
In order to gather data bearing on the effects of individual differences of the "invariant stages" assumption of cognitive-developmental theory, 270 black, Chinese, and white subjects from 5th, 7th, and ninth grades were administered a new, 25-item, spatial reasoning task. It was predicted that age, ethnic group, and sex would significantly influence mean levels of achievement, but that analyses of patterns of item performance would reveal a similar sequence of concept and skill acquisition regardless of ethnic group means. These predictions were called the "fixed-sequence hypothesis". Age and ethnic group were significant influences on performance; sex was not. Scalogram analyses showed that each ethnic group's performance tended to form a scalable item set, but for a somewhat different ordering of the 25 items. Thus, the results were inconsistent with cognitive-developmental theory. It was concluded that individual differences are likely to affect development when sequences are highly task specific, but that the bulk of the evidence still supports a general sequence of stages in cognitive development. Implications for developmental theory, intelligence testing, and curriculum planning are discussed.

(Author/DP)
THE FIXED-SEQUENCE HYPOTHESIS: INDIVIDUAL DIFFERENCES IN THE DEVELOPMENT OF SCHOOL RELATED SPATIAL REASONING

David H. Feldmar
University of Minnesota

Research, Development and Demonstration Center in Education of Handicapped Children
University of Minnesota
Minneapolis, Minnesota

March 1970

The research reported herein was pursuant to a grant from the Bureau of Education for the Handicapped, U. S. Office of Education, Department of Health, Education, and Welfare to the Center for Research and Development in Education of Handicapped Children, Department of Special Education, University of Minnesota. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official position of the Bureau of Education for the Handicapped.

Department of Health, Education and Welfare
U. S. Office of Education
Bureau of Education for the Handicapped
The University of Minnesota Research Development and Demonstration Center in Education of Handicapped Children has been established to concentrate on intervention strategies and materials which develop and improve language and communication skills in young handicapped children.

The long term objective of the Center is to improve the language and communication abilities of handicapped children by means of identification of linguistically and potentially linguistically handicapped children, development and evaluation of intervention strategies with young handicapped children and dissemination of findings and products of benefit to young handicapped children.
Abstract

In order to gather data bearing on the effects of individual differences on the 'invariant stages' assumption of cognitive-developmental theory, 270 5th 7th and 9th-grade Black, Chinese, and White subjects, 90 at each grade level and of similar SES, were administered a new 25 item spatial reasoning (map-reading) task. It was predicted that age, ethnic group, and sex significantly influence mean levels of achievement, but that analyses of patterns of item performance would reveal a similar sequence of concept and skill acquisition regardless of ethnic group means. These predictions were called 'the fixed-sequence hypothesis.' Age and ethnic group were significant influences on performance; sex was not. Scalogram analyses showed that each ethnic group's performance tended to form a scalable item set, but for a somewhat different ordering of the 25 items. Thus, the results were inconsistent with cognitive-developmental theory. It was concluded that individual differences are likely to affect development when sequences are highly task specific, but that the bulk of the evidence still supports a general sequence of stages in cognitive development. Implications for developmental theory, intelligence testing, and curriculum planning are discussed.
THE FIXED-SEQUENCE HYPOTHESIS: INDIVIDUAL DIFFERENCES IN THE DEVELOPMENT OF SCHOOL RELATED SPATIAL REASONING

David H. Feldman
University of Minnesota

The Problem

Kohlberg (1968) wrote that an invariant sequence of stages of cognitive development is essential to the 'cognitive-developmental' point of view. In calling for empirical data bearing on the question of invariant stages, Kohlberg commented that:

...it is extremely important to test whether a set of theoretical stages does meet the empirical criteria... If empirical sequence was not found, one would argue that the 'stages' simply constituted alternative types of organization of varying complexity, each of which might develop independently of the other (p. 1022).

Inhelder (1968), in a revision of her 1943 work on reasoning in the mentally retarded, also emphasized the importance of using objective statistical methods for assessing stages in reasoning development. Inhelder noted that results in the Genevan laboratory tend to indicate that rates of development differ among subjects, but sequences of cognitive development are invariant:

In order to use statistics to determine whether the developmental succession of certain behaviors does follow such a hierarchical process, or whether it is simply a question of fortuitous temporal succession, it is necessary to resort to procedures of hierarchical analysis such as Guttman introduced into social psychology. We owe a debt to L.J. Cronbach for having suggested that we use ordinal methods as far back as 1954. Vinh Bang...
found that the indices of reproducibility for the solutions furnished for various tests are not the same for all ages. This is of great interest...because it promotes understanding of the dynamics inherent in the normal evolutionary process, in that it implies that the diverse types of operational behavior do not always follow the same speed of development... (pp. 319-320).

Thus, two key assumptions of the theory, as expressed by Inhelder, are that development rates may differ among subjects, but developmental sequences do not.

Wallach (1963), in his review of research on children's thinking, wrote: "Concerning the effects of sociocultural factors, variation in absolute age norms has been found but not in over-all developmental sequences (p. 257)." A critical review of the recent evidence bearing on the invariant sequence assumption reveals that this conclusion is only partially supported by the data. While a general sequence of stages in achievement of cognitive levels does seem to hold across tasks and samples, several specific discrepancies between theoretical requirements and empirical evidence have been found.

Dodwell (1960, 1968) and Wohiwill (1960), in their studies of the development of number concepts, found general support for a fixed sequence of acquisition, but noted that several tasks were mastered earlier or later than predicted by Piaget. Dodwell (1968) concluded that "on the basis of the results here reported... the pattern of development is neither as regular, nor as simple,
as Piaget has suggested" (p. 112). Kofsky (1966), in her study of classificatory development, found that individuals vary in the steps by which they master cognitive task, and that the evidence for an invariant sequence supports "at best...a partial order" (p. 200). Uzgiris (1964) studied conservation of size, weight, and volume using several different materials (e.g., metal cubes, plastic balls, etc.) and found that individuals varied in the extent to which they conserved with different sets of materials.

Goldman (1965), investigating the understanding of religious concepts, got general support for Piaget's stages, but found that the order of task difficulty was not always as Piaget's theory requires. Similar results were reported by Larsen (1969) for spatial concepts. To produce a fixed sequence (in operational terms, to produce a scalable set of tasks or items), several researchers have reduced the number of steps in a sequence by combining or dropping tasks and/or changing the criteria for passing or failing items (e.g., Peel, 1959; Larsen, 1969; Siegelman & Block, 1969).

Thus, it appears that the more specifically one attempts to operationalize a sequence of steps in development, the more likely one is to find discrepancies between theory and evidence. One possible explanation for these discrepancies is that experiences such as family practices and educational history have little effect on general sequences of development, but may influence more
specific steps or processes through which a general stage of
cognitive functioning is achieved. Uzgiris' (1964) results for
conservation using different materials may support this interpreta-
tion; Uzgiris found individual differences regarding the materials
used in conservation, although the sequence from nonconservation
to conservation appeared to be invariant.

Very little data bearing on the possible effects of individual
or group differences on sequences of development has been collected.
Bruner, Olver and Greenfield (1966) reported several studies com-
paring Wolof, Mexican, Eskimo, and American children on several
Piagetian tasks. In general, Bruner et al found that schooling
vs. lack of schooling and urban vs. rural environments were often
more significant than cultural differences in affecting cognitive
development levels; sequences of development were not systematically
studied, however. Bruner et al suggested that there may be
different modes of achieving the same cognitive goals; for
example, Wolof and American children both achieved conservation,
but they appeared to achieve it through different reasoning pro-
cesses.

Studies by Kohlberg (1963) with Atayal children and Stodolsky
(1965) with Black children applied sequential scaling techniques
to dream concepts (Kohlberg) and classification tasks (Stodolsky).
Kohlberg reported comparable scalability for the development of
dream concepts in Atayal and White children; Stodolsky reported
comparable Guttman scaling patterns in classification skills for
Black and White lower-class children in Chicago. Both studies confirmed previously reported differences in levels of performance but invariant sequences.

Turiel (1969) summarized several studies of moral development in groups of subjects varying in social class (Kohlberg, 1958), urban-rural environments (Owen, 1968), and cultural background (Kohlberg, 1966; Kramer, 1968). With few exceptions, the data appear to support the invariant stage sequence assumption of moral development. Turiel's own work is especially noteworthy for its substantive findings and for its novel methodology, which sheds light on the processes of transition between stages as well as the sequentiality of stages.

The present study initiates a program of research aimed at further testing of the possible effects of group and individual differences on sequences of cognitive development. Because previous studies have found specific discrepancies between theory and evidence as a function of the level of specificity with which sequences are defined, the present study attempted to further test the invariant sequence assumption against a highly articulated sequence. An additional aim of the study was to base the hypothesized sequence on both theoretical grounds and a conceptual analysis of the criterion concepts to be achieved, as suggested by Kohlberg (1968). A test instrument, validated in a pilot study (Feldman,
1969), was designed to assess a hypothesized specific sequence of skills and concepts requisite to performance of a complex reasoning task. Ordinal scaling techniques were used to analyze developmental sequences (Guttman, 1947; Edwards, 1957).

The instrument used in the study was intended to assess the influence of ethnic background on the acquisition of specific concepts and skills requisite to the proper drawing of a geographic map. The items for the instrument were derived from Salomon's (1968) task analyses of the geographic map. Piaget and Inhelder (1967), among others, have shown that map drawing is a useful task in studying the development of spatial reasoning. The modern map, a unique product of Western culture and an example of a scientific model, has also been shown to be difficult for non-Western children to understand (Salomon, 1968; Dart & Lal Pradham, 1967). Finally, Lesser and his colleagues (Lesser, Fifer, & Clark, 1965; Stodolsky & Lesser, 1967) have found ethnic differences in spatial ability even when social class was controlled. Since social class has often been a confounding variable in studies which compare the performance of different ethnic groups (Stodolsky & Lesser, 1967; Hess & Shipman, 1965; Eisner, 1967), the present study selected samples of equivalent socio-economic level, but differing in ethnicity.

Hypotheses

"The fixed-sequence hypothesis" is comprised of two related propo-
sitions, each of which was tested separately. First, it was hypothesized that with social class held constant there are differences in the level of achievement of the three ethnic groups tested. Second, it was hypothesized that despite differences in the level of achievement exhibited by the ethnic groups, sequences of acquisition of concepts and skills would not significantly differ across ethnic group.

In addition to the two parts of the fixed-sequence hypothesis, four secondary hypotheses bearing on the validity of the experimental task were tested. These hypotheses are presented in detail in the results section.

A key assumption of the study was that developmental sequences of concept and skill acquisition are reflected in patterns of item performance, and that rates of development are reflected in mean scores for each ethnic group. Wohlwill (1960) and Kofsky (1966) have discussed some of the problems with this assumption, several of which will be taken up in the discussion of the results of this study.

**Method**

**Subjects and Sampling Techniques**

Subjects were 270 fifth, seventh, and ninth-grade students attending public schools in San Francisco, California. The difficulty in selecting schools with children of comparable social class but varying ethnic background was considerable, since social class and ethnicity tend to covary. Another difficulty, pointed out by
Deutsch (1967), is that the same data may mean different things for different ethnic groups. An income of $6000 for a Black family may mean something quite different than the same income for a Chinese family. These difficulties notwithstanding, income and education level of parents were taken as indicators of social class (Hess, 1968). Table 1 presents the data for education and income of the populations of the schools selected for inclusion in the study.

Schools were chosen with relatively homogeneous ethnic populations at the elementary level; the junior high school chosen had about the same number of students from each of the three ethnic groups under investigation. Table 2 presents the ethnic mix of each of the schools selected.

The neighborhoods were described as lower-middle class by several informants; direct observation tended to support this description. It was recently estimated by the local Chamber of Commerce that a $10,000 income is required for a family of four to live comfortably in San Francisco; the median income of the families in the study was $6900.

Two additional informal sampling procedures were used in the present study. School principals were advised to exclude children from the testing sample who were from 'upper-middle class' or from 'severely disadvantaged' homes. They were also told to select students from all ability levels and behavior patterns, including 'retarded' or 'disruptive' students. Since
<table>
<thead>
<tr>
<th></th>
<th>Median Income</th>
<th>Median Education</th>
<th>Sample</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary School B</td>
<td>$6300.00</td>
<td>12.1 Years</td>
<td>30</td>
<td>459</td>
</tr>
<tr>
<td>Elementary School C</td>
<td>$7000.00</td>
<td>12.4 Years</td>
<td>30</td>
<td>684</td>
</tr>
<tr>
<td>Elementary School W</td>
<td>$7000.00</td>
<td>11.9 Years</td>
<td>30</td>
<td>303</td>
</tr>
<tr>
<td>Junior High School</td>
<td>$6700.00</td>
<td>12.2 Years</td>
<td>180</td>
<td>997</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$6900.00</strong></td>
<td><strong>12.2 Years</strong></td>
<td><strong>270</strong></td>
<td><strong>2443</strong></td>
</tr>
</tbody>
</table>
### TABLE 2

Ethnic Mix of the Schools Selected for Study

<table>
<thead>
<tr>
<th></th>
<th>%White</th>
<th>%Black</th>
<th>%Chinese</th>
<th>Other</th>
<th>Groups Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary School B</td>
<td>14.8</td>
<td>74.1</td>
<td>4.1</td>
<td>7.0</td>
<td>Black</td>
</tr>
<tr>
<td>Elementary School C</td>
<td>4.0</td>
<td>0.4</td>
<td>94.3</td>
<td>1.3</td>
<td>Chinese</td>
</tr>
<tr>
<td>Elementary School W</td>
<td>71.3</td>
<td>2.0</td>
<td>1.3</td>
<td>25.4</td>
<td>White</td>
</tr>
<tr>
<td>Junior High School</td>
<td>40.7</td>
<td>28.6</td>
<td>17.2</td>
<td>13.5</td>
<td>All (B,C,W)</td>
</tr>
</tbody>
</table>
students were tracked in the junior high school, approximately equal numbers of students were selected from each track.

Although not a random sample from a well-defined population, the subjects probably represent a large segment of the population of many American urban areas. However, the study is not generalizable to disadvantaged populations, since part of the current meaning of the term disadvantaged includes poverty.

As stated earlier, the present study attempted to control for social class while examining variation due to ethnic group. The lower-middle class sample finally selected proved to be the only social-class group on which it was possible to match the three ethnic groups selected for investigation.

Instruments and Scoring

The test instrument (map test) had two parts. A map-drawing task, similar to one used by Dart and Lal Pradham (1967) in Nepal and Hawaii, asked subjects to draw a map of the school and the school grounds as seen from above. The instrument was also similar to Eisner's (1967) task, in which subjects drew a picture of the school grounds and their friends, and to a Piagetian diagram-drawing task (Piaget & Inhelder, 1967). The map-drawing task of the present study put greater conceptual demands on the subject than previous drawing tasks used to assess cognitive development; Goodenough (1924) and Harris (1963) asked subjects to "draw-a-man," "draw-a-woman," or "draw-a-tree," while Piaget and Inhelder (1967) and Salomon (1968) provided a visual model which the subjects were free to use in constructing their diagrams. Eisner (1967) asked children to
draw a picture of the school grounds; which required memory and visualization but did not demand abstract representation. Since performance on this map-drawing task required subjects to draw a map from above with no memory support, it was expected that performance might fall somewhat below that found in previous studies. The purpose of the map-drawing task was to provide a criterion against which to validate a 25-item map-reading test (described below).

A scoring system was developed for the map-drawing task which took much of its content from Piaget and Inhelder (1967) and its scoring procedures from Eisner (1967). Based on Piagetian stages of spatial reasoning in diagrammatic layout construction, a six-category rating scale was developed. Each subject's map was rated independently by two trained judges (both of whom were graduate level psychology students). Judges were trained with written descriptions of the categories and with example maps produced by pilot study children (Feldman, 1969). Initial agreement in ratings between the judges was 50-60%. After four training sessions, inter-judge agreement reached a criterion of above 80%; the judges disagreed in 2% of the cases by more than one category. Following independent rating, the judges were brought together to discuss disparate ratings. Agreement on 97% of the drawings was achieved in this manner, comparable to results reported by Eisner (1967). The remaining 3% of the drawings were arbitrarily categorized by the author.
The second part of the experimental task was a map-reading test. On the basis of results from a pilot study (Table 3, Feldman, 1968), a set of 25 map-reading tasks was arranged in a hypothetical fixed sequence of acquisition. The formal characteristics of the fixed sequence are equivalent to Gagne's (1962) "hierarchical learning sets," and were stated by Gagne as follows:

1. If a higher-level learning set (task) is passed (+), all related lower-level tasks must have been passed.

2. If one or more lower-level tasks have been failed (-), the related higher-level tasks must be failed.

3. If a higher-level task is passed (+), no related lower-level tasks must have been failed (-).

4. If a higher-level task has been failed (-), related lower-level tasks must have been passed.

In his studies of learning sets, Gagne was not attempting to infer developmental sequences from his data. It is the distinction between typical learning and typical developmental sequences that separates the present research from Gagne's work. The actual arrangement of the items in the test instrument was intended to form a 25-stage set of prerequisites to the criterion, i.e., to proper drawing of an abstract, formal map.

Following Salomon (1968), the present map reading test used a fictional island as its visual stimulus. Subjects answered questions
<table>
<thead>
<tr>
<th>Concept or Skill</th>
<th>Item Number (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Identification of:</strong></td>
<td></td>
</tr>
<tr>
<td>Stimulus</td>
<td>1</td>
</tr>
<tr>
<td>Ikons</td>
<td>2, 3, 4</td>
</tr>
<tr>
<td>Signs and Symbols</td>
<td>5, 6, 7</td>
</tr>
<tr>
<td><strong>II. Analysis of:</strong></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>8</td>
</tr>
<tr>
<td>Map Heights</td>
<td>9</td>
</tr>
<tr>
<td>Filled vs. Empty Space</td>
<td>10</td>
</tr>
<tr>
<td>Map Directions</td>
<td>11, 12</td>
</tr>
<tr>
<td>Personal and Map Heights</td>
<td>13</td>
</tr>
<tr>
<td>Personal Heights</td>
<td>14</td>
</tr>
<tr>
<td>Personal Directions</td>
<td>15</td>
</tr>
<tr>
<td>Comparative Directions</td>
<td>16</td>
</tr>
<tr>
<td>Knowledge of Latitude</td>
<td>17</td>
</tr>
<tr>
<td>Map Directions with Logic Problem</td>
<td>18</td>
</tr>
<tr>
<td>Personal Directions with Logic Problem</td>
<td>19</td>
</tr>
<tr>
<td>Symbol Understanding-Abstract</td>
<td>20</td>
</tr>
<tr>
<td><strong>III. Interpretation based on:</strong></td>
<td></td>
</tr>
<tr>
<td>Inferences about Weather</td>
<td>21</td>
</tr>
<tr>
<td>Choice of Capital with No Visual Symbol</td>
<td>22</td>
</tr>
<tr>
<td>Two Smallest Towns</td>
<td>23</td>
</tr>
<tr>
<td>Location for Harbor</td>
<td>24</td>
</tr>
<tr>
<td>Complex Hypothesis; Fishing Village vs. Farming Town</td>
<td>25</td>
</tr>
</tbody>
</table>
based on information in the visual stimulus, a copy of which was given each subject, and in materials provided in a prepackaged booklet. Each of the 25 tasks in the test was designed to reflect subject's level of acquisition of a concept or skill hypothesized to be requisite to proper performance in map drawing. The content of the items was based on Salomon's conceptual task analysis of the map and on Piaget and Inhelder's (1967) research on spatial reasoning (see Table 3).

Of the 25 items, 17 were multiple-choice questions with four alternative answers and a blank space if subject wished to write his own answer. Each distractor in the multiple choice items was designed to reflect one of four different reasoning levels. Designing distractors to give information other than 'correctness' was suggested in a paper by Guttman and Schlesinger (1967) and is discussed in greater detail elsewhere (Feldman & Markwalder, 1970). This procedure allows for the differential scoring of subjects on the basis of the types of wrong answers to which they are attracted. Questions were designed so that each one of the distractors was intended to reflect either (a) tautological or imaginary reasoning (preconceptual thought in Piaget's terms); (b) perceptual or associative reasoning (beginning of concrete operations); (c) concrete reasoning (logical but limited to information in the stimulus); or (d) formal reasoning (not necessarily based on information directly obtainable from the stimulus). The four reasoning levels were extrapolated from Piaget's

Three sample items of increasing difficulty are presented below:

#7 What does the little drawing near the town of Koff tell you?
   a. It tells me something nice about the people who live in Koff (Imaginary.)
   b. It tells me there is a cow near Koff (Perceptual.)
   c. It tells me there is a dairy near Koff (Concrete.)
   d. It tells me that Koff is a place where cows are raised for milk and butter (Formal.)

#8 How would you find out how far it is from one side of the map to the other?
   a. I would ask a man who lives on the island (Imaginary.)
   b. I would drive a car and look at the miles (Concrete.)
   c. I would have to see if the roads are good (Perceptual.)
   d. I would use a scale on the map. (Formal.)

#21 Circle the words which tell you what you think the weather is like on this island most of the time. List the reasons why you think the weather is what you say.
   Rainy (Formal); Icy (Imaginary); Cold (Perceptual); Hot (Formal);
   Foreign (Imaginary); Sticky (Formal); Windy (Concrete-Perceptual);
   Dry (Imaginary or Perceptual);
   Foggy (Concrete-Perceptual); Icky (Imaginary); Cool (Concrete-Perceptual);
   Snows (Perceptual or Imaginary);
   Humid (Formal); Pretty (Perceptual); Dusty (Perceptual);
   Smoggy (Concrete-Perceptual); Nice (Perceptual).

Map-reading test items were first scored for the number 'correct' i.e., in a manner similar to a standard multiple-choice test. This score was referred to as the map-reading test score (MS). MS was used to compare group achievement levels; items answered formally, and in some cases concretely, were considered as passes in the fixed-sequence analysis.
A second score was obtained by computing a mean reasoning level (RL) exhibited on the map-reading test, as follows: an imaginary/tautological answer was scored 1; a perceptual/associative answer, 2; a concrete answer, 3; a formal answer, 4. All 25 items were evaluated in this manner and a reasoning level score was computed by the formula:

\[
\text{RL} = \frac{(\text{Imaginary responses } \times 1) + (\text{Perceptual } \times 2) + (\text{Concrete } \times 3) + (\text{Formal } \times 4)}{25 \text{ (Number of Items)}}
\]

Reasoning level and map-reading test score were related since in general the RL increases as the number of items answered correctly increases. However, two subjects with identical map-reading test scores could exhibit markedly different reasoning levels, depending on the level of reasoning exhibited in the answers they missed, as in the following hypothetical example:

**S1**

Total Score = 10  
Formal: 10 x 4 = 40  
Concrete: 15 x 3 = 45  
Reasoning Level = 90/25 = 3.6 (Formal)

**S2**

Total Score = 10  
Formal: 10 x 4 = 40  
Imaginary: 15 x 1 = 15  
Reasoning Level = 55/25 = 2.2 (Perceptual)

RL was used to test the strength of the relationship between map-drawing categories and Piagetian stages of general cognitive development.
Procedures for Administration

A primary purpose of the testing procedure was to reduce, insofar as possible within practical constraints, the dependence of a child's performance on his ability to read. Another purpose was to reduce the anxiety of test-taking to a minimum so that each child had the best possible opportunity to exhibit his reasoning about spatial concepts. The examiner (E) read all directions, each question and its distractors aloud, proceeding as slowly as was necessary to insure each child the opportunity to think about and complete the items. Second, each examiner was of the same ethnic background as the children with whom she worked. Examiners of the same ethnic background as subjects were used by Lesser, Fifer, and Clark (1965) and Stodolsky and Lesser (1967) for similar purposes.

Subjects were tested in groups of 10 to 15 in library, cafeteria, auditorium, or classroom facilities. Procedures for getting the children out of class and into testing rooms varied from school to school and may have contributed to uncontrolled variations in the testing situation.

As the children entered the testing room, the following instructions were given by Examiner:

Good morning (afternoon). My name is Mrs. _____________.
This morning (afternoon) you and I are going to do a map exercise together. In this exercise, we will be interested in how you draw and how you think about maps. Now I will give each of you a booklet which has some blanks for you to fill in on the front. (Examiner distributed map tests.)
Please fill in each of the blanks; circle if you are a boy or a girl. Now in the blank where it says "age" write your birth date, both the month and the year you were born.

Now we will read the instructions, you to yourself and me out loud. (Instructions were read, making sure everybody understood that the drawing was to be of the school and the school grounds as seen from above.)

On the next page, which is a blank page, draw a map of the school and the school grounds as seen from above. You will have ten minutes to draw your map. Then wait for your next set of instructions. Good Luck!

O.K., finish up your drawings and we will go on to the rest of the exercise. Please don't turn the page until I ask you to do so. Now I am going to give each of you one of these (holds up stimulus map). For the rest of the work we will use these. Does everyone have one? (Examiner kept a map for herself to point out what a question was referring to. For example, when Examiner read the question having to do with the green color, she pointed to the green parts of the map.)

What we're going to do is not a test; it is an exercise to find out what you think are the best answers to some questions about maps. There will be no grade given; this is because we want to know your answers. You will have as much time as you need to do each question. If I go too fast, raise your hand, and we will wait. Remember, we are not interested in how fast you can go. I will read each question out loud as you read it to yourself. If you have trouble understanding the question the first time I read it, raise your hand, and I will repeat the question. Does everyone understand?

Is everyone ready? O.K., let's flip over the page and look at question #1. Notice that there are four possible answers and a space for you to write in your answer if you do not find one which you think is best. (Repeated for each question.)

Testing was done in the morning and early afternoon hours. All groups were tested within ten days of each other (October 7 through
Testing time varied from about one hour with fifth graders to about 45 minutes with ninth-grade children. Examiners reported that in general the children at all three grade levels appeared well motivated to complete the task and did not find it too easy or too difficult. However, in a small number of cases, children were not given adequate time to finish their work, contributing further uncontrolled variance to the results.

Results and Discussion

Map Test Validity Hypotheses. Results bearing on three secondary hypotheses of the study were intended to assess the validity of the map test tasks. These results are presented first because confidence in interpreting the fixed-sequence analysis depends upon adequate validation of the test instrument.

Secondary hypothesis 1 predicted a significant, positive correlation between map-reading scores (MS) and map-drawing categories (MC). If the map-reading test assesses the skills and concepts requisite to proper map-drawing behavior, a significant agreement between the two measures was expected. The correlation between MC and MS was found to be .46 (p < .01) for the entire sample; subgroup correlations did not significantly differ from the total sample. Although in support of the hypothesis, the agreement between the two instruments accounted for only 21% of the variation in performance. Restricted variability in MC (six categories) may have affected the MC x MS
correlation, but on the basis of the present results, only modest support for the hypothesis was provided.

Secondary hypothesis 2 predicted significant increases in MS with grade level. Since developmental inferences were made from the data, it was crucial that performance levels increase with grade level.

Table 4 presents the means and standard deviations for each ethnic group at each grade level. Table 5 presents the results of an analysis of variance testing the effects of grade level on MS performance. The data show orderly increases in MS performance with each grade level and within each ethnic group (there were no significant sex differences in MS). Grade level had a significant effect on MS performance according to the analysis of variance.

Hypothesis 3 predicted that reasoning level (RL) would increase as MC increased. This hypothesis was intended to relate the map test's reasoning measures to Piaget and Inhelder's (1967) stages of spatial reasoning (i.e., to MC). Since RL was intended to be a deeper measure of reasoning level than MS, and since Piaget claims that spatial reasoning is a special case of general reasoning development, RL and MC measures should be in substantial agreement.

Figure 1 shows the median RL for subjects whose maps were categorized in each map-drawing category. RL increased with MC up to category 6, at which point there was a decline in RL. However, out of 270 map protocols, there were only two placed in category 6 by the judges;
TABLE 4

Means and Standard Deviations of Map Test Scores For Each Ethnic Group at Each Grade Level

<table>
<thead>
<tr>
<th>Ethnic Group</th>
<th>Grade Level</th>
<th>5th</th>
<th>7th</th>
<th>9th</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td>11.97</td>
<td>13.40</td>
<td>15.17</td>
<td>13.51</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>(3.81)</td>
<td>(3.37)</td>
<td>(3.51)</td>
<td>(3.57)</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td></td>
<td>14.20</td>
<td>16.17</td>
<td>17.20</td>
<td>15.85</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>(3.57)</td>
<td>(3.15)</td>
<td>(3.67)</td>
<td>(3.47)</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td></td>
<td>14.57</td>
<td>17.40</td>
<td>17.63</td>
<td>16.53</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>(3.07)</td>
<td>(3.43)</td>
<td>(3.39)</td>
<td>(3.30)</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>13.53</td>
<td>15.63</td>
<td>16.63</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>(3.50)</td>
<td>(3.32)</td>
<td>(3.52)</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 1
Mean Reasoning Level (RL) for Subjects Classified Into Each Map-Drawing Category (N = 270)

Reasoning Level

(n = 90)  (n = 71)  (n = 24)  (n = 15)  (n = 2)

I  II  III  IV  V  VI
Map-Drawing Category
the RL estimate for category 6 may therefore have been unreliable. Thus, the hypothesis was generally supported by the data, but further research into the unpredicted results is called for.

**Fixed-Sequence Hypothesis.** Part one of the fixed-sequence hypothesis (Hypothesis 5) predicted that ethnic group is a significant influence on MS levels. This hypothesis followed from the vast literature on differences between means for Black and White subjects on standard tests of IQ and achievement (e.g., Jensen, 1968; Tyler, 1965), even when social class has been equated (Bruce, 1940; McQueen & Browning, 1960). Table 5 presents the results of the analysis of variance for ethnic group effects on MS performance. As shown in Table 5, ethnic group was a significant (p < .01) influence on MS. The mean differences in MS were taken to support the part of the fixed-sequence hypothesis which predicted differing rates of development for each ethnic group.

The fixed-sequence hypothesis (Hypothesis 6) also predicted that despite differences in mean levels of achievement on the map-reading tasks, all ethnic groups would exhibit a fixed sequence of item performance. Table 6 presents the results of scalogram analyses of the hypothetical fixed-sequence of item performance. The first column of figures in Table 6 is based on the original ordering of the set of 25 items of the map test. The Coefficients of Reproducibility (CR = 1 - The number of scaling errors over the total number of
TABLE 5

Analysis of Variance of Effects of Ethnic Group and Grade Level on Map Test Scores (N = 270)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethnic Group</td>
<td>2</td>
<td>226.34</td>
<td>19.02**</td>
</tr>
<tr>
<td>Grade Level</td>
<td>2</td>
<td>223.21</td>
<td>18.76**</td>
</tr>
<tr>
<td>EG x GL</td>
<td>4</td>
<td>5.49</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Error</td>
<td>261</td>
<td>11.90</td>
<td></td>
</tr>
</tbody>
</table>

**p < .01
TABLE 6

Coefficients of Reproducibility for the 25 Items in the Map Test as Originally Ordered and as Reordered According to Their Pass/Fail Frequency in Each Group

<table>
<thead>
<tr>
<th>Ethnic Group</th>
<th>Original Order</th>
<th>Reordered</th>
<th>MMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>.64</td>
<td>.82</td>
<td>.74</td>
</tr>
<tr>
<td>Chinese</td>
<td>.55</td>
<td>.83</td>
<td>.77</td>
</tr>
<tr>
<td>White</td>
<td>.58</td>
<td>.85</td>
<td>.79</td>
</tr>
<tr>
<td>Combined Groups</td>
<td>.82</td>
<td>.76</td>
<td></td>
</tr>
</tbody>
</table>
responses) reported for the original ordering of the items do not approach an adequate Guttman Scale for any of the ethnic groups tested; criterion is usually .85 to .90 (Guttman, 1947; Edwards, 1957). That is to say, the hypothetical ordering of the items did not scale as a fixed sequence for any of the three ethnic groups.

It should be noted that Guttman scaling techniques do not permit analysis based on an arbitrary or hypothetical ordering of a set of items (Wohlwill, 1960). Part of the process of obtaining scalability estimates is an empirical reordering of items according to their difficulty. Therefore, the reported coefficients should be viewed as rough estimates of scalability. Estimates were produced by modifying Gagne's (1962) technique for analyzing hierarchical learning sets; that is, by doubling the number of errors that were computed according to Gagne's technique (Elashoff, 1969). Reproducibility estimates were produced in this manner for the original ordering of the 25-map-reading items for each ethnic group.

The second column of figures in Table 6 shows the Coefficients of Reproducibility for each ethnic group for the set of 25 items in the map test as reordered by the Cornell ranking technique, i.e., ranked according to difficulty and rearranged to produce the fewest scale errors (Edwards, 1957).

The results of the scalogram analyses with item sets and subjects reordered separately within each ethnic group (Table 6) indicated that
the three ethnic groups did not differ in the extent to which their overall performance conformed to the fixed-sequence model. Coefficients of Reproducibility were .82, .83, and .85 for Black, Chinese, and White subjects. Thus, the results of the analyses indicate partial but not complete conformity to the fixed-sequence model. There are several factors which affected the reproducibility coefficients, however, which should be taken into account.

The machine program (BMD05S) which performed the analysis rearranged the items according to the "Cornell Ranking Technique" (Guttman, 1947), which reorders the items and subjects so as to minimize 'errors,' thus giving a somewhat inflated estimate of scalability (Edwards, 1957). It is also true that the Coefficient of Reproducibility is necessary but not sufficient evidence for scalability. Also needed are items which vary in the percent of respondents in a given category. That is, if all subjects answer all questions correctly, the set of questions is scalable but trivial. The parameter describing the extent to which items vary is the Minimal Marginal Reproducibility (MMR) measure (see Table 6), which refers to the mean per cent of subjects whose responses were in the same category. With the present set of items, dichotomized as they were into 'correct' or 'incorrect' responses, the MMR was high, but an inspection of the items indicated that they did have a wide range of difficulty. To some extent the high MMR was a function of dicho-
tomizing the item responses into two categories; therefore, the CR is probably not greatly distorted by the high MMR.

Wohlwill (1960) and Kofsky (1966) have pointed out some of the conceptual and procedural difficulties in using scalogram analysis results to infer developmental sequences. Most troublesome are the assumption that cross-sectional, discrete data reflects development; also troublesome are the possible effects of varying instructions and materials from task to task, and the possibility that the scalogram model may not be appropriate for all—or even most—developmental processes. Recent papers by Leik and Matthews (1968) and TenHouten (1969) are promising steps toward the refinement of developmental scaling techniques.

Because of the way the fixed-sequence hypothesis was stated, the results could still be interpreted as generally supporting it. The prediction that the three ethnic groups would not differ in the extent to which their item performance conformed to the fixed-sequence model was supported by the data. However, the items had to be reordered in order for the hypothesis to be supported.

Although not anticipated in the original testing of the fixed-sequence hypothesis, it became of interest to assess the extent to which the reorderings of items differed among ethnic groups. Table 7 presents the 25 map-test items as reordered according to their difficulty for each group. Because of ties in difficulty, all three
TABLE 7
Reordering of the Map Test Items According to their Difficulty for each Ethnic Group

| Step No. | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17  | 18 | 19 | 20 |
|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Black    | 2  | 4  | 7  | 1  | 8  | 14 | 10 | 13 | 16 | 18 | 20 | 17 | 11 | 9  | 3  | 19 | 24 | 22 | 12 | 23 | 2  |
| Chinese  | 7  | 2  | 4  | 16 | 8  | 9  | 19 | 15 | 20 | 17 | 13 | 6  | 11 | 18 | 22 | 24 | 23 | 12 | 21 | 2  |
| White    | 2  | 1  | 7  | 8  | 16 | 15 | 13 | 5  | 9  | 11 | 20 | 14 | 19 | 3  | 10 | 18 | 12 | 24 | 22 | 23 | 2  |
sequences reduced from 25 to 20 steps. An examination of the reorderings of the items yields little systematic information; in general, it appears that many of the items did increase in difficulty for all groups as hypothesized, but the sequence for each group contains some distinct differences. Twelve of the items appeared to be about as easy or as difficult as predicted, and about equally difficult for each ethnic group (that is, they appear about the same place in the sequence for each group). An additional 7 items were either easier or harder than predicted, but not differentially so for any group. The remaining 6 items ranked in difficulty differentially depending upon the ethnicity of the group. Further study of the items appearing at different points in the sequences is planned.

Implications for Research and Practice. A number of possible implications of the present study should be investigated in future studies. The possible effects of order of item presentation on performance, curriculum sequences, and intelligence test organization are of particular interest.

The set of 25 items in the map-reading test was administered in a standard order to all subjects. If difficulty of items reflects sequences of development, the order in which the items are presented should make little difference in the production of a fixed sequence.
or sequences. Future studies should randomize the items throughout the instrument to test this hypothesis. It is true, however, that the initial ordering of tasks in the test did not yield a fixed sequence for any of the three ethnic groups tested in the study; thus, it is equally possible that the order of presentation had disruptive effects on performance levels for all groups and that empirically derived sequences may facilitate performance of children with various characteristics.

If sequences of presentation are shown to affect performance of children in systematic ways, curriculums may be made more sensitive to such influences as are found to exist. The assumption implicitly made by curriculum planners and teachers is that all children acquire knowledge and skills in the same sequence; curriculums are constructed with this assumption built in. If the results of the present study are replicated and extended to other areas of the school curriculum, basic assumptions about curriculum construction may have to be revised.

Another possible implication of fixed-sequence research is in the area of intelligence testing. Since IQ tests are generally constructed on the basis of item performance by members of a white middle-class normalization group, IQ tests may not be allowing for systematic differences in specific developmental sequences among children from various backgrounds. In a typical IQ test situation,
the session is terminated when a subject misses all items at a
given level. Studies are planned in which IQ tests items are given
in random order, and all subjects respond to all questions. Results
will be analyzed in ways similar to those used in the fixed-sequence
study reported here. It may be that children from differing back-
grounds respond to item sequences in systematic but different ways.
The notion of a culture-fair IQ test may take on new meaning in
the light of these studies.

Conclusion. What light, if any, has the present study shed on
the problem of invariant sequences of cognitive development? The
results appear to contradict the cognitive-developmental assumption
that all children develop skills and concepts in the same fixed
sequence. The present study found that each ethnic group's perfor-
mance approached a fixed sequence, but that the sequence was relatively
distinct for the three groups. Three 'fixed sequences' were found
where a single sequence was predicted. It should be noted again
that the sequence tested in this study was highly specific and
based on a task analysis of a single school task.

The bulk of the evidence is still on the side of invariant
sequences, but a task-specific sequence, as in the present study, does
show variation according to differing ethnic background. In terms
of possible educational significance, more specific sequences may
have greater promise for instructional planning than broad, developmental
stages of cognitive development.
It should be noted that ethnic group is but one ID variable that may affect development of task-specific reasoning. Ethnic group per se will probably turn out to be at best an imperfect indicator of more relevant cognitive variables. When educational treatments are devised which take ID variables into account, it is on the basis of cognitive variables rather than ethnic group membership that treatments are likely to be assigned.

No adequate explanation for the unique reorderings has been proffered or verified, but several possible explanations are currently being investigated. The language of the items and instructions, the nature of the stimulus materials, sex differences, and of course the unanticipated cognitive demands of the questions themselves, are all possible sources of variation in the order of acquisition.

There are a number of possible reasons why the results do not completely support the fixed-sequence hypothesis. It may be, for example, that the measures of reasoning development in the study were superficial and do not represent true developmental stages. The skills requisite to map drawing may better reflect differences in learning histories and educational experiences than basic stages in cognitive development. There is, however, inherent difficulty in the argument over whether or not a given set of measures of fixed sequences is basic enough. The argument can be carried on ad absurdum, with the critic at each turn accusing the investigator of being superficial, while
the investigator accuses the critic of holding out until the data are consistent with his theoretical position.

In the present study, an attempt was made to infer a single developmental sequence by testing children at three age levels. Since the level of performance increased as age increased, it was assumed that development took place with age. The process of development was assumed to be reflected in the pattern of responses made by the subjects and in the relationships predicted among various scores. Although by no means unique to the present study, problems in measuring development deserve greater attention.
References


Bruce, M. Factors affecting intelligence test performance of whites and negroes in the rural south. *Archives of Psychology*, 1940, 36, 252.


Hess, R.D. Personal communication, 1968.


Kohlberg, L. Stages in children's conceptions of physical and social objects in the years 4 to 8-- a study of developmental theory. Unpublished monograph, Yale University, 1963.


McQueen, R., & Browning, C. The intelligence and educational achievement of a matched sample of white and negro students. School and Society, 1960, 88, 327-329.


Support for the research reported here was provided by the Stanford Center for Research and Development in Teaching (OE-6-10-078), the Graduate School of the University of Minnesota, and the Research and Development Center for Education of the Handicapped, University of Minnesota (OEG-0-9-332189-4533-032). The helpful contributions of the following persons are acknowledged: E. W. Eisner, P. S. Sears, R. D. Hess, L. J. Cronbach, R. E. Snow, R. D. Bridgham, G. Salomon, M. Linn, R. W. Heath, J. Burke, and K. Baker. Requests for reprints should be addressed to the author, Pattee Hall, University of Minnesota, Minneapolis, Minnesota 55455.


<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Publication Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. Knowles</td>
<td>Recent research on manual communication</td>
<td>Occasional Paper F7, April 1971. (Keynote Address, Division of Communication Disorders, Council for Exceptional Children Annual Convention, Miami Beach, April 1971).</td>
</tr>
</tbody>
</table>


