c.1

INDIVIDUAL DIFFERENCES IN PROCESSES AND STRATEGIES IN ANALOGICAL REASONING

by

(C) NANCY LOUISE WILSON

B.Ed., McGill University, 1976

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Educational Psychology and Special Education

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1980

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Educational Psychology and Special Education

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date August, 1980

ABSTRACT

The need was presented for further research of the quantitative and qualitative nature of individual differences in achievement ability. The componential theory of analogical reasoning was used to identify differences in processes and strategies used to solve pictorial analogies by students of low, average, and high achievement ability.

Subjects were 60 boys and girls from nine grade four classes in four schools in the Lower Mainland. One third of the group were high in achievement ability, one third were average, and one third were low. The criterion used to determine achievement ability was the Canadian Test of Basic Skills. The average age in the low group was 9 years, 9 months, in the average group 9 years, 8 months, and in the high group 9 years, 8 months.

These subjects performed a series of forced-choice pictorial analogy tasks of the standard form A is to B as C is to D_1 or D_2. The analogies were presented in booklets. Each booklet contained 16 analogies, four per page. Subjects were given 64 seconds to work on each booklet. The booklets were administered over two sessions.

Total time spent on an analogy booklet was decomposed into estimates of the time spent on each component (process) used in solution. Response times for number of items correct and number of items completed for each booklet were predicted from independent variables representing variations in the complexity of analogy items over the 24 booklets.

Seven models were fitted to the 24 booklet scores at each ability level. The models differed in the components hypothesized to be used in solution and in the mode of component execution, exhaustive, or self-terminating. The model which best accounted for the variance in the data was designated as the preferred model.

Multiple regression results suggested that there were qualitative differences in analogy solution for the three groups. The same model was preferred by the high and average groups, but a more exhaustive mode of execution was preferred in the low group.

Significant quantitative group differences were found in a univariate analysis of variance which indicated that the high ability group had significantly shorter latencies correct than did the average and low groups. The average group had lower latencies correct than did the low group, but this difference was not significant.

These results were subject to certain limitations in that there was evidence, especially in the low group, that the preferred models were not necessarily the complete models, and that additional factors such as nonlinear processing, speed, floor, and ceiling effects may have affected the results.

Findings were discussed in terms of the above limitations and Sternberg's theory of analogical reasoning.

Implications of these results for future research in individual differences were drawn.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

In the execution of this study and the prepartion of this thesis I was assisted by many people. I thank all those who contributed for their support and guidance.

I am especially indebted to the committee members: my major advisor, Dr. Ron Jarman, and Dr. Todd Rogers and Dr. Nancy Suzuki. Their expert and detailed comments and suggestions contributed immeasurably to the thesis.

I gratefully acknowledge Dr. Robert Sternberg for permission to use the Schematic Picture Analogies; the Educational Research Institute of British Columbia for support through the Discretionary Grant Program; Dr. Gary Holmes for preparing Figure 3; Mrs. Nina Thurston for her expert typing; and the school board officials, principals, teachers, and students for participating in the study.

The love and encouragement I have received from my parents, Darrel and Teresa Wilson, and from my friends, especially Frank, is easily acknowledged but my gratitude for their constant support far exceeds these words of thanks.

CHAPTER I

INTRODUCTION

The nature of mental ability has been the subject of psychological investigations throughout this century. A number of research paradigms have been employed over the years including factor analytic and experimental methods. Recently Robert Sternberg (1977b) proposed the method of componential analysis as an alternative to factor-analytic and classical experimental methods for research on the nature of intelligence. With this method, it is possible to identify processes and strategies used by individuals in solving a variety of tasks, although the method appears to be more suited for use with analogical reasoning problems.

Factor analysts during the first half of the century made important contributions to test theory and proposed theories detailing the structure of mental abilities. Unfortunately, they were able to describe, but not explain, individual differences in intelligence. Indeed, there was a lack of consensus among correlational psychologists as to the factorial structure of mental abilities. Misuse and limitations of the factor-analytic method led to a proliferation of factor theories, and lack of process explanations.

Experimental psychologists, originally unconcerned with individual variation, attempted to identify universal stimulus-response laws of learning. Later, information processing and cognitive research developed, with a renewed emphasis on cognitive processing. Again, methodological problems limited these investigations. Task specific process theories lacked generalizability, and theoretical structures were unparsimonious.

Eventually, due to dissatisfaction with both methods, a unification of differential and experimental psychology in the study of individual differences was suggested (Cronbach, 1957). During the sixties many disillusioned researchers called for process investigations of mental abilities as a means of understanding the nature of ability factors beyond a superficial level (McNemar, 1964; Messick, 1972).

In the seventies, a consensus was reached that new methods should combine the differential and experimental approaches in isolating sources of individual differences in performance. A number of investigators attempted to address these concerns. Distinctions were made between capacity (structural) and strategy (process) components of intelligence (Campione & Brown, 1978; Hunt & Lansman, 1975). Distributive memory models were proposed and adopted (Hunt, Frost & Lunneborg, 1973). Research focused on correlating individual differences as measured by standardized tests with basic information-processing performance. Factors were characterized according to cognitive process models (e.g., Carroll, 1974). A series of instructional studies on retarded and normal subjects focused on the use of strategies and the training of those strategies (Campione & Brown, 1978).

While it was hoped that the new breed of experimental studies of factors would successfully explain the nature of ability differences, this approach lacked explanatory power due to the task-specific nature of the findings which lacked generalizability. An overall framework was needed.

Sternberg (1977b) outlined a new method suited to the new era in intelligence research. The method, componential analysis, incorporated the views of differential and cognitive psychology. Rather than attempting to understand intelligence through intercorrelations among many tasks, or through task manipulations, this method focused on individual differences within a single task.

This method had a number of advantages that were lacking in previous methods. Sources of variance that were confounded in other methods were identifiable in the componential method. Rotational decisions, a weakness of the factor analytic method, were not required in componential analysis, resulting in increased inferential power. Data in componential analysis were intraindividual and not interindividual as in factor analysis, thus permitting more exact interpretations of individual differences. The nature of the technique permitted study of individual differences at a number of levels, rather than study of differences on overall scores or factor loadings. The model or theory was specified <u>a priori</u> and not post hoc as in most factor analyses, and provided a framework which was absent in many experimental studies. Tasks were chosen which were of theoretical interest and were expected to correlate with other tasks and ability measures, thus avoiding the triviality of some information processing approaches.

Essentially, componential analysis accomplishes these improvements through the general method of regression analysis. A task is chosen which is of theoretical interest. Next, models which specify the components (or information processes) involved in analogy task solution are hypothesized. These models also specify the order and mode of component execution. The tasks are administered to subjects. Then the models are used to decompose overall solution latency (or error rate) into estimates of time spent on each component operation. This is accomplished through multiple linear regression for each of the hypothesized models using the complete least squares approach. The proposed models are then evaluated as to their ability to account for the variance in the overall score for the analogy problem. Thus the mental processes involved in task execution are hypothesized and tested, and subjects' scores can be compared not only on the basis of total score, but also at more elementary subtask levels, allowing multi-level isolation of individual differences. Individuals or groups could differ in latency or difficulty of components, in strategies used to combine components (i.e., mode and sequence of combinations) and in the degree of 'fit' of their scores to the predicted scores (all subjects would not necessarily use all of the component processes hypothesized by the theory). This procedure is essentially one of establishing internal validity for the task.

Regression analysis also allows the researcher to establish external validity for the tasks. This is accomplished through correlation of the subjects' scores (both total and component scores) with 'reference ability' tests. Subjects' task scores should predict scores on ability tests which are supposed to measure the same thing as the task is measuring. If they do, this provides evidence for convergent validity. Similarly, task scores should not predict (i.e., have low correlations) scores on reference ability tests which are supposed to measure something different from what the task is measuring. If predictions are poor in this case, evidence for discriminant validity is provided.

Thus, componential analysis appears to be a powerful alternative to existing methodologies for process and strategy research on individual differences.

One application of componential analysis was Sternberg's componential theory of analogical reasoning (1977b). A pictorial analogy task was broken down into five hypothesized components: encoding, inference, mapping, application, and response (Sternberg & Rifkin, 1979). The rule hypothesized for the combination of components was additive; response time for solution of the analogy equaled the sum of the time spent on each component. Models, or strategies, for combining components were developed which specified the order and mode of component execution. The theory was validated both internally and externally (Sternberg & Rifkin, 1979). The theory was successfully employed to identify strategy as well as component and theory differences between the subjects who were second, fourth, and sixth grade students, and adults. Other tasks have also been decomposed through componential analysis (Sternberg, 1977b, 1978c, 1979f), including syllogisms and series completion problems.

In the present study, the relationship between ability level as measured by standardized tests and reasoning ability as measured by pictorial analogies was investigated in order to identify the underlying sources of differences measured by the standardized ability tests. To this date, no previous study had examined the use of componential analysis as a means for investigating this relationship. Students in the fourth grade were chosen to participate and were classified as high, average, or low ability on the basis of their performance on a standardized achievement battery. These students were asked to complete the pictorial analogy tasks in order to identify underlying sources for their variation in ability.

CHAPTER II

REVIEW OF LITERATURE

Historical Antecedents

What do intelligence tests measure? Since the advent of standardized intelligence tests, psychologists have attempted to answer this question. Throughout the first half of this century the dominant paradigm for the enquiry was factor analysis. Despite concerted effort by differential psychologists over the years, we seem no closer to an understanding of the nature of mental abilities than was Boring in 1923 (Boring, 1923/1961):

> If we agree, then, to define intelligence as what the tests of intelligence test, there is a good deal that we can say about it. We can say everything that has been experimentally observed. We can say that it is a "common factor" in many abilities, that it is something like power, that it can be measured roughly although not very finely, that it is only one factor among many in the mental life, that it develops mostly in childhood, that it develops little or not at all in adult life, and that it is largely predetermined at five years of age. Only with more observation and less inference shall we eventually know much more about both intelligence and the special abilities. (Boring, 1923/1961, p. 214)

The concept of intelligence as popularized by Binet, in the form of the mental test, was first received with enthusiasm, followed by skepticism and general disagreement among psychologists (Spearman, 1927/1961). The original enthusiasm was largely due to the hope that individuals could be objectively characterized on the basis of these tests. This collective enthusiasm had degenerated by the early twenties into disagreement as to what exactly was being measured by the tests (Estes, 1975; Resnick, 1976a; Spearman, 1927/1961; Sternberg, 1977b).

Factor Analysis

Despite conceptual disagreements, factor analytic views of intelligence played an important role in the first half of the century. Although this line of research was unsuccessful in explaining individual differences in mental abilities (Jarman & Das, 1977), it did provide a tool for investigating mental abilities at a time when experimental and behavioral psychologists were occupied with overt, non-mental behavior (Sternberg, 1977b). Most of all, factor analysis led to the development of a set of highly predictive tests of intelligence, long considered one of the most important of the products of psychological research.

Limitations of factor analysis. One of the basic weaknesses of factor analysis was that it did not lead to a unique solution in the description of intelligence. As there was disagreement on the definition of intelligence so was there disagreement on the factorial structure of mental abilities. Through the years a variety of theories were in vogue at different times, including Spearman's (1927/1961) two-factor theory, Thurstone's theory of seven primary mental abilities (Sternberg, 1977b), Burt, and then Vernon's, hierarchical model (Jensen, 1970), and Guilford's cube consisting of 120 independent factors (Carroll, 1968; Sternberg, 1977b).

Sternberg (1977b) identified many sources of difference among the various factorial theories of intelligence. He also identified the weaknesses that restricted the explanatory concepts available through this method. Differences stemmed both from misuse of the method and from inherent limitations of the method. For example, the choice of a model in most factor analytic theories was dependent on the following decisions: the number of factors to extract, test selection for analysis, choice of subjects, and psychological definition of the obtained factors. All of these contributed to different interpretations of intelligence. Other factors cited by Sternberg (1977b) in the eventual failure of the method included the lack of specification of <u>a priori</u> models, and non-unique rotation of axes. Most importantly, mental processes of intelligence, imbedded in intraitem data, were not elucidated since factor analysis relied on interitem data.

> The components of intelligence are intra-individual--they exist within individual subjects. Factor analysis, however, is generally interindividual--it analyzes patterns of individual differences across subjects. Since individual differences are meaningless in the context of one individual, it is not clear how factor analysis could enable us to discover what the components within an individual are. (Sternberg, 1977b, p. 33)

<u>Experimental Psychology</u>

Factor analysis was not the only method to fail to explain individual differences. During the years when factor analysis dominated differential psychology, a different paradigm was central to the experimental psychologist. Traditionally, experimental psychologists were concerned with universal laws of learning and not with individual variation (Resnick, 1976a). The stimulus-response (S-R) models in vogue from 1920 to 1960 were reflections of this attitude. In response to the oversimplification of the S-R models, a new computer-based model of information processing was developed in the early sixties (Sternberg, 1979d). An influential book by Miller, Galanter, and Pribram (1960) served to spark an interest in computer simulations of human mental processing. Another series of studies based on the distributive memory model (Hunt, Frost, & Lunneborg, 1973) were characteristic of the renewed emphasis on mental processing which gave rise to the discipline of cognitive psychology.

These new methods avoided some of the weaknesses of most previous factor analyses. The model was usually specified *a priori*, and not subject to post-hoc determination. Rotations were not required, processes were delineated, and the data base was intraindividual (Estes, 1975; Sternberg, 1977b).

Limitations of experimental psychology. Despite these advances, the methods of the experimental psychologists in the sixties were not without limitations. The computer theories were complex and lacked parsimony. Sternberg (1977b) pointed out another major weakness:

> Information-processing methodology does not provide a means for systematically studying correlates of individual differences in performance. If one wants to examine either consistencies in patterns of individual differences across different task parameters, or between task parameters and external measures of performance, the correlational methods of differential psychology are needed to accomplish this goal. (pp. 60-61)

The verbal learning investigations of the sixties and seventies attempted to distinguish between process and capacity (or structure) by either controlling for the capacity variables and measuring the effects of the process variables, or vice versa. These approaches were reviewed by Campione and Brown (1978):

> One is to devise extremely simple tasks in which there is little room for LTM variation to be important (e.g., to attempt to guarantee that the information is equally familiar to all subjects), a second is to experiment in such a way that strategies are likely to be either difficult to implement or unlikely to be of help, and a third is to develop a precise mathematical model of the task in question in which specific parameters reflecting different processes can be readily estimated. (pp. 284-285)

At least one serious drawback to these classical main effect or interaction experiments, wherein one variable was manipulated, was the concomitant limitation of generalizability. The results were overly specific conclusions. Furthermore, without a theoretical framework, the choice of independent variables to be manipulated became a problem. "In many cases it is not clear where training attempts should be aimed, because we do not know exactly how to specify the problem we are trying to remediate" (Campione & Brown, 1978, p. 294). Manipulation of one or two variables was an inadequate approach to the problem of isolating multilevel sources of individual differences.

Call for Unification

Cronbach, as early as 1957, saw the need for a unification of the increasingly divergent differential and experimental disciplines in the study of individual differences; an attempt to right each other's wrongs. "Individual differences have been an annoyance rather than a challenge to the experimenter" (Cronbach, 1957, p. 674). On factor analysts, "his sophistication in data analysis has not been matched by sophistication in theory. The correlational psychologist was led into temptation by his own success, losing himself first in practical prediction, then in a narcissistic program of studying his tests as an end in themselves" (Cronbach, 1957, p. 675).

Dissatisfaction with Existing Methods

Cronbach's suggestions did not result in an immediate confluence of opinion. Still, many theorists during the sixties and early seventies echoed his remarks in voicing their dissatisfaction with existing methodologies.

Humphreys (1962b) criticized the proliferation of factors as well as the practice of interpreting factors as basic and primary no matter how narrow they might be. He proposed a hierarchical factor model along the lines of Vernon's theory (Jensen, 1970) and saw factor analysis more as a means for hypothesis formulation than for hypothesis testing. He advocated a "facet theory" involving task analysis of the format and content of tests, and the development of a priori models for factor analysis, to guide in task selection.

McNemar (1964), in discussing the decline of "g" (general intelligence) as a useful theoretical concept, concluded that while "g" was still important in psychology, its nature was still not understood in spite of years of investigation. He described two types of individual difference studies. The first type (factor analytic) had to do with studies which assumed that patterns of intercorrelations among tests represented the structure of the intellect. The second type (experimental) searched for nontest correlates of test performance.

> Both types of studies certainly force one to stress the overwhelming diversity exhibited among the organisms.
> But these 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. (McNemar, 1964, p. 881)

Messick (1972) joined the others in advocating an augmentation of factor analytic techniques with experimental and observational techniques to derive functional relationships among the constructs of factor analysis, and as a means for understanding the nature of factors beyond a superficial level. He suggested the following:

> Conceptualizations of complex learning processes . . . include not only components of information-processing abilities but also higher-order information-processing heuristics such as plans and strategies, which in turn may implicate variables of personality and cognitive style. One important possibility in this regard is that higher-order traits may enter into sequences not only as components (a simple sequential model) but as organizers of components (a hierarchical personality model). (Messick, 1972, p. 369)

Estes (1974) felt that intelligence should not be characterized in terms of a sampling of item and subscale performances (the correlational approach), but in terms of learning processes. This would allow for identification of the processes that 'cause' intellect. Learning process theory would provide a firmer ground on which to base intervention and remediation programs, although it would not necessarily improve on the predictive powers of existing test batteries. As an example of this procedure, Estes (1974) identified processes hypothesized to underlie subtests such as the digit span and vocabulary tests, with the goal of improving on the diagnostic power of current test batteries. In a later paper (Estes, 1976) he emphasized the need to combine experimental and theoretical techniques in order to explain the complexities of problem solving and development.

The New Wave

In the mid-seventies, intelligence was still very much a major concern of psychologists. The voices of dissatisfaction of the sixties mushroomed into symposia and papers calling for process rather than structural interpretations of intelligence and advocating the unification of experimental and differential psychology (Campione & Brown, 1978; Carroll, 1978; Carroll & Maxwell, 1979; Estes, 1976; Glaser & Pellegrino, 1978; Hunt & Lansman, 1975; Resnick, 1976b; Snow, 1978; Sternberg, 1978b).

> There seems to be widespread concurrence among theoreticians and methodologists alike that new approaches to studying intelligence should somehow combine the differential and cognitive (information-processing) approaches that have been used in the past, and that the combination should somehow enable the investigator to isolate components of intelligence that are elementary (at some level of analysis). (Sternberg, 1978b, p. 196)

The need to isolate a variety of sources of individual differences both at general and more elementary levels of processing was recognized by the "New Wave" intelligence researchers. This new approach reaffirmed the need to explain as well as to predict individual differences in performance.

Existing differential and cognitive methodologies for research were not considered appropriate for the in-depth task analyses necessary to isolate elementary processes and sources of differences. Factor theories had focused on the structure of mental abilities and ignored the underlying processes whereas information processing approaches had demonstrated process effects without incorporating the processes into a theoretical framework or structure (Sternberg, 1979c). The former approach resulted in theories lacking explanatory power, while the latter approach spawned a proliferation of task specific processes, not connected to any overall theory and hence with limited predictive power.

The new wave researchers reiterated the pleas from the sixties, but also took steps to embody these concerns in their research. A start on the problem was made in suggesting necessary conceptual distinctions. Campione and Brown (1978) specified four levels for sources of individual differences on any task. The first term, architecture, was used to refer to the hardware, or the major stores, including short-term memory, intermediate, and long-term memory. Properties of the architecture, including capacity, durability, and efficiency were also defined. Capacity referred to units of storage space, durability referred to the degree of retention of stored information, and efficiency referred to speed of manipulation of stored information.

The other levels involved the contents of memory rather than the architecture, and were thus less fixed.

> First we use the term _knowledge base_ to refer to the existing semantic networks and data structures, the individual's organized knowledge of the world. The term _scheme_ is used to refer to Piagetian rules of thinking, both figurative and operative. Finally, we use _control processes_ to mean the rules and strategies available to the thinker for memorizing, understanding, solving problems, etc. (Campione & Brown, 1978, p. 284)

Hunt and Lansman (1975) made similar distinctions between architecture and control processes such as rules and strategies.

New Wave Research

Hunt, Frost, and Lunneborg (1973) were among the first to put into practice these recommendations.

> The individual differences assessed by an intelligence test provide useful reflections of cultural and biologic differences among men, but the development of those tests has taught us little about the nature of these differences. As a result, intellectual assessment is all too often descriptive rather than prescriptive. (Hunt, Frost, & Lunneborg, 1973, p. 89).

Their goals were two-fold: to demonstrate a substantial relationship between their distributed memory model and cognitive tasks, and to prove that individual differences in task performance were not due to measurement error (as previously assumed), but due to reliable individual characteristics.

These hypotheses were supported. Subjects, classified on the basis of verbal and performance ability scales, differed reliably in task performance.

Qualitative differences as well as quantitative differences were evident; there was some indication that subjects obtaining equivalent results were doing so on the basis of different strategies.

> Hunt begins with theories of memory and then attempts to deduce situations in which individual differences in performance should be observed. His analysis leads, for example, to the

> demonstration of large individual differences in a simple comparison task. The next step is to show that these differences are related to scores on standard tests of verbal intelligence. This . . . sets the stage for research which is now needed in order to close the gap and show through what . . . sequence of processes, individual differences in simple tasks that were predicted on the basis of cognitive theory come to be reflected also in test performance. (Estes, 1976, p. 297)

In the same vein, Carroll (1974) outlined a detailed procedure for characterizing factors according to a model of cognitive processes. He believed that his new "structure of intellect" model would provide a better definition of what the tests were testing. Adopting a modified version of Hunt's distributive memory model (Hunt et al., 1973) he analyzed 48 tests from the Kit of Reference Tests for Cognitive Factors. Each test was categorized according to a number of dimensions, including the following: the type of memory demands, the modality or contents of memory involved, the operations or strategies employed in a 'central processor' and the potential ranges of individual differences on the task.

A number of researchers entered the new era in individual difference research by investigating processes of intelligence in retarded subjects. This research aimed to identify the processes of intelligence, study the remediation of retardate deficits, and to investigate the developmental aspects of intelligence. Of the two schools of research, factor analysis and experimental, the experimental psychologists have shown more interest in the study of differences between normal and retarded subjects of equal chronological age. These researchers concentrated on mental processes and strategies rather than mental capacity. Processes and strategies were viewed as more amenable to remediation and instruction than the relatively fixed architecture.

The identification and training of differences in strategies for task solution was emphasized (Belmont & Butterfield, 1971; Brown, 1974, 1975; Brown & Barclay, 1976; Brown & Campione, 1977; Brown & Lawton, 1977; Kail, 1979; Rohwer, 1973). The majority of these strategy deficits identified in retarded subjects were attributable to production, and not mediation deficiencies (Flavell, 1970). In other words, the retarded subjects, while unlikely to produce strategies spontaneously, were induced to use the strategies with prompting or instructions. Thus recent studies have emphasized not only training of strategies, but also transfer of the training to other tasks (Brown & DeLoache, 1978).

Campione and Brown (1978) adopted the view that by training specific strategies, improvement in executive control processes, and thus transfer, would result. They were able to induce transfer of training in only one strategy. So, despite much research (Brown & Campione, 1977; Brown, Campione, Bray, & Wilcox, 1973; Brown & Lawton, 1977), durability and generalizability of strategy training have remained elusive goals for this group of investigators.

Belmont and Butterfield (1971) attempted to induce transfer of strategy training through training of executive processes. They also failed to find durability or generalizability of training. Sternberg (1979a) concluded that durability and transfer of training was possible, but in order to have long-term changes, the interactive effects of performance components would have to be taken into account.

Despite a large degree of research activity attempting to understand the processes of intelligence, and recognition of the flaws in earlier attempts, this new body of research was not entirely faultless.

Carroll (1978) outlined a number of problems in cognitive research, including the problems particular to multiple regression analysis such as collinearity among predictor variables making regression weights difficult to interpret. He also identified more general problems such as processes identified being task-specific, the lack of distinction between optional and required processes, and the problem of circularity, meaning that the processes identified were dependent on the theoretical model proposed.

Despite these caveats, Carroll (1978) remained optimistic about individual difference research in the cognitive field:

> Even if individual differences are inextricably linked with processes, individual difference methodologies should enable us to narrow down the kinds of processes associated with particular tasks, and to investigate the generality of those processes over different tasks. (Carroll, 1978, p. 110)

In a later paper (Carroll & Maxwell, 1979) this view was reiterated in the hope that experimental studies might help to determine the nature and developmental characteristics of primary abilities.

A further issue in the recent literature was the distinction between qualitative and quantitative differences in intelligence (Jarman, 1980). There was at least suggestive evidence that the same task may measure different cognitive processes or strategies at different developmental and cognitive levels (Jarman & Das, 1977). Others have also found different factor loadings for groups varying on ability (Humphreys & Taber, 1973; Stevenson, Parker, Wilkinson, Hegion, & Fish, 1976). In the classical psychometric and learning studies, quantitative changes were assumed, but qualitative changes were generally ignored (Jarman, 1980).

Snow (1978) indicated that future research must identify individual differences from a variety of sources including process differences, strategy differences and differences in the sequencing of processes.

In summary, an emphasis was noted on not only identifying cognitive processes, but also identifying how these processes were organized into specific strategies (Estes, 1975, 1976). These recent trends in intelligence research underlined the need for research focused on elementary processes of intelligence and multiple sources of individual differences including capacity, content, and strategy. It was hoped that these new trends would give rise to new conceptions of intelligence which "will foster the development of educational possibilities that increase individual accomplishments" (Glaser & Pellegrino, 1978, p. 318).

While accumulated evidence indicated that strategies and processes in task performance were important components of human ability, most of the recent research was in the experimental tradition in which only a few variables were manipulated at any time. Typically, if strategy was being investigated, then capacity components were controlled for rather than included in the analysis. Task by subject interactions were often ignored. Hunt and MacLeod (1978) warned that when one particular strategy model was hypothesized, the parameter estimates became model-specific. Because of these and other limitations, recent experimental research was not much more successful than were the earlier experimental and factor-analytic attempts in identifying the nature of intelligence and in identifying multilevel sources of process and strategy differences. In-depth task analyses within a theoretical framework, and multivariate methodologies were advocated for future research.

Componential Analysis

Pellegrino and Glaser (1979) summarized potential methodologies suited to the new era in intelligence research. Among several suggested methods, one of the most promising was componential analysis, developed by Robert Sternberg (1977b) for the analysis of analogical reasoning.

> The overall purpose of componential analysis is to identify the component mental operations underlying a series of related information-processing tasks and to discover the organization of these component operations in terms of their relationships both to each other and to higher order constellations of mental activities. (Sternberg, 1977b, p. 93)

This method incorporated the views of differential psychology by providing a method for elaborating the underlying traits of intelligence tests in terms of mental operations (processes). The views of cognitive psychology were incorporated by providing a method for discovering elementary information processes and their organization in intelligent behavior and relating these processes to reference abilities.

Sternberg (1977b) provided a detailed account of the method of componential analysis and its use in constructing and validating a theory of analogical reasoning. The method was successfully validated on a variety of reasoning tasks, including verbal, pictorial, geometric, and animal-name analogies (Sternberg, 1977b), linear, categorical, and conditional syllogisms, and classification and series completion problems (Sternberg, 1978c). The following summary of the theory and method is limited to the analysis of analogies of the standard form, A is to B, as C is to D (A:B::C:D). Criteria for task selection and methods for task decomposition were outlined in Sternberg (1978c, 1979f).

Origins of Componential Analysis

Componential analysis was developed as an answer to the concerns being raised in individual difference research. Sternberg reviewed the existing methodologies for research on intelligence and found them lacking in ways mentioned in earlier sections of this review (Sternberg, 1977b, 1979d).

The goal of componential analysis was to develop a research method that would isolate individual differences at a number of levels, mainly intraindividually, and to avoid the predeterminism of factor analytic methods, as well as the specificity of information processing paradigms.

Sternberg, in choosing analogical reasoning as the domain of tasks for investigation, was influenced by the ubiquity of analogy items on a variety of ability tests, as well as in everyday use.

> Reasoning by analogy is pervasive in everyday experience and would seem to be an important part of what we commonly refer to as intelligence. . . . Analogical reasoning is of the utmost importance in a variety of intellectual disciplines. . . . Analogical reasoning also plays an important part in the law, where it may be called reasoning by example. . . . Analogical reasoning has been the subject of a relatively small amount of psychological research. (Sternberg, 1977b, pp. 99-100)

Whitely and Dawis (1974) also commented on the centrality of analogy items in the measurement of general intelligence.

Sternberg (1977b) reviewed a number of cognitive and differential theories dealing with analogical reasoning (Spearman, 1923; Johnson, 1962; Rumelhart & Abrahamson, 1973). He concluded that most of the existing theories of analogical reasoning were incomplete, accounting for limited portions of analogical reasoning. They had insufficient data bases, lacked generality, and none adequately accounted for individual differences in processing. To remedy some of these weaknesses, Sternberg borrowed from the strengths of the earlier differential and experimental approaches.

The componential theory of analogical reasoning specified detailed process models for the steps in analogical reasoning. The theory was specific in describing components, or processes, but also general in that it was applicable to various types of analogy problems. The method of componential analysis was also useful in analyzing a fairly wide variety of tasks (Sternberg, 1978c). It was parsimonious in that only "psychologically significant operations" (Sternberg, 1977b, p. 146) were specified, avoiding triviality. The theory was fairly widely tested and had a substantial data base.

Componential investigations were not limited to analogical reasoning; however, only the theory of analogical reasoning will be treated in this review.

> From a psychometric point of view, componential analysis may be viewed as a detailed algorithm for construct validation . . . from an information processing point of view, componential analysis may be viewed as a set of procedures for discovering the identity and organization of a set of elementary information processes. (Sternberg, 1978a, p. 277)

A component, the basic unit of analysis, was defined as "an elementary information process that operates upon internal representations of objects or symbols" (Sternberg, 1977b, p. 93). A component was a non-optional process in most cases.

Intensive Task Analysis: Internal Validation

Theory

The first step in the construction of a componential theory involved the selection of a task for analysis. The task to be analyzed in this review consisted of a pictorial analogy called a Schematic Picture Analogy of the standard form, A is to B as C is to D (A:B::C:D). The analogy item in Figure 1 is an example of a Schematic Picture Analogy which will be referred to throughout this discussion.

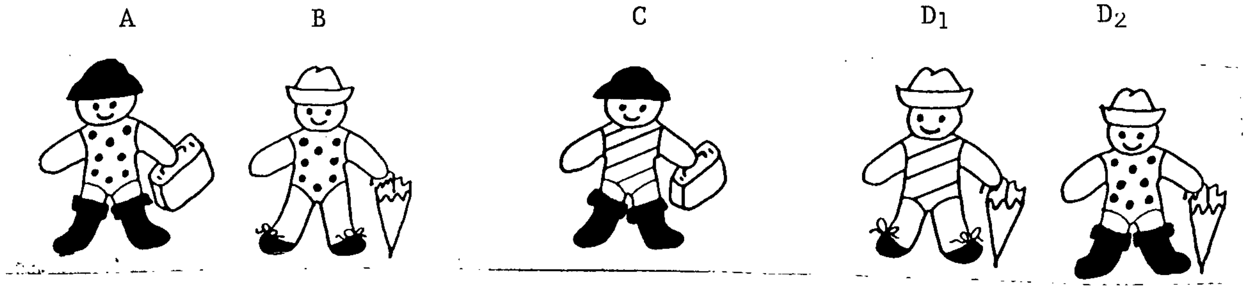
These schematic picture analogies were developed by Sternberg and Rifkin (1979). The goal was to choose the answer option (D_1 or D_2) which correctly completed the analogy. The first two terms, or figures, of the analogy (labelled A and B) established the relationship to be completed.

The next step, at the theory level, involved decomposition of the task into the components (processes) believed to be necessary in solution. Five components were hypothesized for the solution of schematic picture analogies: encoding, inference, mapping, application, and response. While other components may be involved, for parsimony and interpretability only components of theoretical interest were included in a theory.

The third step in the componential theory was specification of a rule (or algorithm) for combining the components. This rule was either an additive one, multiplicative, or a combination thereof. In the pictorial analogy theory, the components were hypothesized to be combined according to a linear, additive model. Thus the total latency for solution of the analogy was equal to the sum of the time spent on each component (see Figure 1). The time spent on each component was a function of the number of times the component was executed, multiplied by the duration of the component (an estimated parameter). The theory also specified that components were analytically executed and therefore separable.

Components

Encoding. Encoding involved storage in short term memory of all of the possible relevant attributes of the analogy terms as well as storage of a value



(from Sternberg & Rifkin, 1979)

Components

1. Encoding:

Attributes	Values
hat color	(black, white)
suit pattern	(striped, dotted)
footwear	(shoes, boots)
handgear	(suitcase, umbrella)

2. Inference: (A-B relation) — hat color (black to white), suit (no change) footwear (boots to shoes), handgear (suitcase to umbrella)

3. Mapping: (A-C relation) — hat color (no change), suit (dotted to striped) footwear (no change), handgear (no change)

4. Application:
   - $C-D_1$: hat (black to white), suit (no change) footwear (boots to shoes), handgear (suitcase to umbrella)
   - $C-D_2$: hat (black to white), suit (striped to dotted), footwear (no change), handgear (suitcase to umbrella)

5. Response: written answer on response sheet

Basic Rule for Component Combination

Total Time = encoding time + inference time + mapping time + application time + response time

Figure 1. Schematic Picture Analogy

corresponding to each attribute.

In the schematic picture analogies (e.g., Figure 1) each term (or figure) had four attributes with two values each. These were, hat color (black or white), suit pattern (striped or polka-dotted), footwear (shoes or boots), and handgear (suitcase or umbrella).

Inference. The inference component involved discovering and storing the relation between the A and B terms of the analogy. In the sample item, the A to B relation for all four attributes was: hat color (black to white), suit pattern (no change), footwear (boots to shoes), and handgear (suitcase to umbrella).

Mapping. The mapping component involved discovering and storing the relation between the A and C terms of the analogy. The first half of the analogy was thus linked to the second half. The full mapping relation for the sample item in Figure 1 was: hat color (no change), suit pattern (polka-dotted to striped), footwear (no change), and handgear (no change).

Application. The fourth component, application, involved applying a relation analogous to the inferred A to B relation from the C term to the answer options. In the sample item the relation from C to D_1 was: hat color (black to white), suit pattern (no change), footwear (boots to shoes), and handgear (suitcase to umbrella). The relation from C to D_2 was: hat color (black to white), suit pattern (striped to polka-dotted), footwear (no change), and handgear (suitcase to umbrella).

Therefore when the A to B inference relation was applied from C to D_1 and C to D_2, option 1 was correct as it permitted the C to D term relation to be analogous to the inferred A to B relation.

Response. The final component, response, involved communication of the chosen option. Thus the analogy task was decomposed.

Component Execution

Components could be executed in an exhaustive or self-terminating fashion.

> If processing is exhaustive, then whenever a component is used in solution of an item, it is executed the maximum possible number of times for that item type. If processing is self-terminating, the component need not be executed the maximum possible number of times. (Sternberg, 1978a, p. 283)

For example, if encoding was exhaustive, the subject would store all of the relevant attributes and values for all five analogy terms in the sample item. If encoding was self-terminating, the subject would perceive and store attributes and values for the terms as they were needed. Thus on the inference step, only the A and B terms need be encoded. When working on mapping, only the A and C terms need be encoded.

Exhaustive inference, mapping, and application components were described in the component section of this review. When each component was defined, the relation was completed for all four attributes at once.

In self-terminating execution attribute values were executed one by one until a unique solution was found. Order of selection of attributes was assumed to be random.

Consider the case where a subject executed the inference, mapping, and application components in a self-terminating fashion. If the subject inferred the relation for hat color first, the inference relation (A to B), would be simply: hat color (black to white). Next the mapping relation (A to C), for hat color would be determined: hat color (no change). The A to B relation for hat color would then be applied from C to D_1 and C to D_2 relations. Unfortunately the correct option would be indistinguishable because the A to B relation for hat color (black to white) applies correctly to C to D_1 and C to D_2. The subject would then have to return to the inference step, choose another attribute (e.g., footwear), and proceed through the inference, mapping, and application components again.

Thus components could be executed from one to four times in the self-terminating mode, whereas with exhaustive component execution the components would be executed once. The number of times a self-terminating component was executed was a function of the number of attributes that had the same values in the two answer options.

Models

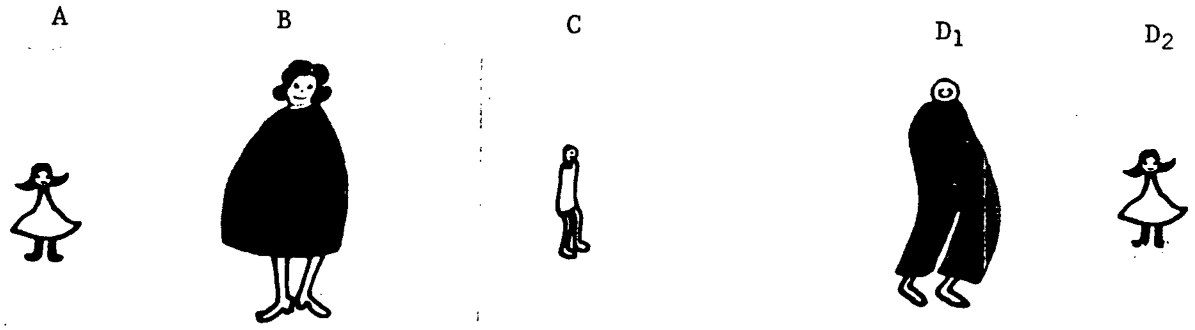
Seven plausible models for solution of pictorial analogies were hypothesized by Sternberg and Rifkin (1979). These models specified the mode and order of component execution: exhaustive or self-terminating. The components were assumed to be combined according to the linear additive rule and a serial mode of processing was assumed in all seven models.

The four models outlined in Table 1 were developed by Sternberg and Rifkin (1979) to explain solution of People Piece Analogies. Figure 2 illustrates a typical people piece analogy. These analogies, like the schematic picture analogies, had four attributes with two values each: height, weight, garment color, and sex. These are known as integral attribute analogies. Sternberg and Rifkin (1979) defined integral attribute stimuli as those in which attributes cannot be nullified without destroying the intactness of the figure. "For example, to portray the sex of a person (or picture of a person), the person must be drawn at some height and at some weight . . . similarly shading in

Table 1

Models 1, 2, 3, 4

Model	Components				
1 =	encoding (exhaustive)	+ inference (exhaustive)	+ mapping (exhaustive)	+ application (exhaustive)	+ response
2 =	encoding (exhaustive)	+ inference (exhaustive)	+ mapping (exhaustive)	+ application (self-terminating)	+ response
3 =	encoding (exhaustive)	+ inference (exhaustive)	+ mapping (self-terminating)	+ application (self-terminating)	+ response
4 =	encoding (exhaustive)	+ inference (self-terminating)	+ mapping (self-terminating)	+ application (self-terminating)	+ response



(from Sternberg & Rifkin, 1979)

Components

1. Encoding:

Attributes	Values
height	(short, tall)
weight	(fat, thin)
garment color	(black, white)
sex	(male, female)

2. Inference: (A-B relation) height (short to tall), weight (thin to fat), garment color (white to black), sex (no change)

3. Mapping: (A-C relation) height (no change), weight (no change), garment (no change), sex (female to male)

4. Application: $C\text{-}D_1$: height (short to tall), weight (thin to fat), garment color (white to black), sex (no change)
   $C\text{-}D_2$: height (no change), weight (no change), garment color (no change), sex (male to female)

5. Response: record answer

Basic Rule for Component Combination

Total time = encoding time + inference time + mapping time + application time + response time

Figure 2. People Piece Analogy

clothing can only be shown if the person wearing the clothing has both height and weight" (Sternberg & Rifkin, 1979, p. 199).

In contrast, schematic picture analogies (Figure 1) had separable attributes. Separable attribute stimuli are those in which attributes may be nullified without destroying the intactness of the stimuli.

The first four models all hypothesized exhaustive encoding (Table 1), but differed in the mode of execution for the remaining components.

Model 1. In Model 1, all of the component operations are exhaustive. The subject encodes the terms of the analogy exhaustively (i.e., all attributes and their values are stored). Inference, mapping, and application are also exhaustive.

> The subject infers all possible relations between encoded attributes of the first two terms of the analogy; next the subject maps all possible relations between encoded attributes of the first and third analogy terms. Finally, the subject applies all possible relations between the third term and each option. (Sternberg & Rifkin, 1979, p. 201)

The processes would be executed as illustrated in the earlier component section of this review.

Model 2. In Model 2, the same steps are followed up to the application procedure, that is, encoding, inference, and mapping are exhaustive and thus executed once. Application is self-terminating. "The subject only applies as many attribute values as are needed to choose a unique answer . . . we assume that order of selection of attributes is random" (Sternberg & Rifkin, 1979, p. 201). For example, if the subjects chose hat color as the first attribute for application in the schematic picture item in Figure 1, they would apply the relation: hat color (black to white) from C to D_1 and C to D_2, but would not be able to distinguish the correct option because both options permit the analogy for hat color to be completed. Another attribute must be chosen and the application component re-executed, until a unique solution is found. Thus the application component may be executed from one to four times in Model 2, dependent upon which attribute is chosen for application.

Model 3. Encoding and inference components are exhaustive in Model 3. Mapping and application are self-terminating.

> He or she maps one attribute value from A to C and then applies the corresponding attribute value from C to D. If the chosen attribute is sufficient to distinguish between the correct and incorrect answer options, the subject responds. Otherwise, the subject maps and then applies another attribute, again trying to select a unique response. (Sternberg & Rifkin, 1979, p. 202)

For example, if the subjects chose to map the footwear attribute first, they found that the A to C relation for footwear was (no change). Knowing that the A to B relation for footwear was (boots to shoes), and the C to 1 and C to 2 relations for footwear were (boots to shoes), and (no change), respectively, they could distinguish the correct option as option 1. Thus mapping and application were executed once in this example. Had the handgear attribute been selected first for mapping, the correct option would have been indistinguishable and the components would have been executed more than once.

Model 4. Encoding is exhaustive, but inference, mapping, and application are self-terminating in Model 4.

> The subject first infers one attribute value from A to B, then maps the corresponding attribute value from A to C, and finally applies the attribute value from C to each answer option. If the subject is able to distinguish the correct from the incorrect answer option on this basis, the subject responds. (Sternberg & Rifkin, 1979, p. 202)

Thus the repeat loop now leads all the way back to the inference component if the correct option is not distinguishable.

Each model, then, differed in the number of times a component operation was executed.

Sternberg and Rifkin (1979) proposed three additional models (Table 2) to describe component execution for the schematic picture analogies in Figure 1. Because of the nature of the schematic picture analogies, Sternberg and Rifkin (1979) hypothesized that the mapping component (as illustrated in Models 1-4) could be by-passed. They felt that stimulus attributes were easily identifiable and thus easily manipulated, removing the necessity for establishing the higher-order A to C relation which linked the domain to the range of the analogy. Subjects would proceed directly from inference to application.

Table 2

Models 1M, 2-3M, 4M

Model		Components		
1M =	encoding (exhaustive)	+ inference (exhaustive)	+ application (exhaustive)	+ response
2-3M =	encoding (exhaustive for A,B terms; self-terminating for C,D terms)	+ inference (exhaustive)	+ application (self terminating)	+ response
4M =	encoding (self-terminating)	+ inference (self-terminating)	+ application (self-terminating)	+ response

The models were labelled 1M, 2-3M and 4M to distinguish them from the first four models. In addition to the absence of the mapping component, these models differed from the first four in that encoding was not always exhaustive. The combination rule remained the linear additive one.

Model 1M. Model 1M differs from Model 1 only in that it lacks a mapping component. Encoding, inference, and application are exhaustive.

Model 2-3M. Model 2-3M involves exhaustive inference with self-terminating application. The A and B terms of the analogy are exhaustively encoded (because inference is exhaustive) but the C term and D term are encoded in a self-terminating mode. Models 2 and 3 were combined in this model because mapping, the distinguishing element between Model 2 and 3 (self-terminating in 3 and exhaustive in 2), was absent in the modified models.

Model 4M. Model 4M involves self-terminating inference and application and all of the terms are encoded in a self-terminating fashion.

Component Estimation

The final step in the theory involved estimating the latencies and difficulties of the individual components. This will be described in detail in the methodology section, but, in general,

> In order to make these estimates, the investigator must quantify the information processing model or models, predicting subjects' latencies and error rates from a set of independent variables. Corresponding to each independent variable is an estimated parameter that represents the latency or difficulty of a single component process. (Sternberg, 1978a, p. 284)

This completes the internal validation phase of the intensive task analysis: specification of the theory, models, components, combination rule, and parameter estimation.

Intensive Task Analysis: External Validation

The second phase of intensive task analysis involved external validation, or determining how the hypothesized components related to individual difference patterns in performance on reference ability tests, or other external tasks. This was a process of demonstrating generalization of the effects of components to other tasks. Subjects' component scores were correlated with their scores on reference ability tests which were hypothesized to measure the same thing as the component scores (e.g., inductive reasoning tests). A high correlation indicated convergent validity. Similarly, discriminant validity was demonstrated by showing low correlations between component scores and unrelated tests, for example, perceptual speed tests (Sternberg, 1978a).

Extensive Task Analysis

Finally, in extensive analysis, the goal "is to integrate the findings of a series of interrelated intensive analyses. In particular, it is designed to demonstrate the psychological reality and generalizability of the identified components" (Sternberg, 1977b, p. 71).

In summary, according to Sternberg (1979f), componential analysis allows for an understanding of the determinants of performance, impossible in listings of tasks and scores (e.g., factor analysis and information processing methods) and provides a framework for investigating the content and structure of mental abilities and for analyzing differences within and across age levels.

> This framework for task analysis could encompass individual differences at several different levels. At the theory level, there could be individual differences in the components people use in performing a task, or in the way component parameters are combined. At the model level, individual differences in the sequence of components or in their mode of operation are possible. And at the component level, individual differences

> in the speed or power of each component could occur. A major virtue of Sternberg's method is that it allows one to attend to all of these levels at the same time. (Pellegrino & Lyon, 1979, p. 170)

Testing the Componential Model

Sternberg (1977b) presented a comprehensive discussion of componential analysis, including the theory, a review of related literature, and four studies supporting the theory. More recently (Sternberg, 1979b, e) he extended the theory of analogical reasoning to one of general intelligence, classifying types of components according to their function and level of generality. The more general application of the theory will not be discussed in this review.

The initial validation of the componential theory of analogical reasoning consisted of a series of experiments on college age subjects. Type of analogy and presentation format varied in each experiment.

Each of the experiments consisted of an intensive task analysis phase including internal and external validation. Four models were tested (Sternberg, 1977a, b). Four types of analogies were used, one in each experiment; people piece analogies (pictorial), verbal and geometric analogies, and animal-name analogies. In most cases the analogies were presented via tachistoscope using the precueing method (Sternberg, 1978c, 1979c) to allow estimation of component parameters and latencies. Analogies differed in the amount of information given (precueing). In some cases subjects viewed the A, B, C, and D terms simultaneously. In some they previewed the A, B, and C terms, or the A and B terms, or only the A term, before being presented with the full analogy. Some analogies involved true or false responses and others involved a choice between two options.

The primary dependent variables were solution times and the primary independent variable was item difficulty, defined as the number of attribute value transformations from the A to B terms, A to C terms, and C to answer terms. Parameters for each component were estimated by fitting linear regression equations to the solution times for different items under the different precueing conditions (Reed, 1977).

Reasoning, perceptual speed, and vocabulary ability tests were administered as reference ability tests to the subjects in the people piece and verbal analogy experiments and to half of the subjects in the animal name experiment. Subjects in the geometric analogy experiment and half of those in the animal name experiment received the Card Rotations Test from the French Kit, and a word grouping test. Twelve subjects who scored between the fifth and twenty-fifth percentiles and 12 who scored between the seventy-fifth and ninety-fifth percentiles on word grouping were selected for the geometric analogy study.

Subjects who took the reasoning and perceptual speed tests were classified into four groups and four subjects were selected from each group, for a total of sixteen. Group 1 consisted of subjects scoring above the eightieth percentile on reasoning and perceptual speed. Group 2 subjects scored high (above eightieth percentile) on reasoning and low (between the tenth and thirtieth percentiles) on perceptual speed. Subjects in group 3 were low on reasoning and high on speed, and subjects in group 4 were low on both measures.

The models hypothesized for component execution were similar, although not identical to those in Table 1. The theory hypothesized that six components were necessary in solution: encoding, inference, mapping, application, justification, and response, and that the combination rule was linear and additive. The additional component, justification, was used when an exact solution was not available.

Justification permits comparison of the options. The option which permits the closest approximation (differing in the fewest elements) to the A to B relation is chosen. This process was unnecessary in the people piece and schematic picture analogies as there was always an exact solution available.

The componential theory of analogical reasoning was supported across the four experiments. The combination rule for components was additive and six components were used to solve analogies (although the justification component was optional).

Model 3 was the preferred model for component execution on all dependent measures for verbal analogies for both high and low reasoners. Model 3 combined exhaustive encoding and inference, with self-terminating mapping and application. On some types of analogies both Model 3 and 4 fit the data, but Model 3 was designated as preferred.

Response times were generally constant across analogy types but absolute times spent on the other components varied. Relative times also varied across analogies, with encoding always taking the most time and application the least.

Error rate data were less conclusive but similar to the latency data.

There was no evidence for individual differences in model choice across these experiments, but the response component latencies were highly correlated with reasoning ability. Longer encoding time on verbal analogies was associated with lower response latency and with success on the reasoning test. This

> was interpreted as evidence that slower and more thorough encoding may pay off in increased ability to compare attributes rapidly or to perform efficiently the numerous bookkeeping operations involved in problem solution. (Sternberg, 1977b, p. 253)

Consistency in strategy use (better model fits) was associated with higher ability scores on the verbal analogies.

> These correlations suggested that subjects higher in reasoning ability tend to be more systematic in their solution of analogy problems and the system they use is that specified by the componential theory. (Sternberg, 1977b, p. 253)

In analogies where discovery of relevant attributes was difficult, the inference, mapping, application, and justification component latencies were positively related to scores on the general ability tests.

Evidence suggesting that reasoning was a good measure of general intelligence (Sternberg, 1977b), included: the correlation of the response component with reference ability tests; the occasional relation of inference, mapping, and application to the reference tests; the relationship between encoding and the other operations; and the relationship between model fits and reasoning scores.

Other studies have followed these initial attempts to validate the theory. Many have involved other types of reasoning, particularly deductive reasoning ability as represented in: linear syllogisms, classifications, series completion, categorical, and conditional syllogisms (Sternberg, 1978c, 1979f).

The method has been used in construct validation of aptitude tests (Sternberg, 1979d). The componential theory of intelligence was also applied to the training of intelligence in the retarded (Sternberg, 1979a). Sternberg concluded that it was possible to train aspects of intelligence in the retarded, and suggested a greater focus on durability and generalizability in the training of strategies as well as emphasis on the motivational and interactive aspects of behavior.

Componential Analysis and Developmental Research

Sternberg extended the componential theory to developmental as well as individual difference research. In an investigation of the development of linear syllogistic reasoning (Sternberg, 1980) no evidence was found for a change in strategy with age, as the same model was preferred by subjects in grades 3, 5, 7, 9, and 11. There was evidence that consistency in strategy use increased with age as the model fits improved with age.

Sternberg and Nigro (1980) studied developmental patterns in the solution of verbal analogies. Twenty subjects in each age group, grades 3, 6, 9, and college, were involved. A strategy shift was indicated in that the third and sixth grade subjects were incomplete reasoners and tended to use word association to solve verbal analogies. The ninth grade and college subjects did not rely on verbal association, but on verbal reasoning. Subjects increased exhaustive processing and had less self-terminating processing with increasing age.

Another developmental study (Sternberg & Rifkin, 1979) investigated the generalizability of the componential theory to second, fourth, sixth grade students, and college-age students. Two experiments were described, one involving people piece analogies, and one in which subjects solved schematic picture analogies during three sessions.

The models described in Table 1 and Table 2 were developed to account for the data.

Developmental differences were predicted at three levels: theory, model, and component levels. At the theory level it was hypothesized that a) subjects might differ in the availability of component operations, specifically that the mapping component might be acquired later than the other components, and b) subjects might differ in the combination rule employed, specifically that older children and adults might use serial processing and the additive rule, while younger children might process holistically.

At the model level, differences were predicted in a) choice of model; older children and adults would choose models that required fewer repetitions of the component processes (i.e., Models 3 and 4), while younger children would follow more repetitive models (i.e., Model 1), and b) consistency in use of model; younger children might use a different model for each problem instead of using a general strategy or model.

At the component level, it was hypothesized that subjects would differ in speed of component operations; older children would be faster. Subjects would also differ in error rate; younger subjects would be less accurate.

Results supported the prediction of differences in use of components. Subjects in grade 2 did not use a mapping component on either analogy type. Subjects in grades 4 and 6 and adults did not use the mapping component in solving separable attribute problems, but did use mapping on integral attribute solutions. Subjects showed no differences in combination rule. All subjects processed serially and conformed to the additive algorithm.

Sternberg (1980) suggested that the absence of the mapping component with separable attribute stimuli was due to the nature of the stimuli. With separable attribute stimuli it was not necessary to extract the attributes one by one. There was some evidence that this component was unavailable to younger children. Mapping required recognition of a second order relationship, a capacity not well developed until the age of 11 or 12.

At the model level, differences were found both with age and type of analogy. In analogies with separable attributes, Model 4M was preferred by all age groups. For integral attribute analogies, subjects in grade 2 preferred Model 4M, grade 4 students preferred Model 4, and grade 6 students and adults preferred Model 3. Thus results from the earlier study (Sternberg, 1977b) predicting Model 3 as the preferred strategy were generalizable only to grade 6 subjects and adults. Young children preferred models with more self-terminating than exhaustive operations. Evidence also indicated an increase in consistency of strategy choice with age.

Brown and DeLoache (1978) have suggested that although exhaustive processing minimized errors, increased use of exhaustive information processing was a general characteristic of cognitive development.

Finally, with respect to component operation latencies and error rates, error rates decreased across age levels as did most component latencies, with the exception of the encoding latency. Encoding times decreased from grade 2 to grade 4, then increased from grade 4 to grade 6 and from grade 6 to adulthood.

As for separable and integral analogy problems, differences in solution algorithms were found. Differences were attributed to differences in encoding strategy. Subjects were believed to employ self-terminating encoding for the separable attribute items and exhaustive encoding for the integral attribute items.

Older subjects were less willing to trade off accuracy for speed. "The more sophisticated strategy, then, is to lengthen one's encoding latency in order to shorten one's comparison latency" (Sternberg & Rifkin, 1979, p. 230).

While evidence has been presented to support the validity of componential analysis for strategy and process research, results on younger populations showed that consistency in use of strategy and choice of strategy was not as unequivocal as research on college age subjects indicated. Developmental differences were found at the theory, model, and component levels.

Pellegrino and Lyon (1979) suggested that although Sternberg (1977b) found little evidence for individual differences at the model (strategy) level in college students, research aimed at a wider range of item difficulty and subject ability may reflect greater variance at the model level.

> We wonder how much of reasoning ability over its entire range might be due to differences in the ability to assemble and monitor highly complex algorithms such as those embodied in Sternberg's models. It seems reasonable that many Stanford undergraduates already have available to them such an algorithm and thus individual differences at this level of ability may not rest in the sheer speed of processing materials. However, other individuals with lower measured abilities and those at an earlier maturational level may manifest differences not so much in the speed of executing each process, but in the likelihood that the program to execute the task can be assembled given meager amounts of practise in the task. (Pellegrino & Lyon, 1979, p. 183)

Thus the method of componential analysis appears to be a promising tool, when compared to previous methods, for investigating differences in processes and strategies.

Summary

A variety of methods and approaches to the study of intelligence have been reviewed. During the first half of the twentieth century factor analytic views predominated, but were unsuccessful in determining the processes underlying the structure of mental abilities. A number of reasons for the failure of this method were outlined, including rotation dilemmas, problems in test selection, and the lack of intraitem analysis.

Experimental psychologists during the sixties and seventies attempted the study of individual differences primarily through information processing models. This approach was also less than satisfactory, resulting in task-specific conclusions lacking external validity.

A solution to the problem was suggested as early as 1957 by Cronbach through the unification of differential and experimental disciplines. Dissatisfaction with existing methods grew during the sixties and by the mid-seventies the consensus was that process explanations of individual differences were necessary and obtainable through a combination of the experimental and factorial viewpoints. Early attempts at convergence included those of Brown (1974, 1975), Campione and Brown (1978), Carroll (1974), and Hunt et al. (1973). Cognitive processes and strategies for organizing the processes were the focus of attention.

One of the most thorough and promising methods for uniting factorial and experimental approaches was that of Sternberg (1977b). This approach, componential analysis, allowed for multi-level investigations of individual differences, and avoided the weaknesses of factor analytic and information processing paradigms by relying on regression analysis in validating detailed theories.

The discussion of componential analysis focused on the componential theory of analogical reasoning. A detailed description of the steps involved in componential analysis followed. A task was chosen, and broken down into a set of components which comprised the theory for that task. A rule was specified for the combination of the components and models were developed which specified the mode and sequence for component execution. Internal validation of the theory consisted of predicting the total score from the subjects' estimated component scores. External validation consisted of determining how the components related to individual differences in performance on reference ability tests. These steps were illustrated by the decomposition of two types of pictorial analogies.

Evidence to support the componential theory of analogical reasoning was presented, based on a college sample and also a developmental study of younger subjects. In the older sample, quantitative differences in latency and error rates were found. Subjects did not differ in model preference but did differ in consistency of model use. Some component scores and model fits were related to reasoning ability as measured by the reference test. In the developmental study individual differences were found at the theory, model, and component levels, in contrast to relatively fewer differences in the college sample.

It was concluded that for the purposes of this study, componential analysis would be the most appropriate investigative tool, especially for dealing with individual differences in processes and strategies as measured by standardized achievement tests.

CHAPTER III

PROBLEM

Statement of the Problem

The review of the literature illustrated the need for further research on the nature of individual differences in ability in elementary school children. The method of componential analysis (Sternberg, 1977b) was chosen in order to isolate sources of individual differences in performance on Schematic Picture Analogies. With this method it was possible to identify the processes and strategies used by individuals in problem solution.

The problem addressed in this study centred around the identification of processes and strategies underlying differences in achievement ability. More specifically the following questions were asked: Do students of different ability levels use different processes (components) in pictorial analogy task solution? Do students of different ability levels differ in the mode of process execution and process combination (i.e., strategy) used in pictorial analogy solution? Do students who differ in ability level also differ in consistency of strategy use?

Rationale

> Major advances in education may well have to wait upon our achievement of deeper understanding of the cognitive processes which the child brings to bear on the tasks, and the way in which these processes come to be organized as a function of different kinds of experience. (Estes, 1975, p. 13)

In recent years individual differences have, more often than not, been viewed within a framework of process, and particularly of strategic differences. Hunt and Lansman (1975) proposed that the ability to produce and use strategies was an important and stable subject characteristic, and might have some relation to degree of schooling. Brown (1975) and her colleagues long emphasized the importance of strategic differences as a source of variation in the performance of normal and retarded subjects. Jarman and Das (1977), despite small sample size and restriction of range on IQ, found evidence for strategy differences in information processing between a high IQ group and the low and average groups. While all the subjects seemed to use simultaneous and successive modes of processing there were qualitative group differences in the methods used to solve the same task. Thus there was evidence suggesting the relationship between processes and the ability to combine the processes into a strategy, or plan, and individual differences as measured by general intelligence tests.

It was decided to further investigate this relationship through the method of componential analysis (Sternberg, 1977b) of pictorial analogy tasks. Analogy solution was chosen as the task for investigation for a number of reasons. The componential theory of analogical reasoning was the most complete and well-documented application of the general method of componential analysis, a relatively new technique. Furthermore, the method was powerful in identifying underlying sources of individual differences on a task at more than one level (e.g., theory, model, and component level differences were isolated).

In addition to the qualities of the analysis itself, analogies have a long history of important roles in psychological theories of intelligence (e.g., Spearman, 1923). As pointed out by Sternberg (1977b), reasoning by analogy is pervasive in everyday life. The measurement of analogical reasoning is a major component in many standardized intelligence and achievement measures including the Standard Progressive Matrices, Miller Analogies Test, Lorge-Thorndike Intelligence Test, and the Graduate Record Examination. A better understanding of what these and other tests measured was to be gained through an in-depth analysis of the processes and strategies of analogical reasoning.

Because scores on standardized group achievement tests were more common measures of individual differences in elementary schools than were standardized intelligence tests, this study adopted a group achievement test as a criterion measure for categorizing students in terms of ability. Group achievement tests were in wide use in elementary schools for a variety of purposes including diagnosis for remediation and enrichment, and grade placement. Despite this widespread use, relatively little was known about the nature of academic achievement as measured by these tests. This study attempted to provide additional information as to the nature of achievement as measured by standardized tests.

One disadvantage to the use of an achievement rather than intelligence measure as the criterion for selection was that achievement measures tend to be related more to school learning than to general reasoning ability. Thus the relationship between scores on the analogy task and group membership on the ability factor could be expected to be somewhat more attenuated than if a measure of general intelligence and reasoning ability (e.g., Standard Progressive Matrices) were used as the criterion.

On the other hand, this disadvantage may be an advantage; because achievement tests tap school learning more than intelligence tests, the results based on an achievement test criterion will also be more closely tied to the educational process with clearer implications for teaching and remediation (Humphreys, 1962a).

Across a number of longitudinal studies, the correlation between achievement and intelligence tests has been assessed. Bloom (1964) reported an average correlation of + .85. Tyler (1974) reported correlations ranging from .40 to .60. Some factors attenuating the correlations included restriction of range, long range predictions, and suitability of the tests for different age groups (Tyler, 1965).

Most researchers assume a consistent, moderate but dependable relationship between intelligence and achievement scores (Tyler, 1965; Vernon, 1970). Humphreys (1962a) found no evidence that achievement differed from intelligence and equated the two. He found that scores on IQ and achievement tests correlated as highly as scores on two different IQ tests.

Thus it was acknowledged that investigation of the relationship between achievement and analogical reasoning would tap one important factor, but not other factors involved, such as motivation, language, SES, and ethnic background.

With these limitations in mind, fourth grade students, approximately 10 years of age, were chosen to participate in the study. Different patterns of strategies and processes were felt to be clear and fairly stable at this age, while these patterns were still in a period of flux in younger students in grades 1 and 2. Furthermore, both intelligence and achievement test scores become increasingly stable with age and by the fourth grade, the scores are reliable in predicting future scores (Bloom, 1964; Tyler, 1974).

Research Questions and Hypotheses

Five research questions were posed. There was sufficient information to state hypotheses associated with the last two questions. Given the ambiguity in the literature surrounding the first three questions, they were not stated as formal hypotheses, but as questions of an exploratory nature.

Theory Level

Question 1. Do high, low and average ability students use the five components (encoding, inference, mapping, application, and response), hypothesized by the componential theory of analogical reasoning in solving separable attribute analogy items?

Sternberg and Rifkin's (1979) investigation indicated that students in grades 2, 4, and 6, and adults did not use mapping on schematic picture analogies, but they did use the remaining components: encoding, inference, application, and response. While their study showed no evidence of developmental differences in the components (processes) used, given the evidence in other studies for individual differences in strategies used in task solution (Brown, 1975), it is possible that Sternberg and Rifkin's (1979) results were sample specific and that students in the present study will differ in the components used in solution.

Question 2. Do the high, average, and low ability groups differ in the extent to which the linear additive combination rule accounts for their performance in schematic picture analogy solution?

In the previous studies using pictorial analogies (Sternberg, 1977b; Sternberg & Rifkin, 1979), all subjects used the additive combination rule and processed in an analytical rather than holistic fashion.

Despite these findings, Sternberg and Rifkin (1979) acknowledged that parallel processing or nonlinear component combination rules were possible in these analogy tasks, thus it is possible that the groups will differ in the degree to which the linear additive combination rule accounts for the data.

Model Level

Question 3. Do high, average, and low ability students differ in the rule they use for combining multiple executions of the same component (i.e., in preferred model choice)?

Components were assumed to be serially executed if the combination rule was linear and additive (Question 2), but the mode of execution within components may differ. Mode (self-terminating versus exhaustive) of execution was specified in the seven hypothesized models (Tables 1, 2).

Sternberg and Nigro (1980), and Sternberg and Rifkin (1979), found that younger students tended to use more self-terminating and less exhaustive operations than did older subjects on people piece analogies. Brown and DeLoache (1978) suggested that increasing use of exhaustive processing was a general developmental characteristic. Thus if low ability students are assumed to be developmentally immature as compared with equal CA average and high ability students, they may use fewer exhaustive operations, on some tasks. However, in Sternberg and Rifkin (1979) all subjects used only self-terminating operations (Model 4M) on the schematic picture analogies, thus no differences in component execution may be evident on these particular tasks.

Question 4. Does consistency in strategy use vary as a function of level of ability on separable attribute analogy tasks?

In previous investigations older subjects were found to be more systematic than younger subjects in strategy use (Sternberg, 1980; Sternberg & Nigro, 1980; Sternberg & Rifkin, 1979) and high reasoners were more consistent in choice of strategy than low reasoners (Sternberg, 1977b).

Hypothesis 4.1. Consistency in model choice will increase as a function of level of ability.

Component Level

Question 5. Do component latencies and error rates vary as a function of level of ability?

Jarman and Das (1977) suggested that speed of central processing may vary with age and covary with intelligence level (among other variables). In Sternberg and Rifkin (1979) older children and adults had fewer errors than did younger children. Older subjects executed most component operations more rapidly than did younger students.

Hypothesis 5.1. Error rates and latency scores for components and overall task on separable attribute analogies will vary as a function of ability level.

An additional area of interest, while not tested directly in hypothesis form was the degree of correspondence between fourth grade students' performance in this study, and the performance of fourth grade students in a similar study by Sternberg and Rifkin (1979).

CHAPTER IV

METHOD

Subjects

The sample pool consisted of 155 fourth grade students from nine classes in four elementary schools. The schools were all within the same school district located in a metropolitan area in south western B.C. The students in the schools represented a range of socio-economic levels and came from a variety of ethnic backgrounds.

Formation of Subgroups

From this sample of 155 students, three equal size groups were selected: students of high achievement, average achievement, and low achievement. The students were selected on the basis of their existing scores on the Canadian Test of Basic Skills (King, 1976), a group administered standardized achievement test. The Canadian Test of Basic Skills (CTBS) had been administered to the students by classroom teachers the previous year, while the students were in grade three.

The CTBS, in wide use across Canada, is essentially a Canadian adaptation of the Iowa Test of Basic Skills. Students are tested on vocabulary, reading comprehension, language skills, work-study skills, and mathematics skills. The test measures generalized academic skills rather than achievement in specific content areas. The test provides grade equivalent, percentile rank, and stanine subtest and composite scores.

The CTBS was standardized in 1973 on a stratified random sample of English schools across Canada (King, 1976). The test manual reported the following split-half reliability estimate for the grade 3 level composite scores, r = .98. Intercorrelations among subtests and between subtests and composite scores ranged from r = .49 to r = .93 for the third grade sample. Stability data were not presented. Validity data were not available in published test materials.

Of the 155 fourth grade boys and girls, 27 either had no existing CTBS scores, or were absent during the experimental sessions, thus reducing the potential sample pool to 128 students. The sampling pool was further reduced to 124 when 4 students were dropped from one session after creating a disturbance.

The remaining 124 students had grade-equivalent CTBS scores. Because the tests had been administered in different months of the year in the four schools, these grade-equivalent CTBS scores were converted to percentile scores (King, 1977). The percentile scores were in turn converted to Normal Curve Equivalent (NCE) scores (Tallmadge & Wood, 1976). The NCE is a "normalized standard score that has been linearly transformed to match the percentile distribution at values of 1, 50, and 99" (Tallmadge & Wood, 1976, p. 2). The scores range from 1.00 to 99.00, with a mean of 50.00 and a standard deviation of 21.06. The scale is assumed to be equal interval, and thus permits numerical calculations, in contrast to percentiles.

Table 3 summarizes the means, standard deviations, and range for the sample of 60 students at three ability levels. Twenty students (12 male, 8 female) had scores between one and three standard deviations below the sample mean and were classified as the low achievement group. Twenty students (9 male, 11 female) who had scores between one and three

Table 3

Descriptive Statistics: NCE Scores

Group	n	Mean	SD	Range	Mean Age
Low	20	36.51	7.68	17.30-44.10	9 yrs 9 mos
Average	20	60.92	9.46	44.70-78.20	9 yrs 8 mos
High	20	89.66	7.71	79.60-99.00	9 yrs 8 mos

standard deviations above the sample mean were classified as a high achievement group. The remaining 84 students, who had scores between one standard deviation above or below the sample mean, were classified as the average achievement group. Twenty students (9 male, 11 female) were drawn from the average group to maintain equal sample sizes at each ability level. The average range of the NCE scores was divided into four equal intervals and students drawn randomly but proportionately from each interval to yield the desired sample of size 20. The standard deviation in the average group was higher than in the low and high groups, but since the data analysis was group oriented, this difference should not influence the results.

Instruments

Schematic Picture Analogies

The Schematic Picture Analogies (Figure 1) were of the form A is to B as C is to D_1 or D_2. Each term of the analogy varied on four separable attributes. Separable attributes are those which can be nullified without destroying the intactness of the stimulus. Each attribute had two possible values: hat color (black or white), suit pattern (striped or polkadotted), handgear (suitcase or umbrella), and footwear (shoes or boots).

The analogies were presented in 24 booklets, each of which contained

16 analogies, 4 per page (see Appendix A). Subjects were given 64 seconds to complete each booklet. The 16 items within each booklet were homogeneous in the number of attribute values transformed from the A to B terms, A to C terms, and D_1 to D_2 terms. For example, in Figure 1 there were three attribute value transformations between the A and B terms: hat color, footwear, and handgear. One attribute value was transformed between the A and C terms: suit pattern. The two transformations between the D_1 and D_2 terms were suit pattern, and footwear. Thus, if this analogy item was in a booklet of 16 items, all the items in that booklet would have three attribute value transformations between A and B, one between A and C and two transformations between D_1 and D_2. The identity of the attribute values transformed varied across items within a booklet, but the number of transformations was constant. Students recorded the 16 responses for each booklet on a separate answer sheet (see Appendix B).

Sets of 24 booklets were numbered, and contained in separate folders. Order of pages within each booklet and order of booklets within each set of 24 were random, as a control for practice and session effects.

Procedure

A pilot administration of the procedure was performed, followed by the main data collection.

Pilot Study: Initial Procedure

An initial test administration procedure, adapted from the procedure developed by Sternberg and Rifkin (1979) was pilot-tested with a group of 35 fourth grade students.

The students were asked to participate in a research project and were reassured that their performance would not affect their school grades. The students were then introduced to the Schematic Picture

Analogy problem. Sample sheets of four schematic picture items were distributed and presented simultaneously on an overhead projector. The four relevant attributes and their values were identified by the students in response to the experimenter's question "How are these clowns alike? In what ways are they different?"

Once the students were familiar with the attributes and values, the goal of the analogy problem was explained. The students were told to choose the answer option (D_1 or D_2) that was the same as and different from C in the same ways that A was the same as and different from B. They were also told to use the four attributes and their values to guide them in their answer choice for the first sample item.

After the students had recorded their responses on the answer sheet, the correct option was indicated and students were told that the option was correct because it was the same as and different from the third analogy term as the second term was from the first. The students were then asked to solve the three remaining sample analogies; feedback was provided and questions were answered for the sample items.

The 24 analogy booklets were then administered. Students were reminded that they were not expected to complete all of the analogies, but to work as well as they could within the allotted time of 64 seconds. The session was to take approximately 1 hour. No feedback was given for the booklets.

Despite the success of this pilot procedure with other children in grades 2, 4, and 6 reported by Sternberg and Rifkin (1979), the lack of understanding and confusion observed in the group of 35 students dictated a revision of the procedure.

A number of problems were identified:

1. Students did not understand the goal of the task even after completion of the sample items;
2. students found the answer sheets confusing;
3. one hour was insufficient time to complete all 24 booklets. Only ten were completed in the first session;
4. the group was too large to permit the experimenter to deal with questions individually; and
5. the group was too large for the experimenter to monitor behavior and prevent disruptions.

In order to reduce the task related confusion, the training, or practice session in the first hour was augmented, to increase familiarization with the task and materials. The time problem was solved by allotting two 1-hour sessions for completion of the 24 booklets. Group size was reduced to allow for more individual attention to questions and to permit better management of the students. And, to reduce the memory demands of the task, the attributes and their values were listed on the classroom blackboard.

Pilot Study: Revised Procedure

The first session of the revised procedure was pilot-tested on a second group of 15 fourth grade students. Since the procedure for session 2 was very similar to session 1, it was not pilot-tested.

Again, the students were asked to participate in a research project which was an attempt to find out how children learn at school and were reassured that the activities were not in any way related to their school grades. They were then told the activities involved cartoon figures (Schematic Picture Analogies), but before beginning they had to learn how to do the activities.

Each student was given two sample analogy booklets numbered 1 and 2, answer sheets, and a pencil. The sample items were also displayed on the overhead projector. The list of relevant attributes and their values were elicited through questioning and written on the board where it remained in view for the duration of the session.

Next the goal of the analogy was introduced, to choose the correct option to complete the analogy. Students were told:

> "Look at the first two clowns. How are they the same? How are they different? Now look at the third clown, this one. You must choose a partner for this clown. But his partner must be the same as and different from him (clown 3) in the same ways as the first two clowns were the same and different."

Next, the students were shown how to record their chosen option on the answer sheets. Feedback was given. The first three pages of booklet 1 were completed in this manner, and the experimenter dealt with any confusion related to the task and ensured that students were recording their responses properly.

The second sample booklet was used to introduce the timed nature of the task. Students were told they had 64 seconds to work on a booklet but that accuracy, and not speed, was important. Following the completion of booklet 2, any further questions were dealt with. The practice session took approximately 25 minutes.

The students then worked on the first nine analogy booklets in their set for the remainder of the hour. No problems were observed using these revised procedures.

Main Data Collection

Data were collected from nine classes. Classes were usually divided into two groups for testing. The smallest group tested consisted of 11 students and the largest group had 22 students. The sessions were conducted in the school library, an empty classroom, or the staff room. Teachers were not present during either session.

The first session, which lasted for 1 hour, consisted of an introduction to the materials and task followed by administration of 9 of the 24 timed analogy booklets. Timing was done by the experimenter with a stopwatch.

The second session, which lasted approximately 40 minutes, consisted of a brief review of the attributes and values which were again listed on the board. Next the goal of the task, task instructions, and response format were reviewed. Then the students completed the remaining 15 booklets.

Design

Dependent Variables

Scoring. Each analogy item was scored 1 if correct, 0 if incorrect. Items not attempted were not scored. The correct option was that which completed the second half of the analogy so that the A to B term relation was the same as the C to D relation, and in the same direction. Three scores were then derived for each of the 24 analogy booklets: number of items answered correctly in a booklet (maximum = 16); number of items completed in a booklet, both correct and incorrect (maximum = 16); and number of items incorrect (maximum = 16). These three scores were used to calculate two latency scores and an error rate score for each subject on each booklet.

Dependent variable 1: latency correct. The score for number of items answered correctly in a booklet was used to calculate the first dependent variable, solution latency for correctly answered items. The time allotted for a booklet, 64 seconds, was divided by the number of correct items on a booklet. Any student with a score of 0 for number of correct items on a booklet was automatically assigned a score of 1, to permit division. In many studies, students with worse than chance performance are dropped. In this study, a major emphasis was the investigation of the performance of the low ability students, thus discarding these subjects with 0 correct scores was unjustified. In lieu of dropping the students, the correction was made. This correction was only necessary for three students and only on dependent variable 1.

Thus, the maximum latency for correct items was $\frac{64}{1}$ = 64 seconds per correct item. The minimum latency was $\frac{64}{16}$ = 4 seconds per correct item.

<u>Dependent variable 2: latency completed</u>. The second raw score was the number of items completed in a booklet, both correct and incorrect. This score was used to calculate the second dependent variable, solution latency for all answered items. This score was calculated by dividing 64 seconds by the total number of completed items. As for solution latency for correct items, the maximum latency for completed items was $\frac{64}{1}$ = 64 seconds per completed item and the minimum latency for completed items was $\frac{64}{16}$ = 4 seconds per completed item.

<u>Dependent variable 3: error rate</u>. The third raw score was the number of items incorrect in a booklet. This score was used to calculate the third dependent variable, error rate. This variable was computed by dividing the number of incorrect items by the total number of items completed in a booklet. The maximum score was 1.00 and the minimum score was 0.00.

Parameter Estimation and Models

The next step in the analysis consisted of deriving component scores from the booklet scores. Component scores were derived by decomposing total time spent on an analogy item into estimates of the time spent on each component. This was accomplished through multiple linear regression using the complete least squares approach.

Seven models were hypothesized by Sternberg and Rifkin (1979) to account for performance on pictorial analogy problems (Tables 1, 2). Table 4 represents the components in the seven models for which latency and error estimates were derived. Some of the component estimates were confounded due to the nature of the task materials (see pp. 66-68). Encoding and response component estimates were confounded in Models 1 to 4 and Model 1M. Inference and application component estimates were confounded in Models, 1, 4, 1M, and 4M.

The basis of the componential method of analysis was that response times (or error rates) for each booklet, i.e., the criterion scores, were predicted from independent variables representing variations in the complexity of analogy items in the 24 booklets.

Criterion variables: booklet scores. The criterion scores were generated by collapsing booklet scores on a dependent variable across subjects at each ability level. The data frame used is summarized in Table 5. There were 20 students in each group. Each of those students had 24 latency correct scores, 24 latency completed scores, and 24 error rate scores; a score for each of 24 booklets on the 3 dependent variables. The criterion scores for the regressions were not subjects' scores, but booklet scores. The 20 subjects within each group served as replications to ensure the reliability of each booklet score. Scores

Table 4

Models for Regression

Model	Components			
1	*encoding-response	*exhaustive inference-application		exhaustive mapping
	$\hat{Y} = b_0$	$+ \ b_1X_1$		$+ \ b_2X_2$
2	*encoding-response	exhaustive inference	exhaustive mapping	self-terminating application
	$\hat{Y} = b_0$	$+ \ b_1X_1$	$+ \ b_2X_2$	$+ \ b_3X_3$
3	*encoding-response	exhaustive inference	self-terminating mapping	self-terminating application
	$\hat{Y} = b_0$	$+ \ b_1X_1$	$+ \ b_2X_2$	$+ \ b_3X_3$
4	*encoding-response	*self-terminating inference-application		self-terminating mapping
	$\hat{Y} = b_0$	$+ \ b_1X_1$		$+ \ b_2X_2$
1M	*encoding-response	*exhaustive inference-application		
	$\hat{Y} = b_0$	$+ \ b_1X_1$		
2-3M	response	exhaustive inference	self-terminating application	self-terminating encoding
	$\hat{Y} = b_0$	$+ \ b_1X_1$	$+ \ b_3X_3$	$+ \ b_4X_4$
4M	response	*self-terminating inference-application		self-terminating encoding
	$Y = b_0$	$+ \ b_1X_1$		$+ \ b_4X_4$

* = confounded components

Table 5

Data Matrix Used to Calculate Criterion Variable Scores

Group	Subject	Booklet 1 DV_1	DV_2	DV_3	Booklet 2 DV_1	DV_2	DV_3	...	Booklet 24 DV_1	DV_2	DV_3
Low	1										
	2										
	.										
	.										
	.										
	20										
Sum		$\sum_{n=1}^{20} DV_1$	$\sum_{n=1}^{20} DV_2$	$\sum_{n=1}^{20} DV_3$							
Mean (Criterion Variable)		$\frac{\Sigma DV_1}{20}$	$\frac{\Sigma DV_2}{20}$	$\frac{\Sigma DV_3}{20}$							
		Mean latency correct	Mean latency completed	Mean error rate							
Average	1										
	2										
	.										
	.										
	.										
	20										
Sum											
Mean (Criterion Variable)											
High	1										
	2										
	.										
	.										
	.										
	20										
Sum											
Mean (Criterion Variable)											

on each booklet were averaged across subjects within ability level for each dependent variable, to yield three sets of 24 mean booklet scores: mean latency correct scores for booklets 1 to 24, mean latency completed scores for booklets 1 to 24, and mean error rate scores for booklets 1 to 24.

Predictor variables. Each of the 24 booklets was structured so that the complexity (defined as the number of attribute-value transformations between relevant analogy terms) varied systematically across the booklets, as shown in Table 6. Each column represents the number of attribute value transformations across the 24 analogy booklets for particular terms. Variation in complexity of these columns, when entered as a predictor variable in the regression, should predict variation in the overall booklet score if that component was used in solution.

For example, the numbers in column 1 of Table 6 represent the 'distances' (numbers of transformations) between the A and B terms of the 24 analogy booklets. The values range from one to three. This column was used as the predictor variable to estimate exhaustive inference latencies and error rate. Thus inference latency and difficulty, and consequently total latency and difficulty should increase as the number of A to B values transformed increases, if exhaustive inference is used in solution.

There are two varieties of estimates, those for exhaustively executed components and those for components executed in a self-terminating mode. Exhaustive component estimates were based on the objective numbers of attribute value transformations. For example, exhaustive inference latency and difficulty were defined as a function

Table 6

Predictor Variables for Regression

Analogy Book	Inference A-B		Mapping A-C		Application $C-D_T$		Encoding		Encoding		$D_T = D_F$	Multiplier
	1	2	3	4	5	6	7	8	9	10	11	12
	e	st	e	st	e	st	e	st	e*	st†	(h)	$\frac{5}{20-4h}$
1	1	.25	3	.75	1	.25	5	1.25	3	.75	0	5/20
2	1	.63	2	1.25	1	.63	5	3.13	3	1.88	3	5/8
3	1	.25	1	.25	1	.25	5	1.25	3	.75	0	5/20
4	1	.63	3	1.88	1	.63	5	3.13	3	1.88	3	5/8
5	1	.63	1	.63	1	.63	5	3.13	3	1.88	3	5/8
6	1	.42	1	.42	1	.42	5	2.08	3	1.25	2	5/12
7	1	.42	2	.83	1	.42	5	2.08	3	1.25	2	5/12
8	1	.25	2	.50	1	.25	5	1.25	3	.75	0	5/20
9	1	.31	2	.63	1	.31	5	1.56	3	.94	1	5/16
10	1	.42	3	1.25	1	.42	5	2.08	3	1.25	2	5/12
11	1	.31	1	.31	1	.31	5	1.56	3	.94	1	5/16
12	1	.31	3	.94	1	.31	5	1.56	3	.94	1	5/16
13	2	.83	1	.42	2	.83	5	2.08	3	1.25	2	5/12
14	2	.50	1	.25	2	.50	5	1.25	3	.75	0	5/20
15	2	.63	1	.31	2	.63	5	1.56	3	.94	1	5/16
16	2	.50	2	.50	2	.50	5	1.25	3	.75	0	5/20
17	2	1.25	2	1.25	2	1.25	5	3.13	3	1.88	3	5/8
18	2	1.25	1	.63	2	1.25	5	3.13	3	1.88	3	5/8
19	2	.83	2	.83	2	.83	5	2.08	3	1.25	2	5/12
20	2	.63	2	.63	2	.63	5	1.56	3	.94	1	5/16
21	3	.75	1	.25	3	.75	5	1.25	3	.75	0	5/20
22	3	1.88	1	.63	3	1.88	5	3.13	3	1.88	3	5/8
23	3	.94	1	.31	3	.94	5	1.56	3	.94	1	5/16
24	3	1.25	1	.42	3	1.25	5	2.08	3	1.25	2	5/12
Mean	1.67	.67	1.67	.67	1.67	.67	5	2.01	3	1.21		

* partially exhaustive
† partially self-terminating

Note: e = exhaustive execution
st = self-terminating execution

of the number of attribute values transformed from the A to B analogy terms. These values for the 24 booklets are listed in column 1 of Table 6, and ranged from one to three. Column 1 was the predictor variable for exhaustive inference estimates in all seven models.

Exhaustive mapping latency and difficulty were defined as a function of the number of attribute values transformed from A to C analogy terms. These values for each booklet, listed in column 3 of Table 6, ranged from one to three. Column 3 was the predictor variable for exhaustive mapping component estimates in all seven models.

Exhaustive application latency and difficulty were defined as a function of the number of attribute values transformed between the C and D_{True} terms (always equal to the A to B distance). Column 5 of Table 6, summarizes these transformations across the 24 booklets. The number of attribute values transformed from C to the correct answer option had to be equal to the number of values transformed from A to B, to permit correct completion of the analogy. Thus columns 1 and 5 of Table 6 are identical.

Encoding was defined as the number of analogy terms (figures) to be encoded. Since in exhaustive encoding all of the terms are encoded, column 7 of Table 6 represents the values for the predictor variable for exhaustive encoding. The value is five for all 24 booklets. In order for a component to be independently estimated, it must be represented by a predictor variable which varies across booklets. Therefore exhaustive encoding and response (which is also constant across booklets) components were confounded and estimated as the regression constant in Models 1 to 4 and in Model 1M (see Table 4).

Independent or predictor variables for self-terminating components are also based on the number of attribute-value transformations but in addition they are a function of the distance between the two answer options, D_T and D_F. The fewer values the options have in common, the easier it is to distinguish the correct option when in a self-terminating mode.

Columns in Table 6 representing predictor variables for self-terminating components were derived by multiplying the predictor values for the corresponding exhaustive component by the following multiplier (see column 12, Table 6):

$$\frac{N + 1}{[N(N - h + 1)]} ,$$

where N = the number of attributes that could be encoded (always four), and h = the number of D_F values that were the same as D_T values (column 11).

Thus the predictor variable (column 2) for self-terminating inference was derived by multiplying column 1 by column 12.

Self-terminating mapping was estimated by using column 4 as a predictor variable, and self-terminating application estimates were based on column 6, which was identical to column 2.

When encoding was self-terminating (Models 2-3M and 4M, Table 4), it was possible to derive separate estimates for encoding latency and difficulty. In Model 4M, encoding was self-terminating for all five terms of the analogy, thus column 8 in Table 6, the predictor variable for self-terminating encoding, was generated by multiplying column 7 by column 12. This created variation across booklets whereas column 7 had no such variation.

In Model 2-3M, encoding was hypothesized to be exhaustive for the first terms and self-terminating for the last three terms. The first two terms were thus confounded with the response component, but the predictor variable for encoding for the last three terms was column 10 of Table 6, computed by multiplying column 9 by the self-terminating multiplier, column 12.

Because the predictor variables for exhaustive inference and exhaustive application were identical (columns 1 and 5 in Table 6), when a model hypothesized both exhaustive inference and application, the component estimates were confounded. Model 1 and 1M in Table 4, are examples of this confounding. Similarly, the predictor variables for self-terminating inference and application were identical (columns 2 and 6 of Table 6), so that self-terminating inference and application estimates were confounded in Model 4 and Model 4M in Table 4.

Thus the multiple regressions for each model in Table 4 consisted of the criterion scores (mean booklet latencies and error rates, see Table 5), and the relevant predictor variables from Table 6.

What were estimated in this regression procedure were the component coefficients. These can be interpreted as estimates of component latencies and difficulty. This differed from the usual regression case in that task variation, and not subject variation was being predicted.

In summary, ability, a three-level factor (high, average, low) was crossed with task materials, a repeated measures factor consisting of 24 schematic picture analogy booklets (see Table 5). At each of three ability levels (low, average, high), three sets of 24 criterion scores were derived by collapsing booklet scores across the 20 subjects within each ability level. These sets of 24 criterion scores for each group were entered into seven regressions, one for each hypothesized model for a total of $3 \times 3 \times 7 = 63$ regressions (ability × set of criterion scores × models = 63). Once each model was fitted to the data, the model which best accounted for task variance was determined for each ability level on all three criterion variables. The criteria for selection of the best model will be discussed in the next chapter together with the presentation of the results.

The program used in the regression analyses was UBC TRP (Le & Tenisci, 1978), and was run on the Amdahl 470 V/6, Model II computer under the Michigan Terminal System (MTS).

CHAPTER V

RESULTS

Preliminary Analysis

Before beginning the regression analyses, the criterion variables described in Chapter IV, Table 5, were evaluated to decide whether or not all three criterion variables should be included in subsequent analyses. The first step was to ensure that the data were analyzable; that is, to determine whether there was sufficient variance across the mean booklet scores to permit regression analysis. The second step involved examination of the variance-covariance matrices of the criterion variables for each ability group. If the covariances of the three variables were similar, then there would be no need to analyze all three criterion variables in a group. Subsequent analyses would therefore be performed on only one of the three criterion variables. The three criterion variables at each ability level were calculated as described in Chapter IV, (see Table 5). The three sets of scores were: mean latency correct, mean latency completed, and mean error rate scores for the 24 analogy booklets at each ability level.

The means, standard deviations, ranges, and variance of the three criterion variables at each ability level are summarized in Table 7. Inspection of criterion variable 3, mean error rate, revealed that in

Table 7

Descriptive Statistics: Criterion Variable

Group	Criterion Variable	Mean	SD	Variance	Range
Low	1	14.51	3.52	12.39	9.58-23.71
	2	8.45	1.73	2.99	6.74-13.73
	3	.34	.07	.00	.22- .48
Average	1	10.51	2.83	8.01	6.41-18.43
	2	7.44	1.18	1.39	5.76- 9.44
	3	.17	.07	.00	.08- .33
High	1	7.71	1.57	2.46	5.49-11.88
	2	7.42	1.35	1.82	5.34-10.71
	3	.03	.02	.00	.00- .08

Note: These values were rounded to two decimal places. Values of .00 were results of rounding.

all three ability groups there was insufficient variance to permit subsequent analyses to be carried out.

The remaining criterion variables, mean latency correct and mean latency completed, were included in regression analyses for all three ability groups.

The second step in the preliminary analysis, examination of the variance-covariance matrices, was not necessary, given the decision to drop variable 3 in subsequent analyses.

Regression Analysis

The results of the regression analyses conducted for each model and which were used to determine model preference and values for component latency estimates are presented separately for each ability group and each criterion variable which was analyzed.

Determination of Model Preference

The seven models hypothesized by Sternberg and Rifkin (1979) for the solution of pictorial analogies were evaluated in two phases. The models described in Table 4 are reproduced in Table 8 for reference.

The first phase involved the evaluation of the two sets of models, models 1 to 4 and 1M to 4M. These two sets of models differed in that a) mapping components were included in Models 1 to 4, but not 1M to 4M, and b) encoding was exhaustive in Models 1 to 4, but could be self-terminating in Models 1M to 4M. Thus, if mapping estimates were not significant in Models 1 to 4, the models were rejected, for unless a component occupied non-trivial time, it was assumed that it was not used in solution. Further evidence for the rejection of Models 1 to 4 would be the significance of self-terminating encoding estimates in the second set of models.

Table 8

Models for Regression

Model	Components			
1	*encoding-response	*exhaustive inference-application		exhaustive mapping
	$\hat{Y} = b_0$	$+ b_1X_1$		$+ b_2X_2$
2	*encoding-response	exhaustive inference	exhaustive mapping	self-terminating application
	$\hat{Y} = b_0$	$+ b_1X_1$	$+ b_2X_2$	$+ b_3X_3$
3	*encoding-response	exhaustive inference	self-terminating mapping	self-terminating application
	$\hat{Y} = b_0$	$+ b_1X_1$	$+ b_2X_2$	$+ b_3X_3$
4	*encoding-response	*self-terminating inference-application		self-terminating mapping
	$\hat{Y} = b_0$	$+ b_1X_1$		$+ b_2X_2$
1M	*encoding-response	*exhaustive inference-application		
	$\hat{Y} = b_0$	$+ b_1X_1$		
2-3M	response	exhaustive inference	self-terminating application	self-terminating encoding
	$\hat{Y} = b_0$	$+ b_1X_1$	$+ b_3X_3$	$+ b_4X_4$
4M	response	*self-terminating inference-application		self-terminating encoding
	$\hat{Y} = b_0$	$+ b_1x_1$		$+ b_4X_4$

* = confounded components

In the second phase, the remaining models were evaluated in terms of the following six interdependent criteria:

a) Values of R^2 in competing models.

b) Significance of increase in R^2. Because some models hypothesized one more component than a competing model, the significance of the increase in R^2 due to the additional component was evaluated. If the increase was not significant, then the additional component was considered not necessary, and the more parsimonious model was preferred.

c) Significance of regression coefficients.

d) Values of regression F for the competing models. The F ratio takes into account the number of predictor variables in an equation. If an additional predictor variable increased the value of F this indicated support for that model. If the additional predictor variable in a model decreased the value of F, then the more parsimonious model was preferred.

e) The proportion of decrease in the standard error of estimate due to an additional component in competing models. If the decrease was small, the additional parameter was considered unwarranted and the parsimonious model was given priority.

f) The nature of the component estimates. If a coefficient was negative, but small and not statistically significant, the negative value was attributed to sampling error and interpreted as nonsignificant. If a negative coefficient was large and/or statistically significant, the models were rejected as this was a sign that the assumption of serial processing and the linear rule for component combination may have been violated.

Criterion Variable 1: Latency Correct

Low ability. The raw regression coefficients and indices of model fit for the seven models on latency correct scores for the low ability students are summarized in Table 9. Using the phase 1 criteria discussed above, Models 1 to 4 were rejected. First, none of the mapping latency estimates differed significantly from zero in Models 1 to 4. Second, the self-terminating encoding latency estimate in Model 2-3M was significant ($p < .01$).

Models 1M, 2-3M, and 4M were then evaluated using the phase 2 criteria. The first criterion was the values of R^2 in competing models. Model 2-3M accounted for the largest proportion of variance ($R^2 = .54$), followed by Model 4M ($R^2 = .34$), and Model 1M ($R^2 = .28$).

The second criterion involved the significance of the increase in R^2 due to an additional parameter in one of the two models evaluated. Since inference and application were unconfounded in Model 2-3M, but were confounded in Model 4M, it was possible to evaluate the change in R^2 between Model 2-3M and Model 4M. The difference was significant ($F_{1,20} = 9.33$, $p < .01$), thus the additional component in Model 2-3M was warranted. Model 4M had an additional component over Model 1M because encoding and response estimates in Model 4M were unconfounded. The change in R^2 between Model 4M and 1M was not significant ($F_{1,21} = 2.00$, $p > .05$), thus Model 1M and 2-3M were favored on this criterion.

The third criterion, the significance of the regression parameters, indicated Model 1M and 4M were preferred over Model 2-3M. Inference-application and encoding-response coefficients were significant in Model 1M. Inference and self-terminating encoding coefficients were significant in Model 2-3M but application and response coefficients were not.

Table 9

Model Fits for Low Ability Group:
Criterion Variable 1

					Component Estimates					
Model	Inf.	Inf.-Appl.	Appl.	Map.	Enc.	Enc.-Resp.	Resp.	R^2	F	σ est
1		2.47*		.08		10.27**		.28	4.12*	3.14
2	1.00		3.67	.08		10.26**		.35	3.55*	3.04
3	1.85		2.33	1.77		8.68**		.38	4.10*	2.98
4		5.01**		.35		10.92**		.34	5.48*	3.01
1M		2.43**				10.46**		.28	8.48**	3.07
2-3M	6.19**		-9.39		8.68**		0.01	.54	7.96**	2.56
4M		4.34*			.63		10.33**	.34	5.48*	2.99

Note: Estimates are in terms of seconds spent on a component per analogy item.

* = $p < .05$
** = $p < .01$

Finally, inference-application and response coefficients were significant in Model 4M but self-terminating encoding was not.

The fourth criterion was the value for the regression F. In this case, the F for Model 2-3M ($F = 7.96$) was higher than that for Model 4M ($F = 5.48$). The F for Model 1M ($F = 8.48$), was higher than that for Model 2-3M or 4M. Despite these differences, all three F ratios were significant ($p < .01$), so no model was designated as preferred on this criterion.

The proportion of decrease in the standard error of estimate (σ est) from Model 4M to Model 2-3M was .15, and the decrease from Model 1M to 4M was .03. Neither decrease was considered sufficient to warrant the additional component, therefore Models 1M and 4M were preferred.

Finally, Model 2-3M had a negative coefficient for the self-terminating application component, but this coefficient was not significant and thus was attributed to random error. Neither Model 1M nor Model 4M had negative component coefficients.

Model 2-3M was rejected after being evaluated on the preceding criteria. Model 1M was designated as marginally preferred over Model 4M for the latency correct data for low ability students. However, in subsequent analyses of criterion variable 1, both Model 1M and 4M were evaluated wherever possible due to the equivocal nature of the latency correct data in this group. In two of the subsequent analyses Model 4M was used as the preferred model to maintain comparability across the groups. This was only for convenience of those two analyses, and was no indication that Model 4M was a better model than 1M for explaining the low group's latency correct data.

<u>Average ability</u>. The same criteria were applied to the latency correct data for average ability students, presented in Table 10. Models

Table 10

Model Fits for Average Ability Group:
Criterion Variable 1

					Component Estimates					
Model	Inf.	Inf.-Appl.	Appl.	Map.	Enc.	Enc.-Resp.	Resp.	R^2	F	σ est
1		2.04*		-.37		7.74**		.37	6.17**	2.36
2	.43		4.01*	-.37		7.74**		.51	6.94**	2.13
3	.69		3.89	.15		6.65**		.50	6.68**	2.14
4		4.90**		-.38		7.49**		.49	10.21**	2.11
1M		2.22**				6.81**		.36	12.41**	2.32
2-3M	2.67		-1.11		3.40		2.71	.55	8.13**	2.05
4M		4.81**			.07		7.16**	.49	10.21**	2.12

Note: Estimates are in terms of seconds spent on a component per analogy item.

* = $p < .05$
** = $p < .01$

1 to 4 were again rejected as estimates for the mapping component were not significant in any model.

In terms of values of R^2, the first criterion, Model 2-3M was preferred ($R^2 = .55$), followed by Model 4M ($R^2 = .49$), and Model 1M ($R^2 = .36$).

The difference in R^2 between Models 2-3M and 4M was not significant, indicating the additional parameter in Model 2-3M was unwarranted ($F_{1,20} = 2.67$, $p > .05$). In comparing the difference in R^2 between Models 4M and 1M, the additional parameter was warranted due to the significant difference in R^2 ($F_{1,21} = 5.42$, $p < .05$). Model 4M was preferred therefore, on the second criterion.

In terms of the third criterion, the significance of the regression coefficients, Model 1M was the preferred model. None of the component estimates were significant in Model 2-3M. The inference-application and response coefficients were significant in Model 4M, but self-terminating encoding was not. All of the component estimates were significant in Model 1M.

The fourth criterion was the values of regression F. Again, all three F ratios were significant, so the models were equivalent on this criterion.

The fifth criterion, proportion of decrease in the standard error of estimate, indicated that Model 1M was preferred. While the proportion of decrease from Model 4M to Model 2-3M (.04) indicated the additional parameter in Model 2-3M was unnecessary, the proportion of decrease from Model 1M to 4M (.09) indicated the additional parameter in Model 4M was also unnecessary. However, since Model 4M was preferred on criterion 2 which was a more objective measure of the value of the additional parameter than was criterion 5, the preference of Model 1M on criterion 5 was outweighed by the preference of Model 4M on criterion 2.

In terms of the nature of the regression coefficients, none of the estimates in Models 1M or 4M were negative. Model 2-3M had a negative estimate (self-terminating application = -1.11) but the estimate was small, and attributed to sampling error, thus no model was preferred over the others on this criterion.

Model 4M and 1M were similar on most of the criteria, but the increase in the value of R^2 of Model 4M over 1M was significant. Model 4M thus was designated the preferred model for latency correct data in the average ability group.

High ability. The latency correct results for high ability students are summarized in Table 11.

Models 1 to 4 were rejected since only one of the four mapping estimates was statistically significant, whereas self-terminating encoding estimates were significant in both Models 2-3M and 4M. Model 1M was also rejected as it accounted for only 7% of the variance in the data.

Model 2-3M had a higher R^2 than Model 4M (.90 and .89 respectively), however this difference in the R^2 values was not significant ($F_{1,20}$ = 2.00, $p > .05$), so Model 4M was preferred on this criterion.

All of the regression coefficients were significant in Model 4M. The inference estimate was not significant in Model 2-3M, thus Model 4M was preferred on the third criterion.

The overall regression F value was larger in Model 4M than in Model 2-3M, but both F values were significant at the .01 level, thus no model preference was established by this criterion.

The proportion of decrease in the standard error of estimate from

Table 11

Model Fits for High Ability Group:
Criterion Variable 1

					Component Estimates					
Model	Inf.	Inf.-Appl.	Appl.	Map.	Enc.	Enc.-Resp.	Resp.	R^2	F	σ est
1		.53		-.05		6.92**		.07	.80	1.59
2	-1.59**		5.28**	-.05		6.91**		.87	44.62**	.61
3	-1.41**		5.05**	.31		6.48**		.87	44.62**	.60
4		3.00**		1.40**		4.77**		.75	31.25**	.82
1M		.55				6.80**		.07	1.67	1.55
2-3M	-.60		2.88*		1.60*		4.86**	.90	60.00**	.53
4M		1.54**			1.41**		3.86**	.89	89.00**	.54

Note: Estimates are in terms of seconds spent on a component per analogy item.

* = $p < .05$
** = $p \leq .01$

Model 4M to Model 2-3M was .02. This was not sufficient to warrant the additional parameter in Model 2-3M so Model 4M was preferred on this criterion.

The nature of the regression coefficients did not lead to any model preference. All of the coefficients were positive in Model 4M. Inference was negative in Model 2-3M but was small (-.60) and thus interpreted as nonsignificant.

Model 4M was designated the preferred model for the first criterion variable, latency correct, for high ability students.

In summary, Model 4M was the preferred model for average and high ability groups for latency correct data. Model 1M was marginally preferred over Model 4M in the low ability group. The decisions regarding model preference on this variable were made with greater confidence in the high ability group than in the low and average ability groups. This corresponded with higher values for the standard error of estimate in the low and average groups than in the high ability group.

The regression weights estimated in the preferred models were interpreted as estimates of the latency of a single execution of a component for one analogy item. For example, in the average ability group, a single execution of self-terminating inference-application component was estimated to take 4.81 seconds of the total time spent on an item. The self-terminating encoding estimate, .07 seconds, indicated that relatively little of the time spent on an item was spent on encoding in the average group.

Assessment of Equivalence of R^2 Values

Once the preferred model for each ability group was designated for a particular criterion measure, the values of R^2 for the preferred model were compared across ability groups following a procedure outlined by Hakstian (1978). The procedure tested the significance of the differences in values of R^2 for the three ability groups.

$$\text{The test statistic } M = \sum_{k=1}^{k} W_k(\hat{\theta}_k - \hat{\theta}_0)^2$$

was distributed as a χ^2 with $k - 1$ degrees of freedom, where k = number of groups.

$\hat{\theta}_k$ = arctanh R^*, where R^* was an unbiased estimate of the population R, calculated by

$$R^{*2} = 1 - [(n-1)/(n-p)](1 - R^2),$$

where p = number of predictor variables, and n = sample size.

$$W_k = \frac{1}{\mathrm{var}(\hat{\theta}_k)} \text{ where } \mathrm{var}(\hat{\theta}_k) = \frac{2(n-1)}{(n-3)(2n-3p)}, \text{ and}$$

$$\hat{\theta}_0 = \frac{\sum_{k=1}^{k} W_k\hat{\theta}_k}{\sum_{k=1}^{k} W_k}$$

Hakstian's procedure was developed for large sample sizes. Since the sample size in this study was small ($n = 20$), the results from this analysis were interpreted with caution. A second limitation was related to the choice of model preference in each group. The preferred models, established in the preceding section were: Model 1M marginally preferred over 4M in the low group, and Model 4M in the average and high groups. Model 4M was used as the preferred model for the low group in this analysis to maintain comparability across groups.

The values of R^2 for Model 4M on criterion variable 1 were: low ability $R^2 = .34$, average ability $R^2 = .49$, and high ability $R^2 = .89$. The overall test statistic was significant ($M = 10.78$, $p < .01$).

Simultaneous 95% confidence intervals were calculated about the simple pairwise contrasts according to the procedure outlined by Marascuilo (1966) where, if

$$\hat{\theta}_{k_1} - \hat{\theta}_{k_2} \pm \sqrt{.95\ \chi^2_{k-1}}\ \sqrt{\text{var}(\hat{\theta}_{k_1}) + \text{var}(\hat{\theta}_{k_2})}$$

spanned zero, the contrast was not significant.

The proportion of variance in latency correct scores accounted for by the preferred model (4M) increased as ability increased. The increase in R^2 values between the low and high groups and between the average and high groups was significant, but the increase between the low and average ability groups was not significant.

Component Latencies for Criterion Variable 1

The total amount of time spent by each ability group on a component across booklet types was estimated by multiplying the raw regression coefficient for that component in the preferred model, by the mean of its predictor variable as shown in Table 6, Chapter IV. This mean represented the average number of times a component was executed over the 24 booklets. These values for the three ability groups on criterion variable 1 are summarized in Table 12 and shown in Figure 3. As expected, the composite values in Table 12 correspond with the means of criterion variable 1, reported in Table 7. In order to permit comparisons between the groups, Model 4M was again substituted for Model 1M as the preferred model in the low ability group. The values for Model 1M were: inference-application, 4.06 seconds per item, and encoding-response, 10.46 seconds per item for a composite of 14.52 seconds per item.

Table 12

Composite and Component Latencies for Correct Responses

	Ability Group		
	Low	Average	High
Encoding	1.27	.14	2.83
Inference-Application	2.91	3.22	1.03
Response	10.33	7.16	3.86
Composite	14.51	10.52	7.72

Note: The units represent seconds per analogy item.

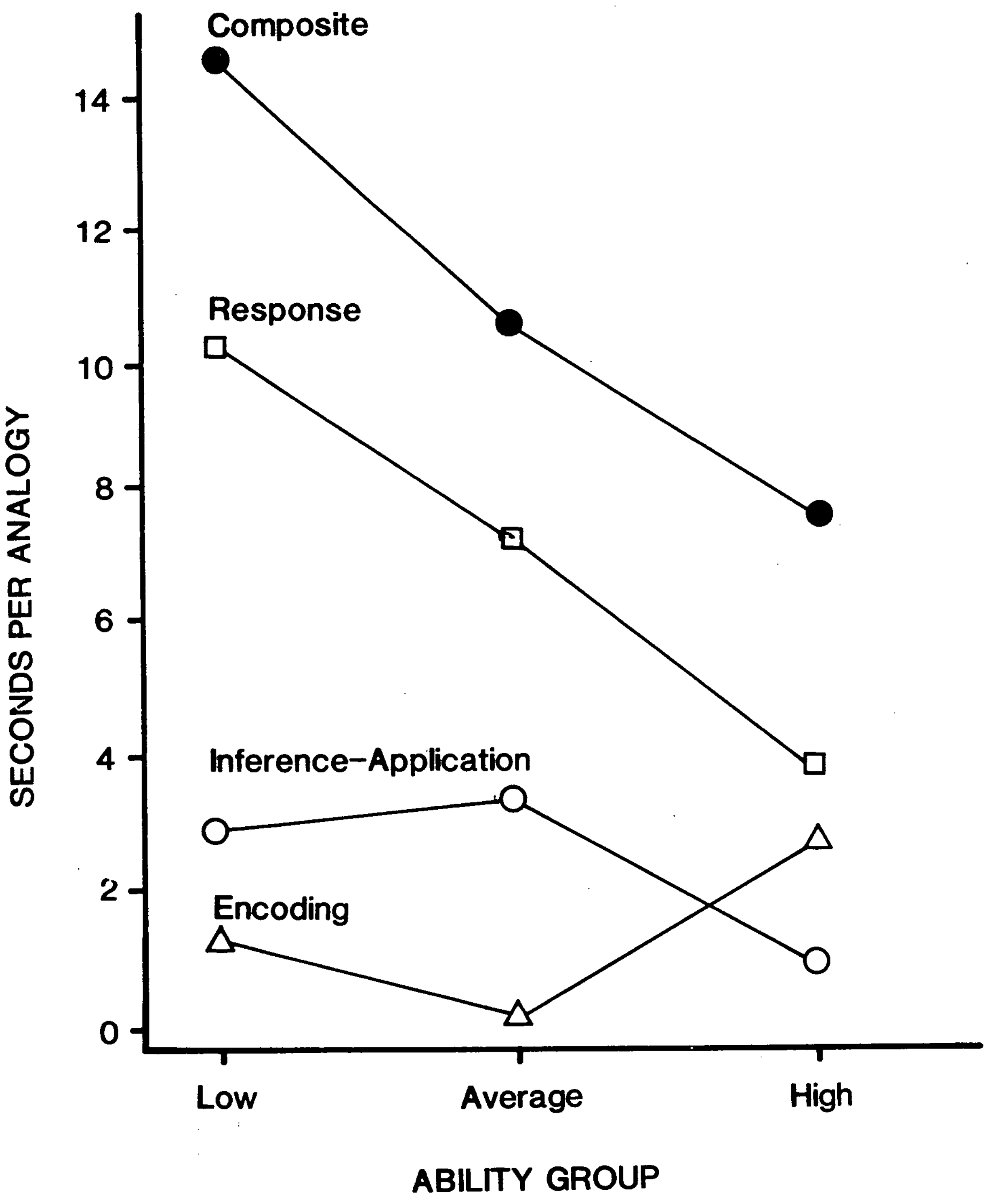


Figure 3. Composite and component latencies for correct responses on Model 4M.

As shown in Figure 3, overall latency per item decreased monotonically as ability increased, as did response latency. The response component took most time at all ability levels. Encoding decreased from low to average ability groups, while inference-application estimates were similar for these two groups. The high group spent more time encoding than either the average or low group and less time on inference-application than the average and low groups.

Criterion Variable 2: Latency Completed

Preliminary inspection of the results for the low, average, and high ability students on criterion variable 2 revealed that the results were difficult to interpret in terms of the criteria used in preceding analyses, possibly due to the low variance in these data. The results for these groups on criterion variable 2 are presented in Appendix C, and will not be discussed further.

Residual Analysis

An analysis of the residuals was conducted to determine whether the variance unexplained by the preferred models contained a systematic as well as a random component. The preferred model was 4M for the average and high groups on criterion variable 1. In view of the ambiguous results in the low group on criterion variable 1, the residuals of both Model 1M and 4M were analyzed in this group.

The criterion variables for the regression analyses were calculated, as shown in Table 5, by computing the mean booklet scores for the 20 students within each ability level. For the residual analysis, the sample of 20 at each ability level was arbitrarily divided into two half-samples of 10 students each. Then the criterion variables were recalculated for the half samples in each group by computing mean booklet scores for 10 students within each ability level. This yielded two sets of criterion variable 1 scores at each ability level. The residual analysis was not conducted for criterion variables 2 and 3.

Once the criterion variables were calculated for the half samples, the preferred models at each ability level were refitted to the data of the half samples. These regressions were calculated in order to derive two sets of residual booklet scores (observed from predicted booklet scores) within each ability level. Then these residuals for each pair of half samples (within ability level) were correlated with each other. The residual booklet scores from the half samples should be uncorrelated if the preferred model is the 'complete' model, that is, if there are no components being used by the students which were unspecified in the models. If the preferred model is not a complete model, that is, some systematic factor unspecified by the model contributes to variance in the data, then the residuals will be significantly correlated because the systematic portions of residual variance in the half samples will be correlated.

Correlations for all ability groups on criterion variable 1 are presented in Table 13. The only significant correlations were for the low ability group ($r_{1M} = .50$, $p < .05$, $r_{4M} = .46$, $p < .05$). This suggested that one or more additional systematic factors were contributing to the variance of latency correct data for the low ability students, and these factors were not tapped by the preferred model 1M, nor the competitive model 4M.

In summary, the preferred models were indistinguishable from the 'complete' models of performance in all cases, except for the latency correct data in the low group. These results are subject to the

Table 13

Reliabilities of Residuals from Preferred Models

Group	Preferred Model	Criterion Variable 1
Low	1M	r = .50*
	4M	r = .46*
Average	4M	r = .17
High	4M	r = -.13

* = $p < .05$

limitation that determination of model preference was equivocal in the average and low groups on criterion variable 1.

Dependent Variable 1

Finally, the quantitative performance of subjects on latency correct data was evaluated. In Chapter IV, Table 5, mean booklet scores were derived by summing subjects' scores within ability level and dividing by 20. This was done for each booklet on each criterion variable to yield three sets of 24 criterion scores at each ability level. In contrast, for the present analysis, subjects' scores were the focus, not booklet scores. Mean latency correct scores for each subject were calculated by summing the dependent variable 1 scores on the 24 analogy booklets and dividing the sum by 24. Thus each of the 60 subjects had a mean latency correct score.

Descriptive statistics for the scores are presented in Table 14. A one-way analysis of variance indicated that latency correct scores differed significantly between the ability groups ($F(2,57) = 9.88$, $p < .01$). Simple pairwise comparisons indicated students in the high and average ability groups had significantly shorter latencies than the low ability students ($p < .05$). The high ability group also had shorter average latency correct scores than did the average group, but this difference was not statistically significant.

Table 14

Descriptive Statistics: Mean Latency Correct

Group	n	Mean	SD
Low	20	14.50	6.91
Average	20	10.51	3.90
High	20	7.71	2.70
Total	60	10.91	5.54

CHAPTER VI

DISCUSSION

The purpose of this study was to investigate the relationship between analogical reasoning, as measured by Schematic Picture Analogies, and ability, as measured by a standardized achievement test, in fourth grade boys and girls. The aim was to identify sources of individual differences in ability in terms of both qualitative and quantitative performance differences on an analogy task. Sternberg's componential analysis (1977b) was chosen as the methodological and theoretical paradigm for the study.

Performance of fourth grade students of low, average, and high ability on schematic picture analogies was analyzed for ability-related differences at five levels:

1. the components (processes) used in solution of pictorial analogies,
2. the rule for combination of these components,
3. the mode of component execution,
4. consistency in strategy (model) use in solution, and
5. quantitative differences in component scores.

The discussion focuses on the two main issues in the study: first, the underlying reasons for differences in achievement, and second, the utility of componential analysis in individual difference research. Prior to discussion of these issues, the limitations of the study are outlined.

Limitations of the Study

The study was designed as a modified version of Sternberg and Rifkin's (1979) developmental study, the main difference being subjects varied in ability rather than age. Unfortunately, the procedure adopted by Sternberg and Rifkin (1979) was unsuccessful with the fourth grade students in the present study. This failure was attributed primarily to sample differences. Sternberg's subjects were from a high SES Hebrew day school, whereas this sample was drawn from an area with a range of economic and ethnic backgrounds. In addition to these sample differences, the answer format for the analogies was different in the current study, which increased the complexity of the analogy task in the initial introduction phase. Thus, the procedure for data collection described in Sternberg and Rifkin (1979) was modified for use in this study. Specifically, the introductory phase of the task was augmented.

These changes had two effects relevant to the discussion. First, due to the apparent sample differences and the procedural differences between the present study and Sternberg and Rifkin's (1979) experiment, cross-study comparisons were severely limited. A second effect of the procedural differences was the risk of influencing the subjects' choice of model or strategy in task solution through augmentation of training on the task. Original group differences or similarities in model preference may have been affected by the extra practice and feedback given in the present study.

A possible confounding variable in the study was response style. The task was timed. Although accuracy and not speed was stressed in the instructions, there was no direct control for individual differences in response style (speed versus accuracy). A nontrivial number of students completed all 16 items in some of the booklets, and these students may have been those with an impulsive response style. These students, while completing more items might have higher error rates than a more reflective, analytic respondent.

The reliability of the data may have been weakened since students who had worse than chance performance were not discarded. Some of the students with worse than chance performance may have been guessing if they misinterpreted the goal of the task as completion of as many items as possible in the allotted time. Since it was not possible to distinguish between a 'guesser' and a student with low scores for more substantive reasons, none of the low scorers were discarded, which may have weakened the reliability of the data particularly in the low ability group.

While the main goal of this study was to explore the underlying nature of differences in achievement ability, the conclusions were limited by the fact that only one aspect of achievement, general reasoning ability, was involved. While analogical reasoning ability may be an important factor in determining scores on standardized achievement tests, it is still only one of many contributing factors, including motivation, SES, ethnic background, and various personality factors. Thus this study did not attempt to explain the whole range of related factors, but focused on one factor. Furthermore, any relationships identified between analogical reasoning performance and achievement ability were correlational and not causal.

Finally, whenever one samples from the upper and lower ranges of the ability spectrum, one may encounter floor or ceiling effects in task performance. The preliminary analysis of criterion variable 3 indicated there was a lack of variance in error rates across the 24 booklets in all three ability groups. Similarly the lack of interpretability in criterion variable 2 model fits for the three groups was attributed to a lack of variance.

Inspection of the data suggested that the lack of variance in the high group was due to a ceiling effect, as scores were generally high in accuracy and students completed all 16 items in a number of booklets. Since the pictorial analogy task was used successfully with sixth grade students and adults by Sternberg and Rifkin (1979), it is possible that the ceiling effect observed in the present data for the high ability group may be linked to the augmentation of training and instruction.

The lack of variance on the error rate criterion in the low ability group suggested a floor effect. The variation in the complexity of the booklets, which was insufficient to affect the performance of the high ability students, appeared to be too great to affect the performance of the low ability students. The ratio of number of items incorrect to number of items completed was even and high across the 24 booklets in this group.

As for the average group, the variance was low but did not seem to represent a floor or ceiling effect. The students seemed to complete a similar number of items across booklets and the ratio of incorrect to completed items was also similar across types of analogies. Students did not have highly accurate or highly inaccurate performance as in the high and low groups, but the variation in complexity of the task across the booklets was not reflected in the variance in booklet scores on criterion variable 2 and 3.

Criterion variable 1 was a measure of the quality, not quantity, of subjects' performance and more variation across the booklets was obtained, although the variance in the high group was still relatively low.

Thus the results of the analyses were interpreted with these limitations in mind.

Analogical Reasoning and Achievement Ability

The results of the model fitting permit the isolation of achievement related differences in analogy solution at three levels, the theory, model, and component levels.

The first research question, at the theory level, asked whether students of high, average, and low ability used the five components hypothesized by the componential theory of analogical reasoning in solving the analogy tasks. If one assumes that the components specified in the preferred model are the components used by the group, then according to this criterion, the high, average, and low ability students in this study did not differ in the components they used to solve schematic picture analogies. Model 1M, the marginally preferred model in the low ability group, hypothesized inference, application, encoding, and response components as did Model 4M, the preferred model in the average and high ability groups. Mapping was not used by these subjects. Sternberg and Rifkin (1979) also reported that subjects of different ages did not differ in the components used, which were: encoding, inference, application, and response. None of the subjects used mapping. Thus, mapping did not appear to be necessary in solution of schematic picture analogies in either study, and students who varied in ability and age all used encoding, inference, application, and response components in solution, according to this criterion.

However, if one applies a more stringent criterion to this question, namely that a component estimate must be significant if the component is assumed to be employed in solution, then the evidence is weakened. Only in the high ability group were all the component estimates significant. According to this criterion, subjects differing in ability used different components in the solution of schematic picture analogies. Because the validity of individual component estimates were relative to the degree of model fit and model fits were less than perfect in the current study, the first criterion was adopted. Mapping was ruled out as a component necessary to the solution of schematic picture analogies for students at the ability levels included in this study. Mapping is considered to be necessary in most analogy tasks, but not with separable attribute stimuli (Sternberg, 1980), and these results lend further support to that conclusion.

According to the first criterion, students of differing ability levels in this study did not differ in the components they used in solving the analogies. These components were: encoding, inference, application, and response.

The second research question, also at the theory level, concerned the appropriateness of the linear additive rule for combination of components. The criterion used to evaluate this question was the amount of variance accounted for by the preferred models. If the values of R^2 are high, then the linear combination rule is supported. The only R^2 value high enough to warrant support of the linear additive rule was in the high ability group. The low R^2 values in the average group and low group could be attributed to: violation of the linear assumption, lack of variance in the data, or missing components in the models.

Since the variance in the average and low groups was higher than that in the high group, but the values of R^2 were lower, the lack of variance explanation was ruled out for these two groups. The fact that the residual analysis of low ability group data indicated a systematic portion of variance in the residuals unaccounted for by both Model 1M and 4M suggested that the low values of R^2 in the low group could be due to the inappropriateness of the hypothesized models for this group's performance. One factor which may have contributed to systematic variance in the residuals was speed, or response style. Data inspection revealed that many students in the low group who had very few items correct had many items completed. The latency correct criterion variable is a measure of quality, not quantity of performance and the model itself does not hypothesize a speed component, thus if speed was relevant in this group's data, it would explain the inappropriateness of the hypothesized models, and thus the poor fit. In addition to these explanations, the low R^2 could have been due to violation of the linear additive assumption.

The poor model fits in the average group were not attributed to a missing component since the correlation of residuals was not significant for this group. The lack of variance explanation was ruled out. It appears more likely that the low R^2 in this group was due to the violation of the linear assumption, or to measurement error.

A related issue is that of the negative regression coefficients in certain models. While the small negative weights were attributed to measurement error and dismissed as nonsignificant, larger and significant negative weights, or consistent patterns of negative weights were less readily dismissed. These negative weights may be a further indication of violation of the linear, additive assumption. Application in Model 2-3M for low ability group latency correct data was large and negative. This may be an indication that components were executed in a holistic rather than a serial fashion, which would invalidate the independent estimates.

One possible holistic strategy involves working backwards through the analogy. When the two answer options share very few attributes and are obviously discrepant, the student could work backwards, encoding the attribute which distinguished two answer options and then matching that attribute relation in the A to B terms from C to D_1 or D_2. Thus only the relevant distinguishing elements would be encoded and applied.

In summary, the linear additive component combination rule was supported by the data in the high group. The appropriate rule for component combination in the average and low groups was not clearly established and may have been linear or nonlinear. These results contrast with those in previous componential studies (Sternberg, 1977b; Sternberg & Rifkin, 1979) in which the linear additive combination rule was unequivocally supported.

The third question concerned the model preference of the three ability groups, or specifically, the preferred mode of component execution. Sternberg and Rifkin (1979) found that students of different ages did not differ in model preference on the schematic picture analogies although they did differ in model choice on other types of analogies. All of the subjects preferred model 4M.

While determination of model preference was difficult in the low and average ability groups in the present study, of the seven hypothesized models, 4M was clearly the preferred model for the high group. The preferred model in the low group was 1M but only marginally, and in the average group, 4M. In view of the difficulty of the decision of model preference in the low and average groups, as well as the question of the appropriateness of the linear assumption on which the model preferences were based, the nature of the preferred models was investigated. The only model to account for large portions of the variance was Model 4M in the high ability group. Furthermore, the residual analysis indicated that the best fitting models in the low group, Model 1M and 4M, were incomplete in that some systematic variance remained in the residuals, unaccounted for by the model. In addition, the linear additive assumption may have been violated in the low and average groups. Thus, it was concluded that while the model preferences indicated that low ability students may prefer exhaustive component execution, and the average and high groups preferred a self-terminating execution mode, these observations were confounded by the limitations previously mentioned.

Sternberg and Rifkin (1979) found no age-related differences in model preference on schematic picture analogies, but differences were found on the people piece analogies. The tendency on the people piece analogies was towards increasing use of exhaustive processing with increasing age, which is a general characteristic of cognitive development according to Brown and DeLoache (1978). It may be that students actually differed in the mode of component execution in the present study, but this result may also have been an artifact of the low variance or the inappropriatness of the linear additive assumption.

Hypothesis 4.1 stated that consistency in model choice would increase as a function of ability. While the values of R^2 were significantly higher in the high group than in the average and low groups, the increases were not necessarily due to an increase in consistency with ability. Other factors, discussed in relation to question 2, which may have influenced the increase include:

1. violation of the linear additive assumption in the low and average groups, and
2. a missing component in the low ability group (response speed).

In view of these confounding factors it was concluded that while the differences in R^2's suggested that the consistency with which subjects used the preferred strategy increased as a function of ability, other factors may have inflated or deflated the R^2's, and thus no conclusions as to the consistency of strategy use across ability levels were made. Sternberg and Rifkin (1979) observed that while R^2 values for the age groups were similar, when the preferred models were fitted to the individual subject's latencies, the values of R^2 increased with age, which he interpreted as a sign of increase in the consistency of strategy use.

The final hypothesis, 5.1, stated that latency and error rate scores would vary as a function of ability. The error rate and latency completed data were not analyzed, but the three groups were found to differ in the latency for number of items completed correctly. The fact that the low and average ability students had fewer correct items than the high group may be because they adopted an inappropriate strategy of responding impulsively in order to finish all of the items in their booklet. Accuracy of response may have been of lower priority to them.

Thus quantitative differences in performance were found between the high ability group and the low and average group. The high group was more accurate than the average and low groups.

As for group differences in individual component scores, in view of the low R^2's, and violation of assumptions, these component estimates may be invalid and at best must be interpreted with caution. In the present study, composite latency and response latency decreased as ability increased. Sternberg and Rifkin (1979) found composite and response latencies also decreased as age increased.

The three groups spent more time on response than on the other components. Response was always estimated as the regression constant; future studies should attempt to specify component processes subsumed under the response component, as it seems unlikely that the act of recording the response would take as much time as it did. It is possible that the latencies of more than one process were estimated as a single component latency.

In summary, students of different achievement ability had quantitative performance differences on an analogical reasoning task and the data suggested evidence of qualitative performance differences in the components used, in the way the components were combined and executed, and in the consistency of these executions. This contrasted with Sternberg and Rifkin's (1979) results wherein students of different ages differed in quantitative but not qualitative performance on the analogy task. Their results may have been sample specific, as students of different ages differed qualitatively on another analogy task reported in the same study (Sternberg & Rifkin, 1979).

In terms of the underlying nature of achievement, the evidence presented did not contradict the hypothesis that students of different achievement levels may use different processes, and different rules or strategies for combining and executing processes, however it was not possible to show conclusive support for the hypothesis. The tentative evidence for qualitative differences was similar to other research which indicated that individual differences were not solely of a quantitative nature. Jarman and Das (1977) also suggested that subjects of different intelligence levels may use different modes of processing on the same task. Brown (1974, 1975) implicated strategy differences as an important source of individual differences in performance. Hunt, Frost, and Lunneborg (1973) found that subjects of different verbal and performance ability also had qualitative and quantitative performance differences on a task. Other supportive evidence included that of Humphreys and Taber, 1973, and Stevenson et al., 1976, who found that individual differences in subjects' ability were reflected in different factor loadings for the same set of tasks.

The inconclusive results of this study restrict the discussion of whether or not students of different ability use different processes and strategies on tasks, but the question remains an important one for future investigations.

Componential Analysis

The strength in the method of componential analysis lies in its ability to directly test hypothesized models of task performance, as well as provide measures of processes used in the solution of the tasks. The method is suited to some tasks such as analogies, but would not be applicable to all tasks (e.g., figure completion).

Sternberg originally developed the method for use in an individual testing situation (1977b), and only later devised the group administered procedure adapted for use in the present study (Sternberg & Rifkin, 1979). Evaluation of the difficulties encountered in the present study suggested that the individual data collection method was preferable to the group administration. Fewer component estimates are confounded using the pre-cueing method (individually administered), and it seems likely that the reliability of the data would be maximized in an individual situation. In group testing situations, measurement error is likely more prevalent. While it may be argued that group data collection is a common element of educational research, and that the logistics of individual testing are prohibitive, the sample sizes required for componential analyses are small enough to permit individual data collection.

Thus, the componential approach to individual differences research (Sternberg, 1977b) was implemented with some difficulty in this study. It was concluded that while the componential analysis technique is well-suited to isolating different sources of individual differences on a task, it is more amenable to some tasks than to others, and it is more manageable in an individual, rather than a group testing situation.

Implications for Future Research

Suggestions for future research in individual differences developed from this investigation.

The suggestion that students of different ability levels may differ in their strategies for component combination and execution should be pursued. This would involve the development and testing of different models to explain individual differences in analogical reasoning. These models should attempt to identify components at a more specific level than the current models, and also test nonlinear rules for component combination. The task performances should be related back to reference ability scores in order to establish external validity.

A second concern is that the current study did not attempt to explain all of the variance in achievement scores. Future investigations should attempt to control for other aspects of achievement ability such as response style and socio-demographic factors.

Given the evidence for strategy differences, the question of remediation and training of these strategies is important. It is necessary not only to determine why children are different, but also to find ways of remediating these differences. The question of the stability of deficits over time is also of interest. A design incorporating both age and ability differences is needed.

Reference Note

1. Hakstian, R. A. Inferential procedures with multiple correlations. Paper presented at the Annual Meeting of The Society of Multivariate Experimental Psychology, Cincinnati, Ohio, November 1978.

REFERENCES

Belmont, J. M., & Butterfield, E. C. Learning strategies as determinants of memory deficiencies. Cognitive Psychology, 1971, 2, 411-420.

Bloom, B. S. Stability and change in human characteristics. New York: John Wiley & Sons, Inc., 1964.

Boring, E. G. Intelligence as the tests test it. In J. J. Jenkins & D. J. Paterson (Eds.), Studies in individual differences: The search for intelligence. New York: Appleton Century Crofts, 1961. (Reprinted from The New Republic, 1923, June 6.)

Brown, A. L. The role of strategic behavior in retardate memory. In N. R. Ellis (Ed.), International review of research in mental retardation (Vol. 7). New York: Academic Press, 1974.

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

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

Brown, A. L., & Campione, J. C. Training strategic study time apportionment in educable retarded children. Intelligence, 1977, 1, 94-107.

Brown, A. L., Campione, J. C., Bray, N. W., & Wilcox, B. L. Keeping track of changing variables: Effects of rehearsal training and rehearsal prevention in normal and retarded adolescents. Journal of Experimental Psychology, 1973, 101, 123-131.

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

Brown, A. L., & Lawton, S. C. The feeling of knowing experience in educable retarded children. Developmental Psychology, 1977, 13(4), 364-370.

Campione, J. C., & Brown, A. L. Toward a theory of intelligence: Contributions from research with retarded children. Intelligence, 1978, 2, 279-304.

Carroll, J. B. Review of The nature of human intelligence, by J. P. Guilford. American Educational Research Journal, 1968, 5, 249-256.

Carroll, J. B. Psychometric tests as cognitive tasks: A new "Structure of Intellect" (Technical Report No. 4). Princeton, New Jersey: Educational Testing Service, 1974.

Carroll, J. B. How shall we study individual differences in cognitive abilities?--Methodological and theoretical perspectives. Intelligence, 1978, 2, 87-115.

Carroll, J. B., & Maxwell, S. E. Individual differences in cognitive abilities. Annual Review of Psychology, 1979, 30, 603-640.

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

Estes, W. K. Learning theory and intelligence. American Psychologist, 1974, 29, 740-49.

Estes, W. K. The state of the field: General problems and issues of theory and metatheory. In W. K. Estes (Ed.), Handbook of learning and cognitive processes (Vol. 1). Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

Estes, W. K. Intelligence and cognitive psychology. In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1976.

Flavell, J. H. Developmental studies of mediated memory. In H. W. Reese & L. P. Lippsitt (Eds.), Advances in child development and behavior (Vol. 5). New York: Academic Press, 1970.

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

Humphreys, L. G. The nature and organization of human abilities. In M. Katz (Ed.), The 19th yearbook of the National Council on Measurement in Education. Ames, Iowa: National Council on Measurement in Education, 1962. (a)

Humphreys, L. G. The organization of human abilities. American Psychologist, 1962, 17, 475-483. (b)

Humphreys, L. G., & Taber, T. Ability factors as a function of advantaged and disadvantaged groups. Journal of Educational Measurement, 1973, 10(2), 107-115.

Hunt, E., Frost, N., & Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In G. 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 (Vol. 1). Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

Hunt, E., & MacLeod, C. M. The sentence-verification paradigm: A case study of two conflicting approaches to individual differences. Intelligence, 1978, 2, 129-144.

Jarman, R. F. Comments on John B. Carroll's "How shall we study individual differences in cognitive abilities?--Methodological and theoretical perspectives." Intelligence, 1980, 4, 73-82.

Jarman, R. F., & Das, J. P. Simultaneous and successive syntheses and intelligence. Intelligence, 1977, 1, 151-169.

Jensen, A. R. Hierarchical theories of mental ability. In W. B. Dockrell (Ed.), On intelligence. London: Methuen & Co., Ltd., 1970.

Johnson, D. M. Serial analysis of verbal analogy problems. Journal of Educational Psychology, 1962, 53(2), 86-88.

Kail, R. Use of strategies and individual differences in children's memory. Developmental Psychology, 1979, 15(3), 251-255.

King, E. M. (Ed.). Canadian tests of basic skills: Manual for administrators, supervisors, and counsellors. Don Mills, Ontario: Thomas Nelson & Sons (Canada) Limited, 1976.

King, E. M. (Ed.). Canadian tests of basic skills: Teacher's guide. Don Mills, Ontario: Thomas Nelson & Sons (Canada) Limited, 1977.

Le, C., & Tenisci, T. UBC TRP: Triangular Regression Package. Vancouver, British Columbia: Computing Centre, University of British Columbia, 1978.

Marascuilo, L. A. Large-sample multiple comparisons. Psychological Bulletin, 1966, 65(5), 280-290.

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

Messick, S. Beyond structure: In search of functional models of psychological process. Psychometrika, 1972, 37, 357-375.

Miller, G. A., Galanter, E., & Pribram, K. H. Plans and the structure of behavior. New York: Holt Rinehart & Winston, 1960.

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

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

Reed, S. K. Review of Intelligence, information processing and analogical reasoning, by R. J. Sternberg. Intelligence, 1977, 1, 331-334.

Resnick, L. B. Introduction: Changing conceptions of intelligence. In L. B. Resnick (Ed.), The nature of intelligence. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1976. (a)

Resnick, L. B. The nature of intelligence. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1976. (b)

Rohwer, W. D., Jr. Elaboration and learning in childhood and adolescence. In H. W. Reese (Ed.), Advances in child development and behavior (Vol. 8). New York: Academic Press, 1973.

Rumelhart, D. E., & Abrahamson, A. A. A model for analogical reasoning. Cognitive Psychology, 1973, 5, 1-28.

Snow, R. E. Theory and method for research on aptitude processes. Intelligence, 1978, 2, 225-278.

Spearman, C. The nature of "intelligence" and the principles of cognition. London: MacMillan, 1923.

Spearman, C. The abilities of man. In J. J. Jenkins, & D. G. Paterson (Eds.), Studies in individual differences: The search for intelligence. New York: Appleton Century Crofts, 1961. (Reprinted from The Abilities of Man, 1927.)

Sternberg, R. J. Component processes in analogical reasoning. Psychological Review, 1977, 84(4), 353-378. (a)

Sternberg, R. J. Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977. (b)

Sternberg, R. J. Componential investigations of human intelligence. In A. Lesgold, J. Pellegrino, S. Fokkema, & R. Glaser (Eds.), Cognitive psychology and instruction. New York: Plenum, 1978. (a)

Sternberg, R. J. Intelligence research at the interface between differential and cognitive psychology: Prospects and proposals. Intelligence, 1978, 2, 195-222. (b)

Sternberg, R. J. Isolating the components of intelligence. Intelligence, 1978, 2, 117-128. (c)

Sternberg, R. J. Cognitive-behavioral approaches to the training of intelligence in the retarded (Cognitive Development Technical Report No. 6). New Haven: Department of Psychology, Yale University, 1979. (a)

Sternberg, R. J. Components of human intelligence (Technical Report No. 19). New Haven: Department of Psychology, Yale University, 1979. (b)

Sternberg, R. J. Six authors in search of a character: A play about intelligence tests in the year 2000. Intelligence, 1979, 3, 283-293. (c)

Sternberg, R. J. The construct validity of aptitude tests: An information-processing assessment (Technical Report No. 20). New Haven: Department of Psychology, Yale University, 1979. (d)

Sternberg, R. J. The development of human intelligence (Cognitive Development Technical Report No. 4). New Haven: Department of Psychology, Yale University, 1979. (e)

Sternberg, R. J. The nature of mental abilities. American Psychologist, 1979, 34(3), 214-230. (f)

Sternberg, R. J. The development of linear syllogistic reasoning. Journal of Experimental Child Psychology, 1980, 29(2), 340-356.

Sternberg, R. J., & Nigro, G. Developmental patterns in the solution of verbal analogies. Child Development, 1980, 51(1), 27-38.

Sternberg, R. J., & Rifkin, B. The development of analogical reasoning processes. Journal of Experimental Child Psychology, 1979, 27, 195-232.

Stevenson, H. W., Parker, T., Wilkinson, A., Hegion, A., & Fish, E. Longitudinal study of individual differences in cognitive development and scholastic achievement. Journal of Educational Psychology, 1976, 68, 377-400.

Tallmadge, G. K., & Wood, C. T. User's Guide, ESEA Title I evaluation and reporting system. Mountain View, California: RMC Research Corporation, 1976.

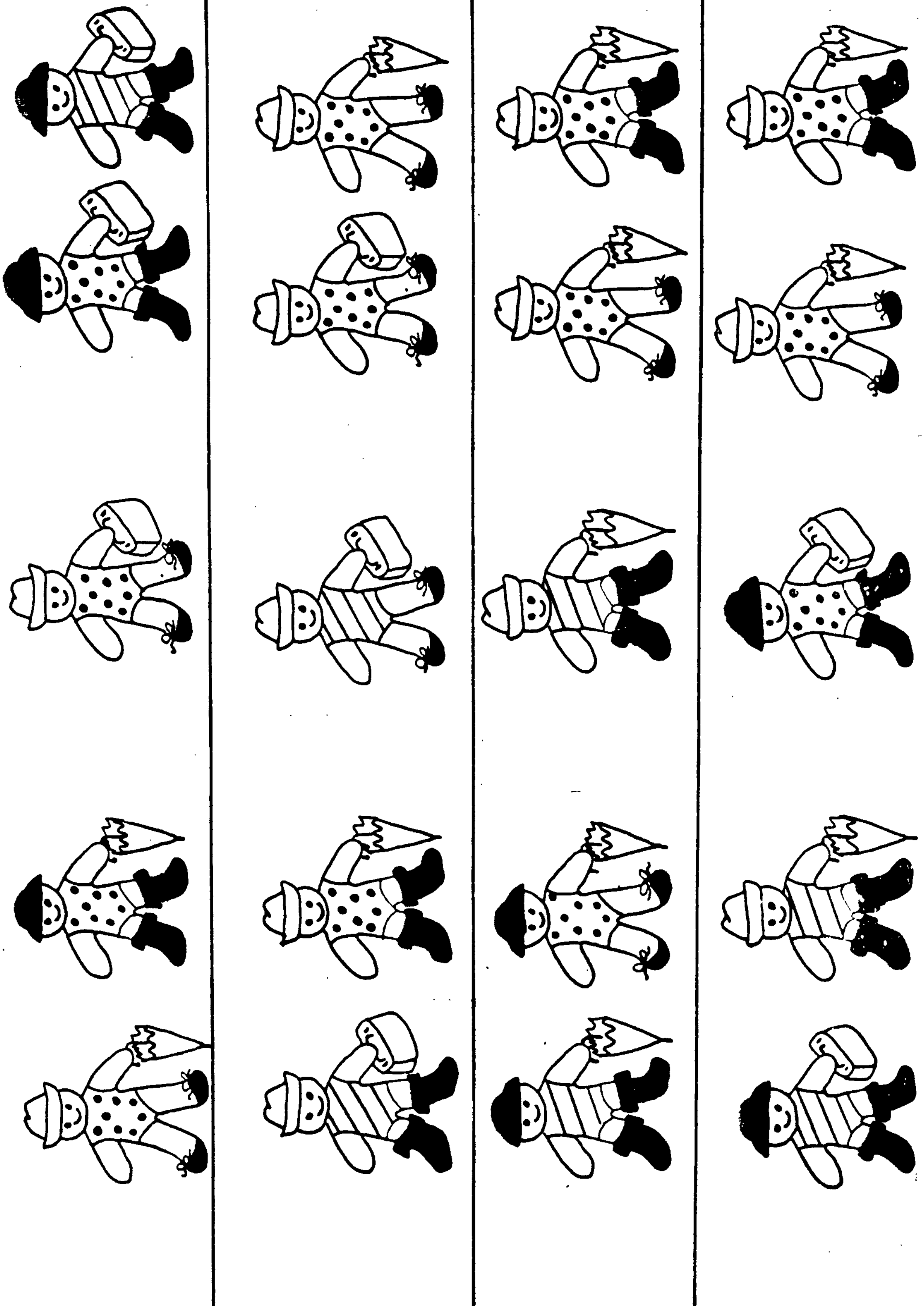
Tyler, L. E. The psychology of human differences. New York: Appleton-Century-Crofts, 1965.

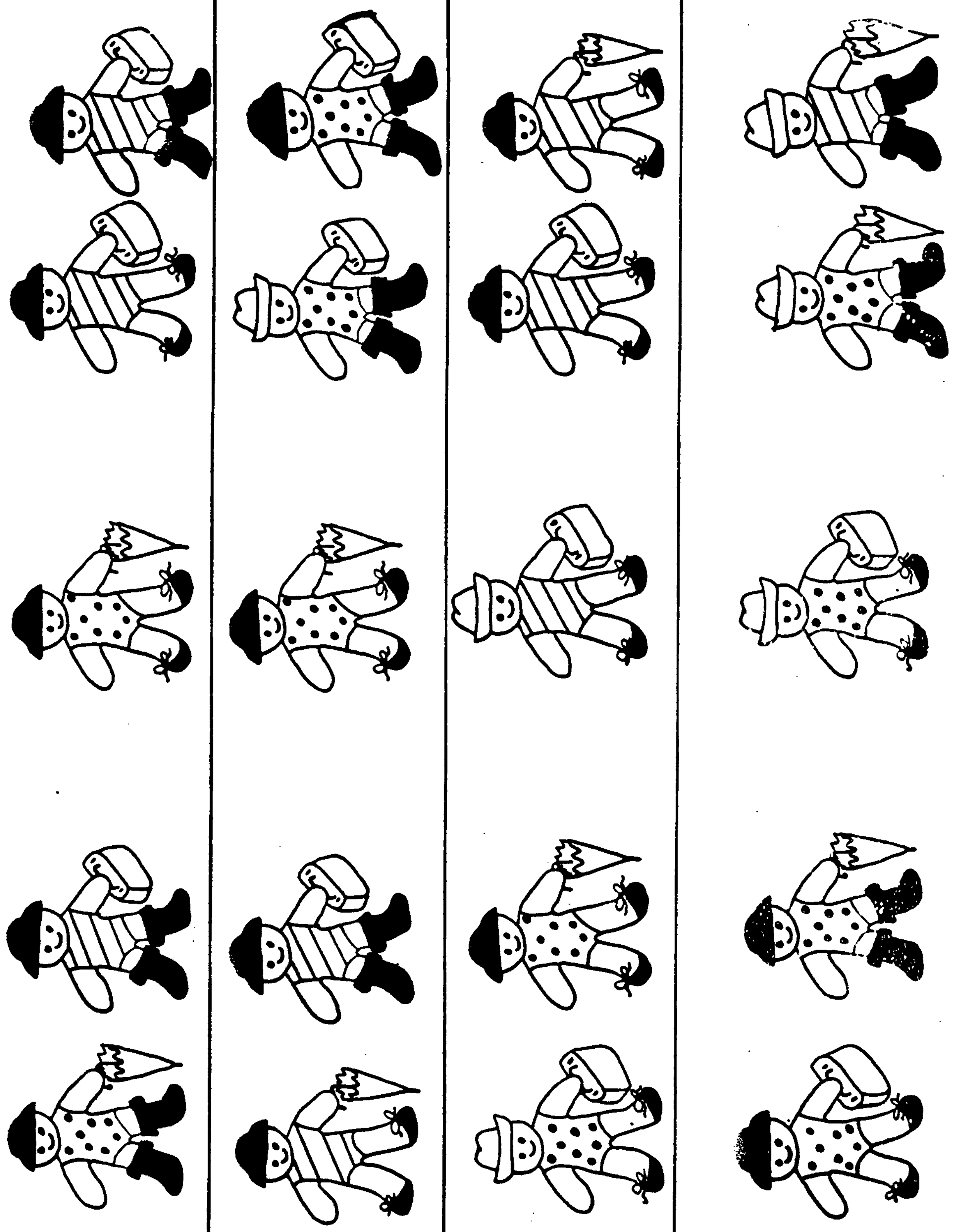
Tyler, L. E. Individual differences: Abilities and motivational directions. New York: Appleton-Century-Crofts, 1974.

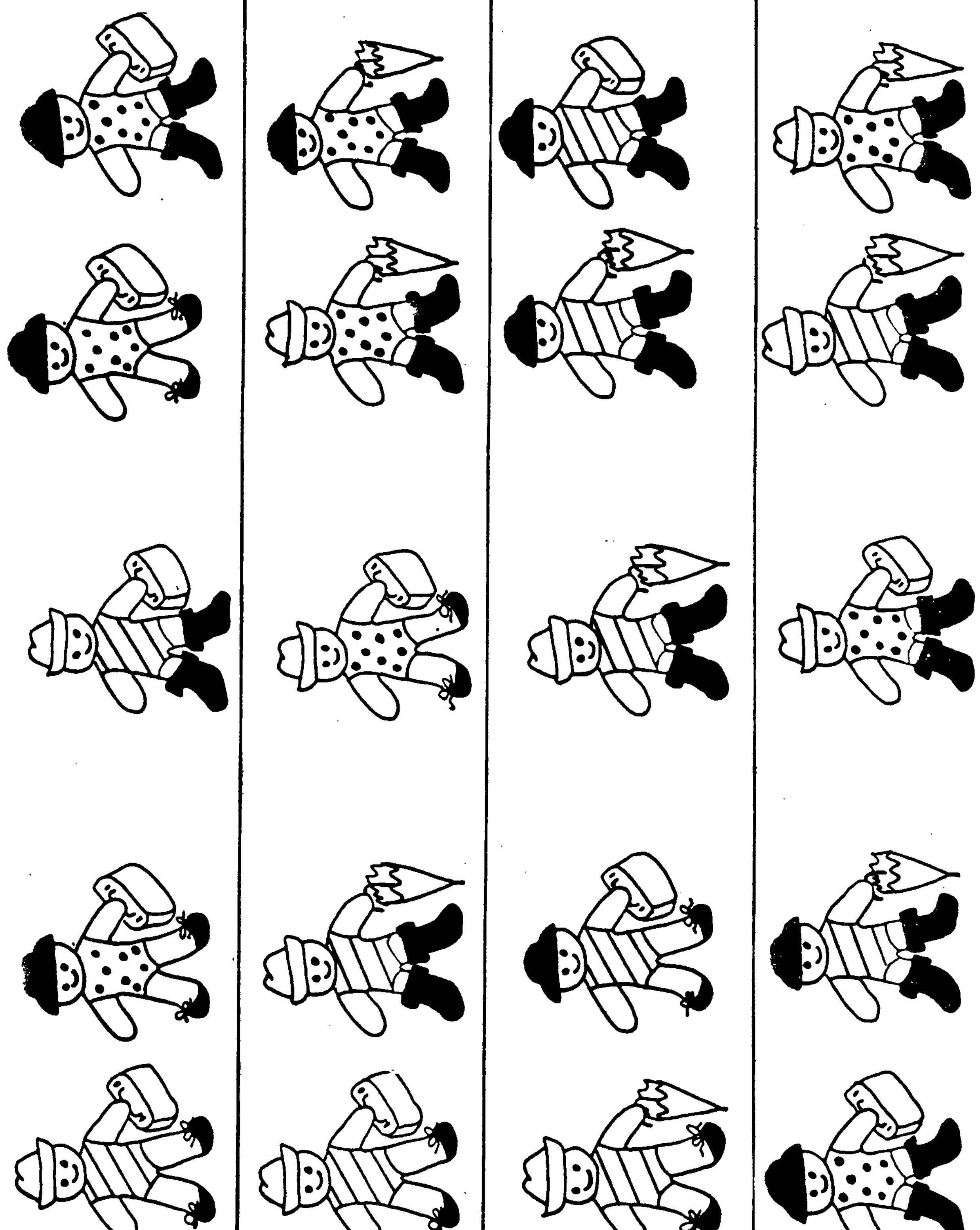
Vernon, P. E. Intelligence. In W. B. Dockrell (Ed.), On intelligence. London: Methuen & Co. Ltd., 1970.

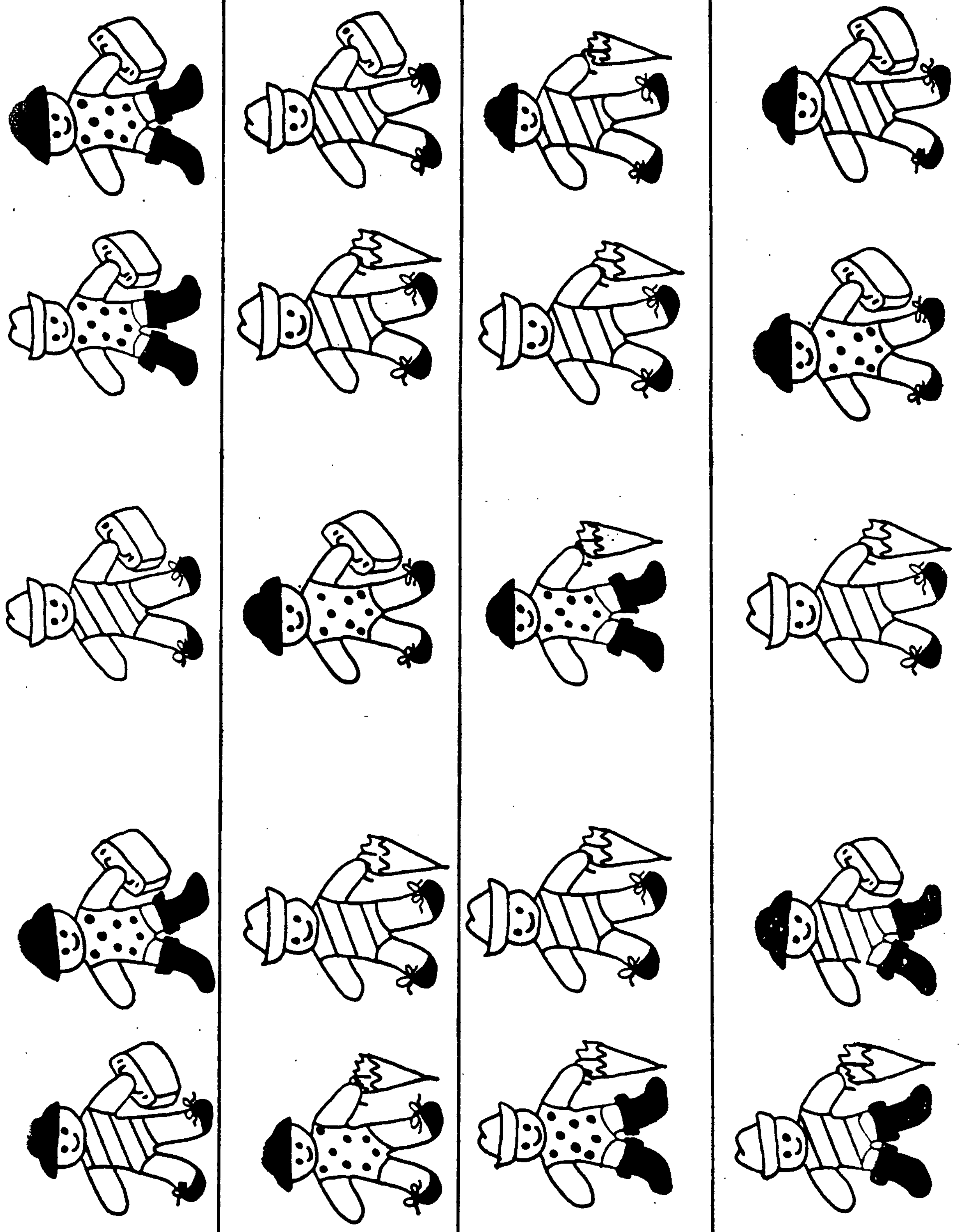
Whiteley, S. E., & Dawis, R. V. Effects of cognitive intervention on latent ability measured from analogy items. Journal of Educational Psychology, 1974, 66(5), 710-717.

Appendix A: Schematic Picture Analogy Booklet







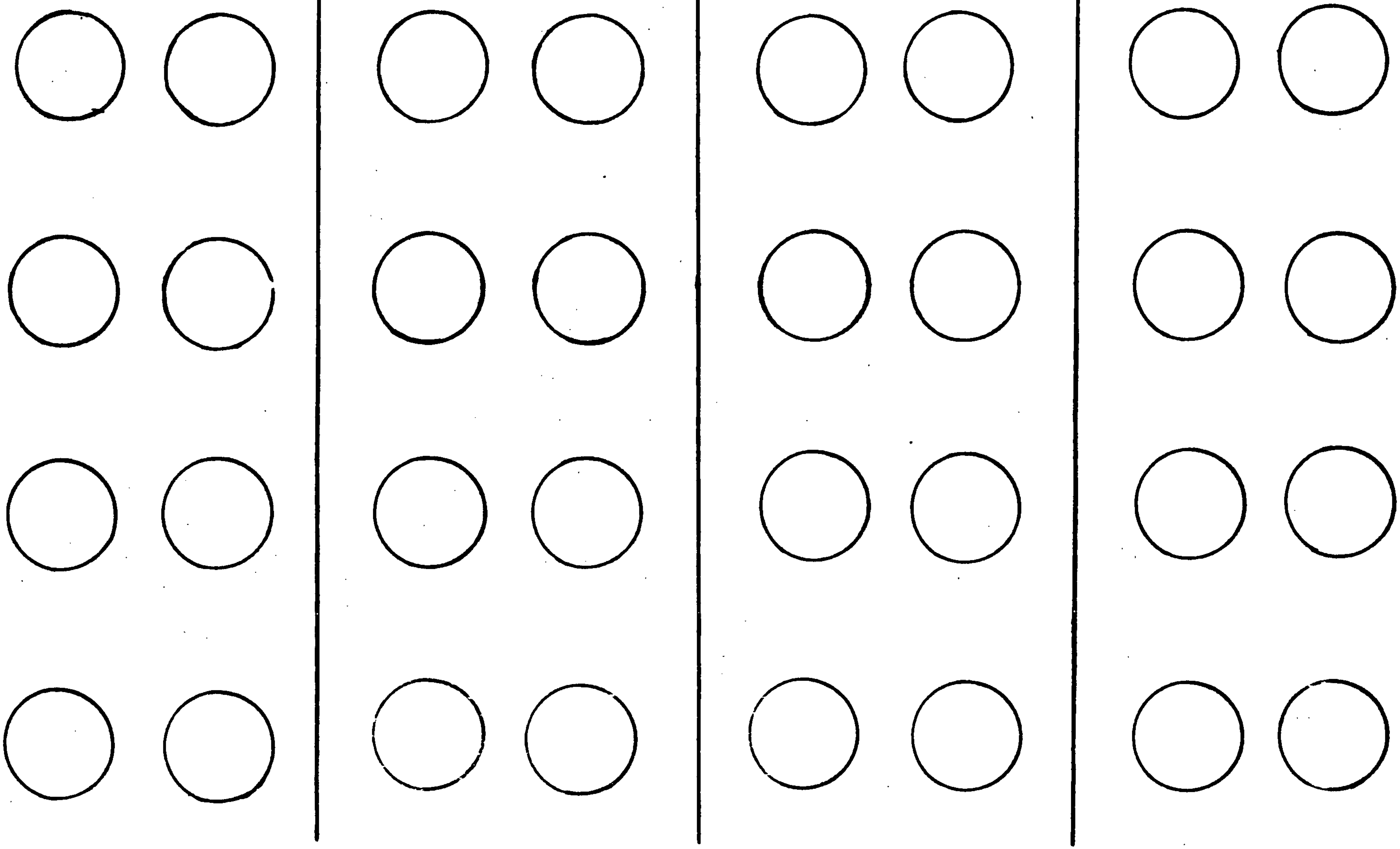


Appendix B: Answer Sheet for Schematic

Picture Analogy Booklet

Page 1 | Page 2 | Page 3 | Page 4

1.
2.
3.
4.



Appendix C: Model Fits for Ability Groups: Criterion Variable 2

Model Fits for Low Ability Group:
Criterion Variable 2

					Component Estimates					
Model	Inf.	Inf.-Appl.	Appl.	Map.	Enc.	Enc.-Resp.	Resp.	R^2	F	σ est
1		-.03		-.23		8.88**		.01	.11	1.80
2	-1.74*		4.26**	-.23		8.87**		.44	5.25**	1.39
3	-2.21**		5.14**	-1.16		9.47**		.48	6.15**	1.34
4		1.94*		.54		6.79**		.23	3.11	1.59
1M		.08				8.31**		.00	.02	1.77
2-3M	-1.65		4.31		-.04		8.35**	.43	5.02**	1.40
4M		.66			1.20*		5.60**	.38	6.33**	1.43

Note: Estimates are in terms of seconds spent on a component per analogy item.

* = p < .05
** = p < .01

Model Fits for Average Ability Group:
Criterion Variable 2

						Component Estimates				
Model	Inf.	Inf.-Appl.	Appl.	Map.	Enc.	Enc.-Resp.	Resp.	R^2	F	σ est
1		-.21		-.25		8.21**		.02	.32	1.22
2	-1.60**		3.45**	-.25		8.21**		.63	11.35**	.77
3	-1.44**		3.41**	.06		7.51**		.61	10.41**	.79
4		1.32*		1.17*		5.77**		.38	6.44**	.97
1M		-.08				7.58**		.00	.04	1.21
2-3M	-.12		.09		2.24*		4.88**	.72	17.14**	.67
4M		-.18			1.43**		4.69**	.72	27.69**	.65

Note: Estimates are in terms of seconds spent on a component per analogy item.

* = $p < .05$
** = $p < .01$

Model Fits for High Ability Group:
Criterion Variable 2

	Component Estimates									
Model	Inf.	Inf.-Appl.	Appl.	Map.	Enc.	Enc.-Resp.	Resp.	R^2	F	σ est
1		.42		-.11		6.90**		.07	.80	1.36
2	-1.36**		4.44**	-.11		6.89**		.84	35.00**	.58
3	-1.21**		4.30**	.19		6.43**		.84	35.00**	.58
4		2.55**		1.12**		4.96**		.72	27.69**	.76
1M		.48				6.63**		.07	1.67	1.33
2-3M	-.32		1.97		1.64*		4.65**	.88	48.83**	.50
4M		1.27**			1.22**		4.12**	.88	73.33**	.49

Note: Estimates are in terms of seconds spent on a component per analogy item.

* = p <<.05
** = p < .01