DOCUMENT RESUME

ED 233 040 TM 830 448

AUTHOR Carlson, Jerry S.
TITLE Applications of Dynamic Assessment to Cognitive and Perceptual Functioning of Three Ethnic Groups. Final Report.
INSTITUTION California Univ., Riverside.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Jan 83
GRANT NIE-G-81-0081
NOTE 100p.; Some tables may be marginally legible due to small print.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Anglo Americans; Blacks; *Cognitive Ability; Cognitive Measurement; Elementary Education; *Ethnic Groups; Hispanic Americans; Perception Tests; *Perceptual Development; Performance Factors; Predictive Validity; Racial Differences; *Test Bias; *Testing; Testing Problems; Test Use
IDENTIFIERS California (San Bernardino); *Dynamic Assessment; *Testing Conditions

ABSTRACT

This study assessed the usefulness of the dynamic testing approach to optimize testing procedures by reducing or eliminating bias, conceived of as error in measurement attributable to factors entering into performance which were not the target of the assessment. The study examined: (1) whether the dynamic assessment approach yields information which is a more appropriate indicator of cognitive competence than standard approaches based on static test theory; (2) why and for whom the dynamic assessment approach is effective; and (3) for what reason the procedures should be used. The sample of 147 fourth- and fifth-grade students from 15 elementary schools in the San Bernardino, California, area included 67 Anglo Americans, 37 Black Americans, and 43 Hispanic Americans randomly assigned to standard, verbalization, or elaborated feedback test conditions. Measures of cognitive and perceptual functioning, cognitive style and personality (e.g., impulsivity-reflectivity, motivation and planning, and school achievement) included Raven's Coloured Progressive Matrices, the Matching Familiar Figures Test, the Cattell Culture Fair Test of "g," Harter's Perceived Competence Scale for Children, and Trail Making and Visual Search Tasks. It was found that the testing condition which seemed to be most effective in increasing performance of all groups was verbalization; a type of compensatory effect for initially observed differences was brought about by subject verbalization. (CM)

APPLICATIONS OF DYNAMIC ASSESSMENT TO COGNITIVE AND PERCEPTUAL FUNCTIONING OF THREE ETHNIC GROUPS

Jerry S. Carlson

University of California, Riverside

Final Report for the National Institute of Education Grant

NIE-G-81-0081

January 1983

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

APPLICATIONS OF DYNAMIC ASSESSMENT TO COGNITIVE AND PERCEPTUAL FUNCTIONING OF THREE ETHNIC GROUPS

Jerry S. Carlson

University of California, Riverside

Preface

The following is a final report for National Institute of Education Grant, NIE-G-81-0081. The project was carried out during the 1981-1982 academic year with the cooperation of the San Bernardino Unified School District. Data analysis and final report writing were done during the Fall of 1982.

Without the cooperation and help of several people, this project could not have been carried out. Thanks are due to Dr. Neal Roberts, Superintendent of the San Bernardino Unified School District, the principals, teachers, and especially the children in the cooperating schools. We were always made to feel welcome in the almost 20 elementary schools in which we worked; this in spite of the responsibilities and heavy work-load which the principals and their staffs have. Having the opportunity to work with our sample of fourth and fifth grade children gave us our greatest reward. They deserve special thanks.

Excellent work in data collection and preparation was carried out by Julie Furuta, Marcia Johnson, Caroline Hearn, and C. Mark Jensen. Their help is greatly appreciated. In addition to primary responsibility for data collection, Mark Jensen proved to be an invaluable co-worker, aiding in data analysis and writing up the final report of this project.

Chapter I

The Problem

1.0 Introduction

A significant concern for our era relates to the conceptualization of what constitues appropriate assessment of cognitive and perceptual processes. The debate has focused on key terms such as "test bias" or "test fairness." Several different approaches to this issue have been taken. These range from the classical test fairness concept of equal prediction elaborated by Cleary (1968), Cleary, Humphreys, Kendrick, and Wesman (1975), and Eichorn and Bass (1971), to criteria involving group membership and cultural variables (Darlington, 1971), and to the incorporation of sociological data in arriving at performance level (Mercer & Lewis, 1975).

Recently, Flaugher (1978) has cautioned that test bias can result from many factors and influences. Focus on just one aspect may be myopic and ignore the breadth of the issues involved. He did suggest, however, that the development of "tailored testing" would perhaps provide a solution to one very important factor: bias related to the atmosphere in which the test is administered. The notion of tailored testing can be extended to the use of a learning oriented approach to assessment (Resnick, 1979). Through such an approach (referred to here as dynamic testing) analyses can be made of how putative motivational, personality, and cognitive style factors interact with assessment approaches to yield performance data. This allows for analysis of bias between and within ethnic and/or cultural groups and

clarification of individual and group differences in terms of psychological and information processing variables. In addition, dynamic assessment provides a basis for modifications in the classroom to incorporate learning characteristics of the child in the teaching strategy and/or teaching methodology used.

1.1 Bases of Dynamic Testing

An assumption often made in assessment of cognitive functioning is that test scores obtained in the traditional, standardized manner serve as adequate estimates of the cognitive competence of the person or group of persons tested. The validity of the assumption is based on the notion that performance is a veridical measure of competence. This suggests that intraindividual variations in information processing, resulting primarily from intellectual and/or personality factors, play relatively unimportant roles in test outcome. The relationship between subject- and task-independent variables is ignored (cf. Detterman, 1979).

The static test approach and the assumptions behind it has been challenged by several authors (Brown & French, 1979; Feuerstein, 1979; Guthke, 1977; Schmidt, 1971). It is argued that testing procedures must be developed and applied which adequately take intraindividual variability into account and reduce the potential discrepancy between measured cognitive functioning and cognitive competence (Bortner & Birch, 1970). This calls for a dynamic testing approach, where modifications in testing procedures can be analyzed as they affect performance levels and interact with sources of individual

variation without diminishing predictive validity. In this approach, variations in testing procedure can be built directly into the test situation itself or implemented through training outside of the actual testing. The latter requires subsequent assessment of the effects of the training on a transfer test and is thereby more difficult for the practitioner to implement than the former approach. We favor the former, incorporating the modifications in the procedures into the testing situation, for pragmatic and methodological reasons. Pragmatically, it is convenient and directly applicable for use by psychological practitioners. Methodologically, it avoids problems related to the measurement of change (Cronbach & Furby, 1970).

1.2 Testing Conditions Leading to Increased Levels of Performance

A series of studies were carried out to isolate and assess the effect of several types of commonly used variables which can be incorporated into testing procedures and be independently manipulated (Carlson & Wiedl, 1976; Carlson & Wiedl, 1979; Wiedl & Carlson, 1977).

The testing procedures isolated involve different techniques of optimizing cognitive functioning. Thus, controlled comparison between types of approaches often used, but heretofore not systematically analyzed in the dynamic testing approach, was made. The dependent measure was the Raven Coloured Progressive Matrices (CPM), a test of nonverbal intelligence (Raven, 1976). The CPM was chosen as it is considered to be a relatively "culture fair" test and a good measure of general intelligence. The subjects tested were 434 second- and fourth-grade children from the general population. The testing conditions (C) employed were:

C_1. Standard Instruction, according to the procedures outlined by Raven.

C_2. Verbalization During and After Solution, which requires the child to describe the main stimulus pattern prior to searching for the correct answer and then, after a particular alternative is chosen, to explain why he had the choice he did.

C_3. Verbalization After Solution, which involves the child describing the reasons for his choice after the choice is made.

C_4. Simple Feedback, where the child is informed only if he is correct or incorrect after his choice.

C_5. Elaborated Feedback, which involves, in addition to simple feedback, an elaboration by the experimenter of the reasons why the chosen answer was correct or incorrect. The principles involved in the task were pointed out.

C_6. Elaborated Feedback plus Verbalization During and After Solution, which is a combination of conditions two and five. That is, it involves the child's verbalization of the pattern to be completed, followed by solution and the child's explanation for the reasons for solution, and elaborated feedback by the experimenter informing the child of the correctness of his response and explaining the principles involved in the task.

In all cases the child's initial response, regardless of condition, was scored.

For the second-grade children, Conditions 2, 5, and 6 led to higher levels of performance than were found for Conditions 1, 3, and

4. The most effective conditions were 5 and 6. For the fourth-grade children, Conditions 2 and 6 led to improved performance. A trend in this direction was noted under Condition 5. Thus, three conditions for younger children and two procedures for older children were found which appeared useful for improving test scores, i.e., bringing performance closer to actual competence.

Beyond ascertaining the main effects of the testing conditions on total score CPM performance, closer analysis of the results was obtained by analyzing the interactive effects of performance with differentiations which the CPM allows in administration and scoring. The test can be administered in either the normal booklet form or in a puzzle, board-form version. In the former, covert processes must be used by the subject as he searches for solution. In the latter, overt trial-and-error procedures can be used. In scoring the CPM, three distinct subgroupings of items can be used. These were distinguished through factor analysis in a previous study (Wiedl & Carlson, 1976). the factors isolated were: simple pattern completion, pattern completion through closure, and reasoning by analogy. (See Carroll & Maxwell, 1979 for a review and Carlson & Jensen, 1980 for further analysis.)

Results of analysis of these differentiations showed that the puzzle version led to higher levels of performance for the second-grade but not the fourth-grade children. The salient item groups were simple pattern completion and pattern completion through closure. The effects of the optimizing testing conditions could be traced mainly to those items requiring abstract thinking.

On the basis of these results, two types of studies were conducted. In one type, the goal was to replicate and extend the above findings by investigating whether or not the salient testing conditions have the effect of optimizing performance for children from special populations. In the other type, in addition to extending general applicability, the goal was to assess the compensatory effect which the procedures could have in reducing differences between particular groups of children.

1.3 Children with Learning Difficulties

In studies involving learning handicapped, low IQ children (Carlson & Wiedl, 1979; Wiedl & Carlson, 1980), basically the same testing paradigm as described for the first study was used. Space limitation prevents full description of the investigations. By way of summary, however, it was demonstrated that for learning handicapped subjects testing procedures C_5 and C_6 are appropriate and the effects could be demonstrated for items requiring abstract reasoning. This result is of special significance as the requirements of tasks of logical thinking are particularly difficult for low IQ children.

The most important results described in the preceding section are summarized in Table 1 on the following page.

1.4 Personality and Cognitive Style Variables

The above presentation of the dynamic assessment approach has been concerned primarily with the validation of the dynamic assessment procedure in terms of effects on mean performance scores. Beyond this, however, it is important to investigate how and to what extent

Table 1

Summary of Salient Testing Conditions with Different Groups of Subjects

Groups of Subjects	Testing Conditions: Differentiated C_2 Verbalization		Elaborated C_5 Feedback		Verbalization C_2 + C_5 + Feedback		Board-Form			
	Item Requirements ++									
	T	p c r	T	p c r	T	p c r	T	p	c	r
Regular School										
Children 2nd grade	o	o	●	●	●	●	●	●	●	
4th grade	●	●			●	●				
Children with Learning Difficulties (avg. 4th grade)			o	o	●	●	●	●	●	

\+ o = slight
● = marked

++ T = total test score
p = pattern completion items
c = closure items
r = reasoning items

learning characteristics (personality and cognitive style) of the child interact with dynamic testing approaches.

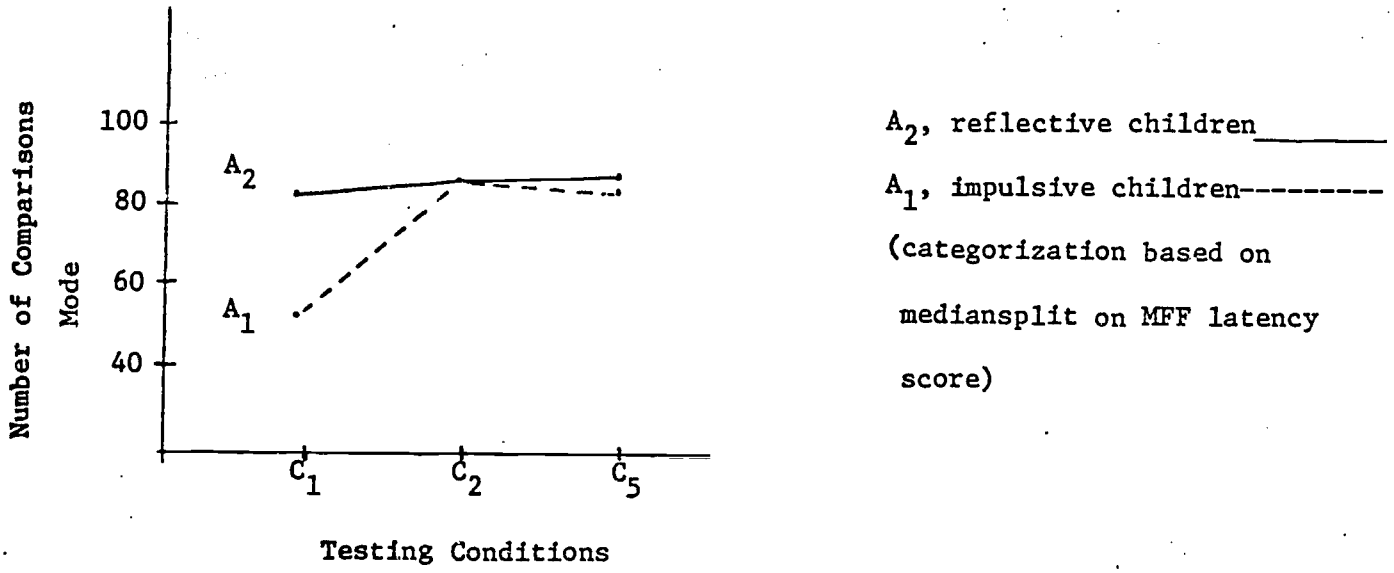
Several studies have stressed the interaction of learning and personality characteristics. Three investigations will be mentioned here that relate to one particular personality (style) characteristic: impulsivity-reflectivity. Bush and Sweck (1975) have concluded that task demands influence strategies children use who display reflective information processing while impulsive children seem to be unaware of the strategy demands of the tasks.

In a study relating the dynamic assessment approach to impulsive children, Wiedl (1979) hypothesized that the performance of impulsives could be made equal to that of reflectives by modified, more appropriate testing procedures. The subjects were 150 second-grade chidlren, average age 7.6 years. The Matching Familiar Figures test (MFF) was administered to all children to assess degree of impulsivity- reflectivity. Based on their MFF performance the subjects were divided into two groups, impulsives or relectives. The classification was obtained by median-split on the error score of the test. The CPM was administered under testing Conditions 1, 2, 5, and 6.

The results showed that impulsives scored lowest on the CPM under the standard testing condition (C_1). No differences between the impulsives and reflectives were detected under Conditions 2 and 5. That is, significant gains in performance were noted for the impulsives, but not for the reflectives when conditions of verbalization (C_2) and elaborated feedback (C_5) were employed. Under

Condition 6, no compensatory effect was detected for the impulsive children. In addition to the MFF error score classification, children were also categorized as impulsives or relectives by a median-split on latency to response. For this classification, similar compensatory effects attributable to testing condition were detected in matrices performance.

In a study designed to ascertain how differences in test performance were reduced under the verbalization (C_2) and elaborated feedback conditions (C_5), Wiedl and Bethge (1980) analyzed eye movements for impulsive and reflective children defined by MFF latency scores. Several types of eye-movement measures were made. But regardless of eye-movement measures taken, the same basic result was found: Under conditions of dynamic assessment, C_2 and C_5, the eye movement patterns of the impulsive and reflective children were about the same. Such was not the case with the Raven Matrices administered under C_1. The following exemplified the general findings.



1.5 Compensation with Regard to Ethnic/Racial Differences: Anglo-Americans, Black-American, Hispanic-American Children.

The potential compensatory effects of the dynamic testing approach in assessing cognitive capabilities of three different ethnic groups, Anglo, Black, and Mexican-American, were studied (Dillon & Carlson, 1978). Several investigations have been carried out comparing performance of these groups on tests like the Raven matrices and various Piaget tests. Most results indicate that Black children tend to perform approximately one standard deviation below Anglos, with Mexican-American performance half-way between. It was hypothesized that the testing conditions shown to be salient in the previous studies would compensate for factors which may contribute to the relatively low Black and Mexican-American performance.

The tests used were Matrices and Order of Appearance taken from a recently developed Piaget battery (Winkelmann, 1975). The testing procedures involved three conditons: Standard (C_1), Verbalization During and After Solution (C_2), and Verbalization During and After Solution plus Elaborated Feedback (C_6).

The tasks were administered to 189 children, representing three American ethnic groups: Anglo, Black, and Mexican-American. The subjects ranged from five to ten years of age. Randomization and assignment to testing conditions was carried out for three age levels: 5-6, 7-8, and 9-10 years. This resulted in a 3x3x3 randomized block design, with three levels of test administration, three age levels, and three ethnic group classifications. There were seven subjects in each cell.

Table 2

Means and Standard Deviations of Matrices for Three Age Groups, Ethnic Groups and Testing Conditions (C).

	age: 5-6			age: 7-8			age: 9-10		
	C_1	C_2	C_6	C_1	C_2	C_6	C_1	C_2	C_6
Anglo									
Mean	1.29	.71	2.43	3.43	2.43	4.29	4.14	6.00	6.29
SD	.95	1.11	2.23	2.37	1.51	.95	2.80	1.73	1.60
Mex-Am									
Mean	.57	1.71	2.43	2.14	2.71	4.00	4.29	5.71	6.14
SD	.98	1.70	2.23	.90	1.50	1.41	2.29	1.38	1.07
Black									
Mean	.57	.86	2.00	1.43	2.71	4.14	2.71	5.72	6.00
SD	.79	.85	1.83	1.40	1.60	.90	1.38	1.25	1.00

— difference not significant

___ difference significant

Since results for the Matrices and Order of Appearance tests turned out to be essentially the same, only those for the Matrices will be given. Inspection of the means and standard deviations reveals a marked decline in differences in performance between the groups. The differences ranged from approximately two-thirds of a standard deviation between Anglo and Black performance under Condition 1, to essentially zero under Condition 6. Although this could be attributed to a ceiling effect for the older group (means of just over 6 from an 8-item test), such an interpretation cannot be made for the younger and middle groups (means around 2 and 4 respectively). The results are summarized in Table 2 on the following page.

1.6 Criterion-Related Validity

The issue of test fairness is closely related to the criterion-validity of the test. The most commonly used application is predictive validity. If a test, i.e., testing procedure which yields the scores, is "fair," the relationship between the predictor and criterion contains minimum random error and the random error is constant between specific groups tested.

In a previous study (Carlson & Wiedl, 1979), differing patterns of corrlations were detected between CPM performance obtained through modified testing procedures and measures of school achievement. Similarly, Carlson and Dillon (1979) found that prediction of school success was enhanced through the application of dynamic testing procedures. Unfortunatley, specific comparisons of predictive validity of several measures of Level II intelligence and school success have

not been made employing dynamic assessment approaches with children of different ethnic/racial background.

1.7 Summary

The approach of applying optimizing testing procedures to the normal population as well as to special groups of children was demonstrated. Restrictions and differentiations in the effects of testing procedures, test format, and task demands were found. The applicability of certain testing procedures for purposes of assessing intellectual performance of normal and special school (learning handicapped, low IQ) children was shown. This indicates that specific forms of performance optimization can lead to more accurate, thus fairer, assessment of intellectual functioning. It was demonstrated that specific dysfunctional individual characteristics such as impulsivity can be compensated for under certain testing conditions. This underscores the necessity of critically analyzing how variations in testing procedure can interact with certain individual differences. The implication here is that the practitioner must be aware of these interactions and under which circumstances performance may be increased for some children, while decreased for others.

The indication that compensatory effects can be obtained in assessing between-ethnic group differences is of practical and theoretical significance. Extension of the approach by employing other measures of Level II intelligence (Jensen, 1973) to larger samples with control over factors such as socio-economic status is necessary before firm conclusions can be reached, however.

Chapter II

The Approach

2.0 Specific Objectives and Methods

The basic objective of the project was to assess the usefulness of the dynamic testing approach as a way of reducing or eliminating test bias. Bias here has two definitional bases: (1) where a significantly large proportion of variance in test performance can be attributed to other than "cognitive" factors; and (2) where differential prediction, i.e., over- or underprediction, of academic performance for any particular group occurs. Bias is not defined in terms of mean differences in performance between groups. Rather, it is conceived of as error in measurement attributable to factors entering into performance which are not the target of the assessment, e.g., when noncognitive and/or style variables unduly affect cognitive performance.

2.1 Research Questions

The research was designed to shed light on three interrelated questions addressing issues related to assessment of cognitive functioning and learning characteristics of children.

1. Does the dynamic assessment approach yield information which is a more appropriate indicator of cognitive competence than that obtained by traditional, standard approaches based on static test theory?

   Corollary A: Do specific testing conditions differentially affect the performance of children

of different social and/or ethnic background?

If, as expected, the answer to the above questions are positive, the following question becomes:

2. Why and for whom is the dynamic assessment approach effective? Stated differently: What are the sources of intraindividual variation, and how may they be accounted for by personality and cognitive style variables?

The third question addressed the issue of predictive validity and can be stated in the following way:

3. For what reason should dynamic assessment procedures be employed?

The objective was to assess the impact of dynamic assessment on the very real problem of predicting school and curricular success.

2.2 Methods

The organization of this section will be along the following lines. First, the sample will be described. Second, the general procedures followed will be outlined. Third, the instruments used will be given and procedures followed in their administration specified.

2.2.1 The Sample

Originally the plan of the study was to attain a sample of 225 fourth-grade children: 75 Anglo-American, 75 Black-American, and 75 Hispanic-American. The children were to be matched on socio-economic level. Our goals in sample size and selection were not fully realized due to logistic and "political" reasons.

The final sample consisted of 147 fourth- and fifth-grade children from 15 elementary schools serving the San Bernardino, California area. The schools were selected from low-, middle- and high-SES areas in order to insure that the widest possible SES range would be included in the sample. There were 67 Anglo-American children (29 males, 38 females), mean age 10 years, 4 months, standard deviation 9 months; 37 Black-American children (20 males, 17 females), mean age 10 years, 8 months, standard deviation 11 months; and 43 Hispanic-American children, (18 males, 25 females), mean age 10 years, 5 months, standard deviation 10 months.

One the factors involved in not obtaining the total sample size we had originally planned on was due to the unwillingness of minority parents to allow their sons or daughters to participate in this project. Written permission, in the form of a signed informed-consent letter, was required before a child was included in the study. The San Bernardino Unified School District left it to the discretion of the principals at the individual schools whether or not a given school would participate in the project. For various reasons, the principals of schools that had large minority populations were less willing to cooperate than principals in schools serving a more heterogeneous group of children. Nonetheless, we were able to obtain sufficient numbers of minority students to carry out most of the analyses planned.

From each ethnic group, subjects were randomly assigned to one of three testing conditions: C_1 standard; C_2 verbalization; C_3

elaborated feedback. The numbers of subjects in each condition are shown in the following table

	Test Condition			
Ethnic Group	C1	C2	C3	Total
Anglo	23	22	22	67
Black	11	12	14	37
Hispanic	14	17	12	43
Total	48	51	48	147

2.2.2 Procedures

Due to the large amount of time required for testing and the limited testing time available per testing session at most schools, a minimum of three sessions was necessary for complete testing of one subject. Total testing time per subject was approximately 2 1/2 to 3 hours. The majority of this time was used to administer the two tests of cognitive and perceptual functioning: Raven's Matrices and the Cattell Culture Fair Test "g" (CFT). Each of these measures took approximately one hour to administer. The Raven Matrices were always administered before the CFT, and usually during the first session. Typical testing sessions were as follows:

Session 1	Session 2	Session 3
Raven Matrices	CFT	Harter Perceived
Matching Familiar	Trail-Making	Competence Scale
Figures Test (MFFT)		Visual Search

All testing was done individually by trained graduate students. Specific administration protocols for each measure will now be presented.

2.2.3 Measures of Cognitive and Perceptual Functioning

1. Raven's Matrices.

Raven's Coloured Progressive Matrices (CPM) consists of three sections (A,Ab,B) of 12 items each. Each item consists of a main stimulus pattern with a piece missing from the lower right hand corner. The subject's task is to select from the alternatives presented underneath the main stimulus pattern the "piece" that correctly completes the pattern. Factor analytic studies of the CPM have revealed item groupings that require simple pattern completion, pattern completion through closure, and concrete and abstract reasoning by analogy (Weidl & Carlson, 1976; Carlson & Jensen, 1980).

Reliability and validity estimates of the CPM are generally reported to be high. Carlson and Jensen (1981) found the CPM to have equally high reliability for each of three ethnic groups: Anglo, Black and Hispanic.

In this study, CPM items were supplemented by section C (12 items) of the Standard Progressive Matrices (SPM). The format of the SPM items is essentially the same as the CPM items; although the items are not in color and 8 alternatives are presented rather than 6. Section C of the SPM was

administered after section A, Ab, and B of the CPM had been completed. Thus the test consisted of a total of 48 items. All 147 subjects completed the Raven's CPM and section C of the SPM. The test conditions and specific administration protocols associated with them are presented below:

C_1 - Standard Instruction - The standard protocol set forth in the CPM manual (Raven, 1965) was used. The tester turned to the first problem in the test booklet (item A1), pointing out to the subject the main stimulus pattern, noting that it had a piece missing from the lower right hand corner. The child's attention was then directed to the six answer alternatives below. He/she was told that only one of the six alternatives was correct and completed the pattern above. He/she was asked to say which one he/she thought was the correct one. If the child chose the correct piece, the tester replied "Yes, that's the right one," and explained why that was the right one and why the other five alternatives were incorrect. If the child picked a distractor, the tester said "No, that's not the right one," and pointed out to the child the correct piece. This was followed by an explanation of why it was the correct one, and why the one the child had picked was incorrect.

The child was then told to write down the number of the correct piece on his/her answer sheet, and that on

each page of the booklet there was a similar problem. He/she was asked to work through the booklet individually, and to write down the number of the piece he/she thought was correct in the appropriate place on his/her answer sheet. When the child completed the CPM, the tester showed him/her Section C of the SPM and told him/her that these were like the puzzles he/she just did, except that they were not in color, and he/she now had eight pieces to choose from. No further instruction was given.

C_2 - Verbalization During and After Solution. The tester showed the child item A1 and asked him/her to describe the main stimulus pattern, paying attention to the details of the pattern. E aided S in "sharpening" his/her description. The child was then asked to choose from the answer alternatives below the pattern that he/she thought was correct and would complete the pattern. The child was then asked to explain why he/she thought the chosen alternative was correct. Again E aided S "sharpen" his/her response. For this sample item, the tester informed the child if his/her choice was correct or incorrect and why. The child was then instructed to work through the rest of the booklet, but

very carefully describe to the tester the main stimulus pattern before giving an answer. He/she was asked to always describe why he/she thought the alternative he/she chose was the correct one. After the first sample problem, the tester gave the child no feedback as to the correctness of his choice. The tester recorded the child's responses on an answer sheet. When the CPM was finished, Section C of the SPM was given.

C_3 - Elaborated Feedback. As in conditions C_1, the tester showed the child item A1 and pointed out that the main stimulus pattern had a piece missing from the lower right hand corner, and that his/her task was to select from the six alternatives below the one which would correctly complete the pattern. If the child chose the correct piece, the tester said "Yes, that's the right one," and explained why it was correct. If the child chose a distractor, the tester replied "No, that's not the right one," and explained why it was an incorrect response. E then pointed out the correct piece, and explained why it was correct. This same procedure was followed for the remaining items, except that the experimenter did not point out the correct piece if the child's response was wrong. The tester always made an effort to explain to the child the principles involved

in the solution of the problem. Often the child chose the correct response after the experimenter had explained the principles involved. However, in all three conditions, only the child's initial response was recorded by the experimenter, i.e., the response given before any feedback.

2. Cattell's Culture Fair Test of "g" (Scale 2, Form A)

Cattell's CET was administered in the second or third testing session, and in the same manner as the Raven's Matrices. The CET is composed of four subtests: "Series Completion" (12 items), "Classification" (14 items), "Matrices" (12 items), and "Conditions," or topological relations (8 items). The Series, Matrices, and Conditions subtests have 3 practice items each. Classification has two items.

The Cattell CFT is a well standardized, highly reliable intelligence test. Total scores on the CFT correlate moderately with other intelligence tests, including Raven's Matrices.

The same general procedures as described above for administration of the Matrices under standard, verbalization, and elaborated feedback conditions were followed. The tester worked through the practice items of each test with the child, informing him/her whether the child was correct or not and why. Due to repeated absences, three subjects did not receive the CFT.

2.2.4 Measures of Cognitive Style and Personality

1. Matching Familiar Figures Test (MFF)

The MFF (Kagan et al., 1963) was used to assess individual differences in the cognitive style dimension "impulsivity-reflectivity." The test consists of 12 items. Each item has a picture of a familiar object, e.g., a house. Below the picture, six alternatives, one of which is identical to the picture above, are given. The subject's task is to point to the picture below which he/she thinks is the one identical to the picture "as quickly and correctly as you can." The tester records latency to the subject's first response, and the number of errors made on each item. Test scores used for analysis are mean latency and total errors across all twelve items.

Impulsive children are defined as those who have fast (below the group median) latencies, and who make many (above the group median) errors. Reflective children are defined as having slow (above the group median) latencies, and few (below the group median) errors. A "double median split" on time and errors is usually performed to identify impulsive and reflective children. However, since this procedures has methodological difficulties, and results in a considerable loss of information. For analysis in this study, MFF time and error were treated as separate continuous variables.

All 147 subjects completed the MFF.

2. Harter's Perceived Competence Scale for Children

The Perceived Competence Scale for Children is a measure of motivation recently developed by Harter (1979a; 1979b). The model upon which the scale is based is that a child's perception of his/her competence is an important factor in his/her intrinsic motivation to be effective, curious, prefer optimal challenge, and engage in independent mastery attempts. The scale assesses four domains of perceived competence: (1) "cognitive competence," which emphasizes academic performance; (2) "social competence," which stresses social and peer group relationships; (3) "physical competence," which focuses on physical prowess; and (4) "general self-worth," which refers to being sure of one's self, being happy with the way one is, etc. The scales have been shown to have moderate to high reliability, and have been used with children 6 to 12 years of age.

Each domain consists of seven items. E read each item aloud to the subject. For example, one item from the cognitive competence domain is: "Some kids often forget what they learn BUT other kids can remember things easily." The child is asked to decide if he/she is more like the first child described, or more like the second child described.

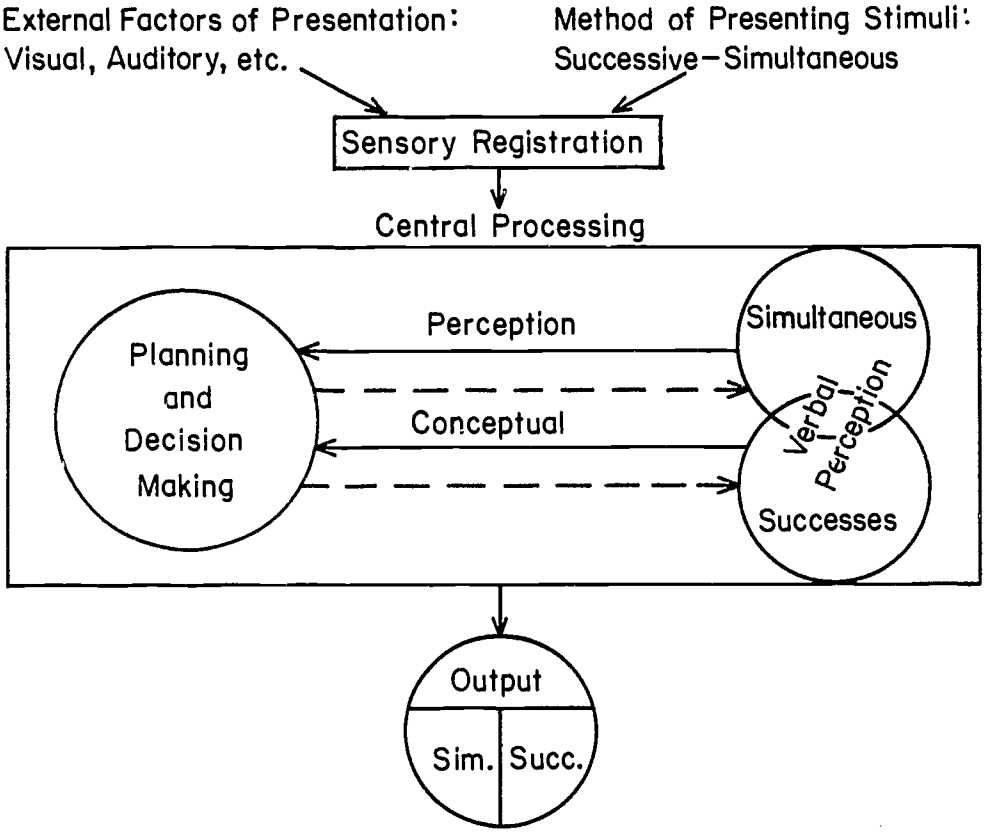
Once the child makes this decision, he/she is asked if the description is "really true," or only "sort of true" for him/her. Each item is scored on a scale of 1 to 4; the lower score reflecting greater perceived competence.

Due to repeated absences, one child was not given the Harter scales.

3. Trail Making and Visual Search Tasks

The Trail-Making and Visual Search tasks are used to measure planning ability. The role of planning and decision making in the performance of cognitive tasks has been established by Luria (1973), and expanded to include psychometric considerations by Das (1980). A good planner exhibits the following characteristics: (1) "looks before he leaps," (2) checks for errors, (3) takes all factors into account before making judgment, (4) evaluates feedback and acts accordingly, and (5) can generally come up with strategies to approach a problem. Das (1980) has found that planning ability affects cognitive performance, but is factorially independent from the impulsivity-reflectivity dimension and simultaneous or successive modes of information integration.

Das (personal communication) presents the following model incorporating planning and decision making in cognitive functioning.



The Visual Search task consists of 48 geometric shapes, letters and numerical sets. Due to limitations in testing time, a subset of 14 items was used in this study. Each item is presented on a transparency, at the center of which is a circle which contains an exact copy of the one "target" figure (or letter or number) contained elsewhere on the

transparency. The subject's task is to find and place his/her finger on the figure which matches the one in the circle. The tester records the elapsed time for the subject to search the transparency and find the target figure (search time), and the time it takes for the subject to move his/her finger and touch the appropriate figure (movement time). Search time is measured from when the transparency is illuminated to when S moves his/her finger from a button on the illumination box. Movement time is the time of the ballistic from the button to the figure on the transparency. Two timers are used. Movement times were not included in any of the analyses, as this variable has not been shown to be central to planning. Search times, on the other hand, are the most important measure on this task. Thus these data were analyzed.

The Trail Making task presents the subject with stimulus sheets containing numbers and letters distributed randomly. The subjects task is to draw lines connecting the letters and numbers in the correct sequence, e.g., 1-A, 2-B, 3-C, etc. There are two parts to the test. The tester records with a stopwatch the elapsed time for the subject to complete each part of the test. Total time on both parts is recorded and used for analysis.

Due to repeated absences, the Trail Making task was not administered to one subject. Due to repeated absences and

equipment failure, complete and usable visual search data were collected or only 131 subjects.

2.25 Measures of School Achievement

In order to assess the "predictive validity" of the cognitive ability measures administered under the different test conditions, two measures of school achievement were obtained. These were standard scores on reading and mathematics from the California Test of Basic Skills (CTBS) administered in April 1982. Reading test scores were available for 130 subjects; math test scores for 122 subjects.

2.3 Summary

A total of 147 fourth- and fifth-grade children representing three ethnic groups participated in this study. Measures of cognitive and perceptual functioning, impulsivity-reflectivity, motivation, and planning were administered. School achievement data, in the form of CTBS reading and math scores, were obtained.

Analysis of the data will address (1) the effect of testing conditions (static vs. dynamic) on total Raven and Cattell scores, as well as on subtests and item groupings of these measures; (2) the relationship of ethnic group classifiction to Raven's and Cattell performance under the different test conditions; (3) the effects of non-target variables (impulsivity-reflectivity, motivation, planning) on Raven and Cattell performance under the different conditions; and (4) the "predictive validity" of the cognitive ability measures administered under standard vs. elaborative testing conditions.

Chapter III

The Results

3.0 Organization

The results will be organized and presented in three sections. The first section will address the effects of testing condition per se on Raven and Cattell performance for the three ethnic groups in the study. The second section will address the relationship between non-cognitive individual difference variables (impulsivity, motivation, planning) and performance on Raven's matrices and Cattell's CFT for each testing condition. The final section will present a comparison of the predictive "power" of Raven's Matrices and Cattell's CFT administered under the different conditions. In Sections 1 and 2, Raven item groupings and the Cattell subtests will be treated as dependent variables in addition to total scores on these two measures.

3.1 Relationship to Test Condition and Ethnic Group Classification to Raven's Matrices Performance

Table 1 presents the means, standard deviations, and coefficent alpha (Cronbach, 1951) reliability estimates for total Raven Matrices scores separately by test condition and ethnic group classification. As will be noted, the reliabilities tend to fluctuate slightly across testing condition and ethnic group classification. Nonetheless, they remain generally high.

Table 1

Raven Matrices Total Score Means, Standard Deviations, and Internal Consistency (Coefficient Alpha) Reliability Estimates as a Function of Test Condition and Ethnic Group Classification

Test Condition	Anglo			Black			Hispanic		
	$\bar{X}$	SD	r_{kk}	$\bar{X}$	SD	r_{kk}	$\bar{X}$	SD	r_{kk}
C_1	36.9	5.4	.82	34.0	5.8	.82	31.4	6.4	.85
C_2	40.8	3.3	.61	36.9	6.1	.86	34.2	5.3	.78
C_3	37.5	4.5	.72	34.2	5.6	.75	32.3	4.7	.66

C_1= Standard Administration; C_2= Verbalization; C_3= Elaborated Feedback

Potential effects of the testing conditions on performance wasr analyzed by analysis of variance. Table 2 presents the results of the 3 (test condition) x 3 (ethnic group) analysis of variance performed on Raven total scores.

Table 2

Test Condition (A) x Ethnic Group (B) Analysis of Variance on Raven Total Scores

Source	df	MS	F
Test Condition (A)	2	135.97	5.20**
Ethnic Group (B)	2	443.28	16.94**
A x B	4	2.99	> 1
Error	138	26.17	

Inspection of Table 2 reveals significant main effects for test conditon (A) and for ethnic group (B). The A x B interaction did not approach statistical significance. Scheffe post hoc comparisons were performed to further analyze the significant main effects for test condition and ethnic group. These comparisons showed that Anglos scored significantly higher than either Hispanics ($p < .01$) or Blacks ($p < .01$), while no significant difference between Hispanic and Blacks was detected. These results are presented graphically in Figure 1.

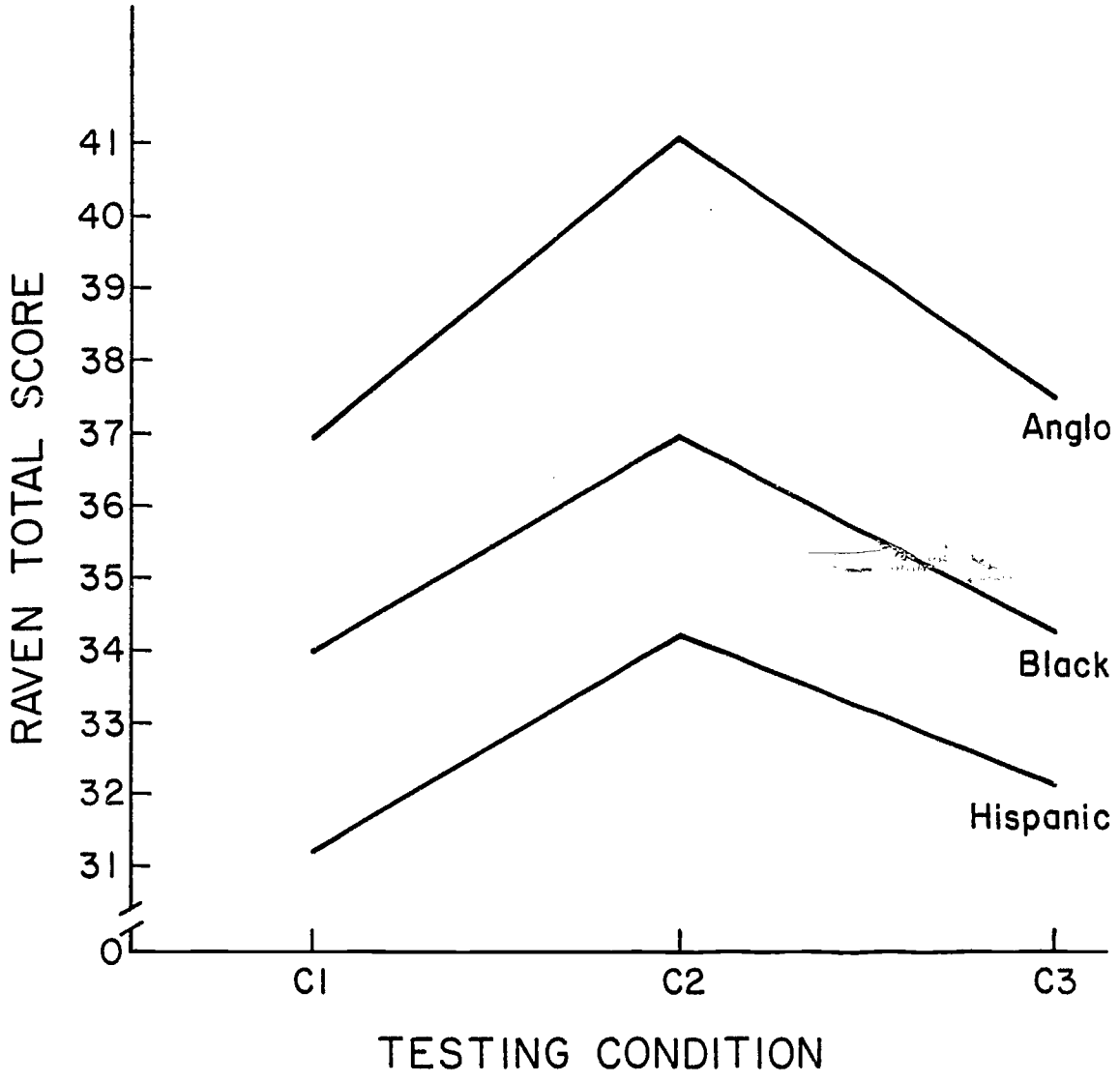


Fig. 1. Performance of Three Ethnic Groups on Raven's Matrices Administered Under Three Testing Conditions.

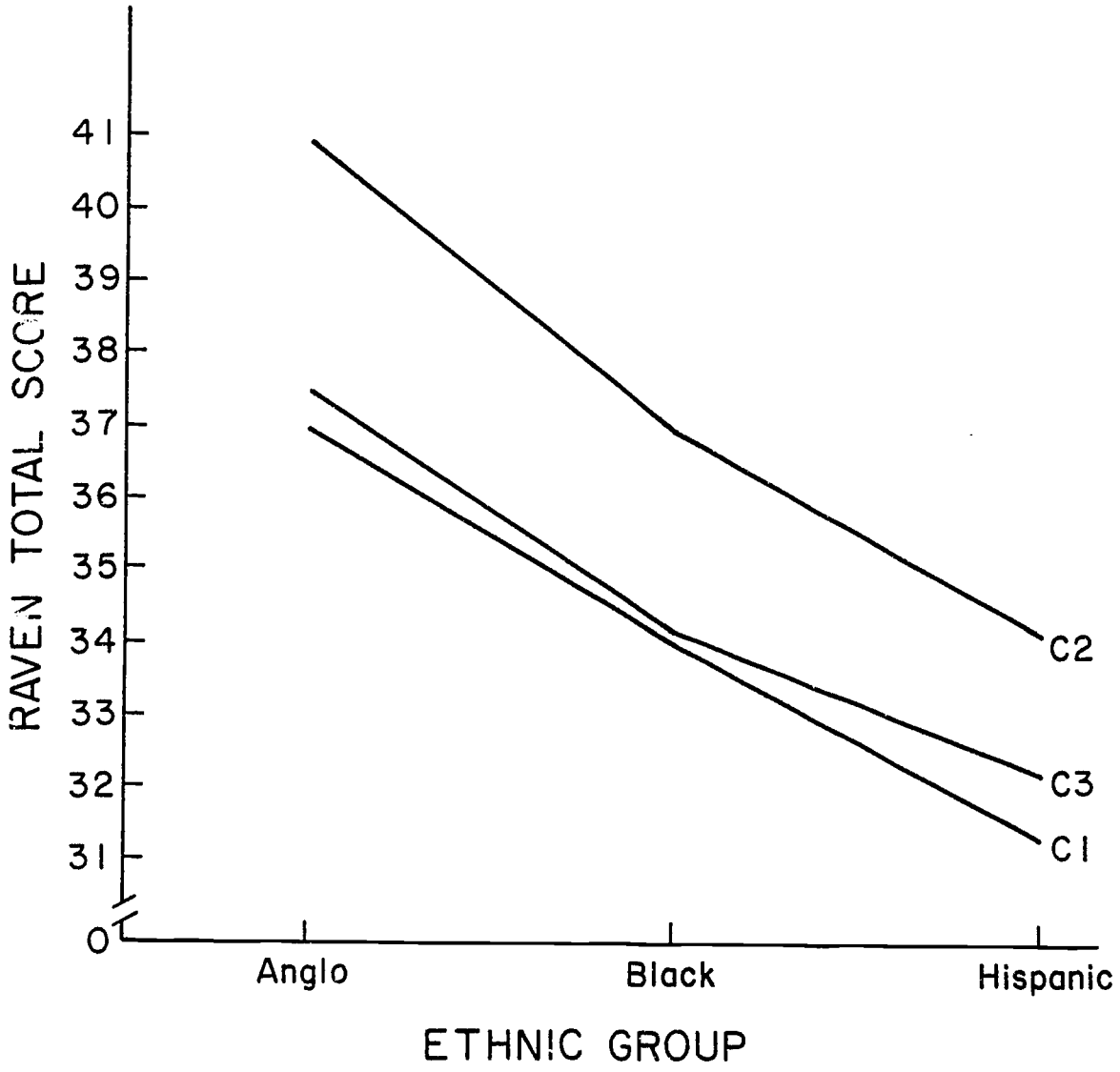


Fig. 2. Effect of Testing Condition on Raven Matrices Performance for Three Ethnic Groups.

The multiple comparisons for test condition revealed that performance under C_2 (Verbalization) was superior to performance under C_1 (standard administration) ($p<.05$), and marginally superior to performance under C_3 (elaborated feedback) ($p = .058$). C_1 and C_3 were not significantly different from each other. Figure 2 graphically presents mean total Raven scores under the three test conditions for each ethnic group.

The results of the test condition x ethnic group ANOVA can be summarized as follows: Regardless of ethnic group classification, performance is highest under C_2 (verbalization), and lowest under C_1 (standard administration). Performance under C_3 (elaborated feedback) is only slightly higher than performance under C_1. Regardless of test condition, Anglos performed significantly higher than both Blacks and Hispanics. Blacks tended to have higher total Raven scores than Hispanics under all three test conditions. However, the difference only approaches statistical significance.

The items of Raven's CPM were divided into two groups which have been shown to require different processes or strategies for solution: Pattern Completion and Closure items and Concrete or Abstract Reasoning by Analogy items. The groupings were based on previous factor analytic studies of the CPM (see Carlson & Jensen, 1980; Wiedl & Carlson, 1976). Section C of the Standard Progressive Matrices (SPM) was analyzed separately as a third subgroup of items. Means and standard deviations for the three item subgroups are presented in Table 3 separately for the test condition x ethnic group categories.

Table 3

Raven Item Subgroup[1] Means and Standard Deviations as a Function of Test Condition and Ethnic Group Classification

Test Condition	Ethnic Group: Anglo		Black		Hispanic	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
(C_1) Pattern Completion	16.2	1.0	15.5	2.1	14.5	2.1
Reasoning	2.8	1.7	2.5	1.9	2.1	1.7
Section C	7.5	1.9	6.2	2.4	5.7	2.5
(C_2) Pattern Completion	16.7	0.6	15.9	1.4	15.8	2.5
Reasoning	4.1	1.1	3.3	1.6	2.8	2.0
Section C	8.4	1.5	7.3	1.9	6.2	2.0

Separate analyses of variance were carried out for each Raven item subgroup. The independent variables were test condition and ethnic group. The results are summarized in Table 4.

Table 4

Summary Table for Results of Analyses of

Variance on Raven Item Subgroupings

Source	Pattern Completion	Reasoning	Section C
Test Condition (A)	xx	x	ns
Ethnic Group (B)	xx	x	xx
A x B	ns	ns	ns

xF value significant, p<.05

xxF value signifianct, p <.01

Significant main effects for test condition were found for pattern completion and reasoning item subgroups. Significant main effects for ethnic group were found for pattern completion, reasoning and Seciton C item groupings. None of the test condition x ethnic group interactions was significant. Scheffe post hoc comparisons were performed to further analyze the significant main effects for test condition and ethnic group classification. These comparisons are summarized in Table 5.

Table 5

Summary of Scheffe Post Hoc Comparisons for Test Condition and Ethnic Group[1] on Raven Matrices Item Subgroups

	Item Subgroup		
Comparison	Pattern Completion	Reasoning	Section C
(A) Test Condition	C2>C3 (p<.01)	C2>C1 (p<.01)	x
(b) Ethnic Group	A>H (p<.01)	A>H (p<.01)	A>B (p<.05) A>H (p<.01)

[1]A=Anglo; B=Black; H=Hispanic

The results of the multiple comparisons for test condition on the item subgroups are similar to those for test condition performed on total Raven scores. The verbalization condition (C_2) was found to lead to higher pattern completion scores than C_3 (p<.01) and higher reasoning item scores than C_1 (p<.01). Since the omnibus F test for test condition on Section C items was nonsignificant, no post hoc comparisons for test condition were performed on this item subgroup.

The results of the multiple comparisons for ethnic group on the Raven subgroup items are also similar to those reported for Raven total scores. Specifically, Anglos performed higher on all three item subgroups than Hispanics (p<.01; p<.05; p<.05, respectively); and higher on section C items than Blacks (p<.05).

Figure 3, on the following page, shows the mean scores on the three item sugroups as a function of test condition and ethnic group.

The graphs reveal additional information concerning trends in the data not revealed by the ANOVA or the multiple comparisons. Specifically, they show that for the pattern completion items, performance tends to be poorest under C_3 for all three ethnic groups; albeit not significantly poorer than performance under C_1. C_2 tends to be the most efficacious condition for pattern completion items for all three ethnic groups and for reasoning and section C items for the Anglo and Black subjects. Interestingly, the Hispanics' best performance on the reasoning and section C items was under C_3. (Again, this is a trend, approaching statistical significance.)

When the results of the Raven item subgroup analyses are combined with the results of the analysis of total Raven scores, the following conclusions can be made concerning the relationships between ethnicity, testing condition, and Raven's matrice performance: (1) Regardless of item type or test condition, Anglos perform better on Raven's matrices than either Blacks or Hispanics; (2) C_2 appears to be the most effective condition for maximizing performance, although C_3 is equally effective or slightly more so than C_2 for Hispanic subjects on the more difficult reasoning and Section C items; (3) both elaborative conditions (C_2 and C_3) lead to enhanced levels of performance on the more difficult reasoning and section C items for all ethnic groups. However, elaborated feedback (C_3) appears to "hinder" performance on the relatively easy pattern completion

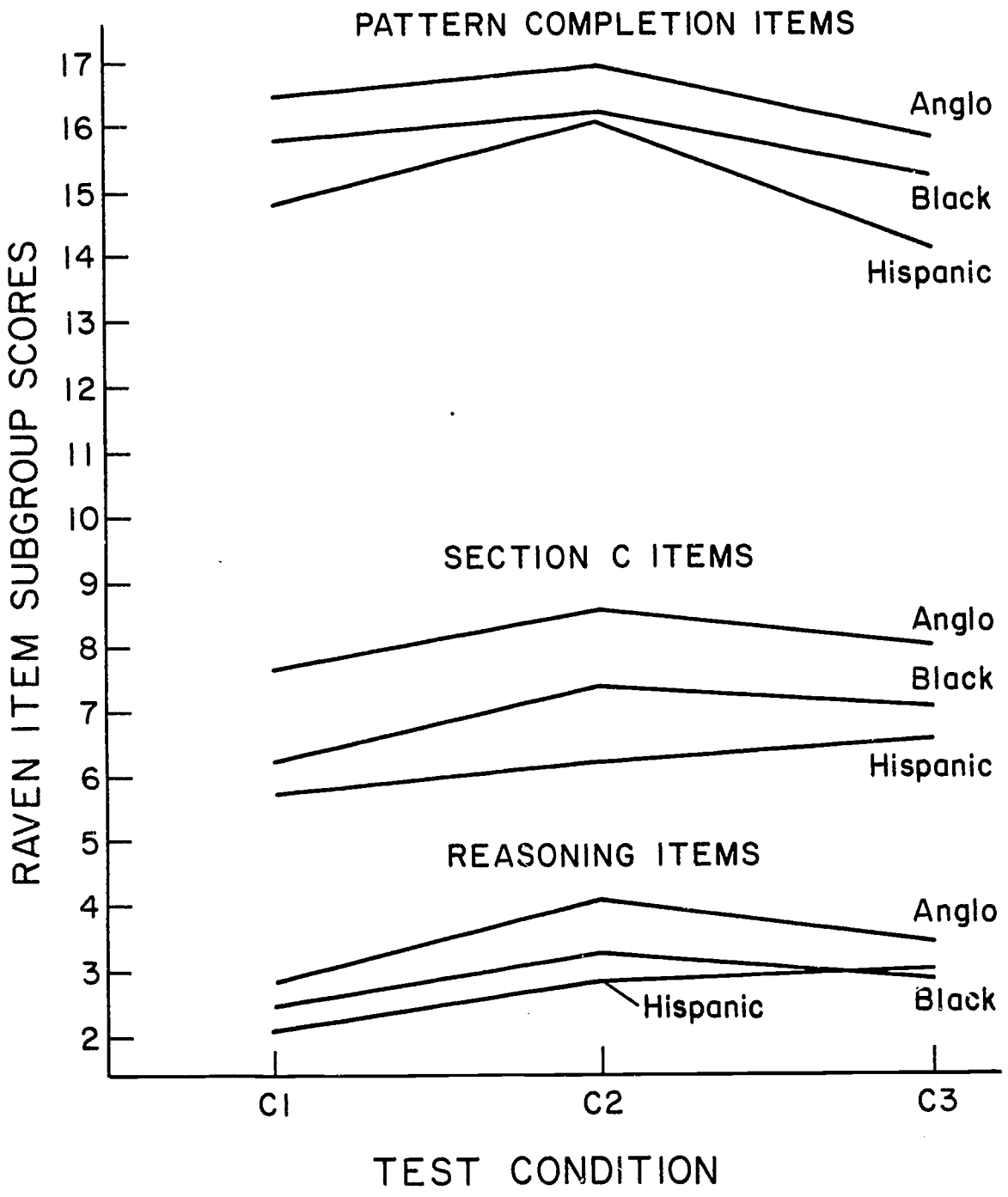


Fig. 3. Performance of Ethnic Groups on Raven Matrices Item Subgroupings Administered Under Three Testing Conditions.

items, since performance on these items tended to be lower than performance under standard administration procedures for all three ethnic groups.

3.1.1 Relationship of Test Condition and Ethnic Group Classification to Cattell CFT Performance

Table 6 presents the means, standard deviations, and coefficient alpha reliability estimates for total Cattell CFT scores separately by test condition and ethnic group classification.

Table 6

Cattell CFT Total Score Means, Standard Deviations, and Internal Consistency Reliability Estimates as a Function of Test Cndition and Ethnic Group Classification

Test Condition	Anglo			Black			Hispanic		
	$\bar{X}$	SD	r_{kk}	$\bar{X}$	SD	r_{kk}	$\bar{X}$	SD	r_{kk}
C_1	27.7	6.1	.80	25.7	5.8	.77	26.8	4.6	.63
C_2	32.8	4.8	.73	28.6	5.8	.79	28.4	5.9	.78
C_3	30.8	4.9	.74	25.9	5.7	.74	26.2	4.5	.61

C_1= Standard Administration; C_2= Verbalization; C_3= Elaborated Feedback

Table 7 shows the results of the test condition x ethnic group analysis of variance performed on Cattell total scores.

Table 7

Test Condition (A) x Ethnic Group (B) Analysis of Variance on Cattell Total Scores

Source	df	MS	F
Test Condition (A)	2	121.55	4.23*
Ethnic Group (B)	2	219.79	7.65**
A x B	4	21.27	< 1
Error	135	28.72	

*p <.05

**p<.01

Inspection of Table 7 reveals results nearly identical to those reported in Table 2 for the ANOVA on Raven total scores. Significant main effects are shown for test condition (A) and ethnic group (B). The A x B interaction is not statistically significant. Scheffe post hoc comparisons for test condition show that performance under C_2 is significantly higher than performance under C_1 ($p < .01$). The differences in mean Cattell scores between C_1 and C_3, and C_2 and C_3 are not statistically significant.

Scheffe post hoc analyses for ethnic group indicate that Anglo subjects scored significantly higher than both Black ($p < .01$) and Hispanic ($p < .05$) subjects on Cattell total scores. Black and Hispanic Cattell scores were not significantly different. Figures 4 and 5 graphically present Cattell total socres as a function of test condition and ethnic group, respectively.

Each of the four subtests of the Cattell CFT was analyszed separately. The means and standard deviations on these subtests are presented separately for teach test condition by ethnic group category in Table 8.

Table 9 summarizes the results of the test condition x ethnic group ANOVAs performed one each of the Cattell subtests.

Inspection of Table 9 reveals significant main effects for test condition on two of the Cattell subtests: matrices and "conditions," or topological relations. A significant main effect for ethnic group was found on the series completion and matrices subtests, with the difference on the "conditions" subtest approaching statistical signifiance ($p = .08$). The test condition x ethnic group interaction on the "conditions" subtest also approached significance ($p = .06$).

Scheffe post hoc comparisons were performed on those subtests for which there were significant main effects. These comparisons are summarized in Table 10.

The comparisons presented in Table 10 show that on the series completion subtest Anglo subjects performed significantly higher than both Black ($p < .01$) and Hispanic subjects ($p < .05$). On the matrices

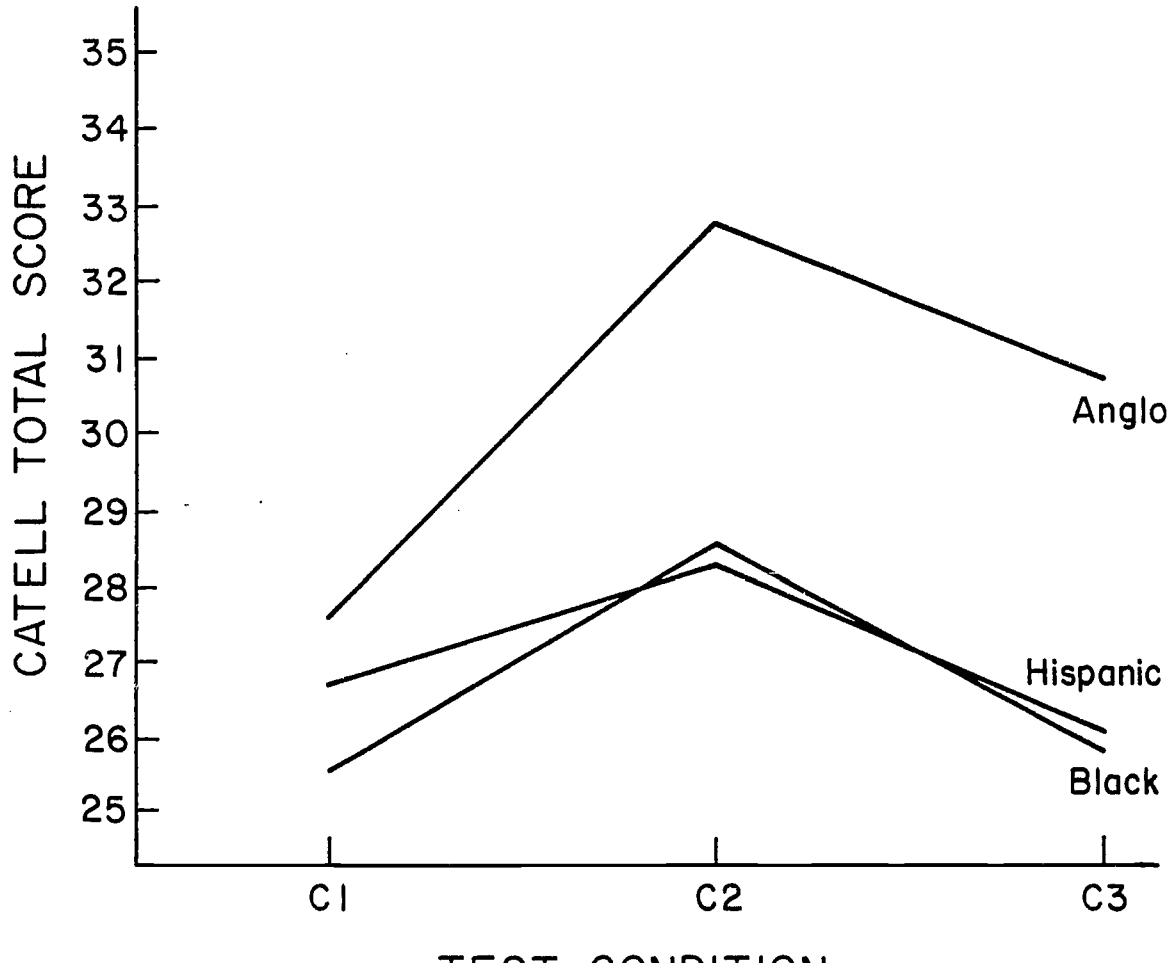


Fig. 4. Performance of Ethnic Groups on Cattell CFT Administered Under Three Testing Conditions.

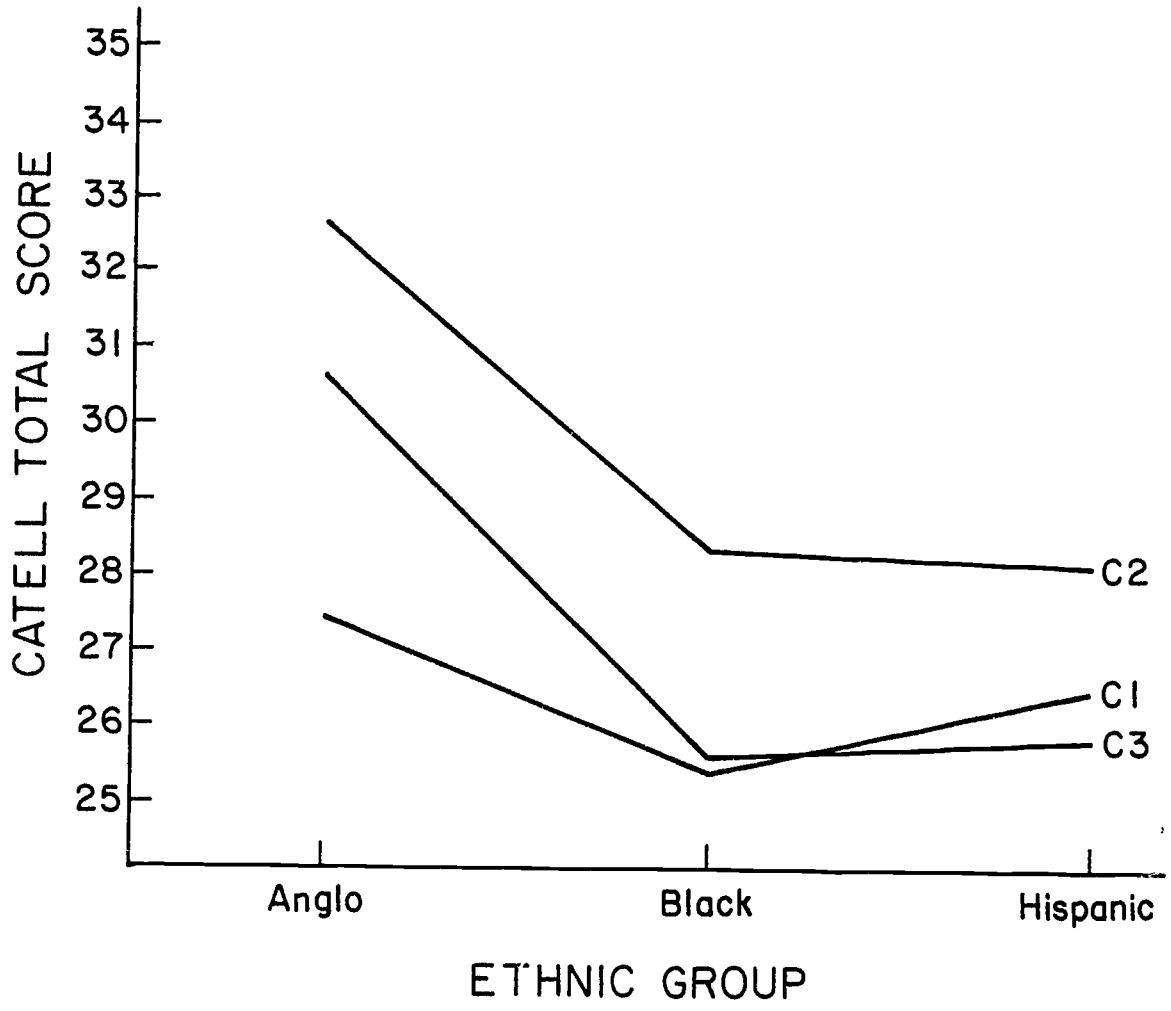


Fig. 5. Effect of Testing Condition on Cattell CFT Performance on Three Ethnic Groups.

Table 8

Means and Standard Deviations for Cattell CFT Subtests as a Function of Test Condition and Ethnic Group Classification

Ethnic Group and Test Condition	Subtest							
	Series		Classification		Matrices		Conditions	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Anglo C_1	8.5	1.5	6.7	2.2	8.1	1.8	4.3	1.9
Anglo C_2	9.3	1.8	7.7	2.1	9.9	1.2	6.0	1.6
Anglo C_3	8.7	1.6	7.6	1.9	9.0	1.8	5.5	1.5
Black C_1	6.9	2.5	6.7	1.5	7.4	2.2	4.7	2.1
Black C_2	7.8	2.0	7.4	2.0	8.3	1.8	5.2	1.5
Black C_3	7.2	2.4	6.9	2.2	7.6	2.2	4.2	1.2
Hispanic C_1	8.1	2.3	6.9	2.0	7.3	1.4	4.6	1.9
Hispanic C_2	7.6	2.0	7.2	1.9	8.0	2.6	5.6	2.0
Hispanic C_3	7.8	1.1	7.0	1.8	7.9	1.8	3.5	1.6

Table 9

Summary Table for Results of Analyses of

Variance on Cattell CFT Subtests

Source	Series	Classification	Matrices	Conditions
Test Condition (A)	ns	ns	x	xx
Ethnic Group (B)	xx	ns	xx	p=.08
A x B	ns	ns	ns	p=.06

^{x}F value, $p < .05$

^{xx}F value, $p < .01$

Table 10

Summary of Scheffe Post Hoc Comparisons for Test Condition

and Ethnic Group on Cattell CFT Subtests

	Cattell Subtest		
Comparison	Series	Matrices	Conditions
(A) Test Condition	x	C2>C1($p \leq .01$)	C2>C1($p \leq .01$)
			C2>C3($p \leq .05$)
(B) Ethnic Group	A>B($p \leq .01$)	A>B($p \leq .01$)	
	A>MA($p < .05$)	A>MA($p < .01$)	x

subtest, C_2 led to higher performance levels than C_1 ($p \leq .01$). Again, Anglo subjects performed significantly higher than both Black ($p \leq .01$) and Hispanic ($p \leq .01$) subjects. On the subtest "conditions", the verbalization test condition, C_2, led to higher performance levels than either C_1 ($p \leq .01$) or C_3 ($p \leq .05$).

Figures 6 through 13 graphically depict performance on the four Cattell subtests as a function of test condition (Figs. 6-9) and ethnic group (Figs. 10-13).

The analyses performed on the Cattell subtests indicate that the main effect for test condition found on Cattell total scores is largely due to the facilitative effect of verbalization on performance on the matrices and topological relations ("conditions") items. The significant main effect for ethnicity on Cattell total scores can be traced to the superior performance of Anglo subjects on the series completion and matrices items. As Figures 7 and 11 show, there was very little difference in performance on classification items attributable to either test condition or ethnic group category.

In general, the elaborative testing conditions, verbalization and feedback, led to significantly higher levels of performance for Anglo subjects. Verbalization led to higher levels of performance for Black and Hispanic subjects as well. However, elaborated feedback did little to improve the performance of the minority groups.

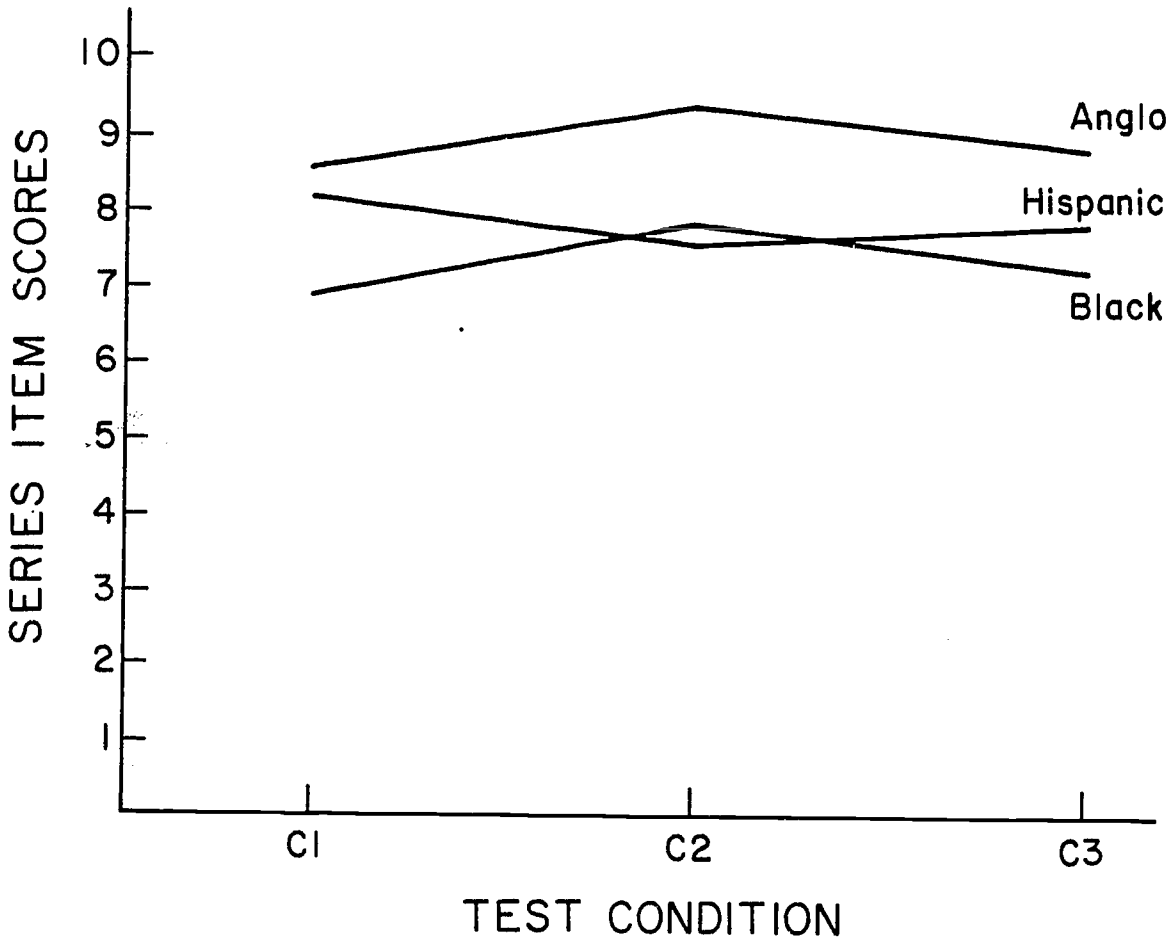


Fig. 6. Performance of Ethnic Groups on Cattell's CFT Series Completion Items Administered Under Three Testing Conditions.

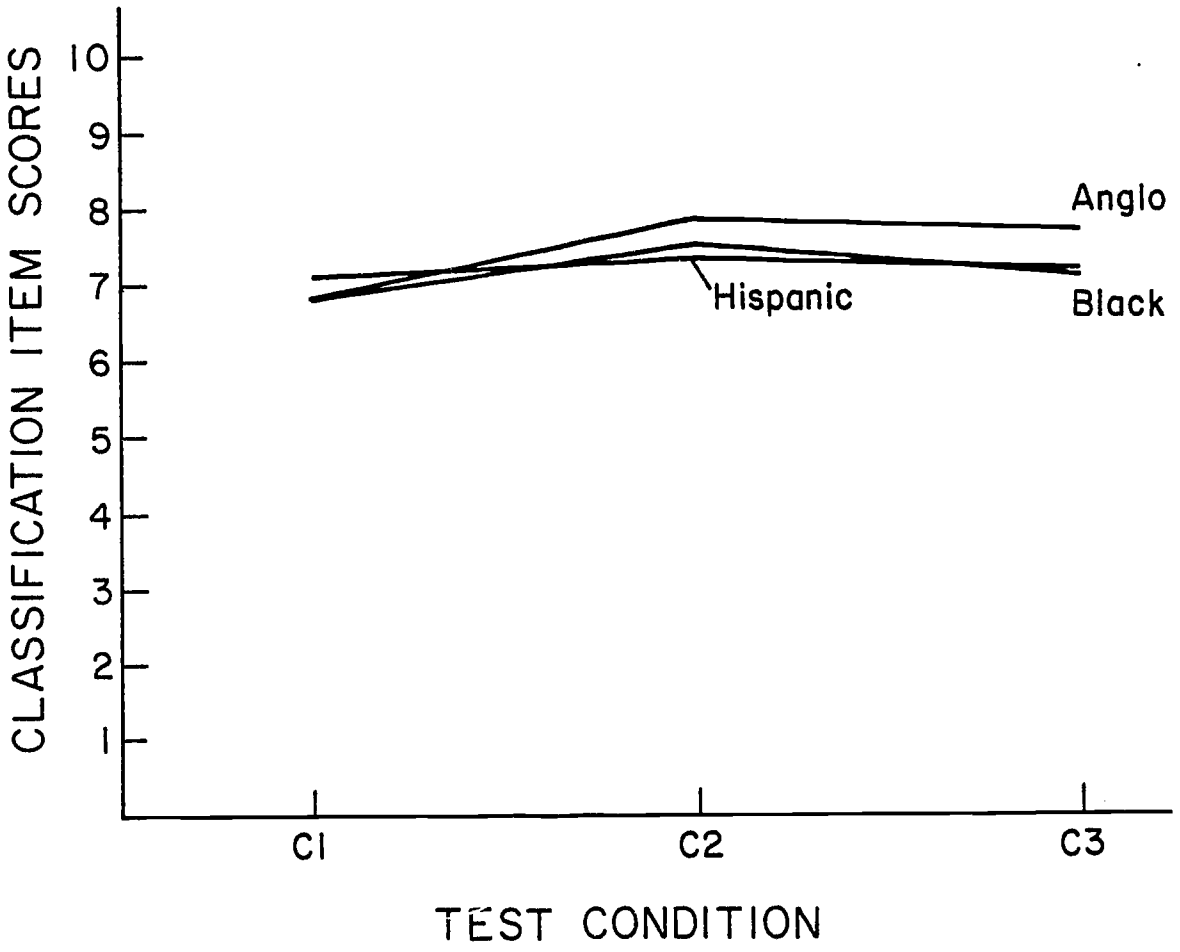


Fig. 7. Performance of Ethnic Groups on Cattell's CFT Classification Items Administered Under Three Testing Conditions.

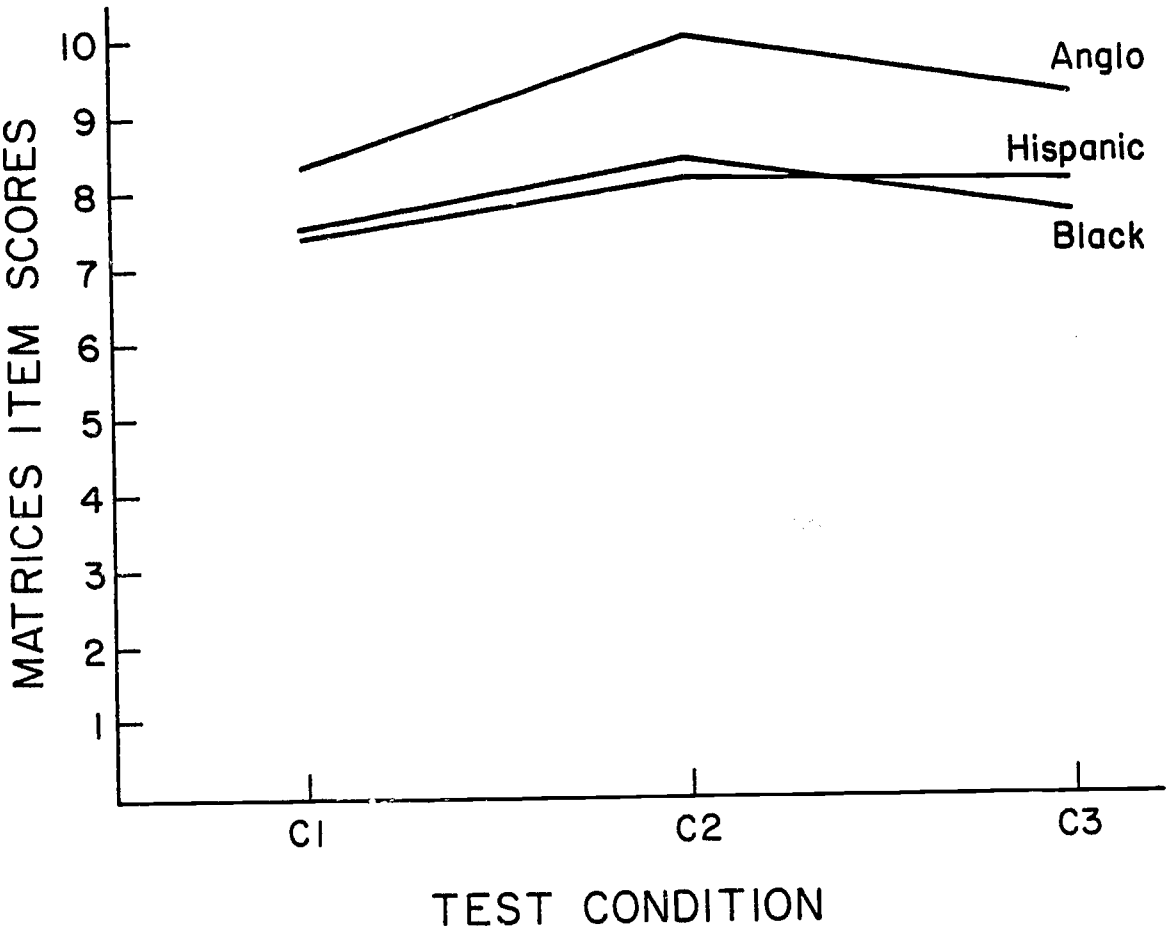


Fig. 8. Performance of Ethnic Groups on Cattell's CFT Matrice Items Administered Under Three Test Conditions.

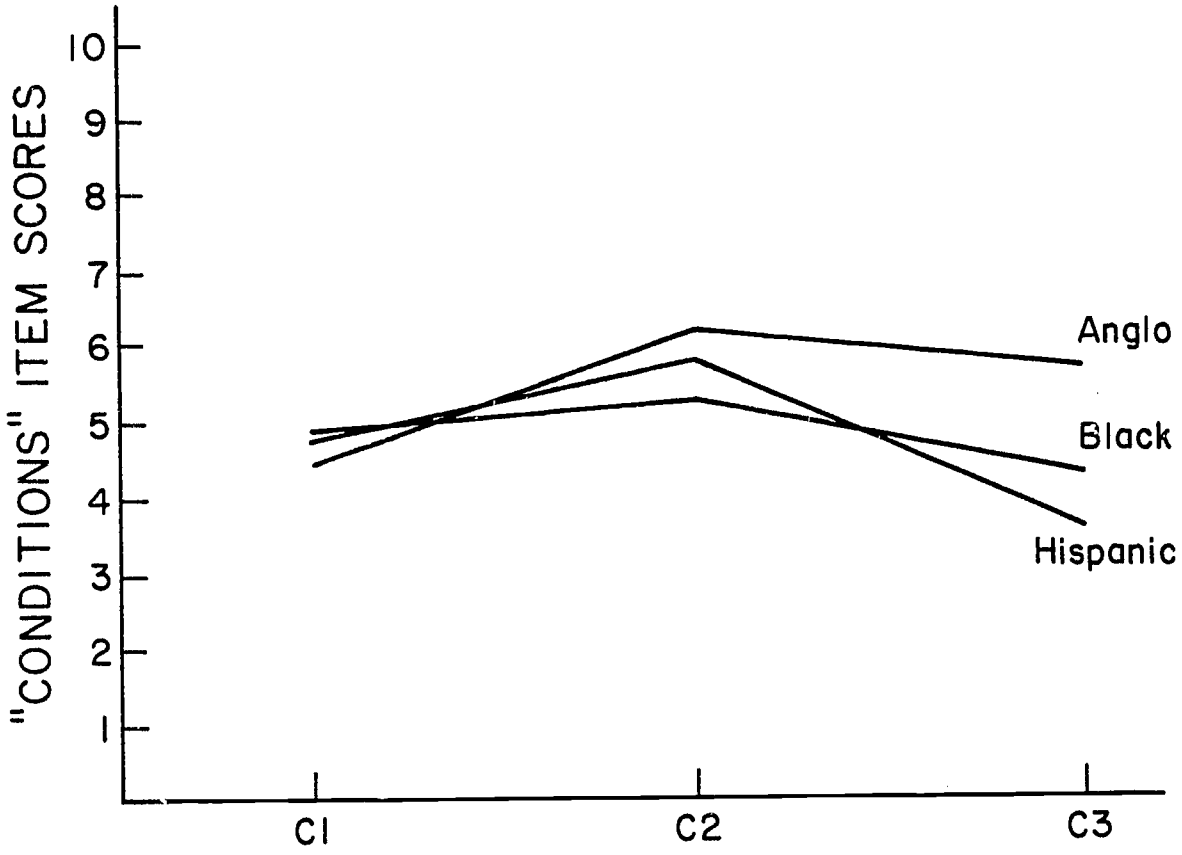


Fig. 9. Performance of Ethnic Groups on Cattell's CFT "Conditions" Items Administered Under Three Test Conditions.

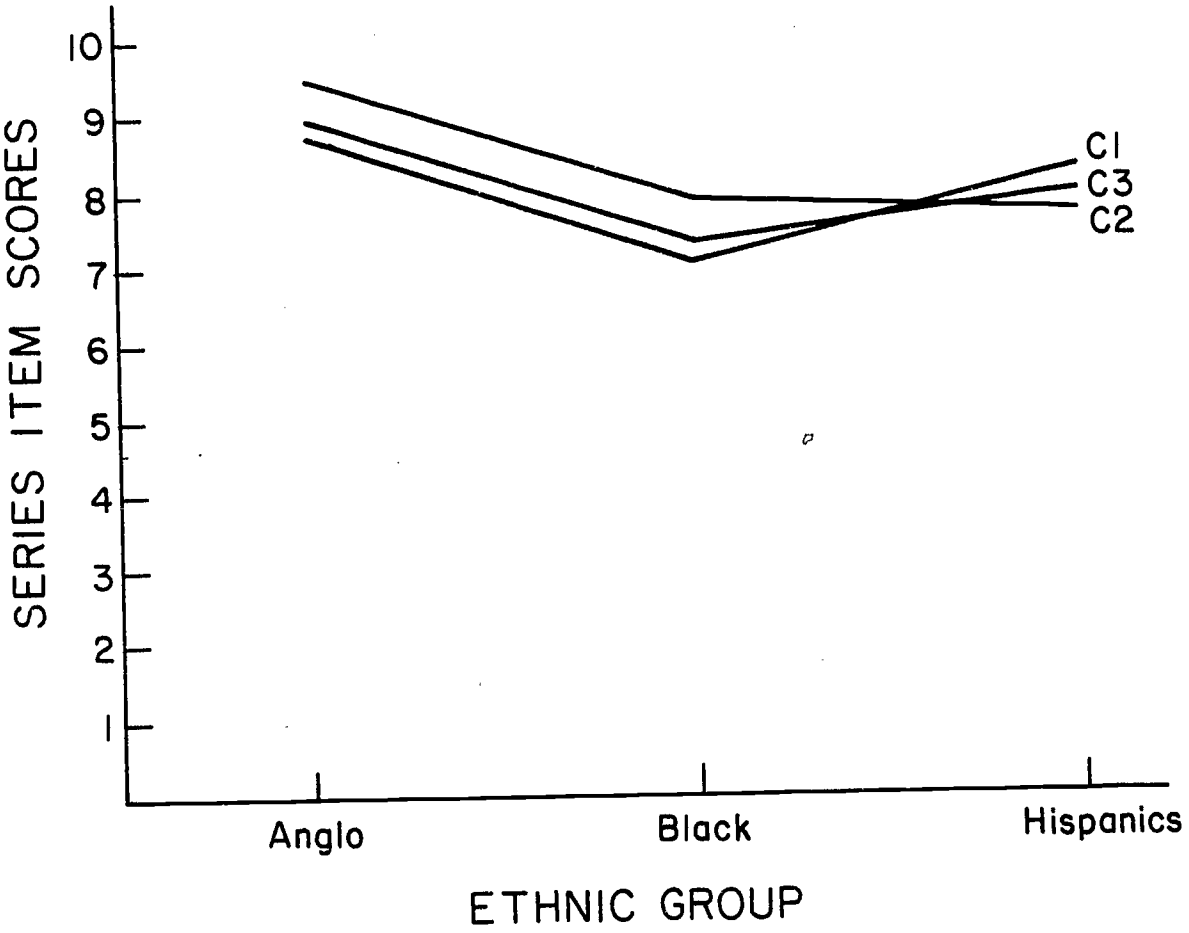


Fig. 10. Effect of Testing Condition on Cattell's CFT Series Completion Items for Three Ethnic Groups.

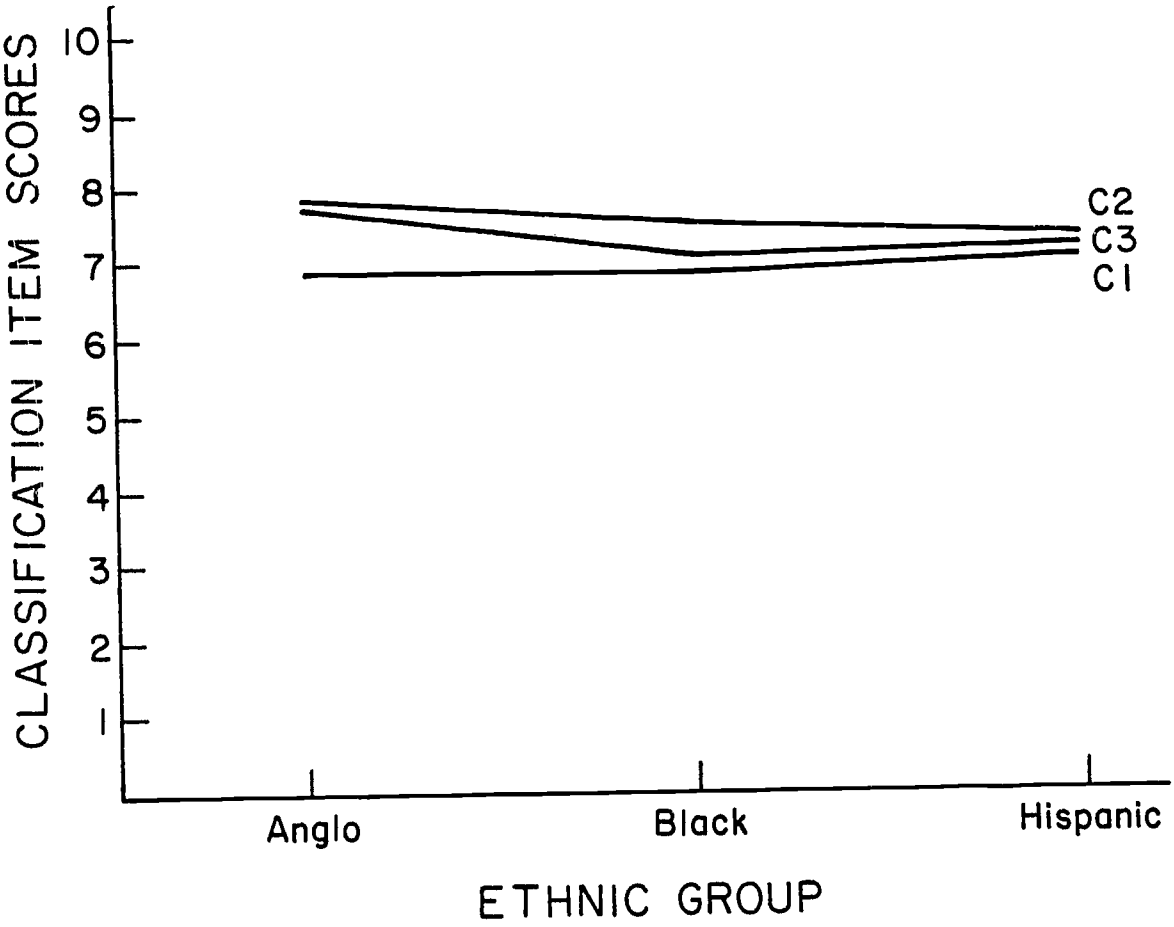


Fig. 11. Effect of Testing Condition on Cattell's CFT Classification Items for Three Ethnic Groups.

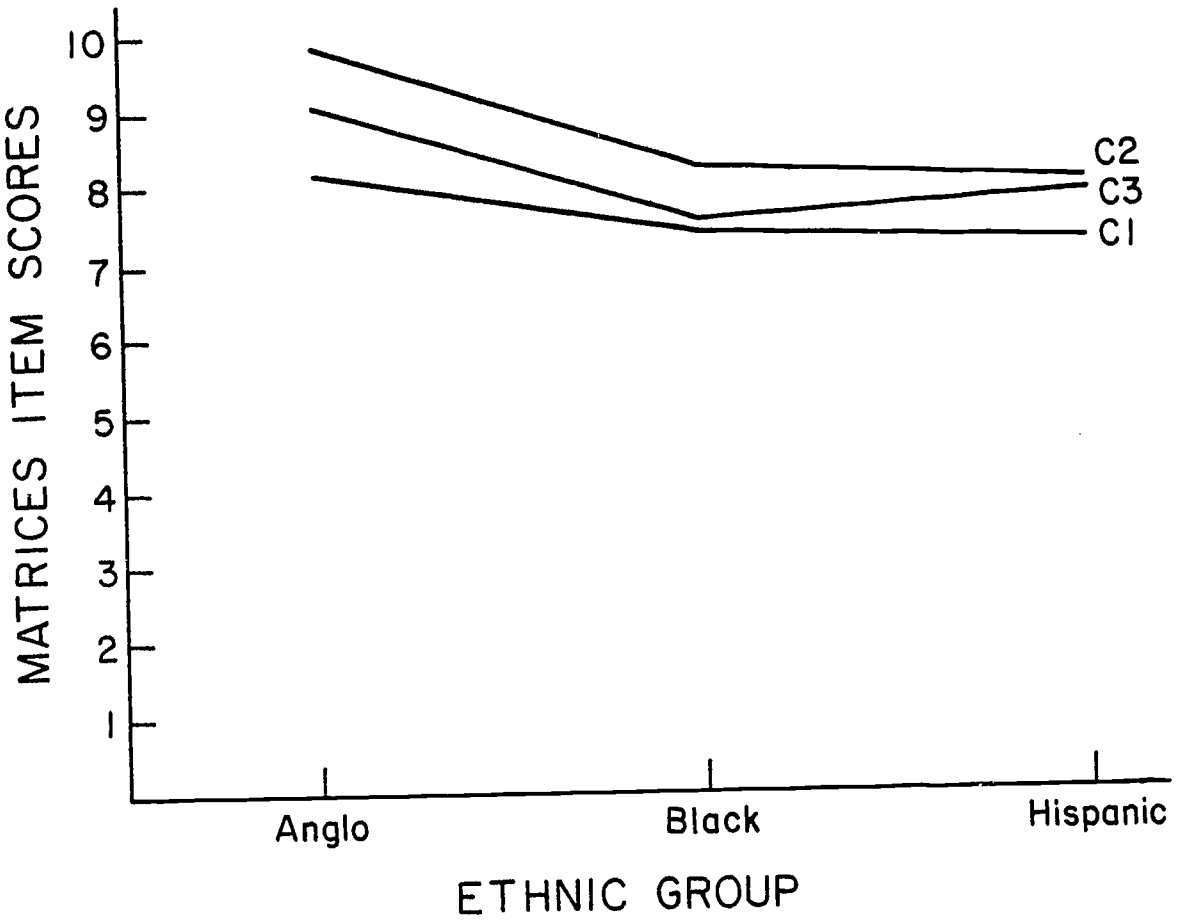


Fig. 12. Effect of Testing Condition on Cattell's CFT Matrice Items for Three Ethnic Groups.

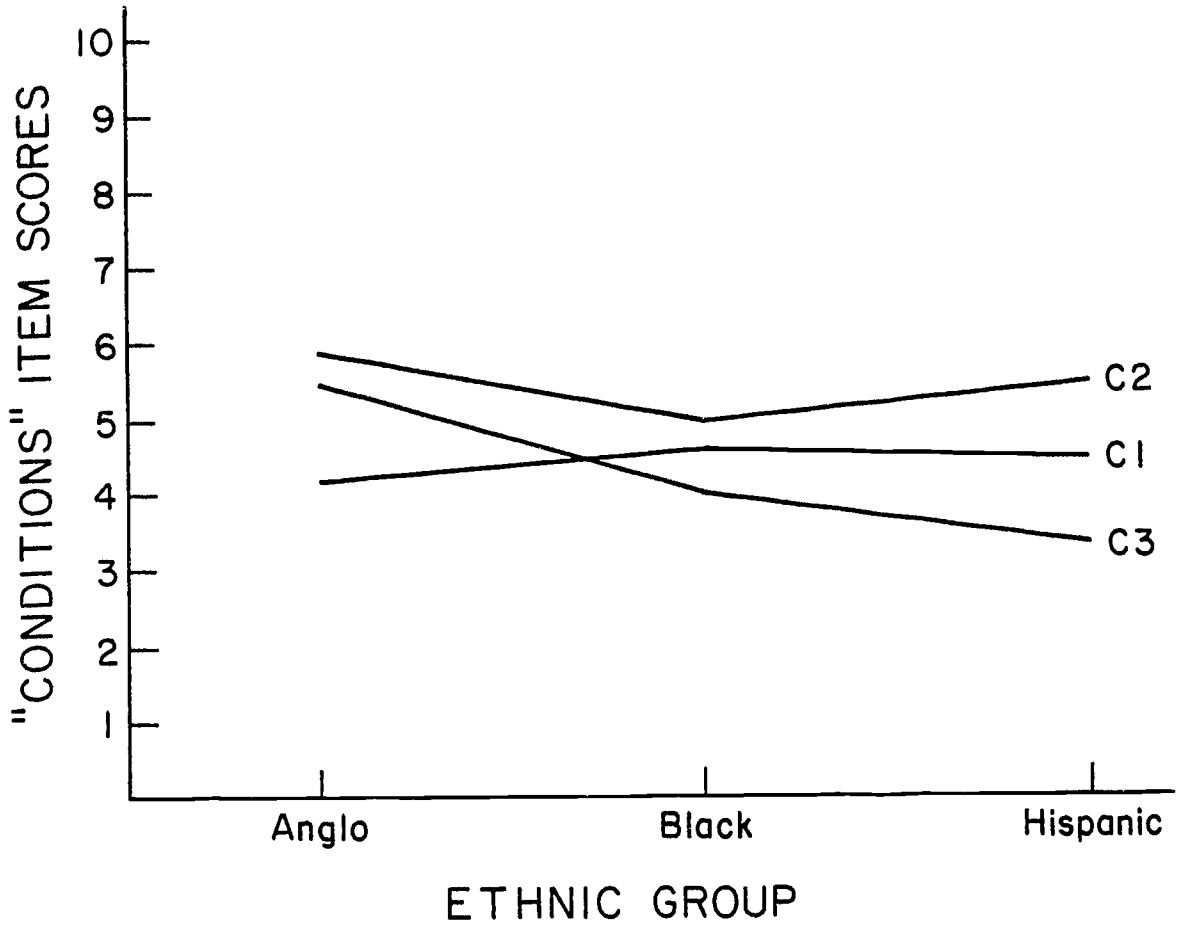


Fig. 13. Effect of Testing Conditions on Cattell's CFT "Condition" Items for Three Ethnic Groups.

3.1.2 Summary

In summary, the results in this section show significant differences in performance on Raven's Matrices and Cattell's CFT. These were attributable to test condition as well as to ethnic group classification. Analysis of the main effects revealed that, for test condition, the verbalization procedure (C_2) was the most efficaceous, leading to the highest performance levels on both Raven's Matrices and Cattell's CFT. Additional analyses indicated that the verbalization test condition led to enhanced performance specifically on the pattern completion and analogical reasoning items of the Raven's, and on the matrices and topological relations items of the Cattell.

Analysis of the main effect for ethnic group classification indicated that, regardless of test condition, Anglos performed at higher levels on both Raven's matrices and Cattell's CFT than either Blacks or Mexican/Americans. This difference held for all three item subgroups of Raven's matrices; but was true only for the series completion and matrices subtests of the Cattell.

3.2 The Relationship of "Nontarget" Variables to Performance on Raven's Matrices and Cattell's CFT Administered under Standard and Elaborative Testing Conditions

Analysis of the effects of impulsivity, motivation, and planning on Raven and Cattell performance will be presented as follows. First, potential ethnic group differences on all the "nontarget" variables will be assessed by one way analyses of variance. Second, the nontarget variables will be correlated with Raven and Cattell scores separately within each testing condition. Where no ethnic group

differences were found, for the nontarget variables, all subjects within each condition will be combined for the correlational analyses. Where significant ethnic group differences are detected, the correlational analyses will be performed separately for the ethnic groups.

The following questions will be addressed: (1) Do ethnic groups differ on any of the impulsivity, planning, or motivational variables? (2) Do any of the impulsivity, motivational, or planning measures correlate significantly with Raven matrices (or with Raven item subgroups) or Cattell's CFT (or with Cattell subtests)? (3) If some of these measures do correlate with Raven and Cattell performance, does the relationship hold for all test conditions? (4) If some of the variables are related to Raven matrices and/or Cattell CFT performance under at least one test condition, are they variables which are differentially related to ethnic group classification?

3.2.1 Ethnic Group Classification, Impulsivity, Motivation and Planning

Table 11 presents the means, standard deviations, and sample sizes on all "nontarget" variables separately for each ethnic group.

One way analyses of variance were performed for each of the nine nontarget variables. The results of these ANOVAS are sumarized in Table 12.

Inspection of Table 12 indicates a significant main effect for ethnic group on only one variable: Trail Making. Ethnic group

Table 11

Nontarget Variable Means, Standard Deviations, and Sample Sizes for Three Ethnic Groups

	Anglo			Black			Hispanic		
Impulsivity[1]	X	SD	N	X	SD	N	X	SD	N
MFFT	10.4	5.8	67	9.4	5.7	37	9.4	5.6	43
MFFE	9.3	4.0	67	10.6	4.6	37	10.9	4.3	43
Planning[2]									
TM	179	43	67	195	43	36	202	57	43
Mean VS	2.79	0.77	61	2.49	0.64	32	2.90	0.72	38
SD VS	1.67	0.78	61	1.39	0.65	32	1.77	0.76	38
Motivation[3]									
CC	15.0	3.8	67	15.6	4.5	36	15.3	4.4	43
SC	14.6	4.9	67	15.2	4.5	36	14.7	4.3	43
PC	16.2	4.8	67	16.8	4.7	36	15.9	4.3	43
GSW	14.4	4.0	67	15.6	4.4	36	15.7	4.2	43

[1]Impulsivity Variables - MFFT, Matching Familiar Figures Test Time (Seconds)
MFFE, Matching Familiar Figures Test Errors

[2]Planning Variables - TM, Trail Making Time (seconds)
Mean VS, Mean Visual Search Time (seconds)
SD VS, Standard Deviation of Visual Search Time

[3]Motivation Variables - CC, Harter's Cognitive Competence Scale
SC, Harter's Social Competence Scale
PC, Harter's Physical Competnece Scale
GSW, Harter's General Self Worth Scale

Table 12

Summary Table for Analyses of Variance on Nontarget Variables

Measure/Variable	df	MS (Ethnic Group)	MS (Error)	F
Impulsivity				
MFFT	2	17.3	32.9	< 1
MFFE	2	41.1	18.1	2.27
Planning				
TM	2	7884	2264	3.48*
Mean VS	2	1.55	0.53	2.92
SD VST	2	1.37	0.56	2.47
Motivation				
CC	2	4.83	17.38	< 1
SC	2	6.24	21.63	< 1
PC	2	6.70	21.29	< 1
GSW	2	26.96	17.18	1.57

*$p \leq .05$

differences approached statistical significance on the other two planning variables (mean and standard deviation of visual search time), and on MFF errors.

Scheffe post hoc comparisons were performed on Trail Making scores, and revealed that Hispanic subjets took significantly more time to complete the trail making task than Anglo subjects ($p \leq .05$). Hispanic subjects did not differ significantly from Black subjects; nor Black subjects from white subjects.

3.2.2. Correlation of Nontarget Variables with Raven and Cattell Performance

Within each testing condition, all ethnic groups were combined and scores on the nontarget variables (exception: Trail Making) were correlated with Raven total scores and item subgroups; and with Cattell total scores and the Cattell subtests (Tables 13 and 14, respectively). Because of the significant ethnic group difference on the Trail Making variable, separate sets of correlations were calculated between Trail Making and the measures of cognitive and perceptual functioning for (1) Hispanic subjects; (2) Anglo subjects; (3) Anglo and Black subjects; and (4) Hispanic and Black subjects. For the sake of clarity of presentation, only statistically significant correlations are reported in Tables 13 and 14.

Inspection of Table 13 reveals that MFF Time is significantly related to total Raven's performance under all three test conditions.

Table 13

Correlations between Nontarget Variables and Raven Matrices Scores as a Function of Test Condition

	Test Condition											
	C_1				C_2				C_3			
Variable	RAVTOT	PC	REAS	SEC C	RAVTOT	PC	REAS	SEC C	RAVTOT	PC	REAS	SEC C
MFFT	.24*	.35*	ns	ns	.23*	ns	ns	.35**	.24*	ns	.43**	ns
MFFE	-.42**	-.41**	-.26*	-.28*	-.30*	ns	-.24*	-.38**	-.38**	-.31*	-.39**	ns
TM (A)	ns	ns	ns	ns	ns	ns	ns	ns	ns	-.39*	ns	ns
TM (H)	ns	ns	ns	ns	ns	ns	-.41*	ns	ns	ns	ns	ns
TM (A + B)	ns	ns	ns	ns	ns	ns	ns	ns	ns	-.38*	ns	ns
TM (H + B)	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Mean VS	ns	ns	-.35**	ns	-.33*	ns	-.37**	-.37**	ns	ns	ns	ns
SD VS	ns	ns	ns	ns	ns	ns	-.45**	ns	ns	ns	ns	ns
CC	ns	.25	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
SC	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
PC	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
GSW	ns	ns	ns	-.26*	-.23*	ns	ns	-.29*	ns	ns	ns	ns

*$p \leq .05$

**$p \leq .01$

However, the tendency to respond quickly appears to be differentially related to performance on different item types depending on the method of test administration. Under standard testing conditions (C_1), "impulsive" (quick responding) children seem to do poorly on the pattern completion items, whereas this variable does not seem to be related to pattern completion performance under C_2 or C_3. Under the verbalization condition (C_2), highest performance was found on Section C items for those subjects who were less "impulsive" (slower responders). Similarly under elaborated feedback test conditions (C_3), performance on the reasoning items were best for the less "impulsive" subjects.

MFF Errors are more systematically related to performance under the three testing conditions. Errors on the MFF are moderately negatively correlated with Raven total scores for all three conditions. Errors are also significantly related to all the item subgroupings under C_1; and with all the item subgroupings except pattern completion (under C_2) and Section C items (under C_3).

The correlations of Trail Making time with Raven's matrices were computed for Anglo and Hispanic subjects separately, as well as for Anglo and Black subjects combined and Hispanic and Black subjects combined. Table 13 reveals that regardless of ethnic group, Trail Making time is not systematically related to Raven performance under any test condition. There are no significant correlations between Trail Making and Raven performance under C_1. For C_2, Trail Making is moderatley correlated ($r=-.41$) with the reasoning items for

the Hispanic subjects only. Under C_3, Trail Making is significantly (r=-.38) related to performance on the pattern compoletion items for the combined Anglo and Black samples. In both of these cases, subjects who had fast trail making times tended to have higher scores on these item groupings.

Planning, as assessed by Visual Search, appears to be related to Raven performance under the verbalization test condition. Mean VST is moderately correlated (r=-.35) with total Raven score under this condition, as well as with the reasoning (r=-.37) and Section C (r=-.37) item subgroups. The standard deviation of VST is moderately (r=-.45) related to the reasoning items here as well.

There appears to be little relationship between the motivation measures and Raven performance. Not surprisingly, the social and physical competence scales do not correlate with Raven total score or with any of the item subgroupings under any test condition. The cognitive competence scale has a single significant correlation (r=.25), that being with the pattern completion items under C_1. However, it is likely that this is a spurious correlation since it is in the wrong direction, i.e., it indicates that individuals who have low perceived cognitive competence tend to have higher scores on the pattern completion items. The "General Self Worth" scale has three significant correlations with Raven Performance. These correlations indicate that those subjects who have a realtively high sense of "self worth" tend to do better on Section C items (under test conditions C_1 (r=-.26) and C_2 (r=-.29) and also higher total Raven scores (under C_2 (r=-.23)).

In summary, the impulsivity variables (MFFT and MFFE) appear to have the strongest and most systematic relationship to Raven performance. This relationship apparently obtains regardless of test condition, although different item groupings tend to be differentially related to MFF Time due to test condition. The planning variables, especially Visual Search, seem to be related to performance under the verbalization test condition, and may relate particularly to the performance of Hispanic subjects on the reasoning items. The motivation variables show little relationship to Raven performance under any condition, except for the "General Self Worth" scale which has several significant, although relatively low magnitude, correlations with Raven performance.

Inspection of Table 14 reveals that MFF Time has only one significant correlation (r=.24) with Cattell's CFT, that being for the motives subtest under C_1. MFF Error, on the other hand, is significantly related to Cattell total score under all three test conditions, as well as to the matrices and topological relations subtests regardless of test condition. MFF Error appears to be the most sensitive predictor of either Raven or Cattell performance of the two impulsivity variables.

Trail Making is more highly related to Cattell performance than to Raven performance. Individual differences on this planning variable seem to significantly effect the performance of Anglo subjects under C_2. The elabaorative conditions seem to have a compensatory effect for those who have the poor planning ability.

Table 14

Correlations between Nontarget Variables and Cattell's CFT Scores[1] as a Function of Test Condition

	Test Condition														
	C_1					C_2					C_3				
Variable	CATTOT	S1	S2	S3	S4	CATTOT	S1	S2	S3	S4	CATTOT	S1	S2	S3	S4
MFFT	ns	ns	ns	.24*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
MFFE	-.35**	ns	ns	-.30*	-.28*	-.37**	-.32*	ns	-.24*	-.38**	-.43**	ns	ns	-.50**	-.41**
TM (A)	-.43*	ns	-.40*	-.67**	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
TM (H)	ns	ns	ns	ns	ns	-.45*	-.56*	ns	ns	-.64**	ns	ns	-.56*	ns	ns
TM (A + B)	ns	ns	-.30*	-.36*	ns	ns	ns	ns	ns	ns	-.42**	-.29*	ns	-.34*	-.48**
TM (B + H)	ns	ns	ns	ns	ns	ns	ns	ns	ns	-.51**	ns	ns	ns	ns	ns
Mean VST	-.28*	ns	-.27*	-.26*	ns	-.38**	ns	-.36**	-.35**	ns	ns	ns	ns	ns	ns
SD VST	ns	ns	ns	ns	ns	ns	ns	-.28*	-.27*	ns	ns	ns	ns	ns	-.31*
CC	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
SC	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
PC	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
GSW	ns	-.28*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns

[1]CATTOT= Cattell Total Score; S1= Series Completion Subtest; S2= Classification Subtest; S3= Matrices Subtest; S4= "conditions", or Topological Relations Subtest.

*$p \leq .05$

**$p \leq .01$

That is, the significant correlations observed between planning and Cattell performance under C_1 (r_s range from -.43 for Cattell total to -.40 and -.67 for the classification and matrices subtests, respectively) become nonsignificant under C_2 and C_3. On the other hand, Trail Making does not appear to be related to Cattell performance for Hispanic subjects under C_1, although it appears to be a factor under the verbalization condition for these subjects. Trail Making seems to be quite strongly related to the performance of Hispanic subjects on both the series completion and topological relations items under C_2. For the combined Anglo and Black sample, Trail Making seems to be primarily related to Cattell performance under the elaborated feedback condition, with significant correlation with Cattell total score (r=-.42) as well as with series completion (r=.29), matrices (r=-.34) and topological relations (r=-.48) subtests. For the Black and Hispanic combined sample, Trail Making is significantly related (r=-.51) only to the topological relations items under C_2. Thus, the relationship between the planning variable, Trail Making, and Cattell performance appears to vary as a function of both test condition and ethnic group classification and is much more sytematically related to Cattell performance than to Raven performance.

Visual Search time is related to total Cattell performance under both C_1 and C_2 (rs=-.28 and -.38, respectively). Both mean Visual Search time and the standard deviation of Visual Search time are correlated significantly with the classification (rs= -.36 and -.28,

respectively) and matrices (rs=-.35 and -.27, respectively) subtests under C_2. Mean visual search time is also significantly related to Cattell total score and to the classification (r=-.27) and matrices (r=-.26) subtests under C_1.

In sum, the planning variables, Trail Making and Visual Search, are mostly related to performance under C_2, and appear to be a factor especially affecting the performance of Hispanic subjects in this condition.

The motivation variables appear to be even less related to Cattell performance than they are to Raven performance. There is only a single significant correlation between any of the motivation scales and Cattell performance, that between the "general self worth" scale and series completion items under C_1 (r=-.28). Since only three significant and interpretable correlations between Raven performance and the Harter scales were with the "general self worth" scale, it would appear that "general self worth" is a fairly global measure of intrinsic motivation and is a better predictor of performance on cognitive ability measures than the other, more narrowly focused, scales.

3.2.3. Summary

The following summarizes the results presented, specifically with respect to the questions posed at the beginning of the section:

(1) Significant ethnic group differences were found on only one nontarget variable (Trail Making), although there were trends toward ethnic group differences on the other planning variable (Visual Search) and on MFF Errors.

(2) Of the three sets of nontarget variables, the impulsivity (MFF Time and MFF Error) and planning (Trail Making and visual search) variables show the strongest relationship to Raven and/or Cattell performance. Specifically, MFF Time is related primarily to Raven performance; MFF Error to both Raven and Cattell performance; Trail Making mostly to Cattell performance; and Visual Search to both Raven and Cattell performance.

(3) Whereas MFF Time (for Raven) and MFF Error (for Raven and Cattell) appear to related to performance regardless of test condition, Trail Making and Visual Search appear to affect performance primarily under the verbalization condition (C_2).

(4) The Trail Making variable, the only one for which significant ethnic group differences were noted, does appear to be differentially related to ethnic group performance on the Cattell CFT. Specifically, Trail Making is significantly correlated with the performance of Hispanic subjects under the verbalization test condition, but is not significantly correlated with the performance of Anglo or Black subjects under this condition. Trail Making is related to the performance of Anglo subjects under standard administration procedures, but not to the performance of Black or Hispanic subjects under this

condition. Trail Making time is related to the perfcrmance of Black subjects under the elaborated feedback condition. It is not related to the performance of Anglo subjects under C_3, and is related significantly to the performance of Hispanic subjects only on the classification subtest.

3.3. Predictive Validity of Raven's SPM and Cattell's CFT

We are primarily concerned in this section with comparisons of the predictive power of the two measures of cognitive ability administered under the different test conditions. The criterion measures used were Reading and Math standard scores on the Comprehensive Test of Basic Skills (CTBS), administered in the Spring of the 1982 academic year. Beyond comparing the predictive validity of Raven's and Cattell as a function of test condition, we were interested in finding out if predictive validity varies as a function of ethnicity and/or cognitive style or personality differences. Unfortunatley, our restricted sample size, due in part to missing data especially on the criterion measures, makes it impossible to perform specific comparisons among groups formed on the basis of test condition and ethnicity and cognitive style or personality variables. That is, our sample size is insufficient to form groups such as "Anglo, impulsive, C_2 or "Black, impulsive, C_2", and perform correlational analyses separately for these groups.

Our approach will be as follows: (1) To estimate and compare the predictive validity of the Raven and Cattell as a function of test condition per se, correlations between the predictor and criterion measures will be made for all subjects combined within each test condition. (2) To estimate the predictive validity of Raven and Cattell as a function of test condition, impulsivity and planning.

Since the analyses performed in the previous section indicated that there is little relationship between the Harter Perceived Competence Scales and either the Raven or Cattell administered under any condition, no comparisons of the predictive validity of these measures were made based on high and low Harter scale groupings.

Correlations were computed between Raven and Cattell scores and the criterion measures separately for each of the above groupings.

The above two stage analysis allows us to address the following questions:

(1) Regardless of ethnic group, cognitive style, or personality differences, does the predictive validity of the Raven and/or Cattell vary as a function of method of test administration?

(2) Is the predictive validity of Raven and/or Cattell different for "impulsive" children (as measured by MFF Time, MFF Error) under different methods of test administration? Is the predictive validity of Raven and/or Cattell different for "poor" vs. "good" planners (as measured by Trail Making and Visual Search) under different methods of test administration?

3.3.1 Predictive Validity as a Function of Method of Test Administration

Table 15 presents the Pearson product-moment correlation coefficients and their squared values between the predictor measures (Raven, Cattell) and the criterion measures (Reading and Math standard scores) separately for each test condition.

Inspection of Table 15 reveals that the correlation of both predictors with both criterion measures is statistically significant for all testing conditions. Within each test condition, Raven and Cattell predict the criterion measures approximately equally well. Between test conditions, however, it can be seen that the predictive power of both the Raven and Cattell appears to be slightly under the elaborative testing conditions than under standard administsration procedures. In terms of percent variance accounted for in the criterion measures by the predictor measures, 10% of the variance between Raven's Matrices and CTBS Reading is shared under C_1. Under C_2, however, this jumps to 23 percent. Similarly, the shared variance between the CFT and Reading is approximately twice greater under C_3 than under C_1. A similar pattern obtains for the prediction of Math scores. The shared variance between the Raven and Math is approximately three times greater under C_2 than under C_1. The CTF predicts Math achievement twice as well under C_2 than under C_1.

Table 15

Pearson Product-Moment Correlations between Predictor Measures of Cognitive Ability and Criterion Measrues of School Achievement

Test Condition and Predictor	Criterion Measures					
	Reading			Math		
	r	r^2 (%)	N	r	r^2 (%)	N
C_1						
Raven	.33*	10.9	40	.33*	10.9	40
Cattell	.35*	12.3	40	.35*	12.3	40
C_2						
Raven	.48**	23.0	45	.54**	29.2	44
Cattell	.42**	17.6	44	.50**	25.0	44
C_3						
Raven	.41**	16.8	44	.44**	19.4	45
Cattell	.50**	25.0	43	.46**	21.2	43

*$p \leq .05$

**$p \leq .01$

3.3.2 Predictive Validity as a Function of Method of Test Administration, Impulsivity, and Planning Ability

Within each test condition, subjects were divided into two groups (above and below the median score) on the two impulsivity variables (MFF Time, MFF Error) and on the three planning variables (Trail Making, and mean and standard deviation of visual search time). Validity coefficients were then computed between the precitor and criterion measures for each of these groups. The validity coefficients for the standard test condition "impulsivity" groups are presented in Table 16.

Inspection of this table shows that for the MFF Time, the validity coefficients tend to be lower for the "impulsive" subjects, i.e., those with below median response times, than for children with relatively longer MFF Times. Both Raven and Cattell predict Math scores better for the above median MFF Error group than for the below median error gorup. It appears that under standard administration procedures (C_1), Raven and Cattell predict the achievement measures best for children who take relatively long to respond and for children who tend to make a lot of errors on the MFF. The predictive validity of Raven and Cattell seems to be poorest for relative fast responders, or "impulsive" children as measured by MFF Time.

Table 17 shows the validity coefficients for the above and below median groupings on the Planning variables for standard administration. Inspection of this table indicates that the validity coefficients are generally higher for those subjects with relatively

Table 16

Pearson Product-Moment Correlatons between Cognitive Predictor Measures and School Achievement Measures for Subject Groupings on Impulsivity Variables - Standard Test Condition

Subject Group and Predictor	Criterion Measures: Reading r	Reading r^2 (%)	Reading N	Math r	Math r^2 (%)	Math N
Above Med MFF Time						
Raven	.35	12.3	21	.38*	14.4	21
Cattell	.49*	24.0	21	.40*	16.0	21
Below Med MFF Time						
Raven	.31	9.6	19	.26	6.8	19
Cattell	.22	4.8	19	.27	7.3	19
Above Med MFF Error						
Raven	.36	13.0	20	.56**	32.4	20
Cattell	.32	10.2	20	.43*	18.5	20
Below Med MFF Error						
Raven	.29	8.4	20	.15	2.3	20
Cattell	.45*	20.3	20	.34	11.6	20

*$p \leq .05$

**$p \leq .01$

Table 17

Pearson Product-Moment Correlations between Cognitive Predictor Measures and School Achievement Measures for Subject Groupings on Planning Variables - Standard Test Condition

Test Condition and Predictor	Criterion Measures: Reading r	Reading r^2 (%)	Reading N	Math r	Math r^2 (%)	Math N
Above Med TM[1]						
Raven	.29	8.4	20	.19	3.6	20
Cattell	.29	8.4	20	.36	13.0	20
Below Med TM						
Raven	.37	13.7	20	.48*	23.0	20
Cattell	.42*	17.6	20	.33	10.9	20
Above Med $\bar{X}$ VST[2]						
Raven	.34	11.6	20	.13	1.7	20
Cattell	.32	10.2	20	.05	0	20
Below Med $\bar{X}$ VST						
Raven	.28	7.8	20	.45*	20.2	20
Cattell	.31	9.6	20	.48*	23.0	20
Above Med SD VST[3]						
Raven	.33	10.9	19	.17	2.9	19
Cattell	.40*	16.0	19	.18	3.2	19
Below Med SD VST						
Raven	.30	9.0	21	.49*	24.0	21
Cattell	.28	7.8	21	.47*	22.1	21

*$p \leq .05$

[1] TM, Trail Making

[2] $\bar{X}$ VST, Mean Visual Search Time

[3] SD VST, Standard Deviation of Visual Search Time

fast Trail Making times as opposed to subjects with relatively slow times on this variable. There is little difference in the predictive validity of Raven and Cattel for the mean Visual Search groupings with respect to Reading. However, both Raven and Cattell seem to predict Math achievement better for children with relatively fast visual search times as opposed to children with relatively slow visual search times. Very similar findings are noted for the above and below median groupings on the standard deviation of visual search time.

To briefly summarize: Under standard administration procedures, there appears to be a slight bias in the predictive power of Raven and/or Cattell against fast ("impulsive") responders, as measured by MFF Time, and against "poor planners," as measured by all three planning variables.

Table 18 shows the validity coefficients for the MFF groupings of subjects who were administered Raven and Cattell under the verbalization testing condition (C_2).

Inspection of Table 18 reveals that the predictive validity of Raven and Cattell is much higher for "impulsive" children as measured by MFF Time than it is for this group under standard administration procedures (C_1). The validity coefficients are also slightly higher, on the average, for the above median MFF Time group than they are for this group under C_1. A similar pattern of coefficients will be noted for the above median MFF Error group here as was noted for this group under C_1. That is, both Raven and Cattell tend to

Table 18

Pearson Product-Moment Correlations between Cognitive Predictor Measures and School Achievement Measures for Subject Groupings on Impulsivity Variables - Verbalization Test Condition

Subject Group and Predictor	Criterion Measures: Reading r	Reading r^2 (%)	Reading N	Math r	Math r^2 (%)	Math N
Above Med MFF Time						
Raven	.57**	32.5	21	.59**	34.8	20
Cattell	.44*	19.4	20	.57**	32.5	20
Below Med MFF Time						
Raven	.42*	17.6	24	.52**	27.0	24
Cattell	.37*	13.7	24	.46*	21.2	24
Above Med MFF Error						
Raven	.22	4.8	20	.51**	26.0	20
Cattell	.20	4.0	20	.45*	20.3	20
Below Med MFF Error						
Raven	.67**	44.9	25	.48**	23.0	24
Cattell	.42*	17.6	24	.41*	16.8	24

*$p \leq .05$

**$p \leq .01$

predict Math achievement well for this group, but not reading achievement. However, for the below median MFF error group, the validity coefficients are much higher for Math achievement and (for Raven) reading achievement than they are for this group under C_1.

Table 19 presents the validity coefficients for the Planning variable groupings of subjects administered Raven and Cattell under C_2. Compared to these groupings under C_1, the validity coefficients are, on the average, much higher for all groups. Especially noteworthy are the higher coefficients reported for the above median Trail Making group (particularly for math achievement), and for the above median mean and SD Visual Search groups (again, particularly for math achievement).

Comparisons of the predictive validity of Raven and Cattell administered under standard and verbalization testing conditions as a function of "impulsivity" and "planning" leads to the following general conclusions: (1) Predictive validity is higher under verbalization than under standard testing procedures, and (2) verbalization appears to reduce or eliminate the predictive bias demonstrated under standard testing procedures against "impulsive" children (as measured by MFF Time) and "poor planners," as measured by the Trail Making and Visual Search Tasks. The improvement in prediction is especially great for math achievement.

Table 20 presents the validity coefficients for the impulsivity variable groupings for subjects in the elaborated feedback condition (C_3).

Table 19

Pearson Product-Moment Correlations between Cognitive Predictor Measures and School Achievement Measures for Subject Groupings on Planning Variables - Verbalization Test Condition

Subject Group and Predictor	Criterion Measures: Reading r	Reading r^2 (%)	Reading N	Math r	Math r^2 (%)	Math N
Above Med TM						
Raven	.37*	13.7	21	.60**	36.0	21
Cattell	.22	4.8	21	.40*	16.0	21
Below Med TM						
Raven	.57**	32.5	24	.53**	28.1	23
Cattell	.53**	28.1	23	.54**	29.2	23
Above Med $\bar{X}$ VST						
Raven	.33	10.9	23	.47*	22.1	22
Cattell	.28	7.8	22	.42*	17.6	22
Below Med $\bar{X}$ VST						
Raven	.61**	37.2	22	.65**	42.3	22
Cattell	.52	27.0	22	.56	31.4	22
Above Med SD VST						
Raven	.48*	23.0	23	.63**	39.7	22
Cattell	.30	9.0	22	.45*	20.3	22
Below Med SD VST						
Raven	.46*	21.2	22	.42*	17.6	22
Cattell	.56**	31.4	22	.54	29.2	22

*$p \leq .05$

**$p \leq .01$

Table 20

Pearson Product-Moment Correlations between Cognitive Predictor Measures and School Achievement Measures for Subject Groupings on Impulsivity Variables - Elaborated Feedback Test Condition

Subject Group and Predictor	Criterion Measures: Reading r	Reading r^2 (%)	Reading N	Math r	Math r^2 (%)	Math N
Above Med MFF Time						
Raven	.37	13.7	20	.61**	37.2	20
Cattell	.54**	29.2	20	.53**	28.1	20
Below Med MFF Time						
Raven	.35*	12.3	25	.09	1.0	23
Cattell	.35*	12.3	25	.20	4.0	23
Above Median MFF Error						
Raven	.40*	16.0	21	.07	0.5	20
Cattell	.36	13.0	21	.22	4.8	20
Below Median MFF Error						
Raven	.31	9.6	24	.57**	32.5	23
Cattell	.51**	26.0	24	.54**	29.2	23

*$p \leq .05$

**$p \leq .01$

Inspection of the table reveals that, as for the subjects in C_1, the validity coefficients for relatively fast responders as measured by MFF Time ("impulsives") are much lower than for relatively slow responders. This is especially true with respect to the prediction of math achievement. Prediction of reading achievement is slightly greater for the relatively slow responding group as well. Unlike the verbalization test condition, elaborated feedback does not seem to compensate for "impulsiveness" as measured by MFF time.

Prediction is greater for the below median MFF error group than for the above medican MFF error group. Again, this obtains particularly for math achievement. Whereas Raven and Cattell under C_1 predict math achievement better for subjects who make relatively many errors on the MFF, under C_3 they predict math achievement best for subjects who tend to make relatively few errors on the MFF. As opposed to the verbalization condition, in which math achievement was predicted approximately equally as well for both high and low error , the standard and elaborated feedback conditions over- or under-predicts for one or the other of these groups.

Table 21 presents the validity coefficients for the planning variable groupings for subjects in the elaborated feedback condition.

The coefficients presented in this table are similar to those presented in Table 19 for the verbalization condition subjects. Prediction is greater for subjects hwo have relatively long Trail Making times ("poor planners") than it is for this group under standard testing procedures, and is higher for the below median

Table 21

Pearson Product-Moment Correlations between Cognitive Predictor Measures and School Achievement Measures for Subject Groupings on Planning Variables - Elaborated Feedback Test Condition

Subject Group and Predictor	Criterion Measures: Reading r	Reading r^2 (%)	Reading n	Math r	Math r^2 (%)	Math n
Above Med TM						
Raven	.33	10.9	24	.43*	18.5	23
Cattell	.44*	19.4	24	.36*	13.0	23
Below Med TM						
Raven	.50**	25.0	21	.46*	21.2	20
Cattell	.54**	29.2	21	.55**	30.3	20
Above Med $\bar{X}$ VST						
Raven	.32	10.2	22	.48*	23.0	20
Cattell	.58**	33.6	22	.54**	29.2	20
Below Med $\bar{X}$ VST						
Raven	.60**	36.0	23	.48*	23.0	23
Cattell	.55*	30.3	23	.49**	24.0	23
Above Med SD VST						
Raven	.30	9.0	22	.57**	32.5	21
Cattell	.58**	33.6	22	.63**	40.0	21
Below Med SD VST						
Raven	.57**	32.5	23	.40*	16.0	22
Cattell	.56**	31.4	23	.50	25.0	22

*$p \leq .05$

**$p \leq .01$

trail making time group ("good planners") compared to this group under C_1 as well.

As for the above median Visual Search time group in the verbalization condition, prediction of math achievement is considerably higher than under C_1. Similarly, prediction of both reading and math achievement is higher, on the average, for the below median visual search time and both SD Visual Search time groupings than it is for these groups under standard administration procedures.

Many of the same conclusions can be drawn concerning the predictive validity of Raven and Cattell as a function of planning ability when these measures are administered under elaborated feedback conditions as when they are administered under the verbalization condition.

3.3.3 Summary

The results of the predictive validity analyses presented here indicate that elaborative testing procedures tend to increase the power of measures of cognitive ability to predict school achievement. In addition, both elaborative conditions appear to compensate for the "bias" in prediction noted under standard administration procedures for specific groups of subjects, i.e., for "poor planners" as measured by the Trail Making and Visual Search Tasks. Only the verbalization test condition appears to compensate for "impulsiveness" as measured by MFF Time, however.

Chapter IV

Summary

4.0 Organization

The rather extensive results presented in the previous chapter will not be reiterated in this section. Rather, the approach will be to highlight and discuss the findings as they relate to the specific questions which guided the research.

4.1 Research Question 1

The initial research question and its corollary was:

> Does the dynamic assessment approach yield information which is a more appropriate indicator of cognitive competence than that obtained by traditional, standard approaches based on static test theory?
>
> Corollary: Do specific testing conditions differentially affect the performance of children of different social and/or ethnic backgrounds?

The results suggest a positive response to the initial question and a negative response to its corollary. The initial expectation was that differences in performance on both the Raven and Cattell tests would be detected between the three ethnic groups under the standard procedure of administering these tests. A further expectation was that under dynamic testing procedures all groups would improve, with lower Black and Hispanic performance improving to the extent that the differences found under standard testing procedures would be reduced. That is, the dynamic assessment apporoach would have a compensating effect. These expectations were only partially fulfilled.

The testing condition which seemed to be most effective in increasing performance of all groups of children was the condition involving verbalization. Under this condition significant improvements on both the Raven Matrices and the Cattell Culture Fair Test were noted when compared with performance under the standard testing procedures. No testing condition by ethnic classification interaction was detected for either the Raven or Cattell. Thus, verbalization was effective in improving performance regardless of ethnic classification. On the other hand, if one compares performance of the Black group with that of the Anglo group on the Raven as well as on the Cattell, it will be noted that Black performance under verbalization is equivalent to Anglo performance under standard testing procedures. The same obtains for the Hispanic group for the Cattell but not for the Raven. One could conclude that a type of compensatory effect for initially observed differences between these groups was brought about by subject verbalization. Using different modes or test administration for some groups as opposed to others in order to bring about this compensation may not be a justified practice, however.

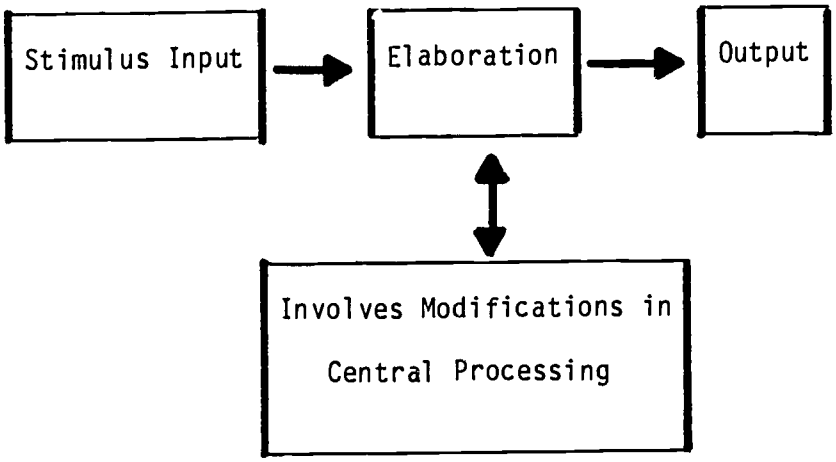
Analysis of the effects of the testing conditions on subgroups of Raven items indicates that performance on the relatively difficult reasoning items was most affected by verbalization. These items require analogical reasoning and cannot be solved by perceptual processes. Performance of all ethnic groups increased on these items and no ethnic classification x testing condition interaction was found. Nonetheless, the type of compensatory effect noted for the

total scores mentioned above was found. Under verbalization, the performance of the Black and Hispanic children reached the performance level found for Anglo children under standard testing conditions.

Analyses involving the Cattell show that verbalization was effective for only certain subgroupings of the test: matrices and conditions. For the former, verbalization performance was higher than performance under standard conditions whereas for the latter it was higher than either standard or elaborated feedback conditions. These results are consistent with those obtained with the Raven matrices as the dependent variable.

The question of why verbalization effectively improves performance over that obtained in either standard testing or in the condition involving elaborated feedback is important. It should be remembered that in the verbalization condition, the subject was not given any feedback concerning the correctness of his/her response. The only difference between verbalization and the procedure followed under the standard testing condition was that the subject was asked to overtly describe the task at hand and his/her thinking processes as the task was solved. Overt relevant verbalization apparently brings about modifications in the central processing of information.

This can be depicted as follows:



Modifications in central processing could be brought about by several factors operating independently or conjointly. These include planfulness, exactness, self-regulation, flexibility, and sharpening of responses. In addition, anxiety could be reduced and position orientation to the test situation increased. Abundant evidence for this has been cited elsewhere (see Carlson & Wiedl, 1980; Carlson, Bethge, & Wiedl, 1981; Dillon, 1981).

4.2 Research Question 2

The second research question was:

> Why and for whom is the dynamic assessment approach effective? Or, what are the sources of intraindividual variation, and how may they be accounted for by personality and cognitive style variables?

The results did not yield clear answers to these questions. Previous research led us to the expectation that the efficacious dynamic assessment approach(es) would be compensatory for individuals whose performance was poor on "nontarget" variables such as impulsivity, planning or perceived competence. More specifically, it was expected

that gains in performance could be at least partially attributable to the reduced negati e effect of the nontarget variable in the efficacious testing condition(s). In addition, it was expected that differences between the ethnic groups in the selected personality/cognitive style variables would be demonstrated and help explain differences in cognitive performance as assessed under standard testing conditions.

General comparisons between ethnic groups on the Matching Familiar Figures Task indicated that Anglos tended (the difference only approached statistical significance) to be more accurate than either the Black or Hispanic children. No significant differences in time taken to solution were found. In planning, Anglos tended to be slightly "better planners" than did Blacks or Hispanics. No significant differences or trends towards differences were found on any of the perceived competence scales, allowing the conclusion that regardless of ethnicity the children's perceptions of their cognitive, social and physical skills were the same.

The most consistent correlations with performance on either the Cattell or Raven measures was for errors made on the Matching Familiar Figures Test. Here, however, the relationships tended to be consistent regardless of testing condition. Hence no interaction between this varible and performance under differing testing conditions was detected. Apparently, at least for the subjects involved in this study, those modifications in central processing which were apparently brought about by the verbalization condition are not systematically related to any of the cognitive style/personalities variables assessed.

4.3 Research Question 3

The third research question guiding the study was:

> For what reason should dynamic assessment procedures be employed?

This question is broad. It can be related to specific issues of predictive validity, indications of how cognitive and perceptual abilities should be assessed, or indications of approaches which might be useful in certain instructional settings. Although the details of the results concerning predictive validity are given in sections 3.3.1, 3.3.2, and 3.3.3 of the previous chapter, further comment is appropriate.

The systematic effect of subject verbalization of increasing performance on both the Raven and Cattell implies that modifications in central processing of information were brought about. Levels of performance were reached which more closely approximate cognitive competence than estimated by either standard assessment or the approach which involved extensive feedback. Generally speaking, the predictive validities of the Raven and Cattell are moderate. Confirmation of this was gained through the correlations of these measures with the mathematics and reading achievement scores. Correlations between both the Raven and Cattell and the achievement measures increased when the former tests were administered under the verbalization condition. Apparently the gains in predictive validity involve certain aspects of individual differences in "impulsivity" and planning. Neither the Raven nor Cattell have better than marginal predictive validity for

"impulsive" responders or poor planners when the tests are administered in the traditional manner. When verbalization is employed, however, the predictive validities for these individuals is increased. Thus, the gains in overall predictive validity noted under verbalization are due largely to the improved prediction gained for those persons classified as "impulsive" and poor planners; characteristics, by the way, for which differences between the ethnic groups was found.

These results suggest that overt verbalization leads not only to more accurate assessment of cognitive and perceptual functioning but that the approach also yields better prediction of school achievement. What would appear to be a bias toward under prediction for poor planners and "impulsives" (again, not unrelated to the ethnic classification of the individual) is compensated for by a testing approach such as that used in verbalization. A further implication of the results is that employing active, overt verbalization in the cognitive school tasks may lead to increased achievements and compensate for certain non-cognitive individual difference factors.

References

Ault, R. L., Mitchell, L., & Hartman, D. P. Some methodological problems in reflection impulsivity research. Child Development, 1976, 47, 227-231.

Banks, C., & Sinha, U. An item analysis of the Progressive Matrices Test. British Journal of Psychology, Statistical Section, 1951, 4, 91-94.

Barratt, E. S. The relationship of the Progressive Matrices (1938) and the Columbia Mental Maturity Scale. Journal of Consulting Psychology, 1956, 20, 294-296.

Birkmeyer, F. The relationship between the Coloured Progressive Matrices and individual intelligence tests. Psychology in the Schools, 1964, 1, 309-312.

Bronfenbrenner, U. The Ecology of Human Development. Cambridge: Harvard University Press, 1979.

Brown, A., & French, L. The zone of potential development: implications for intelligence testing in the year 2000. Intelligence, 1979, 3, 255-273.

Burke, H., & Bingham, W. Raven's Progressive Matrices: more on construct validity. Journal of Psychology, 1969, 72, 247-251.

Carlson, J. S., & Dillon, R. The effects of testing-the-limits procedures on Raven Matrices performance of deaf children. Volta Review, 1978, 4, 216-224.

Carlson, J., & Dillon, R. Effects of testing conditions on Piaget matrices and order of appearance problems: a study of competence versus performance. Journal of Educational Measurement, 1979, 16, 19-26.

Carlson, J. & Jensen, C. The factorial structure of the Raven Coloured Progressive Matrices: A reanalysis. Journal of Educational and Psychological Measurement, 1980, in press.

Carlson, J. & Jensen, C. Reliability of the Coloured Progressive Matrices: age and ethnic group comparisons. Journal of Consulting and Clinical Psychology, 1981, in press.

Carlson, J. S., & Wiedl, K. H. Modes of presentation of the Raven Coloured Progressive Matrices Test: towards a differential testing approach. Trierer Psychologische Berichte, 1976, 3, Nr. 7.

Carlson, J. S., & Wiedl, K. H. The use of Testing-the-Limits procedures in the assessment of intellectual capabilities in children with learning difficulties. American Journal of Mental Deficiency, 1978, 82, 559-564.

Carlson, J. S., & Wiedl, K. H. Towards a differential testing approach: Testing-the-Limits employing the Raven Matrices. Intelligence, 1979, 4, 323-344.

Carlson, J. S., & Wiedl, K. H. Applications of a dynamic testing approach in intelligence assessment: Empirical results and theoretical considerations. Zeitschrift fur Individuelle und Differentielle Psychologie, 1980, in press.

Carroll, J., & Maxwell, S. Individual differences in cognitive abilities. In. M. Rosenzweig & L. Porter (Eds.), Annual Review of Psychology. Palo Alto: Annual Reviews, Inc., 1979, pp. 603-640.

Cleary, T. A. Test bias: prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Cleary, T. A. et al. Educational uses of tests with disadvantaged students. American Psychologist, 1975, 30, 15-41.

Cronbach, L. & Furby, L. How should we measure "change"--or should we? Psychological Bulletin, 1970, 74, 68-80.

Darlington, R. Another look at "cultural fairness." Journal of Educational Measurement, 1971, 8, 71-82.

Das, J. P. Planning: theoretical considerations and empirical evidence. Psychological Research, 1980, 41, 141-151.

Detterman, D. A job half done: the road to intelligence testing in the year 2000. Intelligence, 1979, 3, 295-306.

Dillon, R. & Carlson, J. Testing or competence in three ethnic groups. Educational & Psychological Measurement, 1978, 38, 436-443.

Eichhorn, H., & Bass, A. Methodological considerations relevant to discrimination in employment testing. Psychological Bulletin, 1971, 75, 261-269.

Feuerstein, R., et al. The Dynamic Assesment of Retarded Performers. Baltimore: University Park Press, 1979.

Flaughter, R. L. The many definitions of test bias. American Psychologist, 1978, 33, 671-679.

Guthke, J. Zur Diagnostik der Intellektuellen Lernfahigkeit. Stuttgart: Klett, 1977.

Harter, S. Effectance motivation reconsidered: toward a developmental model. Human Development, 1978, 21, 34-64.

Harter, S. Progress Report HD-09613-03. National Institute of Child Health and Human Development, U. S. Public Health Service, 1979a.

Harter, S. Intrinsic versus extrinsic motivation in the classroom: a new scale. Paper presented at the Society for Research in Child Development, San Francisco, Spring, 1979b.

Haywood, H. C. Alternatives to normative assessment. In P. Mittler (Ed.), Research to Practice in Mental Retardation. Baltimore: University Park Press, in press.

Heller, K. Intelligenz und Begabung. Munchen: Reinhardt, 1976.

Hewitt, F. Education of Exceptional Learners. Boston: Allyn and Bacon, 1977.

Kagan, J., Moss, H. A., & Sigel, I. E. The psychological significance of styles of conceptualization. In J. F. Wright & J. Kagan, Basic Cognitive Processes in Children. Monograph of the Society for Research in Child Development, 1963, 28, 73-112.

Luria, A. R. The Working Brain. Hammondsworth, Great Britain: Penguin, 1973.

Mercer, J. R., & Lewis, J. P. System of multicultural pluralistic assessment: technical manual. Unpublished manuscript, University of California, Riverside, 1975.

National Advisory Council on the Education of Disadvantaged Children. Title I: Expanding Educational Growth. Annual Report to the President and the Congress, 1975. Washington D.C.: U. S. Government Printing Office, 1974.

Pelzer, I. Analysis of three problem solving styles of EMR, LD and normal children in relationship to teacher evaluation. Technical Report of Grant No. GOO-76044-33. Washington, D.C.: Department of Health, Education and Welfare, April 1979.

Raven, J. C. The Coloured Progressive Matrices. London: Lewis, 1976.

Readance, J. E., & Bean, T. W. Modification of impulsive cognitive style: a survey of the literature. Psychological Reports, 1978, 43, 327-337.

Resnick, L. The future of IQ testing in education. Intelligence, 1979, 3, 241-253.

Schmidt, L. Testing the limits im Leistungsverhalten: Moglichkeit und Grenzen. In Duhm, E. (Ed.), Praxis der klinischen Psychologie. Band II. Gottingen: Hogrefe, 1971, pp. 9-29.

Wagner, I. Aufmerksamkeitstraining mit impulsiven Kindern. Stuttgart: Klett, 1976.

Wiedl, K. H. Kompensatorische Interventionen im Rahmen intelligenzdiagnostischer Untersuchunger bei kognitiv impulsiven Kindern. Klinische Psychologie: Forschung und Praxis, 1979, in press.

Wiedl, K. H., & Bethge, W. Intelligenz und Blickbenegungsurhalten fur impulsiver und reflektiver Kinder. Unpublished manuscript, 1980.

Wiedl, K. H., & Carlson, J. S. The factorial strucutre of the Raven Coloured Progressive Matrices Test. Educational and Psychological Measurement, 1976, 36, 409-413.

Wiedl, K. H., & Carlson, J. S. Testdiagnostische Vorgehen und Testleistung beimfarbigen Matrizentest von Raven: ein Beitrag zum "Testing-the-Limits" im Leistungsverhalten. In W. H. Tack (Ed.), Bericht uber den 30. Kongress der Deutschen Gesellschaft fur Psychologie. Gottingen: Hogrefe, 1977.

Wiedl, K. H., & Carlson, J. S. Dynamisches Testen bei Lernbehinderten mit dem Farbigen Matrizentest von Raven. Heipadagogische Forschung, in press.

Wiedl, K. H., & Herrig, D. Okologische Validitat und Schulerfolgsprongnose im Lernund Intelligenztest: eine exemplarische Studie. Diagnostica, 1978, 24, 175-186.