of processing, Holzman was interested in identifying the factors that influence overall solution latency and accuracy, and the extent to which those factors change with age and skill differences within age. The subjects were

a group of college students and two groups of children, one of average aptitude and one of high aptitude. All individuals were given a set of numerical analogy problems that systematically varied a number of features including the type of relationship required to solve a problem, the magnitude of the relationship value, relational ambiguity, and the number of relationships or operations (transformations) required to generate the problem rule.

The data obtained for the performance of the adults was consistent with the latency data obtained by Pellegrino et al. (Note 2) discussed previously. Furthermore, items that showed longer latencies as a function of relational ambiguity, low salience, or relational complexity were also the items with the highest error rates. The overall correlation between problem latency and accuracy was −.93 for the college students. Of greater interest, was the comparison of the factors influencing error rates on the different item types across age and ability groups. Regression analyses were carried out separately for the three groups of subjects. The R^2 values for the individual multiple regressions ranged from .72 to .78. The criterion was average error rate on 60 specific problems, and the predictors were the presence or absence of relational ambiguity in the problem and the number of relationships or transformations that defined the problem rule. For the average IQ children, the only significant variable was the number of transformations. The high IQ children showed an effect of ambiguity, but its relative effect was much smaller than that observed for adults. Ambiguity in these problems refers to the presence of A–B pairs where it was possible for a multiplication inference to take precedence over addition relationships as demonstrated previously in the adults' latency data. The failure to find any influence of such an effect in the average IQ children, its emergence in the high IQ group, and its strong effect in the adult group suggests a shift in the organization and retrieval of certain types of relational knowledge that is associated with age and schooling.

The magnitude of the relational value, i.e., low values such as two, three, and four versus high values such as eight, nine, and larger, was also a significant variable in regression analyses for both groups of children. In addition, the type of simple quantitative relation played a significant role in governing the error rates of the average IQ children, but had no influence in the performance of the high IQ children and the adults. There was a decline in performance when the problem rule involved division. These results can be explained by the models of performance described earlier and by the role of knowledge base factors in influencing processing factors, primarily relational inference. The organization of the knowledge base and the relative ease of retrieving certain types of relational knowledge will determine the extent to which there are relational ambiguities in certain problems, thereby creating the possibility of errors. Similarly, the absolute and relative strengths of certain types of declarative and procedural knowledge will cause relational type : magnitude to play a role in the speed and accuracy of solution. The knowledge base is an important factor in the

solution of numerical analogies, and developmental changes in the knowledge base lead to differences in the factors contributing to processing difficulty, problem error, and individual differences.

1. *Content Knowledge Differences and Problem Solution.* A study by Corsale and Gitomer (Note 3) attempted to characterize the nature of the differences in the knowledge bases of high- and low-ability subjects, and to indicate how these knowledge differences influence subjects' problem solutions. They collected two kinds of data in order to consider the interaction between knowledge representation and use of problem strategies. An initial set of data was used to characterize the knowledge representations of subjects and a second set consisted of protocol data from subjects' problem solutions. The subjects were fifth and eighth grade children. Fifth graders have at least learned multiplication concepts and eighth graders have learned more complex concepts such as fractions, exponential relationships, etc., which permitted analyses of problem solving as a function of different ranges of knowledge.

All children were given a standardized number analogy test to determine their general ability level. This was followed by two tasks designed to detect the salience of mathematical concepts in the knowledge representations of high- and low-ability subjects. The first task was a grouping task in which the child was given a matrix of numbers from 0-32, and was asked to select groups of numbers that went together and justify his or her groupings. On the basis of these justifications, the children's groups were classified into four types. Examples of these types are shown in Table 5.8. Abstract concepts represented mathematically-based groupings with superordinate labels such as the set of primes, multiplicative or exponential relationships. Operational concepts involved the stringing together of numbers into number sentences. Digit-based groupings involved

Table 5.8
Categories Resulting from Analysis of the Number Grouping Task

Abstract Groupings	Operational Groupings	Digit-based Groupings	Nonmathematical Groupings
Multiples	Number sentences	Single digit #s	Idiosyncratic
Exponents	Computation	Decade subsets	No Reason
Primes	Common digit	Orthographic	
Composit			Proximity
Factors			
Odd/Even			

Note. From "Developmental and Individual Differences in Mathematical Aptitude" by K. Corsale and D. Gitomer, paper presented at the annual meeting of the Psychonomic Society, Phoenix, 1979. Copyright 1980 by K. Corsale and D. Gitomer. Reprinted by permission.

numbers that shared common digits, the set of single-digit numbers, etc. Nonmathematical concepts were idiosyncractic groupings or groups based on orthographic similarities.

In the second knowledge representation task, the child was presented with 20 pairs of numbers and asked to state as many relationships as possible for each pair. The total number of different relationships generated by each child for each pair type was recorded as was the number of generated relations that were critical for solving the analogy problems. For example, saying that 28 and 21 were both in the 20's would not be a critical reason in a typical number analogy context whereas "21 is ¾ of 28" would be.

The various measures derived from the grouping task and the pairs task were reduced by means of a factor analysis. This resulted in a three-factor solution, and factor scores were derived for each child. The first factor was readily interpretable as an estimate of the degree of "abstractness," A, found in the subjects' groupings and pair relationships. The second was a nonmathematical factor, called "generativity," G, that seemed to estimate the number of groupings and relations formed. The third factor represented a "preference factor," P, in which operational or computation-based groups and relations were preferred or were more salient than abstract groupings.

Each of the factor scores, A, G, and P, was entered into multiple-regression analyses with analogy test performance as the criterion variable. This was done separately at each grade level. The resulting regression equations are indicated below.

5th graders: $1.26A + .67G$ ($r^2 = .51$)

8th graders: $1.34A + .41G - .41P$ ($r^2 = .46$)

As can be seen, for both grades, the linear combination of these variables accounted for approximately 50% of the variance. These data reiterate previous demonstrations of the importance of knowledge representation in cognitive processing. For fifth graders, only the abstract knowledge factor and the generativity factor significantly predicted performance on the standardized analogy test. For eighth graders, preferred use of operational rather than abstract relationships was also negatively correlated with analogy performance. Apparently, at later stages, low-ability children show a continued use of operational relationships whereas high-ability children combine relationships to form abstract concepts in preference to operational ones. The primary conclusion to be drawn from these regressions is that degree of abstractness in mathematical knowledge is an important predictor of success in analogical problem solving.

Having demonstrated that the form of knowledge representation is a critical variable, Corsale and Gitomer examined the interactions of knowledge representation and strategy usage for children of different abilities. At each grade level, the five highest and the five lowest scores on the standardized analogy test

were selected and engaged individually in a session of oral problem solving. Each individual was presented with a number analogy problem was asked to solve it aloud. Problems were presented using a serial procedure such that the subject was first presented with the A–B pair and asked to generate analogically plausible rules. The subject was then presented with both the A–B and C–D pairs simultaneously so that the hypothesized relationship could be tested and, if necessary, reformulated. The subject was then presented with E and asked to generate the appropriate F. Having generated an answer or indicating that one could not be generated, the subject was given a multiple choice of five alternative answers. Each subject was given 20 problems. For each protocol, three types of measures were obtained: (1) the probability of successfully obtaining the analogical rule after each successive pair of numbers was presented, (2) the use of a backward-interactive strategy in attempting to solve the problem, and (3) the kinds of errors that were made during the course of solution.

The analysis of error data and the interactive strategy data indicate that knowledge representation drives the solution strategy by defining the limits of the problem domain. High-ability subjects, who have abstract, high level mathematical concepts, limit their analogical hypotheses to a few plausible mathematical relationships. Low-ability children, in contrast, have diffuse, lower order mathematical concepts and their analogical solutions indicate that they do not solve analogies with systematic, mathematically-based rules.

When subjects referred to a previous pair following the presentation of a new pair, they were credited with using an interactive strategy. The probability of going backward was high for all subjects, ranging from 50% to 80%, and the use of the backward strategy did not differentiate either grade or ability level. However, a separate analysis was made of only those protocols in which a child was initially incorrect about the A–B relationships and subsequently used a backward strategy when presented with the C–D pair. For these cases, the probability of subsequently arriving at the correct analogical rule was examined. There was a significant relationship at both grade levels between problem solution and ability level. Even though use of an interactive strategy does not differentiate ability level subjects, high-ability subjects use the backward strategy more effectively: 67% of the time, high-aptitude subjects arrive at the correct rule after using such a strategy whereas low-ability subjects are successful only 20% of the time ($p < .001$). Their probability of obtaining the correct relationship when subjects do *not* use an interactive strategy is only about 25%; this demonstrates that the interactive strategy is mediating the effective performance of high-ability subjects.

Qualitative analysis of the error data helps to explain the ability difference in the effectiveness of the interactive strategy. Errors were defined as either violations of mathematical constraints or analogical constraints. Examples of the different error types are presented in Table 5.9. Mathematical violations were of two types: computation errors or digit errors in which the subject treated a number not as a total concept, but as a set of isolated digits. The statement: "64

and 16 go together because they both have a 6 in them" is an example of a digit error. Analogical errors were of six types. Nonrestrictive errors involved the correct characterization of a similarity within pairs that was not specific enough for analogical solution. Series errors involved turning an analogy problem into a series problem. Single-pair errors occurred when the subject adopted a rule to apply to E that was true only of A–B or of C–D, but not both. Children who

Table 5.9
Numerical Analogy Error Types and Examples

Error Type	Definition	Example
	I Mathematical	
A. Computation (MC)	Arithmetic error	"12 x 13 = 169."
B. Digit (MD)	Treating numbers as a set of digits	"64 and 16 are related because both have a 6 in them."
	II Analogical	
A. Nonrestrictive (NR)	Correct characterization of number pairs that is not specific enough to allow analogical solution	"11 and 33 go together because they are both odd."
B. Series (SER)	Stating a serial pattern across pairs	In the problem 28:21 :: 24:18 :: 20:__, the subject may say that the difference between the first and second numbers decreases by 1 with each successive pair.
C. Single Pair (SP)	Using a correct relationship that applies to only the A:B or C:D pair, but not both	In the problem 9:18 :: 6:15 :: 3:__, where the rule is actually + 9, the subject may apply a x 2 relationship derived from consideration of only the first pair.
D. AC – BD (AC – BD)	Looking for relationships across pairs instead of within pairs	In the problem 10:40 :: 6:36 :: 5:__, the subject may notice that 10 – 6 = 4 and that 40 – 36 = 4, yet be unable to come up with the correct analogical rule of A + 30 = B.
E. Nonanalogical computation (NAC)	Computations that are not constrained by identical pairwise relationships	In the problem 2:4 :: 8:64 :: 5:__, the subject may say that the answer is 16 because 2 x 8 = 16.
F. Directional (DIR)	Applying the correct operator, but in the wrong direction	If the rule is to multiply by 3, the subject may divide by 3.

Note. Adapted from "Developmental and Individual Differences in Mathematical Aptitude" by K. Corsale and D. Gitomer, paper presented at the annual meeting of the Psychonomic Society, Phoenix, 1979. Copyright 1980 by K. Corsale and D. Gitomer. Reprinted by permission.

looked for relationships across pairs rather than within pairs were considered to have made an ACBD violation. Nonanalogical-computation violations were analogically inappropriate computations. Finally, direction violations were those in which the subject applied the correct rule but in the wrong direction. Table 5.10 indicates the mean errors for each grade and ability group. The most obvious result is the lack of grade differences in error frequencies. Rather, any differences in error frequency seemed to be due to ability differences.

Analyses of variance indicated a three-way interaction of grade, ability level, and error type. Low-ability children at both grade levels committed more nonanalogic computation, nonrestrictive, and digit errors. They also made more mathematical computation errors although this difference was not significant. The kinds of errors they made indicated that low-ability children do not restrict their hypotheses concerning an analogical rule to mathematical concepts (as noted by the digit errors) or to analogical concepts (as noted by nonanalogic computation). These findings are consistent with the data presented earlier, which indicated that these low-ability subjects had more diffuse, less structured knowledge representations of mathematics.

The parallel between knowledge representation and solution strategy can also be seen in errors of the high-ability subjects. Although the series error was not significant in differentiating between high- and low-ability subjects, it was the only error committed more frequently by high-ability than by low-ability subjects. However, this is a sophisticated kind of error involving the detection of mathematical relationships that follow a constrained rule. Again, the knowledge representation data, which indicated constrained mathematical concepts for these individuals, parallels their use of that knowledge as seen in the kinds of errors they make.

High-ability children, when they could not detect a rule, would "give up" and not select a multiple-choice answer whereas low-ability children would select an answer—usually a wrong one—and justify it post hoc. Such differences

Table 5.10
Frequencies of Errors by Grade, Ability, and Error Type as Defined in Table 9

Grade	Ability	MC	MD	NR	SER	SP	AC-BD	NAC	DIR
5	Low	6.0	9.4	6.8	1.0	3.4	0.2	14.2	0.4
	High	1.8	2.8	1.6	3.2	0.8	0.8	0.4	0.6
8	Low	6.4	3.6	15.6	1.2	3.0	0.0	8.4	1.6
	High	0.6	2.0	1.4	3.0	0.2	1.4	0.2	0.0

Note. From "Developmental and Individual Differences in Mathematical Aptitude" by K. Corsale and D. Gitomer, paper presented at the annual meeting of the Psychonomic Society, Phoenix, 1979. Copyright 1980 by K. Corsale and D. Gitomer. Reprinted by permission.

were found in both grades but were only significant ($p < .05$) for the older children. High-ability children operate within both mathematical and analogical constraints in order to achieve a goal of an analogically correct answer. They give up rather than choose an answer that they know is wrong. Not only do low-ability children choose the wrong answer rather than give up, but the protocol evidence suggests that they are perfectly happy with, and can justify, their choices on nonanalogical and/or nonmathematical grounds.

Considering both the knowledge representation data and the protocol data, the study by Corsale and Gitomer (Note 3) suggests that high-ability subjects have a greater degree of abstract mathematical knowledge and a greater salience of abstract over operational concepts. In addition, this knowledge correlates with and predicts analogy performance. Finally, and most importantly, high-ability subjects use their knowledge of abstract number relationships to constrain the domain of permissible operations. That is, the knowledge representation sets conditions on the appropriate use of strategies. As an illustration of this last point, consider the differences between high- and low-ability fifth graders who were not very familiar with exponential relationships. High-ability children were able to limit their hypotheses to multiplicative ones and often came up with the correct exponential rule by looking for relations between A–B and C–D. Low-ability children, on the other hand, often engaged in nonanalogical computation and nonrestrictive errors. They did not have the highly constrained organizational structure of abstract knowledge that would provide them with constrained rules of operation.

VII. COMPONENTS OF VERBAL ANALOGY SOLUTION

A. Overview

The preceding discussion of numerical analogy solution introduced the problem of knowledge-base influences upon performance. A major issue was the relational ambiguity for pairs of numbers and how such ambiguity relates to what is stored and retrieved in the course of solving number analogies. Problems of representational variability are of even greater concern in the case of verbal analogies. In attempting to define the rule relating a pair of verbal concepts, there can be many different representations that vary in their level of detail or completeness with respect to the correct rule for the problem. The number and type of semantic features that are accessed and utilized in defining a rule is difficult to specify and can vary greatly over individuals and age groups. Representational variability also is involved at the level of encoding the individual terms. Verbal concepts represent rich information sources that can and do vary in their representations given particular contexts and individuals. (The latter problem is less apparent in the case of encoding numbers or figural elements.) Because of this

indeterminancy in the representation of the semantic features defining the individual terms and relationships, there will be a problem in specifying the single best answer for a problem. A number of different answers may show overlap during a match comparison process, and, as a consequence, discrimination among choices and extended feature processing may be the rule rather than the exception in verbal analogy solution.

As in the previous sections on figural and numerical analogy solution, we will first consider a process model for analogy solution as applied to verbal analogy problems. This model provides the context for the discussion of various types of problem features that affect overall item solution. This will be followed by a discussion of various types of verification and forced-choice latency data that show the effects of item feature manipulations on processing. The discussion of data on item-feature processing will also consider skill differences in processing. Finally, we will consider the relationships among item features, errors, and performance.

B. Problem Features and Processing Models

The processes involved in verbal analogy solution and the semantic factors influencing solution have been discussed by a number of individuals (e.g., Gentile, Kessler, & Gentile, 1969; Rumelhart & Abrahamson, 1973; Sternberg, 1977; Whitely, 1976; Willner, 1964; Ingram & Pellegrino, Note 5). The model shown in Fig. 5.14 is a representation of the components of analogy solution as applied to verbal analogy verification. According to this model, there is an initial encoding of the A term followed by an encoding of the B term and an inference of the semantic features of the A–B relationship. The next phase of processing involves encoding of the C term and a possible mapping of the semantic correspondence (if any) between the A and C terms. The final phases of processing include encoding the D term, inferring the semantic features of the C–D relationship and a test of the match between the A–B and C–D relational features. As shown in the figure, such a comparison and decision process can have three possible outcomes based on the extent of match or mismatch between the inferred relationships.

The semantic features that differentiate among items and that influence processing can be divided into two classes according to the locus of their effects. The first is factors associated with the stem or first three items of the analogy and the second with the set of possible completion terms. One factor of the item stem that has been examined previously is the type of semantic relationship represented by a particular analogy problem. Global analyses of verbal analogy problems drawn from standardized tests reveal that the majority of items can be classified by a limited set of relationship types (Ingram & Pellegrino, Note 5; Haynes, Dawis, Monson, Lopex, & Soriano, Note 6). Included among these are: class membership, function, location, part-whole, order in time, and property.

Table 5.11 provides sample analogy stems and definitions for various types of semantic relations found in analogy problems. The reality of such a classification scheme is manifest in sorting data such as that obtained by Whitely and Dawis (1974) where undergraduates were found to group items on the basis of such relational categories.

We have noted elsewhere (Pellegrino & Glaser, 1980) that identifying the semantic relation class as a factor in item differences does little to predict sys-

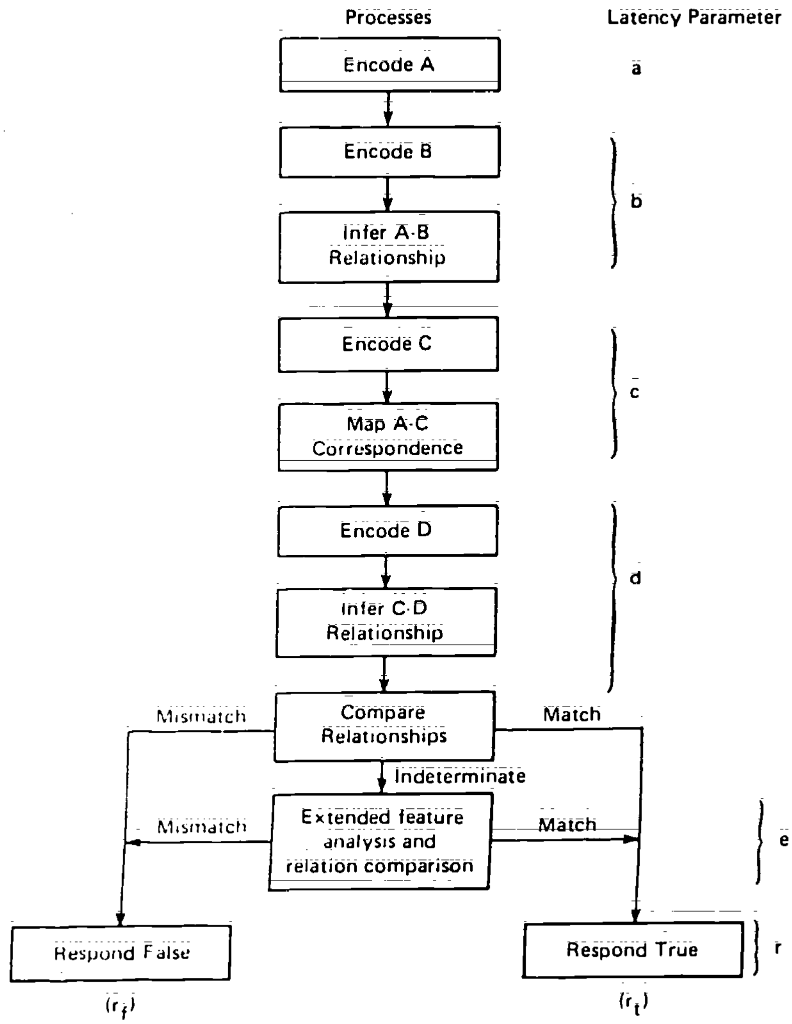


FIG. 5.14. Simplified process model for verbal analogy verification task.

Table 5.11
Examples of Analogy Relation Types

Relation Type	Definition	Analogy Example
Class Member	a) one term is a specific instance of the (more general) other term b) both terms are instances of a more general class	Biology:Science :: Sculpture: _____ Wolf:Dog :: Tiger: _____
Part Whole	one term is a part of the other	Paragraph:Sentence :: Sentence: _____
Order	one term always follows the other in time	Acorn:Oak :: Bulb: _____
Property	one term has the property or quality defined by the other	Green:Emerald :: Red: _____
Function	a) one term performs some function for or action on the other b) one term performs the function defined by the other term	Man:Bread :: Horse: _____ Ear:Hear :: Eye: _____
Conversion	one term is made from or is a product of the other	Sail:Cloth :: Oar: _____
Location	a) one term is located in, on, or about the other often or always b) one term performs some activity in, on, or about the other often or always	Butter:Bread :: Sugar: _____ Train:Rails :: Automobile: _____
Part-Whole & Function	one term is a part of the other and performs a specialized function	Car:Brakes :: Ship. _____

tematic effects of this factor on performance. One reason for this is that, with the exception of class membership versus property comparisons (e.g., Collins & Quillian, 1969), relatively little is known about differences in the inference of different types of semantic features or relations. A second problem is that categorizing the general relationship type captures the most salient relational feature for a pair of terms, but fails to capture the number of specific features that

must also be represented. Third, it is difficult to make predictions about the similarities and differences among items of the same relationship class in terms of the relative ease and likelihood of inferring the appropriate semantic features. These problems require a normative analysis of the extent to which items differ in ease of extraction of appropriate semantic information.

Ingram and Pellegrino (Note 5) obtained normative data on the generation of completion terms for various analogy stems of the same relation type. There were substantial differences in the semantic appropriateness of responses generated to complete the analogies. For some items, only 50% of all the responses were appropriate, while for other items, 99% of all responses were semantically appropriate. Differences were also observed in terms of the size and distribution of the set of answers generated for specific items. Across items, the single most frequently generated answer for an item was given by 10% to 80% of the subjects. These data indicate substantial differences in the constraints on semantic feature processing across items and individuals.

The second locus of effect in semantic feature processing involves the set of possible completion terms. For any given analogy, there is a set of acceptable responses that vary among themselves in generative probability, and that could satisfy the semantics of the item. Normative differences in the production of appropriate completion terms can be considered as representing a "goodness of fit" or semantic distance factor (e.g., Rumelhart & Abrahamson, 1973; Smith, Rips, & Shoben, 1974). This factor should affect decision time in a verification task given processing assumptions of the type represented in the model shown in Fig. 5.14.

C. Data and Theory on Item Processing

1. *Verification Tasks.* A set of verification studies was conducted (Pellegrino & Ingram, Note 4) that examined processing differences among items with different types of completion terms. Four types of completion terms were used. The first two types were acceptable answers that differed in semantic appropriateness (high versus low). Two other types of completion terms were analogically incorrect, i.e., they failed to represent the appropriate relation given in the stem of the analogy. One of these terms was a strong free associate of the C term and the other was a very weak free associate of the C term. College students differing in analogy solution skill (as assessed by the Cognitive Abilities Test (Thorndike & Hagen, 1971, and Lorge-Thorndike Test, 1964) were presented 240–400 separate items for verification, and overall latency and accuracy on the individual items were measured. Presentation of whole analogies and parts of analogies in sequence were used to estimate, in an additive fashion, the component latencies for each of the processing stages of the model presented in Fig. 5.14. In addition, estimates were made of the probabilities that extended feature analysis and relation comparison would be required by each of the four types.

The left column of Table 5.12 shows the parameter estimates obtained for the group mean data. The correlation between observed and predicted values was .98 with an RMSD of 198 msec. As can be seen in Table 5.12, the low-associate items have a very low probability of requiring extended feature analysis before being rejected. In contrast, the high-associate items have higher probability of an indeterminate initial match thus requiring further feature analysis before rejection. As expected, the low-appropriate items have a very high probability of requiring extended feature analysis prior to acceptance. The high-appropriate items, which vary widely in their normative strength, also show some probability of requiring extended feature analysis prior to acceptance. The pattern of results suggests that analogy verification may involve an initial global semantic feature analysis that provides the basis for selecting and rejecting extremely good and bad answers.

Model fitting also revealed skill differences in the precision of the initial, inference, comparison, and decision processes. Subjects were classified as either skilled or less skilled solvers and the data were separately modeled. Table 5.12 shows the parameter estimates obtained from fitting the mean data for each of

Table 5.12
Latency Parameters for Verbal Analogy Verification Task

Parameter	All Subjects (N= 150)	Skilled (N=60)	Less Skilled (N=60)
Encode A (a)[a]	183	77	173
Encode B & Infer A B (b)	326	233	302
Encode C & Map A C (c)	407	308	360
Encode D, Infer C-D, & Compare (d)	1425	1179	1527
Extended feature analysis (e)	1171	1230	1013
Probabilities of extended feature analysis			
α (high appropriate)	.23	.13	.32
α (low appropriate)	.55	.46	.65
α (high associate)	.19	.17	.43
α (low associate)	.07	.03	.25
Response (r)	472	543	489

[a]Small letters in parenthesis refer to Figure 14.

these groups. In both cases, the correlation between observed and predicted values was .97 and the RMSDs were less than 272 msec. The general pattern obtained in both skill groups was the same as that just discussed for all subjects: Both groups seem to be sensitive to the same sets of semantic variables at the level of individual completion terms.

There were, however, differences between the two groups that suggest differences in the quality or precision of the initially inferred semantic information. The low skill group showed higher probabilities of extended feature processing for all four types of completion terms. The largest difference relative to the skilled subjects was with respect to the processing of high associate incorrect completion terms. That less skilled individuals seem particularly susceptible to this type of item suggests an incomplete or sparse feature analysis and representation during the initial stages of processing. The latency estimates for the components of stem processing showed a relatively small difference between the skilled and less skilled subjects up to the point of presentation of the D term. At that point, there emerged a substantial difference. Thus, it would appear that skilled subjects are no faster in the initial phases of processing, but may extract more information that leads to faster and more precise decisions about individual completion terms. Less skilled subjects seem to extract either less information or more global relational features; this makes subsequent decisions more difficult and time consuming. A possible interpretation of such results is that less skilled individuals are forced (or choose) to solve items interactively. The implication that skilled individuals develop higher level and more precise initial problem representations is supported by studies of performance differences on other aptitude tasks (e.g., Egan, 1978; Sternberg, 1977). Thus better representation facilitates the speed and accuracy of final solution.

Our discussion of the influence of item features on processing has been limited to differences in the semantic features of the individual completion terms that were presented. Such results are applicable to all items and permit the testing of the general model in Fig. 5.14. In addition to systematic latency patterns at the level of individual completion terms within items, there also were systematic latency differences between items. Type of semantic relation had a significant effect on overall solution latency as did the degree of semantic constraint as estimated from normative data. When the latency to solve an item was divided into stem processing time versus processing associated with the completion term and the match between relations, there were significant effects of both semantic relation type and semantic constraint in both stem and completion term processing.

2. *Forced-Choice Tasks.* A substantial number of verbal analogies require some form of extended feature analysis to accept the "best" answer and items vary widely in the consensual agreement about what are "best" answers. Thus, forced-choice analogy solution may require considerable amounts of interactive

processing involving the relative match comparison process as discussed earlier. In an attempt to examine such a possibility, normative data were obtained from a group of college students on the semantic appropriateness of each response alternative. Individuals were presented with analogies drawn from the Cognitive Abilities Test, and the alternatives were individually rated with respect to their appropriateness for completing the analogies. The values assigned to each alternative for a given problem were used to derive discriminability indices (a d' type procedure was used). The items showed wide variance in the discriminability values and some items even had negative values, indicating that the college students actually rated incorrect options as being more appropriate than the correct one.

A second group of college students was then asked to solve these items. Those items with a high level of discriminability were solved more rapidly with less repetitive cycling through the alternative set prior to final solution. The overall correlation between the index of discriminability and time spent in processing the alternatives was −.79. Protocol data obtained from yet another group of college students indicated that such items often required interactive solution procedures of the complex type shown earlier in Fig. 5.3.

D. Item Features, Errors, and Performance

1. *Verification Tasks.* The factors that influence the latency of analogy verification also influence the accuracy of solution. Answers that could be considered as less appropriate completion terms take longer to accept, and require extended feature analysis before being accepted. Such completion terms have the highest probability (.26) of being rejected as unacceptable completion terms. The rejection of completion terms of this type can be viewed as an error. Alternatives that are analogically incorrect but high associates of the C term have some probability of requiring an extended feature analysis prior to rejection. Individuals sometimes accept such completion terms and thereby make logical errors (.12). The most approprite completion term has a comparable probability of extended feature analysis and an equivalent error rate (.11). The least appropriate completion, which has the lowest probability of requiring extended feature analysis prior to rejection, has the lowest rate of error (.07).

Skill groups not only differed in latency but also in accuracy in the verification task. Table 5.13 shows one of the significant interaction patterns obtained for the skill group contrast. As can be seen in the table, both groups show the same general pattern across types of completion terms. The most outstanding difference in the pattern of errors is the much lower error rate of the high-skill group on the analogically incorrect completion terms, particularly the incorrect high associates. As was argued before with respect to the model fits for the latency data, it appears that less skilled reasoners have less adequate feature representations that lead to longer times to accept and reject completion terms and

Table 5.13
Error Probability for Verbal Analogy Completion Terms

	D (High Appropriate)	D (Low Appropriate)	D (High Associate)	D (Low Associate)
Skilled	.07	.24	.03	.03
Less Skilled	.12	.32	.15	.10

greater likelihoods of accepting analogically inappropriate answers. A related and consistent finding was that the largest difference between the skill groups in errors was on the items with low levels of semantic constraint. Thus, as semantic feature extraction becomes more difficult, the difference between skill groups in the representation of semantic features becomes more apparent.

2. *Forced-Choice Tasks.* Item difficulty in a forced-choice procedure is a function of both stem and option processing. Evidence for such relationships comes from correlations between item difficulty indices and stem and option processing latencies on standardized test items (Heller & Pellegrino, Note 7). College students solved items that varied substantially in item difficulty and the separate correlations of item difficulty with stem processing latency and option processing latency were .44 and .79 respectively. Both the latency and error data suggest that a substantial proportion of the difficulty in solving standardized test items is associated with choice discriminability and the need to invoke relative match comparison processes. As further support of this conclusion, measures of choice discriminability also correlate at the level of .87 with item difficulty.

3. *Procedural Knowledge Differences and Problem Solution.* In our studies of verbal analogy verification and forced-choice performance, we have been considering the performance of college students who vary in skill. The variations that exist are not at the procedural level and do not seem to reflect differences in a knowledge of task constraints. All our college students seem to reason analogically, albeit with significant differences in the speed, precision, and accuracy of what is represented and processed. However, studies of high school (Heller, 1979) and elementary school (Goldman, Pellegrino, Parseghian, & Sallis, in press) students indicate that an important aspect of performance differences may involve knowledge of task constraints and procedural knowledge that is required to satisfy the goals and constraints of analogy solution. These studies of verbal analogy solution show numerous instances of solution procedures that violate one or more of the constraints of analogical reasoning.

Heller (1979) examined the solution episodes of college and high school students to determine whether they included behaviors that violate task con-

Table 5.14
Sample Solution Protocols

Analogy Elements Presented		*Solver's Response*
Analogical Solution: Protocol 1 Conceptually driven; one option matches initial specifications		
TEA:COFFEE : : BREAD:		Tea is to coffee as bread is to . . . rolls because tea and coffee, they're both drinks, and they're about the same thing, just two different names for two different drinks, and a bread and a roll would be about the same—two different names for the same thing.
MILK	(Reject)	That doesn't fit, it's a drink.
BUTTER	(Reject)	Butter is something you put on bread, that doesn't fit.
ROLLS	(Accept)	That's good.
JAM	(Reject)	It's like butter, something you put on bread. It wouldn't fit because you don't put coffee on tea or in tea.
Analogical Solution: Protocol 2 Interactive; initial failure to identify A-B relation—analogical rule identified during option verification		
ABATE:DECLINE : : WAX:		This is a good one, Oh Christ, I don't know—I can't say anything yet because I don't know what "abate" means.
POLISH	(Accept)	Well, wax and polish mean almost—well they're very close, and maybe abate and decline are very close. I don't know, I'm just gonna put true.
INCREASE	(Reject)	I just don't know.
WANE	(Reject)	To me, decline seems to have something to do with abate, even though I don't know what it means, but wane doesn't have anything to do with wax.
IMPROVE	(Reject)	I was thinking, maybe abate means "to decline" because wax may mean "to improve." And like before, it means "to polish." I like polish better, though.

(Table continued)

Note. From "Cognitive Processing in Verbal Analogy Solution" by J. I. Heller, unpublished doctoral dissertation, University of Pittsburgh, 1979. Copyright 1980 by J. I. Heller. Reprinted by permission.

(Table 5.14 continued)

Analogy Elements Presented		Solver's Response
		Non-Analogical Solution: Protocol 3 Consideration of C·D' relations only
LINE:RULER : : CIRCLE:		Ball. Because a ball is a circle, it's round.
ROUND	(Accept)	Yeah, a circle is round.
DRAW	(Accept)	No, because draw can't be a circle. Oh! Yes, it could be because you draw a circle.
RADIUS	(Accept)	Radius is the numbers in the circle, that's good.
COMPASS	(Accept)	Compass you use to go around—like you put your pencil and it's a circle.
(Which of these do you think best completes the analogy?)		Round, because a circle is round.
		Non-Analogical Solution: Protocol 4 Identification of A·B and C·D' relations; no relational comparison
TELL:LISTEN : : GIVE:		Take. If you tell something, they're like taking it in. If you give something, they take it.
PRESENT	(Accept)	Tell is to listen as give is to present? Yeah, I'd go with that! You give presents?
LOSE	(Reject)	No. Most people find something, they ain't gonna give it back.
GET	(Accept)	Yeah. If you get something, somebody gave it to you.
HAVE	(Accept)	When they give it to you, you have it. Yeah.
(Which of these do you think is best?)		Present. Because you give presents.

(Table continued)

(Table 5.14 continued)

Analogy Elements Presented	*Solver's Response*
Non-Analogical Solution: Protocol 5 Consideration of A-B-C-D interrelations only	
SUBJECT:CITIZEN : : KING:	King—king—queen.
(Could you explain how you got that?)	Well, subject to citizen—like the king is married to a queen so I figured king and queen. They stay together.
(What about subject and citizen made you think you'd need something that went with a king?)	Well, citizen is a person and is like a subject. So I figured that king and queen ought to fit into it. Same as subject and citizen. If I hear you talking about a subject, then it's probably the queen.
RULE (Accept)	This one is a good one here because you're describing the rules. The king and rule is almost like the citizens and rule and I think that, I guess this is a pretty good one. It's kinda hard.
(Could you explain a little bit more what subject has to do with citizen and king has to do with rule?)	Well, the subject is a type of one thing and a citizen is like a person. So the king is a man who's higher and the rule is—the king rules the citizen.
KNIGHT (Reject)	I don't think so. Because knight—I can't really say why.
(What is a knight?)	A knight is a man that guards the king. That's all I can really say.
PRESIDENT (Accept)	This one's all right. President—king's almost the same thing, and both of them are citizens and they're subjects.
(What do you mean "They are subjects?)	Well, it's something—subject to something—I can't explain. King and president are citizens and they're subjects to another person—they're the subject of what other people are talking about.
KINGDOM (Accept)	This one's all right because the kingdom's where the king lives. I guess it's all right—I can't go against it.
(You said that president, kingdom, and rule are possible. Which of those three do you like the best?)	President. I like the king and the president because they're almost the same persons, they both rule in different places.
(And how do they connect with subject and citizen?)	Because they're both citizens and are subject to a person.

straints. Analogical solutions contained no violations of task constraints, and were characterized by: (1) consistent attention to the relations contained in two allowable word pairs, and (2) consistent attention to the match between these pairs of relations. Nonanalogical solutions are characterized by: (1) attention only to relations between "illegal" pairs of elements, and/or (2) consistent attention to the match between inappropriately selected pairs of relations in two word pairs, and/or (3) a consistent disregard for the match between relations contained in two word pairs. "Buggy" solutions (the term is borrowed from the computer programming notion of procedural bugs with missing or faulty subroutines) contained both of the types of behavior described for analogical and nonanalogical solutions. Bugs included suspension of consideration of the match between two relations, consideration of the match between inappropriate word pairs, or "forcing" a match between two relations by stretching the interpretation of relations between word pairs. Sample solution protocols for analogical and nonanalogical solution types are shown in Table 5.14, protocols one through five (Heller, 1979).

Heller found differences in the availability and utilization of analogical solution procedures among a small sample of college students (high ability) and high school students (intermediate and low ability). Table 5.15 shows the proportion of each type of solution in each of the three ability groups where ability reflects overall differences in performance on standardized analogy items. The proportion of analogical solution decreases with decreasing ability level, and the categorization of solution type is independent of the correctness of the final choice.

Table 5.15
Mean Proportions of Anal nical, Nonanalogical, and "Buggy" Analogical Solutions for High, Intermediate, and Low Ability Solvers

	Ability Level[a]		
Solution	High	Intermediate	Low
Analogical	.99	.71	.34
"Buggy" analogical	.01	.13	.15
Nonanalogical	0	.16	.50

[a]High ability subjects were 20 college students
Intermediate ability subjects were 6 tenth grade students in the upper quartile of verbal ability
Low ability subjects were 9 tenth grade students in the lower quartile of verbal ability

Note. From "Cognitive Processing in Verbal Analogy Solution" by J. I. Heller, unpublished doctoral dissertation, University of Pittsburgh, 1979. Copyright 1980 by J. I. Heller. Reprinted by permission.

While the data in Table 5.15 reflect average performance within each of the three ability groups, they are not necessarily indicative of either an individual's consistency in strategic performance or of individual differences within each group. To identify individual patterns in performance, Heller determined profiles of each solver based on the kinds of solutions he or she utilized over the entire set of items. Reliance upon consistently analogical behavior is a characteristic of skilled analogy solvers. Whereas 85% of the high-ability individuals solved *all* items with an analogical solution procedure, none of the intermediate- or low-ability solvers solved all items analogically. While all of the individuals in the intermediate-ability group were capable of solving items analogically, they did not do so consistently. The majority of the low-ability individuals (67%) were also capable of solving items analogically, but again, did not do so consistently. Three individuals within this group, however, solved no items using entirely analogical behavior. These data suggest major differences in the global types of solutions utilized by individuals of different abilities. Not only is the analogical solution procedure relied upon far more consistently by higher-ability individuals, but it appears to be unavailable to some low-ability individuals.

Differences in solution behaviors were also observed in processing activities within analogical solution episodes. All high-ability solvers utilized the additional processes required for interactive solutions in their solutions of some items. However, among those who used analogical solution, some intermediate-ability solvers (17%) and half of the low-ability solvers never solved items using the additional processes required for interacrive solution. Thus, it appears that low-ability solvers are less likely to solve items analogically; when they do reason analogically, they are less likely to demonstrate the complex processing required for interactive analogical solutions. High-skilled individuals, conversely, are characterized by a more adequate knowledge of task constraints and an ability to develop an understanding of the analogical rule using additional information generated from the response options.

Thus, similar to the results in the study of numerical analogies, individual differences in verbal analogical reasoning ability appear to correspond to the differential availability or utilization of the additional processes required in interactive solutions. In items that are solved analogically, low-ability solvers show a high proportion of conceptually-driven solutions. Where an initial inference is easy or possible, the inference is made and maintained throughout solution. Although some low-ability solvers are also capable of modifying rules and solving items interactively, the frequency of their doing so is less than that of higher-ability solvers. On more difficult items, which are less likely to be solvable in the conceptually-driven mode, low-ability solvers exhibit performance that violates task constraints.

The foregoing data argue for the presence of ability differences at the level of procedural knowledge. A major difference in the representation and understand-

ing of task constraints is suggested by the presence of individuals who never reason analogically, and some incidence of nonanalogical behavior in the solutions of both the intermediate- and low-ability solvers. There are a variety of ways in which nonanalogical behavior can be exhibited and these can be classified, at a gross level, in terms of the amount of information that is being attended to in the item. Table 5.16 presents data on the nonanalogical solutions of the intermediate- and low-ability solvers in terms of the proportion of cases showing different violations of analogical constraints.

Three types of nonanalogical solutions were identified: One type in which no attempt was made to identify the A–B relation—attention was paid only to the presence or absence of C–D relations. A second type represented solutions in which all four analogy terms were considered, but attention was paid to the interrelations among three or four terms rather than to the match between two distinct relations within element pairs. A third type represented solutions in which an attempt was made (either successful or unsuccessful) to identify both A–B and C–D relations, but no apparent attempts were made to determine whether any two relations matched. All three of these violated the central constraints of the analogy task—that behavior should be directed toward identifying two distinct relations that are analogous or matching. However, the three conform with task constraints to different extents, i.e., nonanalogical solutions by the low-ability solvers were primarily of the types where no attempt is made to identify the A–B relation or to refer to two distinct relations, while, in contrast, intermediate-ability solvers considered all four terms in most of their nonanalogical solutions, and attended to two allowable relations.

In general, this research suggests that skilled analogy solvers are characterized by greater knowledge of task constraints, and by the ability to develop an under-

Table 5.16
Proportions of Nonanalogical Solutions of Each Type for Intermediate and Low Ability Solvers

Nonanalogical Solution Type	Ability Level: Intermediate	Low
I. Only C D relations considered	.25	.65
II. A-B C D interrelations considered	.17	.10
III. A-B and C-D relations considered but no relational comparison	.58	.25

Note. From "Cognitive Processing in Verbal Analogy Solution" by J. I. Heller, unpublished doctoral dissertation, University of Pittsburgh, 1979. Copyright 1980 by J. I. Heller. Reprinted by permission.

standing of the analogical rule in response to the item stem and set of completion terms. This is accomplished by suspending the top analogical goal momentarily and working on subgoals of the problem structure while maintaining overall task constraints. Conversely, less skilled solvers proceed analogically only when they can easily identify an analogical rule, but if that rule is initially inaccessible, or no C–D relation can be found to match the initially specified rule, they violate task constraints of appropriate analogical syntax.

The Heller study suggests a parallel between sources of developmental differences and individual differences within age. The differences exhibited in the contrast between skilled college students, intermediate-skill adolescents, and low-skill adolescents were pursued further in a set of studies focusing on age and skill differences for children ages 9–12. In two studies, Goldman, Pellegrino, Parseghian, & Sallis, (in press) examined the relationship between developmental and individual differences by assessing performance in two ways: first, by determining the outcomes of the different phases of processing in the analogy task, and second, by considering task understanding and adherence to necessary task constraints.

Third and fifth graders were shown the A–B pair and asked to describe the relationship between the two words. Each child's answer was evaluated semantically to provide a measure of the success of A–B relational inference. The child was then shown the A–B pair and the C term and asked to generate a completion term for the analogy. The child's answer was again evaluated semantically to provide a measure of the success of the C–D' relational application to the A–B pair. Additionally, the child's response and justification were evaluated relative to the previously specified relationship. In this way, it was possible to determine if the child was attempting to meet the constraint of parallel relationships. The final subtask involved a forced-choice test where the child chose the best answer and then verbally justified the choice. The child's performance in this final subtask provided data about two types of processes. First, whether the child could identify the correct response given that he or she failed to generate an acceptable term in the previous task. Second, whether the child had the ability to maintain correct responding in the face of potentially interfering relational information. Finally, the child's verbal justification for his or her choice was evaluated for evidence of adherence to the constraint of parallel relationships in true analogies.

The results of this study showed age differences for all the phases of processing. There was also substantial variability within grades both in overall performance on the forced-choice task and on the more detailed process outcome measures. The pattern of differences in both grades was the same and the distribution of performance for the two grades overlapped substantially. The individual differences within a grade not only reflected developmental differences, but they were far greater than the mean developmental differences. Regression analyses indicated that in both experiments, the measures that assessed processes

relating *pairs* of relationships accounted for individual differences in overall performance.

The data indicated that children varied substantially in effective use of the various kinds of information available to guide responding; a finding similar to Heller's analysis of intermediate- and low-skill individuals. For example, skilled children (as defined by overall forced-choice performance) used semantic information in the A–B pair more effectively to guide the generation of a completion term for the analogy. Then, when faced with a set of alternative answers, they were able to recognize the correctness of their previous efforts and find a more suitable answer when one was available.

Such variability in individual performance can arise because of semantic knowledge differences; differences in the success of process execution, or procedural differences representing a lack of attention to details of analogical syntax. The latter possibility was considered by examining measures that reflect the child's understanding of the task. One such measure involved evidence of relational parallelism in verbal justifications for those items where the child selected the *correct* response in the forced-choice task. This provides an index of the child's understanding of the need for parallel relationships and adherence to that constraint. The skilled children appropriately justified a correct choice over 90% of the time; while less skilled children failed to do so for their *correct* answers half the time. In the two separate studies, this single measure, based only on correct responses, correlated .73 with overall performance.

The different types of measures used in the Goldman et al. (in press) study were derived from the previously stated theory (see Section IV) of the processes necessary for analogy solution and the task constraints that must also be a part of the child's representation of the task. These different sets of measures converge on an explanation of individual differences consistent with that of the Heller (1979) study, i.e., skilled performance is associated with extracting, applying, and organizing sets of semantic information to arrive at unique solutions. Knowledge of task constraints is a relevant feature of the skilled child, adolescent, or adult's problem space that leads to the correct organization and execution of processes of relational inference, comparison, and choice.

However, it is also clear that poor performance is not simply a function of minor violations of an overall correct solution procedure. This is to say that a single process model of correct analogical reasoning is not necessarily the appropriate scheme for understanding errorful performance. When a child shows low or intermediate performance, one cannot assume that this is due to less powerful execution of *the* appropriate set of procedures and processes. Such an interpretation assumes that all children follow the same strategy or have similar process models for task performance. Rather, the data obtained by Goldman et al. and by Heller (1979) indicate that performance is a function of substantial variability in the representation and understanding of task constraints and accompanying procedural knowledge.

VIII. CONCLUDING COMMENTS

A. Discussion of Results

The outcomes of our initial research efforts can be summarized briefly in the context of the general analytic scheme outlined at the beginning of the chapter. We have tried to demonstrate that it is possible to construct information processing models for aptitude test tasks and to show the generality of the theory and models across content and format variants of the same task. In the course of this effort, we have generated a good deal of information about the procedural and declarative knowledge that governs analogy item processing in three different content-symbolic domains—figures, numbers, and words. The analyses reflect some of the complexity of knowledge and skill that presumably underlies performance in an analogical reasoning/problem-solving task. Considerable effort is still needed in theory development and experimental work. We are cautious, however, about the extent to which we wish to become embroiled in detailed efforts at studying the paradigm or task for its own sake; we need instead to make stronger connections with psychometric concepts such as test score differences, item difficulty, test validity, and factor structure.

Our present purpose in studying analogy solution was to select a key aptitude test performance and understand it well enough to apply the analyses to the problems of understanding individual differences in aptitude and differences in item difficulty. As noted in the analytic scheme, a test of both the theory and the modeling of the tasks is the ability to show that these provide a useful basis for characterizing, in cognitive process terms, individual differences in skill. We have made a start in this direction and it appears that this general problem-solving theory of analogy solution and the specific task models have been able to capture some interesting differences in performance.

At the expense of much oversimplification, we infer from this work three interrelated factors that appear to differentiate high- and low-skill individuals. These are the management of memory load, organization of an appropriate declarative (or conceptual) knowledge base, and procedural knowledge of task constraints. We have discussed each with reference to relevant studies of figural, numerical, and verbal analogy tasks, and will subsequently suggest their possible implications for improving skills of learning.

Studies of figural and numerical analogy solution highlight the importance of the management of memory as it is reflected by differences in the speed of performance and the handling of demands on working memory. Our studies of numerical analogy solution show that the structure of the declarative-conceptual knowledge base and the level of representation of this knowledge can also differ as a function of ability: High-skill individuals employ conceptual forms of knowledge that constrain their induction of relations, whereas low-skill individuals encode at more surface levels, which limits their inferential power. In our studies of verbal analogy solution, individuals differed in their knowledge of the

constraints of problem-solving procedures—what we have called the syntax of analogical problem solving. Effective problem solution is characterized by problem-solving steps directed toward the satisfaction of particular goals that are determined by problem-solving constraints. The more constraints the solver is aware of, the more highly constrained will be the goals pursued. Faced with a difficult problem, a skilled individual generates subgoals, pursues them, and can return to higher level goals. For the low-skilled individual, solution difficulty results in violations of problem-solving constraints, the imposition of procedural bugs, and the inability to recover higher level goals when subgoals need to be pursued.

Another of our goals stated in the general analytic scheme was to consider individual differences as a function of development. Our studies of numerical and verbal reasoning differences suggest at least two major developmental trends. First, the development of analogical reasoning ability in part reflects the development of an understanding of the constraints or logical restrictions and requirements of this type of task. For young children, the analogy task is unfamiliar and it often requires considerable explanation and practice before the child even appears to understand what to do. At the other extreme is the college student who has been asked to solve analogies many times in test situations and elsewhere; who is often taught by the use of analogy, and who apparently has a good understanding of the task. Often, no more instruction is required than to simply say: "I want you to solve some analogy problems" at which point the individual is ready to proceed. However, the Heller (1979) data on high school students indicate that even at this age, there are individuals who have an incomplete understanding, and some who have virtually no understanding, of the analogy task requirements.

One aspect of the development of analogical reasoning ability may be the gradual acquisition of the necessary elements of an appropriate task representation causing (or coinciding with) the acquisition of more complete and complex procedures for solving both easy and difficult problems. Younger children often lack the ability to solve items that require a more complex and interactive mode of solution. When difficulties in process execution occur, or ambiguities arise, there is a tendency to relax the constraints of the task rather than actively pursue the appropriate set of goals.

Individual differences within age groups seem to reflect developmental trends. Skilled subjects have an appropriate representation of the task, actively attempt to satisfy the appropriate set of goals by using more complex and interactive solution procedures when necessary, and only relax task constraints when absolutely necessary. In fact, evidence was presented that skilled individuals prefer not to respond at all rather than give an answer that they cannot justify given the constraints of the task.

The theoretical analysis of analogical reasoning and the data on age and skill differences suggest a possible multilevel scheme for the factors that contribute to analogical reasoning performance. These factors are: (1) the individual's prob-

lem space, i.e., representation and understanding of the analogy task; (2) availability and utilization of the processing strategies necessary for analogy solution; (3) organization and coordination of processes for solution of different item types (including the stability and flexibility of these processes); and (4) the automaticity and precision of each component process. Such a scheme is useful when one attempts to understand why an individual performs poorly on a test: The lower the level of performance, the more likely it is that the individual's problem exists somewhere at the top of this hierarchy. This is based upon the assumption that an inadequate problem space—representing a failure to understand the task or a weak understanding with inadequate constraints—will lead to chance or near chance performance. When one understands the task, but lacks proficiency in the component procedures for solving difficult items, then performance may be at an intermediate level, reflecting a failure to deal with complex rules and ambiguous items. At the upper end of the performance range, differences are more likely to be manifest in the stability and coordination of the component performance routines necessary to solve more difficult items. At all levels, declarative knowledge differences may also account for item failures. It is also quite likely that differences in analogical reasoning skill may have quite different sources at different age levels. This issue needs to be pursued in more depth if we are to fully understand the correlations (i.e., predictive validity) of test scores with educational success.

Finally, we have yet to attack the problem of what is "general" across analogy tasks, let alone across inductive reasoning tasks. We would expect that generality would be observed at the level of procedural skill and knowledge of task constraints. We can speculate that at early levels of development and individual skill, performance is limited to specific knowledge domains. As procedural skills and knowledge are exercised in the context of these domains, these skills and knowledge become more abstract knowledge. This eventuates in general inductive reasoning ability that is then available for application to specific problems in various content domains, given that the appropriate declarative knowledge is also available or can be generated.

B. Instructional Considerations

Given what is currently known about analogical reasoning, is it possible to develop diagnostic assessment procedures for the component processes involved, and what then are the implications for instruction? Answering these questions involves some small and large inferential leaps. The simplest case to consider is the teaching of analogical reasoning per se. The more difficult case is instruction in "aptitudes for learning."

1. *Analogy Instruction.* Based on the research we discussed here, we can argue that understanding of the analogy task, i.e., knowledge of the constraints of the task, is a logical instructional target in order to teach analogical reasoning

skills. Correct interpretation of the structural representation "A:B :: C:D" is a prerequisite for solution of even the simplest analogies. The requirement that solvers consider the match between two relations, and that the two relations be chosen according to syntactic rules, can be stressed explicitly during instruction. By doing so, one would be providing basic familiarity with the task to solvers who would otherwise attempt solutions in nonanalogical ways. Such instruction would also refine the knowledge of individuals who have a beginning, but weak, understanding of analogy. Equipped with a strong, appropriately constrained goal, solvers at least have a chance to perform the task of "analogical reasoning."

One conceivable way to teach the concept and constraints of analogy would be to give the student the task of discrimination among positive and negative instances (including "near misses") of analogies. Reasons for decisions could be elicited and corrected as necessary, with explicit emphasis on the proper attributes of analogies. In addition, the "thinking aloud" technique could be adapted for use with instructor and/or peer guidance. Individuals could practice solving simple analogies while verbalizing their thoughts and decision making rationales. By receiving corrective feedback and practicing until syntactic errors were eliminated, individuals may acquire clear and strong knowledge of constraints on this form of reasoning.

Ensuring that individuals understand the essential meaning of analogy does not, however, guarantee that they will be proficient in all the strategic problem-solving activities or component processes required for solution of analogies. Highly skilled solvers are capable of solving analogies using either conceptually-driven or interactive solutions. These solvers can initially identify the analogical rule and evaluate the match between A–B and C–D relations. After identifying C–D relations in all available completion terms, skilled solvers use these relations to help develop an optimal conceptualization of the unifying analogical rule. These behaviors are weak in lower ability solvers, and it would seem that development of instructional methods for teaching these skills would also be a promising area to explore.

It is important to note, however, that the difficulty of analogy items can be manipulated by increasing their processing demands in various ways. Take as an example the case of verbal analogies. If declarative knowledge demands (e.g., vocabulary level) comprise major determinants of item difficulty, increases in syntactic or procedural knowledge will be of limited utility to individuals with an insufficient knowledge of the language, and procedural knowledge will have an inadequate data base. These items remain "vocabulary tests in an analogy vehicle" as Willner (1964) suggested, and would seem to reveal little about individuals' ability to reason analogically. If, however, test designers could create items by increasing semantic processing demands without increasing vocabulary level, then a well developed repertoire of strategic knowledge would be more appropriately assessed. That is, items containing familiar words but complex, abstract, or ambiguous relations are likely to demand solvers' abilities to identify, recon-

ceptualize, and refine relations. Analyses of the kind presented in this chapter could be used by test constructors to ensure that standardized tests actually measure and diagnose appropriate reasoning skills.

Improved performance on standardized tests is not, of course, the major reason for encouraging students to develop analogical reasoning skills. Analogies, metaphors, and similes are all common vehicles of expression and communication in and out of schools (Ortony, Reynolds, & Arter, 1978). Analogical reasoning per se has been referred to as a high level mental activity, and has been equated with intelligence (Spearman, 1923; Thurstone & Thurstone, 1941). There would appear to be an intrinsic value in analogical reasoning as an effective way to think and communicate about the world, thus making analogies valuable as test items. In classroom settings, "teaching by analogy" is a mechanism often used to promote students' understanding of new information. While the relationship between solving multiple-choice analogy test items and learning by analogy has not been ascertained, certain general speculations about their common demands can be offered.

Communicating by analogy involves the presentation of a familiar situation from which one is expected to infer important aspects about an unfamiliar situation. At the same time, discrepancies between the two ideas must be suspended from consideration. That is, the learner must be able to distinguish relevant from irrelevant relations; and it is assumed that relational features common to both situations can be identified while dissimilar relations are ignored. It is critical that *two* matching relations be held in mind simultaneously though the two are instantiated in different domains. In most nontest situations, the individual is not expected to evaluate the match between two relations, but rather to strive for and discover a match that exists and to construct a new representation of relations from a presumably known set. This process is similar to the application of known C-D relations to an unfamiliar A-B relation (or the converse)—an activity highly skilled solvers are capable of performing in this study (Heller, 1979), but one that poorer students exhibit less often and with less sophistication. Explicit study of analogical reasoning in learning and instructional environments is necessary if these processes are to be fully understood. The implications from our studies, however, are that specific and identifiable skills may be required for learning by analogy, and unless students are capable of the necessary processing, common instructional techniques will be relatively ineffective for those students.

2. *Aptitudes for Learning.* The analysis of prototypical test tasks must ultimately generate implications for conceptions of academic learning skills, and means by which they can be fostered. Our speculations on this problem rest upon a consideration of the three differentiating aspects of skill that were mentioned earlier: (1) memory management, (2) procedural knowledge, and (3) content knowledge. The memory management component of skilled performance might suggest that one should focus on processing facility and process training such as

the employment of rehearsal and organizational strategies of the kinds studied in memory experiments. The other two components, however, concemed with knowledge representation and problem-solving procedures, suggest a different emphasis. Emphasis on memory management suggests a focus on the possibility of influencing mental processing skills, e.g., better methods for searching memory and elaborating connections to facilitate storage and retrieval. The literature on the training and transfer of such skills indicates that simple process training approaches to improving the skills of learning are not very promising (Campione & Brown, 1979). An increasing amount of evidence indicates that execution of such "basic processing" skills is dependent on the content and organization of the knowledge base (e.g., Chase & Simon, 1973; Chi, 1978).

In contrast to "basic process" training, if one emphasizes training related to the knowledge base, i.e., conceptual information and knowledge of problem-solving procedures and constraints, then progress is seen in terms of improving the ways in which a knowledge base is activated and manipulated. When highly skilled individuals learn something new or undertake a new problem of induction, they engage a highly organized structure of appropriate facts, relationships, associated procedures, and constraints. Skilled individuals are skilled because of their knowledge of the content involved in a problem and their knowledge of the procedural constraints of a particular problem form such as inductive or analogical reasoning. These two kinds of knowledge interact so that procedural constraints are exercised in the content knowledge base, and the knowledge base enables procedural goals to be attained.

This kind of analysis leads us to suggest that the improvement of the skills of learning will take place through the exercise and development of procedural (problem-solving) knowledge in the context of specific knowledge domains. The suggestion is that learning skills are developed when we teach more than mechanisms of recall and recognition for a body of knowledge. Learning skill ensues as the content and concepts of a knowledge domain are attained in learning situations that constrain this knowledge to serve certain purposes and goals. The goals are defined by uses of this knowledge in procedural schemes such as those required in analogical reasoning and inductive inference.

How this facility could actually be taught is difficult to say at this time. One might teach more of the knowledge base and its high level concepts, or one might teach procedural knowledge such as planning ahead and recognizing when procedural constraints are violated. However, teaching either separately would probably be unsuccessful because each kind of knowledge facilitates the development of the other. Learning skills are probably developed through graded sequences of experience that combine conceptual and procedural knowledge. This is what must take place when a good instructor develops a series of examples that stimulate thinking.

Finally, there is the problem of diagnosing weaknesses in individuals who are unskilled in academic learning. When this is done, we generally find that their

knowledge base is not rich and that their skill in maintaining directed use of this knowledge is not developed. Perhaps a reasonable tactic is to identify some attained knowledge base in an individual where instruction can begin. Knowledge developed in the course of an individual's prior cultural experience can provide knowledge representations and goal-directed behavior that can be exploited. Knowledge structures exist in varying forms in individuals as a result of prior experiences, and this available knowledge can be transferred to domains of related knowledge that approximate more and more closely the formal abstractions and procedural requirements necessary for school learning.

The goals expressed at the beginning of this chapter should be reiterated at its end. The technology of aptitude measurement appears to have reached an asymptote of progress that cannot be changed without further understanding of the details of human cognition. At the present time, scientists are beginning to identify the components of individual differences in terms of modern information processing theory. Educators are seeking more than static measures that predict success in learning; they require knowledge of both the ways that abilities for learning can be influenced in educational environments, and of the limits of this influence. It is possible that analysis of the processes and knowledge required by tests that have correlated with educational achievement can contribute to providing such information.

ACKNOWLEDGMENTS

This research was sponsored by the National Institute of Education, U.S. Department of Education, and was carried out at the Learning Research and Development Center. We are grateful to our colleagues and students, namely Michelene Chi, Kathleen Corsale, Drew Gitomer, Joan Heller, Susan Goldman, Timothy Mulholland, and Patricia Parseghian, whose research is discussed in this chapter.

REFERENCE NOTES

1. Dawis, R. V., & Siojo, L. T. *Analogical reasoning: A review of the literature. Effects of social class differences on analogical reasoning* (Tech. Rep. No. 1). Minneapolis: University of Minnesota, 1972.
2. Pellegrino, J. W., Chi, M. T. H., & Majetic, D. *Ability differences and the processing of quantitative information*. Paper presented at the meeting of the Psychonomic Society, San Francisco, 1978.
3. Corsale, K., & Gitomer, D. *Developmental and Individual differences in mathematical aptitude*. Paper presented at the meeting of the Psychonomic Society, Phoenix, 1979.
4. Pellegrino, J. W., & Ingram, A. L. *Components of verbal analogy solution*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, 1977.
5. Ingram, A. L., & Pellegrino, J. W. *Response generation norms for verbal analogies* (LRDC Publication 1977-2) Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1977.

6. Haynes, J. L., Dawis, R. V., Monson, E. Q., Lopex, F. G., & Soriano, L. V. *Relation eduction norms for 500 picture pairs and 10 relations: High school sample* (Tech. Rep. No. 6). Minneapolis, University of Minnesota, 1974.
7. Heller, J. I., & Pellegrino, J. W. *Cognitive processes and source of item difficulty in the solution of verbal analogies.* Paper presented at the meeting of the American Educational Research Association, Toronto, 1978.

REFERENCES

Atkinson, R. C. Adaptive instructional systems: Some attempts to optimize the learning process. In D. Klahr (Ed.), *Cognition and instruction.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Atkinson, R. C., & Paulson, J. A. An approach to the psychology of instruction. *Psychological Bulletin,* 1972, *78,* 49–61.

Bisanz, J. *Processes and strategies in children's solutions of geometric analogies.* Unpublished doctoral dissertation, University of Pittsburgh, 1979.

Bracht, G. H. *The relationship of treatment tasks, personological variables, and dependent variables to aptitude-treatment interaction.* Unpublished doctoral dissertation, University of Colorado, 1969.

Bracht, G. H., & Glass, G. V. The external validity of experiment. *American Educational Research Journal,* 1968, *5,* 437–474.

Bruner, J. S. Going beyond the information given. In H. Gruber (Ed.), *Contemporary approaches to cognition.* Cambridge, Mass., Harvard University Press, 1957.

Campione, J. C., & Brown, A. L. Toward a theory of intelligence: Contributions from research with retarded children. In R. J. Sternberg & D. K. Detterman (Eds.), *Human intelligence: Perspectives on its theory and measurement.* Norwood, N.J.: Ablex, 1979.

Carroll, J. B. Psychometric tests as cognitive tasks: A new "structure of intellect." In L. B. Resnick (Ed.), *The nature of intelligence.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Carroll, J. B. On the theory-practice interface in the measurement of intellectual abilities. In P. Suppes (Ed.), *Impact of research on education: Some case studies.* Washington, D.C.: National Academy of Education, 1978.

Chase, W. G., & Simon, H. A. Perception in chess. *Cognitive Psychology,* 1973, *1,* 55–81.

Chi, M. T. H. Knowledge structures and memory development. In R. S. Siegler, (Ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior,* 1968, *8,* 240–247.

Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist,* 1957, *12,* 671–684.

Cronbach, L. J. *Essentials of psychological testing* (3rd ed.). New York: Haprer & Row, 1970.

Cronbach, L. J., & Gleser, G. C. *Psychological tests and personnel decisions.* Urbana: University of Illinois Press, 1965.

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington, 1977.

Egan, D. E., & Greeno, J. G. Theory of rule induction: Knowledge acquired in concept learning, serial pattern learning, and problem solving. In L. W. Gregg (Ed.), *Knowledge and cognition.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Estes, W. K. Learning theory and intelligence. *American Psychologist,* 1974, *29,* 740–749.

Evans, T. G. Program for the solution of a class of geometric-analogy intelligence-test questions. In M. Minsky (Ed.), *Semantic information processing.* Cambridge, Mass.: MIT Press, 1968.

Forehand, G. A. Knowledge and the educational process. In L. W. Gregg (Ed.), *Knowledge and cognition.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Gentile, J. R., Kessler, D. K., & Gentile, P. K. Process of solving analogy problems. *Journal of Educational Psychology*, 1969, *60*, 494-502.

Glaser, R. G. The new aptitudes. *Educational Researcher*, 1972, *1*, 5-13.

Glaser, R. Components of a psychology of instruction: Toward a science of design. *Review of Educational Research*, 1976, *46*, 1-24.

Glaser, R., & Pellegrino, J. W. Uniting cognitive process theory and differential psychology: Back home from the wars. *Intelligence*, 1978, 2, 305-319.

Goldman, S. R., Pellegrino, J. W., Parseghian, P., & Sallis, R. Developmental and individual differences in verbal analogical reasoning. *Child Development*, in press.

Greeno, J. G. Natures of problem-solving abilities. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Gregg, L. W., & Simon, H. A. Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology*, 1967, 2, 246-276.

Heller, J. I. *Cognitive processing in verbal analogy solution*. Unpublished doctoral dissertation, University of Pittsburgh, 1979.

Holzman, T. G. *A cognitive-developmental analysis of performance on rule induction tests*. Unpublished doctoral dissertation, University of Pittsburgh, 1979.

Holzman, T. G., Glaser, R., & Pellegrino, J. W. Process training derived from a computer simulation theory. *Memory & Cognition*, 1976, 4, 249-256.

Hunt, E. Quote the Raven? Nevermore! In L. W. Gregg (Ed.), *Knowledge and cognition*. Potomac, Md.: Lawrence Erlbaum Associates, 1974.

Hunt, E. Varieties of cognitive power. In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Hunt, E. Mechanics of verbal ability. *Psychological Review*, 1978, *85*, 109-130.

Hunt, E., Frost, N., & Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7). New York: Academic Press, 1973.

Hunt, E., & Lansman, M. Cognitive theory applied to individual differences. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Introduction to concepts and issues* (Vol. 1). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Jacobs, P. I., & Vandeventer, M. Evaluating the teaching of intelligence. *Educational and Psychological Measurement*, 1976, *32*, 235-248.

Kotovsky, K., & Simon, H. A. Empirical tests of a theory of human acquisition of concepts for sequential events. *Cognitive Psychology*, 1973, *4*, 399-424.

Lorge, I., Thorndike, R. L., & Hagen, E. P. *Lorge Thorndike Intelligence Tests* (Multi-level edition). Burlington, Mass.: Houghton Mifflin, 1964.

McNemar, Q. Lost: Our intelligence? Why? *American Psychologist*, 1964, *19*, 871-882.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. Components of geometric analogy solution. *Cognitive Psychology*, 1980, *12*, 252-284.

Newell, A., & Simon, H. A. *Human problem solving*. New York: Prentice-Hall, 1972.

Norman, D. A., Gentner, D. R., & Stevens, A. L. Comments on learning schemata and memory. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Oppenheimer, J. R. Analogy in science. *American Psychologist*, 1956, *11*, 127-135.

Ortony, A., Reynolds, R. E., & Arter, J. A. Metaphor: Theoretical and empirical research. *Psychological Bulletin*, 1978, *5*, 919-943.

Pellegrino, J. W., & Glaser, R. Cognitive correlates and components in the analysis of individual differences. *Intelligence*, 1979, *3*, 187-214.

Pellegrino, J. W., & Glaser, R. Components of inductive reasoning. In R. E. Snow, P-A. Federico, & W. E. Montague, (Eds.), *Aptitude, learning, and instruction: Cognitive process analyses of aptitude* (Vol. 1). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Pellegrino, J. W., & Lyon, D. R. The components of a componential analysis. *Intelligence*, 1979, 3, 169-186.

Polya, G. *Mathematics and plausible reasoning: Induction and analogy in mathematics* (Vol. 1). Princeton, N.J.: Princeton University Press, 1965.

Raven, J. C. *Progressive matrices: A perceptual test of intelligence* (1938, individual form). London: Lewis, 1938.

Reitman, W. *Cognition and thought*. New York: Wiley, 1965.

Restle, F. Theory of serial pattern learning: Structural trees. *Psychological Review*, 1970, *77*, 481-495.

Rumelhart, D. E. *Introduction to human information processing*. New York: Wiley, 1977.

Rumelhart, D. E., & Abrahamson, A. A. Toward a theory of analogical reasoning. *Cognitive Psychology*, 1973, *5*, 1-28.

Scarr, S. From evolution to Larry P., or what shall we do about IQ tests? *Intelligence*, 1978, *2*, 325-342.

Simon, H. A., & Kotovsky, K. Human acquisition of concepts for sequential patterns. *Psychological Review*, 1963, *70*, 534-546.

Simon, H. A., & Lea, G. Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Smith, E. E., Rips, L. J., & Shoben, E. J. Semantic memory and psychological semantics. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press, 1974.

Snow, R. E. Aptitude and achievement. *New directions in testing and measurement*, 1980, *5*, 39-59. (a)

Snow, R. E. Aptitude processes. In R. E. Snow, P-A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction: Cognitive process analyses of aptitude* (Vol. 1). Hillsdale: N.J.: Lawrence Erlbaum Associates, 1980. (b)

Spearman, C. *The nature of intelligence and the principles of cognition*. London: Macmillan, 1923.

Sternberg, R. Component processes in analogical reasoning. *Psychological Review*, 1977, *84*, 353-378.

Thorndike, R. L., & Hagen, E. *Cognitive abilities test*. Boston: Houghton Mifflin, 1971.

Thurstone, L. L., & Thurstone, T. G. *Factorial studies of intelligence*. Chicago: University of Chicago Press, 1941.

Tyler, L. E. *Tests and measurements*. New York: Prentice-Hall, 1963.

Vitz, P. C., & Todd, T. C. A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, 1969, *76*, 433-449.

Whitely, S. E. Solving verbal analogies: Some cognitive components of intelligence test items. *Journal of Educational Psychology*, 1976, *68*, 232-242.

Whitely, S. E., & Dawis, R. V. Effects of cognition intervention on latent ability measured from analogy items *Journal of Educational Psychology*, 1974, *66*, 710-717.

Willner, A. An experimental analysis of analogical reasoning *Psychological Reports*, 1964, *15*, 479-494.