INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)” . If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in “sectioning” the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.
PARETE, JESSE DAVID, JR.
FORMAL REASONING ABILITIES OF COLLEGE AGE STUDENTS: AN INVESTIGATION OF THE CONCRETE AND FORMAL REASONING STAGES FORMULATED BY JEAN PIAGET.

THE OHIO STATE UNIVERSITY, PH.D., 1978

© Copyright by Jesse David Parete, Jr. 1978
FORMAL REASONING ABILITIES OF COLLEGE AGE STUDENTS:
AN INVESTIGATION OF THE CONCRETE AND FORMAL
REASONING STAGES FORMULATED BY JEAN PIAGET

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Jesse David Parete, Jr., B.S., M.A.

The Ohio State University
1978

Reading Committee:
Professor Jon Higgins
Professor Marilyn Suydam
Professor Thomas Ralley

Approved By
Professor Jon Higgins
(Adviser, Department of Science and Mathematics Education)
VITA

December 5, 1944 . . . Born - Seneca Falls, New York
1966 ............. B.S. In Ed., State University of New York, Brockport, New York
1968 ............. M.A. in Mathematics, Bowling Green State University, Bowling Green, Ohio
1969-1972 . . . . . Instructor, Department of Mathematics, The Ohio State University-Lima, Lima, Ohio
1974-1975 . . . . . Instructor, Campus Middle School, State University of New York, Oswego, New York
1975-1978 . . . . . Lecturer, Department of Mathematics, The Ohio State University-Lima, Lima, Ohio

FIELDS OF STUDY

Major Field: Mathematics Education

Studies in Mathematics Education. Professors Jon Higgins and Marilyn Suydam

Studies in Mathematics. Professor Thomas Ralley
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>VITA</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>iv</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF RELATED LITERATURE</td>
<td>12</td>
</tr>
<tr>
<td>III. RESEARCH PROCEDURES</td>
<td>31</td>
</tr>
<tr>
<td>IV. RESULTS FOR COMPONENT I</td>
<td>70</td>
</tr>
<tr>
<td>V. RESULTS FOR COMPONENTS II AND III</td>
<td>132</td>
</tr>
<tr>
<td>VI. CONCLUSIONS</td>
<td>161</td>
</tr>
<tr>
<td>APPENDICES</td>
<td></td>
</tr>
<tr>
<td>A. The P-test</td>
<td>181</td>
</tr>
<tr>
<td>B. Guidelines for Evaluation of the P-test</td>
<td>204</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>216</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1. Transformation of stages to Numerical Values</td>
<td>55</td>
</tr>
<tr>
<td>2. Level of Formal Reasoning by Score on P-test</td>
<td>55</td>
</tr>
<tr>
<td>3. Distribution of Subjects by Level of Reasoning on the Pendulum Task</td>
<td>71</td>
</tr>
<tr>
<td>4. Distribution of Subjects by Level of Reasoning on the Triangle Task</td>
<td>85</td>
</tr>
<tr>
<td>5. Distribution of Subjects by Level of Reasoning on the Flexible Rods Task</td>
<td>97</td>
</tr>
<tr>
<td>6. Distribution of Subjects by Level of Reasoning on the Proportional Reasoning Task</td>
<td>109</td>
</tr>
<tr>
<td>7. Correlations Among the Four P-test Tasks</td>
<td>121</td>
</tr>
<tr>
<td>8. Distribution of Subjects by Age</td>
<td>122</td>
</tr>
<tr>
<td>9. Distribution of Subjects by P-test Level</td>
<td>122</td>
</tr>
<tr>
<td>10. P-test Score Distribution</td>
<td>123</td>
</tr>
<tr>
<td>11. Results of a Factor Analysis on the P-test Tasks and the Interview Tasks</td>
<td>127</td>
</tr>
<tr>
<td>12. Correlations Among Experimenter (E) and Additional Raters</td>
<td>130</td>
</tr>
<tr>
<td>13. Intercorrelations of Variables Employed in the Regression Analysis for Component II</td>
<td>134</td>
</tr>
</tbody>
</table>
List of Tables (Contd.)

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>15. Correlations Between the P-test and Select Measures of Academic Achievement</td>
<td>135</td>
</tr>
<tr>
<td>16. Distribution of the P-test Scale Within the Six Mathematics Classes</td>
<td>136</td>
</tr>
<tr>
<td>17. Contingency Tables for Comparison of Math 148 with Math 150 on the P-test Scale</td>
<td>140</td>
</tr>
<tr>
<td>18. Intercorrelations of Variables Employed in the Regression Analysis for Group a (Math 100) of Component III - the P-test Investigation</td>
<td>143</td>
</tr>
<tr>
<td>19. Summary Regression Table Used to Test Hypothesis $H_0$: Math 100 Final Exam Variance Accounted for by other Variables Removed</td>
<td>143</td>
</tr>
<tr>
<td>20. Intercorrelations of Variables Employed in the Post-Hoc Regression Analysis for Group a of Component III - the P-test Tasks Investigation</td>
<td>144</td>
</tr>
<tr>
<td>21. Summary Regression Table of Math 100 Final Exam Variance Accounted for by the P-test Tasks with Variance Accounted for by other Variables Removed</td>
<td>145</td>
</tr>
<tr>
<td>22. Intercorrelations of Variables Employed in the Regression Analysis for Group b (Math 102) of Component III</td>
<td>147</td>
</tr>
</tbody>
</table>
### LIST OF TABLES (Contd.)

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>23. Summary Regression Table Used to Test Hypothesis $H_{6b}$: Math 102 Final Exam Variance Accounted for by P-test with Variance Accounted for by other Variables Removed</td>
<td>147</td>
</tr>
<tr>
<td>24. Intercorrelations of Variables Employed in the Regression Analysis for Group c (Math 103) of Component III</td>
<td>149</td>
</tr>
<tr>
<td>25. Summary Regression Table Used to Test Hypothesis $H_{6c}$: Math 103 Final Exam Variance Accounted for by other Variables Removed</td>
<td>149</td>
</tr>
<tr>
<td>26. Intercorrelations of Variables Employed in the Regression Analysis for Group d (Math 148 and 150) of Component III</td>
<td>152</td>
</tr>
<tr>
<td>27. Intercorrelations of Variables Employed in the Regression Analysis for Group e (Math 151) of Component III</td>
<td>154</td>
</tr>
<tr>
<td>28. Summary Regression Table Used to Test Hypothesis $H_{7a}$: Math 151 Final Exam Variance Accounted for by P-test with Variance Accounted for by other Variables Removed</td>
<td>155</td>
</tr>
<tr>
<td>29. Correlations Used for the Test of Hypothesis $H_8$ for each Mathematics Class Group</td>
<td>157</td>
</tr>
<tr>
<td>30. Comparison of P-test Means by Sex for Math 100, 102, 103, and Math 148 Combined with 150</td>
<td>158</td>
</tr>
<tr>
<td>31. Comparison of P-test Results to those Derived by other Researchers</td>
<td>165</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Measurement of Angle A with a Perpendicular</td>
<td>47</td>
</tr>
<tr>
<td>2. Response Made by P2-IIA</td>
<td>87</td>
</tr>
<tr>
<td>3. Response Made by P2-IIB</td>
<td>89</td>
</tr>
<tr>
<td>4. Response Made by P2-IIIa</td>
<td>92</td>
</tr>
<tr>
<td>5. Response Made by P2-IIIb</td>
<td>94</td>
</tr>
<tr>
<td>6. P-test Score Distribution</td>
<td>125</td>
</tr>
<tr>
<td>7. Distribution of Math 100, 102 and 103 on the P-test Scale Level</td>
<td>138</td>
</tr>
<tr>
<td>8. Distribution of Math 148, 150 and 151 on the P-test Scale Level</td>
<td>139</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Statement of the Problem and the Rationale

This study reports the development of a "paper and pencil" test, called the P-test, designed to determine the different levels of reasoning abilities of young adults. The test was developed to provide a set of written tasks that would measure a subject's level of concrete or formal reasoning and obtain results that could be claimed to be equivalent to results obtained from administration of the tasks in the usual Piagetian clinical interview approach.

The motivation for the development of this test was a desire to explore mathematics learning. An underlying assumption of this study is that an extension of Jean Piaget's theory of cognitive development would be a logical avenue to follow in the exploration of mathematics learning. This assumption derives from the nature of Piaget's theory. He deals with intelligence and knowledge from the developmental point of view. He attempts to explain how structures are used by the child to gain and organize knowledge and his theory explains how different structures are related to each other in a developmental sequence. Based on the cognitive structures a child possesses, he predicts what concepts a
child is capable of understanding and explains why a given child is unable to learn other concepts. Applications of such a theory should be useful to the exploration of mathematical learning. However, this dissertation does not pretend to attempt to explain what mathematical concepts a young adult should be capable of or exactly why he or she does not understand a given mathematical concept. Such a goal is a long range objective of the experimenter and this dissertation represents only a first step. Piaget has provided the theory and this study attempts to provide a useful tool to extend and apply that theory.

As another step toward the exploration of mathematics learning, this study will examine the P-test relationship to mathematical achievement in college mathematics courses and to standardized measures of academic achievement. This examination seems like a logical step toward a beginning of the exploration of mathematics learning based on Piaget's theory.

Besides the specific application to mathematics learning, an extension of Piaget's theory should have other obvious applications for Education in general. John Flavell, a noted interpreter of Piaget's work, devotes a section of his book, The Developmental Psychology of Jean Piaget, to the enumeration of these applications. Among these are: the assessment of students' general intellectual development to be used for

proper grade placement, planning curricula so that the level of concepts taught is compatible with the structures students possess for gaining knowledge, and developing effective instructional methods which might result from an understanding of the structures or schemata that children, adolescents, and adults use to organize knowledge.

Piaget used an interview approach to evaluate a subject's level of reasoning. This approach consisted of the individual administration of a problem task to the subject, and levels were based on the subject's reactions. The researcher probed the subject to attain all possible information about his or her ability to reason in the given situation. Piaget called this method the clinical approach and he described it in the introductory chapter of his book, *The Child's Conception of the World* (1960). Since Piaget was more interested in charting intellectual development and constructing a theory, he did not worry about standardization of his problem tasks to make psychometric measurements possible. Flavell stated that a standardization of Piaget's cognitive tasks, "... would surely be a logical extension of his (Piaget's) work. . . ."\(^2\) The P-test is an attempt to standardize four Piagetian tasks. But this test goes one step further than standardizing Piaget's clinical approach. This study attempts to eliminate the time consuming individual administration of the tasks by the development of a "paper-and-pencil" group test to evaluate

\(^2\)Ibid., p. 361.
a subject's level of reasoning. Piaget's own statements suggest that the construction of such a test is possible. In his description of the clinical method, Piaget stated that, "...it is obvious that with sufficient ingenuity, the tests (psychometric measures) can be so varied as to reveal all the components of a given psychological reaction."³ Piaget warned of the risk of falsification of a subject's natural mental inclinations with such tests, but since the items developed for the P-test were taken from descriptions of Piaget's interviews with subjects it was felt that this risk was minimized. Piaget also stated, "Now, the most prominent feature of formal thought is that it no longer deals with objects directly but with verbal elements."⁴ Thus the successful assessment of formal reasoning in the verbal context of a paper-and-pencil test seems plausible.

Piaget also said:

... it is possible to get correct reasoning about simple propositions as early as the 7-8 year level, provided that these propositions correspond to sufficiently concrete representations.⁵

The P-test attempts to include concrete representations for propositions in order to assess subjects' level of concrete 


⁵Ibid., p. 252.
reasoning as well as their ability to reason at the formal stage.

The strongest statement in support for the possibility of a successful construction of a written Piagetian test came from Piaget's statement:

... propositional logic appeared in its most characteristic forms as readily when the subject dealt with our experimental apparatus as when he was confronted with a purely verbal task.  

From the passages quoted it was concluded that a paper-and-pencil test could elicit similar reasoning behaviors as some of those found by Piaget in his clinical approach. That this is the case with the P-test will be demonstrated in this study.

Background for the Study

Piaget's theory postulated four stages of cognitive development. Each stage is defined by the schemata or structures that are characterized by the behaviors used by a subject to achieve cognitive knowledge of his or her world. The four stages are: (1) sensory motor (age 0-2); (2) pre-operational (ages 2-6); (3) concrete operations (ages 6-11); (4) formal operations (ages 11-16). The last two stages are those investigated in this study.

Piaget formulated two substages for each of the stages, concrete operations and formal operations. (Inhelder and

---

6Ibid., p. 253.

At the concrete stage the two substages are labeled IIA, early concrete, and IIB, late concrete. The main difference between the two is that subjects at the IIB, late concrete, level are capable of the formation of categories using two or more criteria. For example, a level IIB subject is able to sort geometric attribute blocks by both color and shape; without difficulty he or she can form a class of blue triangles or classes of large red circles, etc. The level IIB subjects also can find intersection and form unions of such classes. The level IIA, lower concrete, subject is limited to operations of class inclusion and simple serial ordering operations.

The formal substages are labeled IIIA, early formal, and IIIB, late formal. Subjects at these levels have fully mastered the operations described for the concrete stage, and, in addition, they are capable of reasoning hypothetically, possess a system of combinatorical operations which give rise to the hypothetical thought, and have the ability to use proportional reasoning schemes. The primary difference between subjects' behaviors at these two levels is the extent to which their experimentation in a situation takes into account the possible hypotheses that can be entertained. Stage IIIA subjects are tempted to draw conclusions before all possible alternatives that need to be tested are investigated. The need of stage IIIB subjects for definitive proofs derives from their abilities to reason hypothetically and consider all possible combinations definable in a situation; this need is one
of the dominant characteristics of the level IIIB reasoners.

While Piaget listed the age range from 6 to 16 as the time when children and adolescents achieve the concrete and formal levels of reasoning, recent research with American college freshmen indicated that 50 percent or more of these young adults are at or below the concrete stage. These studies include work by Karplus and Karplus (1970) and McKinnon and Renner (1971). Such findings raise serious questions. Do these findings hold true for the majority of college freshmen? If so, what reasoning skills do these students possess? What level of reasoning can they be expected to reach and can this situation be corrected so they can achieve a formal level of reasoning earlier? To answer these questions a great deal of research is required. The development of the P-test is one step in this direction. The P-test should provide an efficient tool for conducting research on Piaget's defined levels of reasoning at the concrete and formal stages.

In addition to the development of the Piagetian measure, a second objective of this study is to investigate the relationship between a subject's level of reasoning and his or her achievement in a college mathematics course. In the development of mathematics placement procedures for The Ohio State University, Crosswhite (1963) tested the predictability of 24 measures of academic achievement on success in freshman mathematics courses. His findings indicated that ACT and SAT scores, both batteries of standardized tests of
achievement, were less predictive of the criterion than was a content placement test and a measure of high school mathematics performance. It seems that the standardized achievement tests, although significantly correlated with the criterion, did not measure anything different from the placement examination and the high school mathematics measure. Since then, Bart (1972), DeVries (1974), and Stephens, et al. (1972), have reported that Piagetian tests, although correlating significantly with standardized tests of achievement, measure more than the standardized tests. Because of these recent results, an investigation will be made of whether the P-test can add to the prediction of achievement in freshman mathematics courses.

Method of Study

There are three components to this study. The first is the development and validation of a paper-and-pencil test to measure the Piagetian level of reasoning in college freshmen. To accomplish this, 231 entering freshmen at The Ohio State University-Lima Branch, Fall 1977, were administered the final version of a previously piloted paper-and-pencil test called the P-test. In addition, a sample of 27 students were also given two Piagetian tasks through the clinical interview procedure. Responses on the P-test are evaluated by comparison of behaviors elicited to those described by Piaget for concrete and formal reasoners. The interview scores and results of the P-test for the sample of 27
subjects are also compared by statistical procedures.

The second component consisted of comparisons between subjects' P-test results and their scores on other standard tests of achievement. Particular attention is given to the relationship between the P-test and the American College Test's mathematics examination, the Math ACT test.

The third component investigates the relationship between the P-test and mathematics achievement in six different freshman-level mathematics courses. Achievement is defined as the score a subject receives on the final examination in a mathematics course. The traditional predictors of success in these courses is a set of mathematics placement examinations combined with a measure of a student's high school mathematics performance. The power of the P-test to add to these variables' predictability of mathematics achievement is tested.

Hypotheses

The general hypotheses for this study are:

(1) It is possible to convert Piagetian tasks into a paper-and-pencil test which will evaluate a subject's level of reasoning; assessment with this instrument would be consistent with results from Piaget's clinical interview method.

(2) The college freshmen given the P-test will be distributed into each substage reasoning level with a large percentage judged to be at the concrete levels.
(3) The P-test is highly related to standard measures of academic achievement.

(4) The P-test will improve the prediction of mathematics achievement provided by the placement test and high school mathematics performance in college freshman mathematics courses at The Ohio State University. Specific hypotheses which can be tested statistically will be distilled from the general hypothesis stated in this section. These hypotheses as stated here outline the broad objectives for this study. They represent a first step toward what is assumed to be a productive approach to exploration of mathematics learning in general and to mathematics learning by young adults in particular.

Outline of the Study

In the preceding pages an effort was made to provide the reader with an understanding of the problems studied in this dissertation. A brief description was given of the methods used in the research and a summary of the study goals was made through the statement of general hypotheses.

The results of earlier studies and other related literature are discussed in Chapter II. Chapter III is a specific description of the methods used in this study. Three components for investigation are described; the populations, instruments and variables measured are discussed. Nine hypotheses are stated and statistical procedures are outlined for the test of these hypotheses. The results of the
first component, an investigation of the development and validation of the P-test, is reported in Chapter IV. In Chapter V, the results are reported for the second component, an investigation of the relationship of the P-test to other standard tests of academic achievement. Also included is the third component in which the power of the P-test to predict mathematics achievement in six mathematics courses is tested. Conclusions, implications and recommendations for further study are explored in Chapter VI.
CHAPTER II
REVIEW OF RELATED LITERATURE

The research and theories of Jean Piaget have been reviewed and discussed by many authors in great detail.\(^1\),\(^2\),\(^3\),\(^4\),\(^5\),\(^6\) Special aspects of his theory as they relate to this dissertation are also discussed in Chapters I, III, IV, and VI. Thus, the literature reviewed in this chapter will be limited to that which deals with or is related to formal reasoning in young adults of college age and late adolescence.

Even for early adolescence, research on formal reasoning abilities has not been extensive. Parsons and Milgram, in


their introduction to Inhelder and Piaget's book, *The Growth of Logical Thinking from Childhood to Adolescence* (1958), state, "... the stage of concrete operations has probably been the more extensively studied ..." than formal operations. In his review of Piagetian research before 1963, Flavell states, "The majority of learning studies deal with acquisitional processes in the two related areas: number and quantity." Both areas deal with concrete operational concepts. In his review of research on logical operations, the studies Flavell cites involve logical class-inclusion, logical classification, sorting behavior, and sorting rationals; all of these topics are operations on classes and form the substance of concrete operations. Flavell brings to light the amount of research that might be expected for formal thought with the following appraisal.

> There has probably been more written about Piaget's earliest, pre-1930 investigations - more reviews, more critiques, more research articles - than about all his subsequent work put together.

Piaget's and Inhelder's texts which review formal reasoning were published after 1950. And since around 1960 some

---


9 Ibid., p. 397.

10 Ibid., p. 379.
increase in interest in Piaget's stage of formal reasoning has been found among researchers.

Confirmation of Piaget's Results

Most of the research about formal operations conducted in the early and late sixties can be categorized as replications of Piaget's work. The most prominent is that by Lovell (1961). Lovell tested 200 British students between the ages of 8 and 18 years on ten of the sixteen formal reasoning tasks presented in *The Growth of Logical Thinking from Childhood to Adolescence* (1958) and confirmed Piaget's results. Lunzer (1965) also replicated several of Piaget's tasks, in particular those on geometric and proportional reasoning. He reported validations of Piaget's work through the testing of subjects on tasks he developed which required the same types of reasoning skills outlined by Piaget. Besides the confirmation of the existence of the psychological structures Piaget postulated, these two studies gave evidence of another significant result. Both researchers found that larger than expected numbers of their subjects had not attained any stable formal reasoning skills. Lovell concluded "... it is only rarely that the average to bright junior high school children reach the stage of formal thinking."¹¹ Lunzer drew the same.

conclusion from his results. Elkind (1961) made similar remarks after testing American secondary school students on conservation of volume tasks. In general, Lovell, Lunzer, and Elkind found 50 to 75 percent of their adolescent subjects at or below a concrete reasoning level.

Other researchers also finding that between 50 and 75 percent of the adolescents they tested were at or below the concrete level were Jackson (1965) and Hughes (1965).

Formal Reasoning and Older Subjects

Elkind (1962) was the first to extend his investigations of subjects' formal reasoning levels to American college students (young adults). Surprisingly, 42 percent of his college subjects had no abstract concept of volume. This finding seemed to be the first to give impetus to investigations of the extent to which young adults as well as adolescents have developed formal cognitive operations.

McKinnon (1976) gave a comprehensive report on the state of college-age subjects' levels of reasoning. He used eight Piagetian tasks, two of which were "Elimination of Contradiction" and "Exclusion of Irrelevant Variables," designed by Inhelder and Piaget (1958) to assess formal reasoning. He

---


selected 185 students from seven different institutions of higher education: a private nonsecular liberal arts college, a private sectarian college, a large public state university, two state colleges, a traditional junior college, and a private city community college. McKinnon found that 51 percent of these subjects' responses were evaluated as concre­
operational. Only one institution, the private nonsecular liberal arts college, had more than 50 percent of its sub­
jects evaluated as capable of formal reasoning but still less than 60 percent.\textsuperscript{14}

Kohlberg and Gilligan (1971) tested American subjects over a wide range of ages. Even though they used only one task, Inhelder and Piaget's pendulum task, the results are interesting because of the age range. The age intervals and percentage of those subjects evaluated at the formal level on the pendulum task were ages 10-15 with 45 percent, ages 16-20 with 53 percent, ages 21-30 with 65 percent, and ages 45-50 with 57 percent.\textsuperscript{15} Tomlinson-Keasey (1972), using five Piagetian tasks, obtained similar results with a sample of women ranging in age from eleven to fifty-four.

Other researchers who have tested random samples of college students and found that about 50 percent are


\textsuperscript{15}Lawrence Kohlberg and Carol Gilligan, "The Adolescent as a Philosopher: The Discovery of the Self in a Post­

Those researchers who tested secondary school students have found even a smaller percentage of subjects operating at a formal level. These include Lawson (1973), Lawson and Renner (1974), Lawson, Nordland, and Kahle (1975), and Lawson and Blake (1976).

Common to all the research reviewed thus far is an agreement with Piaget that there are two qualitatively different levels of cognitive operations and that concrete operations are a prerequisite for the attainment of formal operations. Besides this reaffirmation of Piaget's psychological constructs, they point up the fact that the number of adolescents and adults who seem not to reach the level of formal operations is larger than may be expected and that the majority of those who attain the formal level do so at a much later age than Piaget had previously predicted.

While some researchers are interested in the fact that many adults may not achieve formal operations, Arlin (1976) has proposed the existence of a fifth stage. Arlin describes this stage as the "problem finding" stage as opposed to the "problem solving" stage usually referred to as the formal stage. Arlin reports that while her results do not give conclusive evidence for a fifth stage beyond formal operations,
the results warrant further investigations.\textsuperscript{16}

Formal Reasoning and Intelligence

In a review of literature which examines the relationship between a subject's mental quotient (IQ) and his stage of reasoning, low but statistically significant correlations have been obtained. Evidence has been produced which indicates that Piagetian cognitive level assessments measure more than general intelligence. Studies drawing this conclusion have done so by demonstrating that a Piagetian measure correlates significantly with a number of other psychometric intelligence measures and achievement measures, but that the latter two do not correlate as highly with each other (e.g., Bart, 1972). Stephens et al. (1972), DeVries (1974), and Bart (1971) used factor analyses to conclude that Piagetian tasks measured more than other psychometric intelligence measures.

Two studies, one by Webb (1974) and another by Yudin (1966), give more information about the relationship between IQ and the Piagetian level of cognitive reasoning. Webb tested 25 children who were from 6 to 11 years in age and whose IQ's were in excess of 160. Because their mental ages are beyond 20 years and because only four subjects demonstrated any formal reasoning ability, he concluded that the development of formal cognitive reasoning is tied to

maturity. Yudin tested 36 subjects from 12 to 16 years old and with IQ's from 80 to 130. He reported that the higher the IQ score, the younger the age at which a subject was likely to demonstrate formal reasoning abilities. From these two studies, it may be concluded that there is an interaction effect between age and intelligence in the development of formal cognitive skills. It seems that more research is needed to explore this interaction further in light of the great number of young adults who have not attained a formal level of cognition.

Formal Reasoning and Academic Achievement

Predictions of academic achievement by Piagetian measures of cognitive level are present in five different research reports. Two of these involved reading ability. Lawson, Nordland and Kahle (1975) tested 35 high school subjects 14 to 17 years of age on 10 Piagetian tasks and the STEP-Reading Test. They obtained a correlation of $r = 0.70$ between scores derived from Piagetian tasks and the reading scores. Raven, Hannah, and Doran (1974) obtained

---


$r = 0.406$ between a Piagetian test developed by Raven (1973) and the Iowa Silent Reading Test, Advance Form AM, for a sample of 123 college freshmen.\(^{20}\)

Raven, Hannah, and Doran (1974) also obtained similar correlations between Raven's test, two logic tests, and two science content standardized tests.\(^{21}\) The subjects they tested were all enrolled in an introductory physical science course. No report was given about the subjects' success in the course.

Sayre and Ball (1975) tested how well a score derived from five Piagetian tasks predicted scholastic grades in science for junior and senior high school students. After splitting the samples by those assessed formal or concrete, they found that both junior and senior high school formal reasoners received significantly higher grades than non-formal reasoners. When the sample was divided by grade level the results were mixed. For grades 8 through 11, the Piagetian assessment was a good predictor of science grades, but for the seventh graders and twelfth-grade physics students it was not a good predictor. Sayre and Ball pointed out that the last two grades were at the bottom and top ends


\(^{21}\)Ibid., p. 565.
of the Piagetian scale.  

Lawson and Renner (1975) also studied the relationship between Piagetian levels of cognitive development and achievement in high school science. They prepared an examination for each of three classes, a biology class with 51 students, a chemistry class with 50 students, and a physics class with 33 students. Each examination consisted of 30 content questions for each class. Fifteen of the questions were judged to require formal reasoning skills and fifteen were judged to require only concrete reasoning ability. Correlations between the subjects Piagetian assessments and each part of the examinations were significant at p .05. Upon inspection of the correlations the researchers concluded that, "... there appears to be a more positive correlation between the tasks and understanding of formal concepts than of concrete concepts."

Last, Lawson, Nordland, and DeVito (1975) reported a study on the relationship of formal reasoning to achievement, aptitude, and attitude in pre-service teachers. They used standardized tests covering achievement and aptitude in mathematics, science, and English, and they used a standard attitude measure. They reported correlations which are


significant at the .05 level and above, and which range from .25 (with high school rank) to .53 (with a verbal aptitude test). They concluded from their results that formal-operational reasoning abilities are significantly related to achievement, aptitude, and attitude for their sample of 71 subjects.

Accelerating Attainment of Formal Reasoning

Because Piagetian reasoning level seems to be related to academic achievement and because of the reports that large numbers of adolescents do not achieve formal operational status, a few studies have investigated the possibility of speeding up the transition from concrete to formal operations. The results of these studies have been less than successful. The issues revolve around what is called specific and non-specific transfer. Specific transfer refers to the subjects' ability to solve a task that is very much like one they were trained or taught to solve. For example, the experimenter creates two tasks in which a number of variables cause some reaction. Subjects are trained to control variables in order to solve a problem in one experimental situation and then asked to complete the second experiment. If they are able to complete the second experiment, then


25 Ibid., p. 430.
specific transfer is judged to have taken place. Non-specific transfer testing involves the subjects' ability to solve a task that requires different but equivalent reasoning skills than used in the training sessions. For the example given above, a task requiring proportional reasoning such as the "Equilibrium in the Balance" task developed by Inhelder and Piaget (1958) could serve as the non-specific transfer task. Both the controlling of variables and proportional reasoning are considered formal reasoning tasks. Thus, the test of non-specific transfer is a more general test of whether a subject's level of reasoning has advanced. Studies by Lawson and Wollman (1976), Tomlinson-Keasy (1972), and Lawson, Nordland and DeVito (1974) all reported specific transfer taking place, but no non-specific transfer. Lawson and Wollman reported that although their results for non-specific transfer were not significant, those subjects who were judged to be transitional (post-concrete or early formal) in their level of cognitive reasoning made the most growth from pre- to posttests on the tasks in which no training was given.26 This result seems somewhat consistent with the conservation training studies at the concrete stage.

Subjects' progress in these studies depended on their initial

The best training results for formal operations are reported by McKinnon (1970), using two groups of college freshmen. One group (N=69) took a special science "inquiry" course. A second group (N=62) took a regular introductory science course. The researcher claimed that no reference was made to any of the Piagetian tasks used as part of either the pre- or posttests. He found that the group which took the inquiry course made a significantly larger gain on the posttest than did the group which took the regular science class; both groups scored virtually the same on the pretest.

Sex Differences and Formal Reasoning

Another issue that seems to have surfaced in more recent research on Piaget's work is that of sex differences in performance on Piagetian tasks. The basis for such investigations cannot be traced to any of Piaget's work and that this


is the case is reflected by Brekke and Williams' (1973) opening remark in their report on sex differences. "Sex differences have often been compared incidentally in studies of Piagetian conservation."30 Perhaps the reason researchers have investigated sex differences is because of differential maturation rates found between boys and girls.31

The results on sex differences for subjects eleven years or older are mixed. At the college level, Elkind (1962) reported significant differences in favor of the male subjects, and Stone (1966) reported no significant differences. Elkind used only conservation of volume to make his assessments while Stone used tasks different from, but judged to be equivalent to, Piagetian tasks.

At the high school level, Higgins-Trenk and Gaite (1971) reported that female adolescents outperformed males on a measure of formal reasoning in which subjects responded to open-ended situational dilemmas; there was no difference between male and female responses to a Piagetian task. The 162 subjects in their study were 13 to 18 years old. It is more than just interesting to note that the Piagetian task they used was the conservation of volume task used by Elkind (1962) and it was administered by the same procedures

outlined by Elkind for his study in which he found differences in favor of males. Lawson (1975) also examined sex differences among a sample of 62 high school students. He found that boys significantly outperformed girls when subjects were given the apparatus to be manipulated in the Piagetian tasks, but that girls outperformed the boys when purely verbal solutions were required.

At the junior high school level, Keating, Schaeffer, and Rosalind (1975) reported that among high ability students, younger boys outperformed younger girls on Piagetian tasks. They split their sample of 92 subjects into two age groups, 11 to 12 and 12 to 13 years of age, and into two ability groups, bright and average. The Iowa Test of Basic Skills was used to assess ability. Except for the younger bright group, there were no sex differences found within the subgroups. Roberge (1976) also reported mixed results at the junior high school level. On a test of combinatorial reasoning, sixth grade boys did significantly better than sixth grade girls, but on the same test, seventh grade girls outperformed the seventh grade boys.

Reported from the literature cited were sex differences favoring males, sex differences favoring females, no sex differences and mixed results dependent on ages or type of test. A case cannot be made for sex differences on Piagetian tests from these results. The only conclusion that can be made is that more research is needed to examine the situation.
Written Tests of Formal Reasoning

Evidence has been cited earlier for the case that the body of research on the formal stage is much smaller than that for the earlier stages. This may be an artifact of the population which must be sampled to conduct research on the formal stage. Subjects expected to be capable of formal reasoning should be at the junior high school level or higher. It seems that schedules of students in the secondary school and colleges are less easily disturbed than schedules for elementary school students. To assess subjects level of reasoning by using Inhelder and Piaget's tasks requires individual interviews with each subject; this, together with the inaccessibility of subjects, makes it difficult to test large samples. The need for a large group test which could produce equivalent results to the interview administrations would help alleviate this problem.

Since 1972 reports on four tests have been made in which subjects are administered the measures in large groups. Bart (1972) constructed logic items in three content areas, biology, history, and literature. He examined validity of these tests by comparing results obtained on them with results obtained on a measure derived from four Piagetian tasks and concluded that the correlations were too low to claim that the two tests measured the same psychological variable. His assessment of the state of Piagetian test development at that time was that "there still exists no strictly valid instrument (paper-and-pencil) testing formal
Raven (1973) constructed a paper-and-pencil test with items to test skills in the following areas: classification, seriation, logical manipulation, compensation, probability correlations and proportions. While the topic areas are consistent with those researched by Piaget, he did not attempt to compare his test to Piagetian tasks nor tie directly his scale to the levels of reasoning outlined by Piaget. Because of these limitations it is hard to compare his results to those obtained in other Piagetian studies.

Rowell and Hoffmann (1975) report their successful efforts to convert two Piagetian tasks, the pendulum task and the colorless chemicals task, into a group-test version. But they did not compare their results to actual interviews and their test is not a true paper-and-pencil test. The researchers provided each subject with the required apparatus to perform experiments called for in each task.

The most recent attempt, that by Renner (1977) seems to have met with some success. This researcher examined twelve problem tasks that dealt with familiar objects and events but were similar to tasks given by Inhelder and Piaget. The results for 811 high school subjects on the twelve items were compared to results on Piagetian tasks obtained through the clinical interview procedure. Renner used

a regression analysis to predict the interview scores from scores obtained in the paper-and-pencil tasks. Four of his tasks, three requiring proportional reasoning and one requiring the ability to sort out variables in experimentation, produced a multiple R of 0.62. This prediction may at first seem high, but when one considers that this can be interpreted to mean that their four items account for just a little more than 36 percent of the variance on the interview assessments, the result is disappointing.

In an intensive review of literature on Piagetian research, Cowan (1978) states, "I am aware of no studies which compare a clinical (interview approach) and a psychometric (paper-and-pencil) assessment approach using the same kinds of items." The two studies reviewed in this paper that made such comparisons met with limited success. Thus the need for such a test remains.

Summary

Several observations are easily made from the literature reviewed. First, exactly why large numbers of adolescents and adults have not achieved formal operational thought is a mystery. Second, whether or not educational programs can

---

33 John W. Renner, "Evaluating Intellectual Development Using Written Responses to Selected Science Problems," A Report to the National Science Foundation, 1977 (Number EPP75-19596), p. 120.

foster movement from the concrete to the formal stage is an open question. Third, while there are some reports on the relationship of Piagetian cognitive levels to academic achievement, little research, if any, has been done in the area of mathematics achievement. Fourth, the possibility of a fifth stage beyond formal reasoning should be explored. And, fifth, there is presently no valid paper-and-pencil test which may be used to facilitate research on the formal level of cognitive development. If the gap noted in the fifth observation were filled, then research pertaining to the other four observations could be carried out in a systematic and efficient manner. The study completed for this dissertation attempts to develop a valid paper-and-pencil test and to measure Piagetian cognitive level of reasoning. It also will investigate the relationship between level of cognitive reasoning and mathematics achievement for a college population.
CHAPTER III
RESEARCH PROCEDURES

Introduction

This study has three components: Component I is the development of a test to be called the P-test, which is derived from the work of Jean Piaget; Component II is an investigation of the relationship of the P-test with standardized mathematics achievement; and Component III is an investigation of the power of the P-test to predict mathematics achievement in six freshman mathematics courses.

General Background for the Three Study Components

Populations

The population for this study consists of entering freshmen at The Ohio State University-Lima Campus, Fall quarter, 1977. The P-test was administered to 231 students from a total freshman population of 309 during the first week of the 1977 Fall quarter. These subjects served as the population for Component I. Of the 231 subjects identified for the first component, a sub-population of subjects (N=196) who took the American College Test (ACT) series was identified as the population for Component II. Those subjects in the Component I population who were enrolled Fall
quarter in one of six mathematics courses were identified as the population (N=171) for Component III.

The P-test

The P-test is an instrument with four sections of questions called the P-test tasks patterned after tasks developed by Jean Piaget or researchers who have extended Piaget's work. Piaget identified four developmental stages of reasoning: sensory-motor, pre-operational, concrete operational, and formal operational. This study will investigate the abilities of college freshmen to operate in the formal operational stage. The main interest of this study is to distinguish concrete operational and formal operational reasoning patterns among subjects by means of a paper-and-pencil test. For this purpose, Piagetian tasks which are administered in clinical interviews were examined for their potential conversion to a paper-and-pencil version. These Piagetian tasks are described in The Growth of Logical Thinking from Childhood to Adolescence (Inhelder and Piaget, 1958) and The Child's Conception of Geometry (Piaget, Inhelder, and Szeminska, 1960). The development of the P-test involved the selection of tasks from the two books listed which could be described and responded to on paper. That selection and try-out processes is described in the later section, "Development of the P-test."
ACT scores were available for 196 subjects in the total study population; this number included 156 out of the 171 subjects identified as the population for Component III, the investigation of the predictive power of the P-test in mathematics courses. The ACT scores were obtained from The Ohio State University's testing center through the office of the Associate Director of the Lima Campus. The ACT scores include an English score, a Mathematics score, a Social Science score, a Natural Science score and an ACT composite interval scale score called the ACT level. Standardized scores are used as a subject's score on each of the ACT individual tests.

Two standardized mathematics examinations called the D-placement test and the B-placement test were administered by The Ohio State University-Lima Campus Academic Advisement Office. Both of these tests examine a student's mathematics content knowledge. The D-placement test covers algebra, trigonometry, analytical geometry and complex numbers. The B-placement test includes basic algebra and topics from plane geometry.

The last measure investigated in Component II is an assessment of a subject's high school mathematics performance. This variable will be called H.S. Math Points. Each entering freshman at The Ohio State University is assigned a score, H.S. Math Points, which is computed by the
Academic Advisement Office based on a questionnaire filled out by the student. The student reports his or her high school mathematics courses and the grades received in these courses. Four points are assigned for every A, three for every B, two for every C and one for every D; the sum of these scores is the student's measure of high school mathematics achievement.

The placement tests and H.S. Math Points will also be used in the analyses to be outlined for Component III. They are included in Component II in order to analyze their relationships with the P-test in the larger study populations; in Component III smaller sub-populations will be identified for different investigations.

Mathematics Placement Procedures at The Ohio State University

The placement of entering freshmen in The Ohio State University mathematics sequence is a two-step procedure. A student's Math ACT standard score is first used to determine which of two placement exams a subject will take. If a subject has a Math ACT standard score of .25 points or higher he or she takes the D-placement test; the score he or she receives on this test together with the score he or she has been assigned for his or her High School Math Achievement are used to assign the subject a mathematics placement level of either I, IIA, or IIB. To take Math 151 a subject must have a Math placement level of I; to take Math 150 a subject must have a placement level of IIA; and to take
Math 148 a subject must have a placement level of IIB. Since very few subjects who take the D-placement test fail to achieve a level of at least IIB, any subject who would not attain this level will not be included in the populations for Component III of the study. The reason for this elimination is that a subject who achieved a Math ACT standard score of 25 or better would be different initially from the subjects placed at a level below IIB despite his or her placement examination score.

Those subjects who attain a Math ACT standard score below 25 take the B-placement test. The score subjects received on this test together with their High School Math Achievement Score are used to assign them to one of the mathematics placement levels III, IV, VA and VB. The highest mathematics course considered in this study that could be taken by a subject with a given mathematics placement level is: Math 103 with a level III placement, Math 102 with a level IV placement and Math 100 with a level VA or VB placement.

These placement procedures were developed from the doctoral dissertations of Crosswhite (1964) and Zwick (1964) and are described in greater detail in those sources.

**The Mathematics Courses Offered to Freshmen**

The six Ohio State mathematics courses included an investigation of Component III range in content from beginning elementary algebra to a first course in calculus.
These courses will be called by their Ohio State Mathematics Department's sequential course numbers. Math 100, 102 and 103 are the first three courses in the math sequence; their content consists of a review of the algebra topics normally found in a two-year high-school algebra program. Math 148 is the first part of a two quarter sequence called the 140 series; its content is college algebra. Math 150 is the traditional college algebra and trigonometry course which covers in one quarter the content of the 140 series. Math 151 is the usual first college level course in calculus and analytic geometry. These courses represent each of the different levels at which a freshman could enter The Ohio State University mathematics sequence in the fall of 1977 at the Lima Campus. There are several other courses offered in the fall, but they are not primarily freshman mathematics courses.

Mathematics Achievement in the Mathematics Courses

In Component III, the investigation of the predictive power of the P-test in six mathematics courses, the final examination grades in each of the six courses described above will be taken as a measure of a subject's mathematics achievement.

The experimenter had little actual control over the formation of the final examinations. His input to these examinations consisted of a request that the instructors include at least four questions which required more than rote
learning for a solution. Specifically, each instructor was asked to include problems which were unlike any examples presented during the course but could be solved with a good understanding of the concepts within the course. This portion of the test could be viewed as that part measuring formal reasoning in mathematics and the remainder of the test could be viewed as a concrete component. This additional feature to the construction of the final examination will be explored in Component III. Other than the request for the special problems, it is assumed that the instructors constructed their examinations as they have done in the past.

The experimenter also had no control over the teaching method used by any of the instructors. No suggestions were even made as to which method was preferred by the experimenter. Instructors taught as they usually did.

Component I: Development of the P-test

Selection of the Tasks

Three tasks developed by Piaget and his associates plus one task developed by researchers following Piaget's work were converted to written tasks for the P-test instrument. The criteria for which Piagetian tasks were selected for conversion involved three factors: (1) how easily they converted from clinical interview versions to paper-and-pencil versions; (2) by results obtained from pilot work prior to
the final draft of the P-test tasks; and (3) the desire to include a set of tasks that require a range of the formal operational schemata Piaget described for the stage of formal operations. Description of those tasks selected and the schemata involved in the solutions of the tasks is present in a separate section.

For the discussions that are presented throughout this dissertation the converted written tasks will be referred to as the P-test tasks and other tasks directly or indirectly related to Piaget's work will be referred to as Piagetian tasks.

The pilot work for the development of the P-test tasks was conducted in the Summer of 1977. Volunteer subjects taking courses at The Ohio State University-Lima Campus were given drafts of sample tasks. Sixteen subjects in all were given one or more of the items being considered for the test. Based on a comparison of their responses to those anticipated, changes were made in the tasks. Six of these students also volunteered to be interviewed about their responses and reactions to the P-test tasks. Based on these interviews questions on the tasks were rewritten so as to eliminate ambiguities and to achieve the maximum likelihood that subjects would interpret the questions as intended. The tasks were also examined by colleagues of the researcher to be sure that someone expected to be capable of formal reasoning could answer the questions in the tasks. A chemistry teacher, a physics teacher, a developmental
psychology teacher, and a philosophy teacher all read the items and made suggestions.

After consideration of results of the pilot work, four tasks were selected and rewritten for the P-test. Two of the tasks were selected from those described by Inhelder and Piaget in *The Growth of Logical Thinking from Childhood to Adolescence* (1958). These tasks are titled by their developers as (1) "The Oscillation of a Pendulum and Operations of Exclusion" and, (2) "Flexibility and Operations Mediating the Separation of Variables." In the P-test descriptions these tasks will be labeled the Pendulum Task (P1) and the Flexible Rods Task (P3), respectively. A full description of these tasks is given in Chapters 3 and 4 of *The Growth of Logical Thinking from Childhood to Adolescence* (Inhelder and Piaget, 1958); a brief description of the two tasks is also given when the paper and pencil versions are described in this chapter.

The third Piagetian task which was adapted for the P-test is taken from *The Child's Conception of Geometry* (Piaget, Inhelder, and Szeminska, 1960), Chapter 8, section II. It is titled "Measurement of Triangles;" in the P-test this task will be labeled the Triangle Task (P2).

The last task selected for the P-test involves proportionality. It was felt that Piaget's "Equilibrium in the Balance," a classical task to measure proportional reasoning, would not translate into a paper-and-pencil task for college age students. Therefore, proportional reasoning
tasks developed by Karplus (Karplus, Karplus, and Wollman, 1974) and Zepp (1975) were investigated. Zepp used Karplus' work and developed an eight-item test. Four of Zepp's items were selected for the fourth P-test task on the basis of whether they could elicit the behaviors which Piaget describes in his discussion of the "Equilibrium in the Balance" task. The last task will be labeled the Proportional Reasoning Task (P4) throughout the rest of the study.

To be sure that the correct P-test tasks are associated with each of the levels given above, they are presented together below:

P1 Pendulum Task—The Oscillation of a Pendulum and the Operation of Exclusion.

P2 Triangle Task—The Measurement of Triangle.

P3 Flexible Rods Task—Flexibility and the Operations Mediating the Separation of Variables.

P4 Proportional Reasoning Task—The Proportionality Problems.

Description of the P-test Tasks

Nature of the Pendulum Task (P1)

A multifactor experimental situation is derived through the exploration of those factors which affect the period of a pendulum. Four factors may be considered when a pendulum is set into motion: they are (1) the weight of the bob, (2) the length of the string, (3) the release point or amplitude of the pendulum's arc, and, (4) the force of the push given the bob when it is released. All of these are
obvious potential factors that can be manipulated by changing their degree of measure (weight, length, amplitude, and force). The only operant factor is the length of the string. As the string is lengthened (shortened) the period increases (decreases).

The Schema for the Pendulum Task (P1)

For success in this task a subject must be able to set up experiments in which a single variable is manipulated while all other variables are held constant and in this manner deduce the effects of each factor. Piaget and Inhelder have labeled this schema "all other things being equal," and have characterized it as the act of separating out the variables.¹

The implementation of this schema presupposes the operation of a formal combinatorial system and the use of hypothetico-deductive reasoning.² To understand the formation of this presupposition it is necessary to describe the differences between the behaviors of the concrete operational subject and the formal operational subject.

The concrete operational subjects' behaviors are not guided by the principle of isolating a variable in order to observe the variable's effect. Their plan, if they have one, 


²Ibid., p. 46, p. 79, and p. 277.
is to try something to see what happens. They continue in this manner rather than using information derived from previous trials to guide their experimentation. They seem unable to use previously derived information deductively to draw conclusion based on logical necessity. For example, they may observe the effects of changing both length and weight in one trial and they may change length alone in another trial; both situations result in a change in the period caused by the change in the length of the string. The type of conclusion often drawn by concrete operational subjects is similar to a one-to-many correspondence, i.e., "the period changes when . . ." and "the period does not change when . . ." If they do experiment with variation on weight alone it is usually not done out of need to resolve the logical uncertainty of the information derived from earlier trials; they merely place the result among their one-to-many lists.

Since they are accustomed to operations of classification, serial ordering, and correspondences, they are limited to simple tables of variation and do not conceive of the multiplicity of combinations which can be drawn from them.3

In contrast, formal operational subjects demonstrate that their behaviors are guided by the principle of isolating the variables. They seem to be aware of the multiplicity of possibilities among the four factors considered; they also intuit that certain combinations are crucial to

3Ibid., pp. 277.
the determination of the logical relationships among the variables when the result on the pendulum's period is observed.

It is conceivable that each one of the variables that appear together in the experimental situation plays an independent causal role, in which case the crucial combination is the one in which the factors are varied one at a time, the others being held constant . . . But it is also possible that two or three variables have to work together to produce the observed effect . . . Or two variables might be mutually exclusive in the production of the effect, or one might work in the opposite direction from the other . . . We see that separating out variables may lead to any one of a number of distinct possibilities which can be formulated by means of implications, equivalences, disjunctions, conjunctions, exclusions, incompatibilities, etc., depending upon the particular case. When these possibilities already exist in the subject's mind as expectations - i.e., when as a result of previous acquisitions he possesses a wide enough range of possible combinations - it is evident that they can guide the separation of variables by a feedback effect which we can understand easily.4

This quotation explains why it is that the concrete operational subjects, after observing the effect of a change of both length and weight merely catalog it in their one-to-many associations; it also demonstrates why the formal operational subjects are directed to experiments in which each of the variables is varied by itself. The formal subject's ability to entertain all possible combinations is a necessary prerequisite to the formation of logical associations, and both of these result in the schema "all other things being equal."

4Ibid., pp. 287-88.
The Pendulum Task and General Procedures for Categorizing the Responses

A pendulum is pictured on the first page with an explanation of the period of a pendulum. Subjects are asked to tell what factors they believe influence the period. The information is not used in the evaluation, but the question serves to acquaint the subjects more fully with the pendulum task before they begin making responses which will be used to evaluate their level of reasoning. Next, a printed page of nine pendulums which vary by length and weight is presented to subjects for referral in their responses to the rest of the items. Item number 2 is the first to be evaluated. In this item subjects are asked to select three pendulums which they feel can be ordered by their periods. They are asked to order these pendulums from fastest to slowest. Since they do not have real pendulums to manipulate they will have to focus on the two variables, weight and length, which should be obvious variables from their inspection of the pendulum diagrams. Level IIA subjects usually begin comparisons of pendulums based on one factor (weight or length), then they switch to another factor to make subsequent comparisons. This procedure does not produce a correct serial ordering based on their reasoning. Subjects at level IIB and above should have no difficulty ordering three pendulums by weight or length or both; how they order them depends on how they believe the weight and length variables influence the periods. Thus, in addition
to discriminating between level IIB and IIA subjects, this item serves to give the experimenter a clue to subjects' beliefs about the nature of pendulums.

The next item, number 3, presents subjects with a situation in which two pendulums varying in both length and weight have been found to have different periods. Subjects are asked to conclude exactly the cause of the difference. This question checks whether subjects are able to anticipate the schema "all other things being equal." From their ability to formulate all possible combinations for experimentation they are able to select critical combinations that would yield information and recognize those combinations in which no information can be extracted. The question is used to separate upper-level formal reasoners, level IIIB, from lower-level formal reasoners, level IIIA. The later stage subjects may still be capable of isolating variables in an experimental situation, but may not always anticipate its application. With additional prompting provided in the wording of the questions of this task, they are usually led to the application of the proper schema.

Items 4 and 5 are designed to find which subjects are fully capable of isolating the variables, and, thus, they allow for a sort of subjects into concrete and formal levels. Subjects are required to formulate experiments in the following manner: they must choose from the list pendulums which,

---

5Ibid., p. 75 and p. 288.
when tested, allow them to deduce whether a given variable, length, weight, or amplitude plays a causal role in the pendulum's period.

Nature of the Triangle Task (P2)

This situation requires the subjects to draw an exact copy of a triangle. Subjects are provided with a 3 by 5 index card to be used as a straight edge, and as a means of transferring copies of line segments; it is also used as a tool to form perpendiculars with the corners of the card. The subjects are not told how to use the index card, but it is essential that they use it as described in order to complete the task.

The Schema for the Triangle Task (P2)

Piaget claims that when adolescents demonstrate that they can accept the necessary aspects of a situation as a subset of that which can be imagined to be possible, then they have given evidence that they have achieved the formal stage of reasoning.

... in formal thought there is a reversal of the direction of thinking between reality and possibility in the subjects' method of approach. Possibility no longer appears merely as an extension of an empirical situation or of actions actually performed. Instead, it is reality that now is secondary to possibility. Henceforth, they conceive of the given facts as that sector of a set of possible transformations that actually come about; for they are neither explained nor even regarded as facts until the subject undertakes verifying procedures that pertain to the entire set of possible hypotheses compatible with a given situation.6

---

6Ibid., p. 251.
In this task, the set of "necessities" consists of the three line segments and the three angles which form the triangle. In order to copy the triangle accurately, the subject must be able to frame the triangle in a coordinate plane, i.e., as a subset of a two-dimensional field. More specific to this task, the subject must form a perpendicular drawn from one side of the triangle to an adjacent side (or its extension), in order to measure angular separation. The figure below demonstrates how the perpendicular must be drawn; it is the length of the line segment that forms the perpendicular which is a measure of the separation of angle \( A \) that can be transferred to a copy.

![Figure 1](image)

**Figure 1**

Measurement of Angle A with a Perpendicular

This act of forming the perpendicular is either an overt or a subconscious super-imposition of a coordinate system on the plan containing the triangle. It also gives evidence that the subject has viewed the triangle, the sides and the angles, as a subset of the possible, which includes segments connecting two sides and a perpendicular
to one of the sides.

A more detailed discussion of subjects' integration of concrete schema resulting in the ability to complete this task can be found in The Child's Conception of Geometry (Piaget, Inhelder, and Szeminska, 1960).^7

The Triangle Task and General Procedures for Categorizing Responses

This situation consists of only one item. Subjects are each given an index card and a copy of an obtuse triangle at the top of a paper. At the bottom of the paper a space is indicated as the area in which they are to make an exact copy of the given triangle. They are told to use the index card to draw the triangle and that they may use it in any way except that they are not to trace the triangle on the card.

The behaviors at the concrete stage demonstrate little attention to measurement, especially angular measure. Level IIA subjects approach this problem in a nearly topological fashion. They draw an obtuse triangle and use the index card only to draw straight line segments. Level IIB subjects give attention to length of the sides, but not to the angular separation; i.e., he uses the index card to measure the lengths of the sides.

Formal subjects, on the other hand, try to measure angular separation as well as the length of the sides of the triangle. The difference between IIIA subjects and IIIB subjects is the extent to which they measure accurately an angle with the index card. Level IIIA subjects approximate angles through trial and error, while IIIB reasoners use the corner of the index card to draw the required perpendicular.

Nature of the Flexible Rods Task (P3)

A multifactor experimental situation is derived through the exploration of those factors which effect the flexibility of metal rods. The factors presented in this problem are the lengths, diameters, shapes and metal types of the rods. In this task a rod is attached to the edge of the table and a weight is affixed to the other end of the rod. Subjects are required to investigate the flexibility of the various rods by observing how far the weight causes a rod to deviate from a horizontal position. Unlike the pendulum problem, all the factors have an effect on the rods and one factor may work to complement or negate the effect of the second factor.

The Schema for the Flexible Rods Task (P3)

As in the pendulum task, subjects must be able to isolate variables in order to test their effect on the flexibility of the rods. But the fact that one factor may cancel the effect of another or that two different factors may cause the same result makes this situation somewhat
different than the pendulum task. This additional feature requires a schema for handling non-metric proportionality. 8

Piaget describes this schema as the algebraic group of transformations called the INRC group. This logo, INRC, can best be understood by an explanation of the four transformations which are its elements.

The "R" stands for a reciprocal transformation used when a subject reasons that one variable may be altered to cancel the effect of a second variable. For example, in this problem the flexibility of long rods may be negated by their thickness.

The "C" stands for the correlative transformation used when a subject judges that two variables have the same effect on the flexibility of rods. For example, some long rods are very flexible and some thin rods are also very flexible.

The "I" and the "N" are the identity and negation transformations which are most often present in the behavior of children at the concrete operational stage. The identity property can be summarized as the knowledge that if the relevant characteristics of a substance are not changed then the property under observation is unchanged. The negation property can be described as the awareness that when a variable is altered, it causes a change in the

property under observation; then returning that variable to
its original value will cause the change in the property to
be cancelled.

Each of the transformations in the INRC group are
present during the concrete stage, but they are not inte­
grated into a group structure until the formal stage. When
they are integrated, the subject who possesses this schema
is able to test a variable's role in comparison to other
variables in the situation. For the specific behaviors in
this task Piaget describes the schema as a multiplication
between the variables, e.g., longer times thicker equals
shorter times thinner. The equation represents a subject's
understanding that long, thick rods bend about the same as
short, thin rods if all other factors are held constant.

The Flexible Rods Task and General Procedures
for Categorizing the Responses

Items 1 through 6 test subjects' ability to use the
"multiplication" schema which is derived from the INRC
group. In items 1 through 4, five rods are illustrated.
The subjects are told to focus on a specific rod and to
select a second rod which might bend about the same amount
as the specified one. They are asked to give a reason for
their selection. The reason is next in importance to selec­
tion of the correct rod. In all cases, no more than one rod
will completely fit a multiplication equation with the focus

^Ibid., p. 64.
rod according to the attributes of metal, length, shape, and thickness. Level IIIB subjects should be able to consistently select the correct rod. Level IIIA subjects may select an incorrect rod because they ignore one of the four attributes of the rods, but they should demonstrate in their responses that they are using the "multiplication" schema; e.g. "rod A bends about the same as rod B because, while A is longer, B is thinner."

Items 5 and 6 also have rods pictured as in 1 through 4; however, subjects are asked to consider a pair of rods and to respond whether they believe the two rods possess similar flexibility characteristics. In both cases the pair of rods are matched so that their attributes can be entered into a multiplication equation. The level IIIA subjects may not anticipate the use of the schema in all instances of the first four items, but on items 5 and 6 they are, in a sense, prompted to do so.

The last two items, 7 and 8, ask subjects to select rods from among sixteen pictured which can be tested to determine how shape (round versus square) and metal type (steel versus brass) affects the flexibility of a rod. These two items test the subjects' ability to use the schema "all things being equal" described in the pendulum task.

Nature of the Proportional Reasoning Task (P4)

Four problems are presented which require a proportion between two ratios to be formed and the resulting equation
to be solved for an unknown variable. The ability to set up proportional equations and to reason about situations in which variables are related through proportionality is one of the hallmarks of Piaget's stage of formal reasoning.

The Schema for the Proportional Reasoning Task (P4)

Piaget claims that proportional reasoning ability derives from the INRC group of transformations. In this task the INRC group is applied in a situation in which the variables have numerical values. The group operates in much the same way as described in the rods task (P3); however, this is a metric application rather than a non-metric one.

The Proportional Reasoning Task and General Procedures for Categorizing Responses

The items in this task were selected from among eight problems constructed by Zepp (1975). Zepp's work was guided by the research of Karplus (Karplus, Karplus, and Wollman, 1974). Zepp found that his task separated subjects into categories identified by Karplus on the basis of a subject's raw score on his eight item test. Karplus' categories are an extension of those Piaget described for the concrete and formal stages.

An analysis similar to Zepp's will be used to evaluate subjects on this task. It will be assumed that subjects at level IIB should be able to set up all four proportions in a formal equation and solve them correctly. Level IIIA subjects have a tendency to use a scaling strategy, that is,
they focus on one ratio as a multiplicative comparison and apply it to the second ratio to derive the value of the unknown variable. While this procedure works for many problems it is unlikely that subjects would answer all four problems correctly without a sound understanding of the concept of proportionality. Problems 1 and 2 have a whole number scaling factor with the numerical values in problem 2 much larger than in problem 1. Problems 3 and 4 both have fractional scaling factors with the numerical value in problem 3 larger than in problem 4, but the fractional scaling factor in problem 4 is $7/2$, an improper fraction. A scaling strategy that is not derived from a fully operational base is not likely to succeed in all four problems.

Level IIB subjects should be capable of using an incomplete scaling strategy to solve problem 1, the easiest problem, and perhaps to solve problem 2 or 4, depending on their organization of the problem. Level IIA subjects should be able to solve only problem 1.

Specific instructions for evaluating the P-test tasks are included in Appendix B.

Categorization of Subjects

Based on Task Responses

Each P-test task will be evaluated on a four point scale. This scale is based on Piaget's division of each of the two stages, concrete operations and formal operations, into two substages. A subject whose responses are
characteristic of a given substage will receive points for a task according to Table 1.

Table 1
Transformation of Stages to Numerical Values

<table>
<thead>
<tr>
<th>Stage Level</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Concrete (IIA)</td>
<td>1</td>
</tr>
<tr>
<td>Late Concrete (IIB)</td>
<td>2</td>
</tr>
<tr>
<td>Early Formal (IIIA)</td>
<td>3</td>
</tr>
<tr>
<td>Late Formal (IIIB)</td>
<td>4</td>
</tr>
</tbody>
</table>

A total score on the four items will be calculated for each subject. Based on the total P-test score, subjects will be classified into one of five levels of reasoning ability. These categories listed in Table 2 are made to facilitate descriptive analyses of the study population and to conform to Piaget's theory of stages of cognitive development which the P-test was designed to measure.

Table 2
Level of Formal Reasoning by Score on P-test

<table>
<thead>
<tr>
<th>Stage Level</th>
<th>Score on Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concrete Operational</td>
<td>4-6</td>
</tr>
<tr>
<td>IIA</td>
<td></td>
</tr>
<tr>
<td>IIB</td>
<td>7-9</td>
</tr>
<tr>
<td>Transitional</td>
<td>T</td>
</tr>
<tr>
<td>Formal Operational</td>
<td>11-13</td>
</tr>
<tr>
<td>IIA</td>
<td></td>
</tr>
<tr>
<td>IIB</td>
<td>14-16</td>
</tr>
</tbody>
</table>

A range of total scores for categorization of levels balances out the effects of over- or under-assessment of a
subject's level of reasoning due to the content of a given
task. Piaget's theory discusses how content affects a sub­
ject's responses in his approach to testing.¹⁰

Some subjects could possibly be evaluated at the con­
crete level on half their task responses and at the formal
level on the other half. A fifth category, the Transitional
stage, is included in the scheme outlined in Table 2 to at­
tempt to sort out some of the subjects that may have a half
concrete and half formal response pattern. Subjects with
such a pattern would have a P-test score between 8 and 12.
It is possible to score 10 on the P-test and have a half
concrete and half formal response pattern in a larger number
of ways than it is possible for any other score calculated
from a half concrete and half formal response pattern. The
score of 10 is also the average of those scores that can be
obtained for a total P-test score from a response pattern
with half of the tasks evaluated as concrete and other half
as formal responses. For these reasons a score of 10 is
categorized as a transitional score. With this sorting plan
the estimation of the true portion of concrete and formal
operational subjects in the population should be enhanced.

Procedures for Evaluation of the Validity
Reliability of the P-test

The primary mission of this study and the focus of
Component I is the development and evaluation of a paper-

¹⁰Jean Piaget, "Intellectual Evolution from
and-pencil test, the P-test, intended to be equivalent to Piaget's method of individual interview assessment of a subject's level of reasoning. Thus a comparison will be made between the results obtained on the P-test and those obtained by Piaget for similar tasks. This will be accomplished by the presentation of sample responses given by subjects on the P-test and will be reported in the beginning of Chapter IV. This aspect of the study then is a descriptive evaluation and a validity check. Close attention will be given to whether or not the P-test elicited similar or comparable behaviors to those obtained by Piaget in his research.

The P-test will be checked for validity with Piaget's interview process in a second way that will allow statistical procedures to be applied. This will be done by testing ten percent or more of the subjects in the population for Component I (N=231) on two tasks administered as closely as possible to the manner in which Piaget has prescribed. The two tasks selected for the interview procedures are (1) "The Oscillation of a Pendulum and Operations of Exclusion," and, (2) "Combinations of Colored and Colorless Chemical Bodies." These two tasks will be used primarily because the apparatus required for their administration in clinical interviews was easily available. Both tasks are described in detail in Chapters 4 and 7 of Inhelder and Piaget's (1958) book, The
Growth of Logical Thinking From Childhood to Adolescence.  

Each of these two tasks will be scored on a scale of 1 to 4 depending on the level at which a subject is functioning; the scoring is in accordance with that given in Table 1. The sum of the two derived scores will constitute an "interview score."

Finally, the P-test will be checked to see if the results depend upon the person rating the subjects' responses. In order to do this, two additional raters will be selected: a psychology instructor with a background in Piaget's work and a philosophy instructor with a limited knowledge of Piaget. Written guidelines for rating the test will be revised according to the two raters' suggestions. These guidelines give instructions about which types of responses are to be considered correct, incorrect or partially correct, and what type of responses are especially expected from a subject reasoning on a given level.

To test interrater reliability, the P-test results for 30 subjects will be randomly selected. Both additional raters and the experimenter will independently score these tests. Comparisons among the three separate evaluations of the 30 P-test results will be made and will be reported for Component I.

---

Hypotheses for Component I:  
The P-test Development

As part of the analysis of procedures delineated in the last section, the following hypotheses will be tested:

$H_1$: There will be a significant correlation between subjects' P-test scores and their "interview scores."

$H_2$: There will be a significant interrater reliability coefficient derived for the independent evaluation of 30 subjects' P-test scores by the three raters.

$H_3$: There will be significant correlations between raters' evaluations on 30 subjects' P-test scores.

Statistical Analysis to be Followed for Component I

To test how well the P-test tasks match the subjects' interview scores and to test hypothesis $H_1$, correlations between the P-test scores and the interview scores will be calculated. The test for significance for this procedure will be taken at the .05 level.

To gain additional information about how well the P-test predicts the interview scores, a stepwise multiple regression will be performed using the P-test tasks as the independent variables. While each of the P-test tasks evaluate some similar reasoning skills, they also assess some different skills. Thus a stepwise regression would allow the differences between tasks possibly to add to the multiple correlation rather than be diluted by the summation of task scores obtained as the P-test score used in the simple correlation.
To test hypotheses $H_2$ and $H_3$, an interrater reliability coefficient and correlations among the three independent raters' results in evaluation of subjects' responses will be computed. A significance level of .01 will be used to accept or reject the hypotheses, $H_2$ and $H_3$.

Before any of the above statistical procedures can be considered as yielding reliable information, an examination must be made of the P-test's level of internal consistency. To test the internal consistency of the P-test, an intercorrelational reliability coefficient will be computed with the four tasks as the test items. Ordinarily, reliability on a four-item test would be difficult to assess. But, in the case of the P-test there are theoretical grounds from Piaget's work for the belief that the task should consistently measure the same psychological variable.

In addition to the correlational analysis, a factor analysis will be performed on the P-test and the two interview tasks to assess further the contention that the four measure a single variable, the same variable as in the interview procedure.

Component II: Relationship of the P-test with Standardized Mathematics Achievement

Background for Component II

In the earlier sections describing placement procedures at The Ohio State University, several tests were mentioned. Among these is the Mathematics ACT examination.
Entering freshmen at OSU take this examination as part of their matriculation. Because of the availability of the Mathematics ACT examination scores for subjects in this study these scores will be used as a common measure of subjects' mathematics achievement. An investigation of the P-test relationship to the Mathematics ACT examination will be made to provide a measure of the relationship of the P-test to mathematics achievement. While Component III will also be designed to investigate the relationship of the P-test to mathematics achievement, the investigation will be carried out in a narrower context. Mathematics achievement in Component III will be defined as a subject's score on a final examination given in the mathematics course in which the subject will be enrolled. Since subjects will be enrolled in different mathematics courses and take different final examinations, their mathematics achievement as measured in Component III will be tied to the mathematics course in which they will be enrolled. In contrast to Component III, the investigation of the relationship of the P-test to mathematics achievement in Component II will have a broader nature derived from a common measure of subjects' mathematics achievement.

**Hypotheses Related to Component II**

$H_4$: There will be a significant multiple $R$ derived from a regression analysis with the P-test tasks predicting Math ACT standard scores.
H$_5$: The regression coefficients for each P-test task entered into the multiple regression analysis to predict Math ACT standard scores will be significant.

Statistical Analysis to be Followed for Component II

To test hypotheses H$_4$ and H$_5$, the P-test tasks will be entered into a stepwise multiple regression analysis as the independent variables to predict Mathematics ACT standard scores, the dependent variable. For a test of the significance of the multiple $R$ in this procedure, .05 significance level will be used and this level will also be used for each of the tests of significance of the four P-test variables' regression coefficients. From the latter test a maximum prediction equation for Mathematics ACT standard scores will be derived for the P-test tasks.

If hypotheses H$_4$ and H$_5$ are supported, then it may be interesting to do post-hoc correlations between other measures of mathematics achievement and other measures of academic achievement in the "larger population" of the study. These other measures include English ACT, Social Science ACT, Natural Science ACT, Composite ACT level, the D-placement test, the B-placement test and H.S. Math Points.

Component III: An Investigation of the Power of the P-test to Predict Mathematics Achievement

Background for Component III

As described in the earlier sections, entering freshmen at The Ohio State University are placed in mathematics
courses on the basis of the D- or B-placement exams and a measure of their high school mathematics achievement. At the time (1964) that this placement procedure was developed, there were six courses in that part of the mathematics sequence which were at or below the first quarter calculus course. Today there are nine courses in this range; six of them included in this study are described at the beginning of this chapter.

In Crosswhite's (1964) investigation of the predictability of 24 measures of academic achievement it was found that the D-placement test and High School Mathematics Achievement were the best predictors of achievement in the first quarter calculus courses. Following Crosswhite's study, Zwick (1964) investigated the effectiveness of the placement procedures which used a version of the B-placement test. Although Zwick did not measure the predictability of the placement test against the variables used in Crosswhite's study, he determined that the procedure was successful.

Among the variables in Crosswhite's regression analysis were ACT scores and SAT scores. These scores correlated significantly with Crosswhite's criterion for achievement in mathematics.\footnote{F. Joe Crosswhite, "Procedures for Admission With Advance Standing in Mathematics at The Ohio State University" (unpublished Ph.D. dissertation, The Ohio State University, 1964), p. 91.} It may be speculated that the reason that these variables did not enter a maximum prediction equation along with the placement test and the H.S. Math Points was
that these variables all measured about the same variance on the criterion achievement test.

Recent studies on Piaget's stage of formal reasoning have shown that the Piagetian tasks, although correlating significantly with standardized tests of achievement, measure something different than the standardized tests. Because of this it was felt that a Piagetian measure would account for variance on a measure of achievement in mathematics which is not measured by either The Ohio State University mathematics placement tests or by the variable H.S. Math Points. Thus in this section a test will be carried out to determine whether the P-test can account for variance in mathematics achievement not accounted for by the placement exams or H.S. Math Points.

Procedures for Component III

Students in the study population who are also enrolled in one of the six mathematics courses described earlier, Math 100 to Math 151 will comprise the population (N=171)


for this component. The measure of mathematics achievement in each course will be a subject's final examination score. As discussed earlier, each course will have a different exam; thus standard scores will be used. The amount of variance in final examination standard scores which can be accounted for by the P-test will be measured after the variance accounted for by the subjects' placement scores and their H.S. Math Points has been removed. The additional variance accounted for by the P-test which will be used to assess the P-test predictive power of mathematics achievement.

Because the investigation of the P-test is tied to the mathematics placement tests, a minimum of two separate analyses will be necessary, one analysis with the D-placement test and one with the B-placement test.

Subjects in Math 151, 150, and 148 take the D-placement test and subjects in Math 103, 102, and 100 take the B-placement test. Therefore an initial split of the population into these two mathematics class groupings will be planned. But, further divisions into smaller sub-populations and additional analyses within the two defined groups may be necessary. Final determination of which classes may be combined will have to be made after the P-test results are recorded and analyzed. This contingency strategy is necessary because it may happen that different classes within each of the two placement group divisions may have quite different P-test results. Comparison of
P-test scores to standard final examination scores between such classes would have no interpretation. Whether two or more classes are combined will depend on how similar their distributions are on the P-test.

To simplify the analysis of the P-test distribution in the six mathematics courses, the P-test will be re-coded on an interval scale to be called the **P-test Scale**. The P-test Scale will consist of three scores: high (3), middle (2), and low (1). A subject with a P-test score in the range from 12 to 16 (16 is the highest score attainable) will have a P-test Scale score of 3, those with a score in the range 9 to 11 will have a P-test Scale score of 2, and those in the range 4 to 8 (4 is the lowest score attainable) will have a P-test Scale score of 1.

As described earlier, one part of the final examinations in all six mathematics courses was a special set of four questions worth from 20 to 30 points (depending on the math class instructor's perogative). This special part was designated the "formal portion" of a final examination and the rest of the questions was called the "concrete portion." An analysis of the correlations between the P-test and the two parts of the final examinations will be made. A general hypothesis forming the basis for this investigation is that the P-test, which measures the subjects levels of concrete and formal reasoning, will be more highly correlated with the formal portion of a final examination than with the concrete portion.
Lastly, because of the unresolved question concerning sex differences on Piagetian measures cited in the literature review, an examination will be made of differences between the P-test means for male and female populations with each mathematics class group defined. Should a difference be found, then the analysis outlined for each mathematics class group will be done for separate female and male populations.

Hypotheses for Component III

The following hypotheses will be tested as part of analysis required by the procedures outlined for Component III.

$H_6$: The increase in $R^2$ due to the P-test will be significant after the B-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in the defined mathematics class groups.

$H_7$: The increase in $R^2$ due to the P-test will be significant after the D-placement test and H.S. Math points are entered into the regression equation to predict mathematics achievement in the defined mathematics class groups.

$H_8$: The correlation between the P-test and the formal part of the mathematics examinations will be significantly larger than the correlation between the P-test and the concrete part of the mathematics examinations within the defined mathematics class groups.

$H_9$: The mean on the P-test for male subjects is significantly different from the mean for female subjects in the P-test within the defined mathematics class groups.
Statistical Analysis to be Followed for Component III

To test hypotheses $H_g$, the P-test, the B-placement test, and the H.S. Math Points will be entered as independent variables in a multiple regression analysis to predict mathematics achievement (mathematics final examination scores) in the groups defined. An hierarchical decomposition method described in Statistical Packages for the Social Sciences will be used for the regression. In this method, the B-placement test and H.S. Math Points will be entered first, followed by the P-test. In this manner, the variance in the dependent variable accounted for by the P-test, beyond that accounted for by the other two variables, may be measured with an $F$ ratio. The significance level to be used for this procedure will be .05.

To test hypothesis $H_7$, the same statistical procedure defined for the test of $H_g$ will be followed except the D-placement test will replace the B-placement test among the independent variables.

If hypothesis $H_g$ or $H_7$ is supported within a defined mathematics class group, then it may be interesting to do a post-hoc investigation using the individual P-test tasks. To accomplish the additional analysis, the same hierarchical decomposition of the dependent variable, mathematics achievement, will be used except that the P-test tasks will be included.

---

enter as separate independent variables replacing their sum, the P-test.

To test hypothesis $H_8$, a test will be made for differences between dependent correlations. This procedure is listed by Bruning and Kintz. The significance level for the tests will be set at the .05 level.

The differences between male and female means on the P-test will be tested for statistical differences by means of a t-test. The t-score for differences of the means within each mathematics class group will be calculated and a significance level of .05 will be used to accept or reject hypothesis $H_9$.

Summary

In this chapter, procedures for research on three separate, yet related, components were outlined. In Chapter IV, the results will be reported for Component I, the development of the P-test. In Chapter V, the results will be reported for Component II, an investigation of the relationship of the P-test with standardized mathematics achievement, and also for Component III, the power of the P-test to predict mathematics achievement in six freshman mathematics courses. Interpretations and implications of the results obtained will be discussed in Chapter VI.

---

CHAPTER IV
RESULTS FOR COMPONENT I

Introduction
The description of the P-test in Chapter III includes a general account of the behaviors expected from subjects reasoning at different substages. In the first section of Chapter IV, samples of subjects' responses will be given and analyzed for the four P-test tasks, which are the Pendulum Task P(1), the Triangle Task P(2), the Flexible Rods Task P(3) and the Proportional Reasoning Task P(4). Tables will also appear, listing the distribution of the study population for each P-test task into categories by substages early concrete (IIA), late concrete (IIB), early formal (IIIA), and late formal (IIIB). The second section of Chapter IV will contain the statistical results used to analyze the reliability, validity and the interrater reliability of the P-test.

Pendulum Task (P1)
The distribution of subjects into substage levels is given in Table 3 for the Pendulum Task; each substage had a sizable number of subjects judged to be operating at that level. These figures were computed from the actual
responses of 231 student subjects in the entering Freshman class for the school year 1977-78 at The Ohio State University-Lima Branch.

Table 3

<table>
<thead>
<tr>
<th>Stage Level</th>
<th>Early Concrete</th>
<th>Late Concrete</th>
<th>Early Formal</th>
<th>Late Formal</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>49</td>
<td>81</td>
<td>81</td>
<td>20</td>
</tr>
<tr>
<td>Percent</td>
<td>21.2</td>
<td>35.1</td>
<td>35.1</td>
<td>8.7</td>
</tr>
</tbody>
</table>

Following are the tasks and examples of the responses from subjects reasoning at the given substages; they demonstrate that the behaviors described in Chapter III were actually found for each task.

Items of the Pendulum Task

Listed below are the series of questions and requests for responses presented to subjects on the Pendulum Task in the P-test. For this task Item Number 1 was not used for evaluation of reasoning levels because it served only as a "warm-up" question; therefore, it is not listed.

Item # 2 (Ordering pendulums by their period lengths):

Select three of the pictured pendulums so that your second choice takes a longer time to make one swing than does your first choice and that your third choice takes a longer time to make one swing than does your second choice.
Item # 3 (The comparison of two pendulums which differ by weight and length):

Suppose you check the time for one swing of pendulum No. 4 and also for pendulum No. 9 and find that their times are different. Can you conclude from this what caused the time difference? (Pendulums 4 and 9 differ by both their length and their weight.)

Item # 4 (Designing tests to check the effects of length, weight, and amplitude):

Other people have suggested the factors listed below as those which influence the time of one swing of a pendulum.

(1) The distance that the ball is drawn back before it is released.
(2) The length of the string holding the ball.
(3) The weight of the ball.

You may have listed some of these factors or you may have given others.

Tell which pendulums you would select and how you would use them to demonstrate which factors influence and which do not influence the time of one swing of a pendulum. Be sure your demonstration will leave no doubt as to which factors do and which do not have an influence in this situation.

Item # 5 (Designing a test to eliminate weight as an operant factor)

Tell which pendulums you would choose and what observations you would have to make to conclude that the weight of the ball has no influence on the time of one swing of any of the pictured pendulums. It is a fact that weight of the ball does not effect the time, but this question requires you to tell only how, through experimentation, you could discover this fact. Please take time to detail your experiment, listing which pendulums you would investigate and what observations you would make to discover that weight has no influence and also explain how your experiment would allow you to conclude that weight has no effect on the time of one swing of a pendulum.
Background for the Pendulum Task

In order to understand how subjects reasoned on this task, it is helpful to know how the pendulum operates; these principles were not stated anywhere in the written test or the interview. A pendulum's period is governed by the length of its string. The shorter the string, the shorter the pendulum's period (arc) will be. Weight of the bob and the amplitude of the release point have no effect on the period. Because subjects in the P-test's Pendulum Task did not have actual pendulums with which to experiment, whether they knew that string length is the operant factor or believed that another factor or combination of factors affected the period was not directly important to the evaluation of the responses. The important thing was their ability to describe experiments which would separate out the effects of the variables, weight, length, and amplitude. A response was scored "correct" if the logical pattern was correct even though the physical explanation might be wrong.

Level IIA, Early Concrete - Examples of Responses to the Pendulum Task

The subject whose responses are detailed below will be referred to as Respondent Pl-IIA because on the Pendulum Task (Pl) he serves as the example of a lower concrete reasoner (IIA).

On Item 2 (ordering pendulums by their period length), Respondent Pl-IIA selected his first two pendulum choices on the basis of their weight; the first choice was a one-foot-
long pendulum with a one-pound bob; the second choice was a one-foot-long pendulum with a two-pound bob. To make his third choice, Respondent Pl-IIA switched to string length and chose a two-foot-long pendulum with a one-pound bob.

In response to Item 3 (the comparison of two pendulums which differ by weight and length), Respondent Pl-IIA claimed he could determine what caused the difference and explained,

"Number 9 is heavier and will move faster even though it is longer."

Pendulum Number 9 has both a longer length and a heavier weight than does the other pendulum in Item 3.

In response to Item 4 (designing tests to check the effects of length, weight, and amplitude), Respondent Pl-IIA copied the picture of a one-foot-long pendulum with a three-pound-bob and wrote,

"This one will swing fast because of its weight and it will cover its designated short length of string. The ball can only be drawn back a short distance because of the length of the string."

On Item 5 (designing a test to eliminate weight as an operant factor), Respondent Pl-IIA copied a three-foot-long pendulum with a one-pound-bob and reasoned,

"This pendulum has a very small amount of weight on it. In testing this one the length of the string would determine the time involved. The more string you have when you release the ball the faster it will travel."

There are many typical IIA concrete behaviors in the responses above. Piaget states,
Stage II subjects are able to order the lengths, elevations, etc., serially and to judge the differences between observed frequencies objectively. Thus they achieve an exact formulation of empirical correspondences but do not manage to separate the variables except insofar as the role of the impetus is considered.¹

Relative to the behaviors demonstrated by Respondent Pl-IIA, Piaget states, "At substage IIA serial ordering of the weights is not yet accurate. . ."² In Item 2 (ordering pendulums by their period lengths), Respondent Pl-IIA focused on weight to make a comparison between his first and second choice, but switched to length when comparing his second and third choices. Although he demonstrated the ability to order two pendulums by a given criterion, he had trouble making a complete serial ordering. It was surprising that so many subjects at this age level (college students) should not be able to complete this item successfully. Respondent Pl-IIA believed that both factors, weight and length, influence the period of a pendulum. His inability to separate out the effects of the two factors hindered his ability to form a consistent serial ordering of pendulums.

In Item 3 (comparison of two pendulums which differ by weight and length), Respondent Pl-IIA struggled to sort out the effects of length and weight. Note the phrase "even though it has a longer string." The pendulum to which he


²Ibid., p. 70.
referred had both the heavier weight and the longer string length. He seemed to be saying that length has an effect, but that weight dominates. His whole thinking was dominated by what he believed to be true. He never entertained the thought that experimentation could be used to determine how or which of the possible combinations of factors actually influence it.

One may object that if Respondent Pl-IIA had the chance to actually manipulate physical pendulums, his behavior would be different. This may be the case, but in another portion of this study subjects were interviewed while they manipulated real pendulums. Responses similar to Respondent Pl-IIA's were witnessed. One interviewed subject in particular was also convinced that both weight and length were operant factors and that weight dominated over length. His reaction to observing two pendulums of identical length but different weights was one of disbelief. This subject shortened the pendulum with the heaviest weight which caused it to swing more quickly than the other. It swung faster because of its shorter string. In spite of the earlier result with equal string lengths, he chose to continue his belief that both string length and bob weight influence the period of the pendulum.

One feature missing in the behaviors of both the interviewed subject and Respondent Pl-IIA's written response is the ability to hypothetically consider alternative possibilities in the Pendulum Task. In Item 4 (designing tests
to check the effect of length, weight, and amplitude), subjects are encouraged to entertain alternatives. Respondent Pl-IIA stuck to a description of what he believed to be true about pendulums. In Item 5 (designing a test to eliminate weight as an operant factor), Respondent Pl-IIA reacted to the statement that weight does not play a role in the determination of the pendulum's period by choosing a pendulum with a three-foot-long string and a one-pound weight, the largest and the smallest values of length and weight respectively. He seemed to be hoping that by diminishing weight and increasing string length, he would somehow, without relinquishing his beliefs, satisfy the requirements that the weight play no role.

Respondent Pl-IIA thus qualifies for the early concrete substage by his behaviors which demonstrated his inability to separate out variables, especially weight and length on Items 4 and 5 of the pendulum task, and his difficulty with the formation of a serial ordering of pendulums based on either weight or length in response to Item 2 of the pendulum task.

**Level IIB, Late Concrete - Examples of Responses to the Pendulum Task**

This Respondent will be known as Respondent Pl-IIB because on the Pendulum task (Pl) he will serve as an example of the late concrete reasoner (IIB).

In response to Item 2 (ordering pendulums by their periods lengths), Respondent Pl-IIB selected three pendulums
for which both weight and length are ordered from the smallest value to the largest. His first choice was a one-foot-long pendulum with a one-pound-bob, his second choice was a two-foot-long pendulum with a two-pound-bob; third, he chose a three-foot-long pendulum with a three-pound-bob.

In response to Item 3 (the comparison of two pendulums which differ by weight and length), Respondent Pl-IIB indicated that he could determine what caused the difference in the periods of the two pendulums. He reasoned,

"Number 4 has lighter weight causing it to swing faster, plus with shorter string - less area to cover."

In Item 4 (designing tests to check the effects of length, weight, and amplitude), Respondent Pl-IIB selected the same three pendulums he listed in response to Item 2, one each with both weight and length at the same numerical value. He, like Respondent Pl-IIA, did not begin to describe experiments. He wrote,

"Seven (the 1 ft. long and 1 lb. pendulum) has shortest string length - your pull back only determines how big of a swing you have (referring to the pendulums arc). Length of the string could also determine the same. Weight determines how fast something falls. The heavier something is the more gravity there is pulling it."

Respondent Pl-IIB did not give a response for Item 5 (designing a test to eliminate weight as an operant factor). Many concrete subjects were unable to give a response to Item 5 because of their unshakable belief that weight was an operant factor. One subject, during the P-test, asked the experimenter how he could possibly expect a response to
Item 5; he even needed to be reassured that a response was expected and that it was not a trick question.

It can be seen that Respondent Pl-IIB's ability to order variables in a consistent system was quite different from Respondent Pl-IIA's; on Item 2 (ordering pendulums by their period lengths), Pl-IIB ordered serially both weight and length while Pl-IIA did not. On the other items the thinking of both subjects was dominated by preconceived notions about pendulums.

Respondent Pl-IIB was apparently unable to consider other possibilities. His ability to order serially and his inability to consider alternatives to his beliefs characterize the late concrete reasoner. Piaget observed in his work that

At this stage the subjects lack a formal combinatorial system. Since they are accustomed to operations of classification, serial ordering, and correspondences, they are limited to simple tables of variation and do not conceive of the multiplicity of combinations which can be drawn from them.3

Level IIIA, Early Formal - Examples of Responses to the Pendulum Task

This respondent will be known as Respondent Pl-III A because on the Pendulum task (Pl) he will serve as the example of the early formal reasoner (IIIA).

On Item 2 (ordering pendulums by their period lengths) Respondent Pl-III A chose three pendulums of the same length.

3Ibid., pp. 72-73.
His first choice had a three-pound-bob, his second choice had a two-pound-bob, and his third choice had a one-pound-bob. Thus, he made a serial ordering.

On Item 3 (the comparison of two pendulums that differ by weight and length), Respondent P1-III A claimed to be able to determine what caused the difference in the two pendulums' periods and gave the following reason,

"The difference would be caused by the difference in the weight of the balls and the lengths of the strings."

Thus, he did not anticipate that the effects of each variable, weight and length, must be examined alone while the other is constant for both pendulums.

His response to Item 4 (designing tests to check the effects of length, weight, and amplitude) was,

"I would use pendulums with long strings and balls with different weights to show how timing differs. I would eliminate the different timing between strings of different lengths but the same balls. I would also show how much difference in timing there would be in how high point A was" (Point A is the point at which a pendulum is set in motion, i.e., the release point.)

One Item 5 (designing a test to eliminate weight as an operant factor), Respondent P1-III A chose to compare pendulum 5, a three-foot-long pendulum with a three-pound-bob. The lengths are the same and the weights are different. Respondent P1-III A responded,

"I would choose pendulums 5 and 9 to show that (the) weight of the ball has no influence on the timing. I would demonstrate that point A makes a difference, but not weight."
Respondent Pl-III A's response to Item 3 (the comparison of two pendulums that differ by weight and length) was used as the primary criterion for his placement into level IIIA rather than level IIIB. Because the two pendulums described in Item 3 differ by weight and length, a comparison of these two pendulum periods cannot be used to draw a single conclusion.

There are three possible conclusions: (1) the different weights caused the difference, (2) the different string lengths caused the difference, or (3) both string length and bob weight differences contributed to the unsimilar periods. Using only the facts given in Item 3, there is no way to tell which of these are correct.

Respondent Pl-III A demonstrated from his attempt to account for the effect of variables on the periods that he was unaware of the inclusive nature of the experiment described by Item 3. However, when he was asked to design experiments in Items 4 and 5 that would yield conclusive results, he demonstrated that he did understand that only one variable could be manipulated while the others must remain constant. Thus, while Respondent Pl-III A demonstrated in response to items 4 and 5 that he understood that variables should be isolated to determine their effect, he also demonstrated in response to item 3 that he did not automatically anticipate the application of the schema "all other things being equal."

Piaget has described the level IIIA subject's behavior in the pendulum task.
formal operations are already present in a crude form, making certain inferences possible, but they are not yet sufficiently organized to function as an anticipatory schema.  

It must also be pointed out that although Respondent Pl-IIIIA selected appropriate pendulums for the experiments in Item 4, he stated what results he intended to produce; this feature is somewhat similar to the way Respondent Pl-IIB answered Item 4. But it can be seen that Respondent Pl-IIIIA also isolated variables from the statement, "I would demonstrate the different timing between strings of different length but the same balls." Respondent Pl-IIIIA at the early formal substage indicated by this statement that he was manipulating string length while he held the weight of the balls constant and as such he was separating out the effects of the variable weight from those of string length. No such behavior was demonstrated by Respondent Pl-IIB, at the late concrete level.

Thus, Respondent Pl-IIIIA was labeled an early formal reasoner because he demonstrated the ability to separate out variables in experimentation, but failed to anticipate this procedure when he was not directly asked to use it.

### Level IIIB, Late Formal - Examples of Responses to the Pendulum Task

This respondent will be referred to as Respondent Pl-IIIB because on the Pendulum task (Pl) he placed in the

\[\text{Ibid.}, \ p. \ 73.\]
late formal substage (IIIB).

In response to Item 2 (ordering pendulums by their period lengths), Respondent Pl-IIIB chose three pendulums which all had a one-pound-bob. He ordered their string lengths as one-foot, two-feet, and three-feet.

He correctly responded to Item 3 (the comparison of two pendulums that differ by weight and length) by stating that he could not draw any conclusion. He explained,

"It could be either the weight of the ball or the length of the string used, since Number 4 has 2 feet and 1 pound and number 9 has 3 feet and 3 pounds."

On Item 4 (designing tests to check the effects of length, weight, and amplitude) Pl-IIIB responded

"Use pendulums 2,7,8 (pendulums of the same lengths with different weight values) to show what factor (effect?) weight of the balls has. Time the swing for each with length of string and distance of draw-back as constants, and weight as variable.

Use pendulums 4,5,7 (pendulums of the same weight with different length values) to show what effect the length of string has. Time the swing for each with weight and draw-back as constant and length of the string as a variable.

Use any one of the pendulums and make at least 3 runs with different drawbacks."

On Item 5 (designing a test to eliminate weight as an operant factor), Respondent Pl-IIIB said,

"Pendulum No. 2 Make 3 runs, timing the runs; record data.
No. 7 Make 3 runs, timing the runs; record data.
No. 8 Make 3 runs, timing each swing; record data."

(Pendulums 2,7, and 8 all have the same length but each has a different weighted bob.)
Compare the data for the three pendulums. Observe if the length of time is different by only (in)significant amount. ± 1 sec. If the time differs by less than the above amount, it can be concluded that the weight of the ball has no significant influence on the time of 1 swing of the pendulum."

Piaget describes the fully formal reasoner's behavior as follows:

Hence, conclusions are rigorously deduced from premises whose truth status is regarded only as hypothetical at first; only later are they empirically verified. This type of thinking proceeds from what is possible to what is empirically real.5

Respondent Pl-IIIB's responses were different from those of Respondent Pl-IIIA. They both demonstrated the need to isolate variables in their proposed experiments, but the upper formal respondent (Pl-IIIB), unlike his early formal counterpart (Pl-IIIA), sets up his experiments in order to discover what effect a variable had, if any. This is seen by the phrase of comparison, "to show what effect the length of the string has," made by Pl-IIIB, with the phrase, "I would demonstrate that position of Point A makes a difference," from Pl-IIIA. The fully formal reasoner is capable of placing reality as a subset of the possible that needs to be discovered through experimentation. This ability or lack of it is the prominent difference between the late formal reasoner and the early formal reasoner on the pendulum task.

Respondent Pl-IIIB's behavior on the pendulum task was definitely indicative of formal reasoning. He designed his experiments from the perspective that reality is that which

---

5Ibid., p. 251.
needs to be discovered from empirical results.

The Triangle Task (P2)

The distribution of subjects by stage level is presented in Table 4 for the Triangle Task; each substage had a sizable number of subjects judged to be operating at that level. These figures were computed from the actual responses of 231 students in the entering freshman class for the school year 1977-78 at The Ohio State University-Lima Branch.

Table 4

<table>
<thead>
<tr>
<th>Stage Level</th>
<th>Early Concrete</th>
<th>Late Concrete</th>
<th>Early Formal</th>
<th>Late Formal</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>37</td>
<td>126</td>
<td>24</td>
<td>44</td>
</tr>
<tr>
<td>Percent</td>
<td>16.0</td>
<td>54.5</td>
<td>10.4</td>
<td>19.0</td>
</tr>
</tbody>
</table>

As in the Pendulum Task, there are subjects in each substage. Unlike the Pendulum Task and the other P-tasks, more than 50 percent of the study population have been evaluated as IIB, late concrete, reasoners.

The responses of four subjects will be presented and analyzed. Each subject's responses typifies those of other subjects categorized at his level of reasoning on the P-test; this was the criterion for the selection of a subject's response to be included as an example.
In this task, subjects were required to draw a copy of an obtuse triangle using only a pencil and a 3 by 5 index card. Their responses consisted of reference marks made on the given triangle and the index card as well as on the drawn copy. Therefore, all three, the original triangle, the index card, and the copy will be reproduced just as the subject marked them.

Subjects had to use the index card in two ways in order to draw an accurate copy. First, they had to use the card to copy lengths of the sides of the triangle and to copy lengths of segments connecting other points. Second, they had to use the corner of the index card to form perpendicular line segments. As was explained in Chapter III (see figure 1), the second application of the card was necessary in order to measure the angles of the given triangle and to transfer accurately these measurements to the copy. What follows is an account of how well subjects at each substage were able to make these applications of the index card and, thus, how successful they were in copying the obtuse triangle.
Level IIA, Early Concrete - Examples of Subject's Responses on the Triangle Task

The Original:

The Index Card With Subject's Markings:

The Level IIA, Lower Concrete Response:

Figure 2
Response Made by P2-IIA
The responses in this first example are those of a subject judged to be a lower concrete reasoner, IIA, on this task. This subject will be referred to as Respondent P2-IIA in these analyzations.

The most prominent feature of Respondent P2-IIA's response is the lack of any attempt to measure properties of the triangle. This is most noticeable from lack of markings or erasers on the index card and the shape of the triangle he drew. The triangle he drew was at most topologically equivalent to the given triangle. Respondent P2-IIA's attention appeared to be focused on the obtuse angle, the most prominent feature of the given triangle. Other subjects' drawings placed in the IIA level ranged from triangles which looked less like P2-IIA's example to copies which appeared to be similar in the geometric sense of equal angles. Again the one feature common to all level IIA subjects was the lack of any attempt to measure segments or angles. Based on his observations, Piaget sums up the early concrete reasoner's behavior on this task with the statement, "There is no attempt at measurements."  

---

Level IIB, Late Concrete - Examples of Subject's Responses on the Triangle Task

The Original:

The Index Card With Subject's Markings:

The Level IIB, Upper Concrete Response:

Figure 3
Response Made by P2-IIB
The responses in this second example are by a subject judged to be a late concrete, IIB, reasoner on this task. He will be named Respondent P2-IIB in these discussions.

The bracket marks on the reproduction of Respondent P2-IIB's index card are drawn by the experimenter to refer to measurement marks used by the subject to copy lengths of the indicated sides of the given triangle. The first bracket from the top gives the subject's measurement of side BC; measurement of AB and AC are noted by the respective brackets. Sides AB and AC were reproduced accurately in the subject's copy. But, the angle formed by sides AB and AC was not measured and, thus, it was not reproduced accurately in the copy. When Respondent P2-IIB attempted to draw side BC, he seemed to ignore the fact that it did not fit.

It is observed that Respondent P2-IIB, unlike Respondent P2-IIA, measured the sides of the triangle. Like P2-IIA, he made no attempt to measure angles. When the lack of angular measurement caused him to be unable to fit the sides of the triangle together in the copy, he ignored one of the length measurements he had already made.

To ignore angular measure and not to attempt to resolve the problem created by the sides not fitting together is a characteristic of responses at the late concrete, IIB, level.

Piaget describes the behaviors of level IIB subjects on the Triangle Task in this manner.
In reproducing a triangle these children measure each side of the sides in succession, but fail to reproduce their inclinations and are therefore at a loss to know where to make them meet.\(^7\)

\(^7\)Ibid., p. 186.
Figure 4
Response Made by P2-III A
The response reproduced in the third example was given by a subject judged to be early formal, IIIA, on the Triangle Task, P2. Therefore, he will be referred to as Respondent P2-III A.

On Respondent P2-III A's index card, he indicated by brackets the marks he used to measure sides of the triangle. Also notable are sketched arcs which he drew at angle A on the original and on his copy. The bracket labeled AR indicates the radius length of the arcs. P2-III A apparently used the arcs to help estimate angle A. Once one angle is accurately, or in this case, almost accurately copied, drawing the sides to the correct lengths will suffice to reproduce a nearly accurate copy of the given triangle; from the marks on P2-III A's index card and the near accurate copy he produced, it is assumed that he used the procedure described to make his copy.

It is the attempt to measure an angle that characterized the behaviors of subjects classified a level IIIA on this task. Not all subjects used exactly the same procedure as Respondent P2-III A, but they used equivalent procedures. Some subjects in the IIIA category repeatedly drew one side of the triangle until the angle it made with an adjacent side was near to the measure of the corresponding angle in the original triangle. Piaget describes the attempts to measure angles at this level as a method of "trial-and-error aimed at making the sides meet."^8

^8Ibid., p. 189.
Level IIIB, Late Formal - Example of a Subject's Response to the Triangle Task

The Index Card With Subject's Markings:

The Level IIIB, Upper Formal Response:

Figure 5
Response Made by P2-IIIB
The responses in the fourth example are those of a subject judged to be a late formal, IIIB, reasoner on the Triangle Task, P2. Therefore, he will be referred to as Respondent P2-IIIB.

The brackets and pairs of letters on Respondent P2-IIIB's index card and the on construction of his copy are made by the experimenter in order to refer to measurement marks and lengths used by P2-IIIB.

By the measurement marks and lines drawn by this subject, the outline of a seven step procedure may be inferred to explain his behavior. First, he extended sides AB and AC of the original triangle. Second, he placed the lower left corner of the index card at vertex A with the edge of the card along side AB. Third, with the lower right corner of the card serving as a right angle, he marked point C' (the C' is the experimenter's and the mark next to C' was made by the subject) to note the length of PC', the perpendicular that measures angular separation at angle A. Fourth, he marked the length of BC as indicated by the experimenter's bracket on the index card. Fifth, the subject copied AB and AP onto the area where the copy was to be drawn. He also marked the point labeled C' (circled by the experimenter) to reproduce the length of PC'. Sixth, he drew segment AC through to C' (note that he drew AC' only to the extent needed to complete the construction). Seventh and last, he placed the copy of length BC with point B on the card at point B on the copy and adjusted the card until BC marked on
the card struck the segment AC' and then marked point C on AC'. Thus, AC was automatically given in the copy as a result of the reasonably accurate copy made of angle A.

Some subjects judged to be late formal on this task drew the perpendicular PC' inside the triangle (see figure 1, Chapter III). Except for the necessity to extend the sides of the triangle in order to construct the outside perpendicular, the two procedures required the same kinds of measurements and constructions.

In the response of P2-IIIB it may be seen that he went beyond the obvious properties of the triangle and produced a perpendicular relative to two sides of the triangle. This extension of the given characterizes the behavior of the level IIIB subject, differentiating this stage from the lower stages. Piaget states,

This construction . . . argues a greater freedom from the limitations of what is perceptually given, and as such it is typical of the beginnings of formal thought. . .9

The Flexible Rods Task (P3)

The distribution of subjects by substage level is given in Table 5 for the Flexible Rods Task. These figures were computed from the actual responses of 231 subjects in the entering freshman class for the school-year 1977-78 at The Ohio State University-Lima.

Table 5

Distribution of Subjects by Level of Reasoning on the Flexible Rods Task

<table>
<thead>
<tr>
<th>Stage Level</th>
<th>Early Concrete</th>
<th>Late Concrete</th>
<th>Early Formal</th>
<th>Late Formal</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>77</td>
<td>80</td>
<td>61</td>
<td>13</td>
</tr>
<tr>
<td>Percent</td>
<td>33.3</td>
<td>34.6</td>
<td>26.4</td>
<td>5.6</td>
</tr>
</tbody>
</table>

The distribution of subjects on this task is similar to that found on the Pendulum Task, Pl. Each substage had a sizable number of subjects operating at that level, thus, the behaviors characteristic of each substage were well represented.

As in the previous tasks, responses of four different subjects are presented as examples of the behaviors observed in this task. Each set of responses represents those typical of each substage.

On these tasks subjects were required to compare different rods. The rods varied by four factors. (1) There were three different lengths, long, medium, or short; (2) they were either thick or thin; (3) they were either round or square in shape; and (4) they were made of two metals, either brass or steel. The rods to be compared were represented pictorially in diagrams which accompanied each of the items in this task.

In this task presentation, not all of the eight items will be examined. The first four items are quite similar and
a detailed description of how the differ would not shed any light on the behaviors found. Therefore, only responses for the first item of the four will be discussed. The same is true for items 5 and 6 where only responses to item 5 will be discussed, and for items 7 and 8 where only responses for item 8 will be discussed.

**Items of the Flexible Rods Task**

Only those items which are examined for the sample response sections are listed below. The items not included may be found in the P-test in Appendix A.

**Item 1: (selection of an equally flexible rod)**

To answer question 1 please use diagram 1 on the next page.

(1) If you tested the flexibility of rod No. 5, which rod do you think would bend about the same amount? Choose the best possible answer and please explain the reasons for your choice.

Rod No. 5 is a medium length, thin, square, steel rod, and diagrammed with No. 5 is the correct selection, rod No. 4, a medium length, thick, square, brass rod. These two rods may be related to each other through the compensation phrase, "Thick brass bends like thin steel." This phrase represents the compensation of flexibility between the variables metal type and thickness. Thin rods are more flexible than thick rods, but steel rods are also less flexible than brass rods. Thus, the fact that rod No. 5 is thinner than rod No. 4 is compensated for by the fact that rod No. 5 is made of steel and rod No. 4 is made of brass. This schema will be called the compensation schema.
Item 5: (recognition of rods related by compensation)
To answer question 5, please refer to diagram 5 on the next page.

(5) Will rods 3 and 7 bend about the same amount? Why or why not?

Rod No. 3 is a long, thick, square, brass rod and rod No. 7 is a medium length, thin, square, brass rod. The two rods fit the compensation phrase by "Thin and short bends like thick and long."

Item 8: (selection of rods to test the effects of shape on flexibility)
(8) List all pairs of rods that could be compared to find out if brass rods bend more than steel rods. Explain how you decide what choices to make.

Background for the Flexible Rods Task

In order to understand how subjects reasoned correctly or incorrectly on the Flexible Rods Task, it is helpful to know how the flexibility of a rod is affected by the four factors stated earlier; this was not stated anywhere in the written test. The amount of bend for a rod under the conditions stated for this task is directly proportional to use each of the two factors length, and strength of the metal type; the bend is also proportional to the force (weight) acting on the rod, but in this task the force was defined to be constant. The amount of bend of a rod is inversely proportional to each of the two factors thickness, and "squareness" (shape). Both thickness and "squareness" are generally combined as a single factor, the cross-sectional area of the rod. However, most people react to thickness and shape separately.
In estimating a rod's flexibility, subjects must consider that the effects of the factors that are directly proportional to the amount of bend can be compensated by the effects of the factors inversely proportional to the amount of bend. For a given rod one of each type of factor can be related to each other in a non-metrical ratio, a compensation ratio. The compensation schema consists of the ability to identify two rods that have compensation ratios that cause the two rods to bend alike. Formulation of such relationships between rods is given in the description of the early formal and the late formal subjects' sample responses.

**Level IIA, Early Concrete - Examples of a Subject's Responses on the Flexible Rods Task**

The responses reviewed are those of an early concrete, IIA, reasoner on the Flexible Rods Task (P3). Therefore, this subject will be referred to as Respondent P3-IIA.

Respondent P3-IIA listed rod No. 2, a short, thin, square, steel rod as his choice for Item 1 (selection of an equally flexible rod). He stated as his reason:

"Rod 2 because it is made of the same material and has equal width."

On Item 5 (recognition of rods related by compensation), Respondent P3-IIA did not believe that the designated rods would bend about the same. He responded:

"It is doubtful the rods will bend equally because of the difference in thickness."

The two rods can be related through the compensation schema.
On Item 8 (selection of rods to test the effect of shape in flexibility), Respondent P3-IIA selected pairs of rods that not only differ by the shape but also differ in some cases by all other factors possible. The pair he listed first consisted of a short, thin, square, steel rod paired with a long, thick, round, brass rod. He selected a correct pair that differed by shape alone, but in light of his other choices, this pair was probably selected solely because one was round and the other was square. It did not appear that he was attempting to keep any of the other factors constant.

Respondent P3-IIA did not demonstrate in his responses to Items 1 through 4 that he possessed the ability to equate variables by a compensation phrase. Examination of his responses to Item 1 (selection of an equally flexible rod) is typical of the responses by subjects judged to be early concrete. He chose rod No. 2, a short, thin, square, steel rod. Notice that three of the factors, thickness, shape, and metal type are the same as those of rod No. 5. His reaction to this task was very concrete in nature. It was a classification of rods by their physical properties rather than by how the physical properties affect flexibility. Since he appeared to be unable to deal with the task in any other way, he seemed simply to ignore the fact that the two rods were not the same length. This fact is even more perplexing when his responses are examined for Item 8 (selection of rods to test the effects of shape on flexibility). In Item 8 he correctly listed rods of equal
length in order to control length as a variable. Thus, he demonstrated awareness that length does affect a rod's flexibility.

Subjects at level IIA responded to Items 1 through 4, like Respondent P3-IIA, by categorizing rods by physical properties rather than by the effects of these properties on flexibility. This level IIA method of operation on the rods task causes Respondent P3-IIA to respond incorrectly to Item 5 (recognition of rods related by compensation). The rods that subjects were asked to compare in Item 5 fit the phrase, "long and thick bends like short and thin." Respondent P3-IIA notices only that the rods have different thicknesses; again, this is a typical response based only on observable properties.

Correct responses to Item 8 (selection of rods to test the effects of shape on flexibility), requires the selection of rod pairs that differ only by their shape. As in the pendulum task, all other variables must be kept constant between two rods of different shapes in order to ascribe observed differences in flexibility to the different shapes. Respondent P3-IIA's selections did not control for the metal type in two out of three of his choices. Level IIA early concrete reasoners in general do not seem to understand the necessity of holding all other variables constant in order to test a given variable.
Level IIB, Late Concrete - Example of a Subject's Responses on the Flexible Rods Task

The response described below are those of a late concrete IIB reasoner on the Flexible Rods Test, P3. This subject will be called Respondent P3-IIB.

On Item 1 (selection of an equally flexible rod), Respondent P3-IIB selected a long, thick, square, brass rod to be compared with the given rod, No. 5, a medium long, thin, square, steel rod. He reasoned:

"Rod No. 3 (his selection) is longer than No. 5 but the reason I feel it would bend close to the other is because of the heavier metal it is made of. The longer the rod the more bounce. Also the wider or (heavier) the metal the less bounce. It evens up."

Respondent P3-IIB responded to Item 5 (recognition of rods related by compensation):

"No, because the length and weight of the rods differ too much. The longer the rod the more it bends."

In response to Item 8 (selection of rods that test the effect of shape on flexibility), Respondent P3-IIB selected four pairs of rods. In three of his selections he did not control for the type of metal. One of these three selections was the comparison of a long, thick, round, steel rod with a long, thick, square, brass rod.

The reasoning shown by Respondent P3-IIB on Item 1 (selection of an equally flexible rod) reflected an awareness that the flexibility of rods could be compared by the compensating effects of the rod's physical properties. He wrote, "Rod No. 3 is longer than No. 5 but . . . of the heavier
metal." This is in reference to the fact that rod No. 3 is thicker than rod No. 5. The relation he forms can be equivalent to "long, thick bends like short and thin." This relationship would be valid for rods No. 3 and 5 if they were made of the same metal; rod No. 3 is brass and rod No. 5 is steel. Such a response would have been good enough to place Respondent P3-IIB in the lower formal category, but his answer to Item 5 reveals additional information. On Item 5 (recognition of rods related by compensation), he was asked to compare two rods that fit the same reasoning schema he used to answer Item 1. The two rods can be related by the phrase "long and thick bends like short and thin;" the rods are even made of the same metal. On this item he failed to demonstrate that he possessed the appropriate schema.

While Respondent P3-IIB demonstrated that he was beginning to acquire the mental schemas necessary to reason logically on this task, he also showed that his understanding was not yet very stable.

In this task, as in the pendulum task, responses on different items must be contrasted with each other in order to obtain a more complete picture of the subject's level. This is especially true in the case of Respondent P3-IIB.

That the late concrete respondent's responses were different than the early concrete respondent's is apparent because he did not always use physical properties to relate rods. It was by physical properties alone that early
concrete Respondent P3-IIIA reasoned in the flexible rods task. Also, it will be seen that Respondent P3-IIIB's responses were different from subjects classified as substage IIIA, early formal, because he was not consistent in his use of the compensation schema. In Item 5 (recognition of rods related by compensation), he emphasized physical properties as his means of rod comparison. This is evident in his phrase, "... length and weight of the rods differ too much."

**Level IIIA, Early Formal - Examples of a Subject's Responses on the Flexible Rods Task**

The responses listed below are those of a subject judged to be an early formal, IIIA, reasoner on the Flexible Rods Task, P3. This subject will be referred to as Respondent P3-IIIA.

In this example there is a slight modification in the report of Respondent P3-IIIA's answers on the Flexible Rods Task. Item 6 is worded exactly as Item 5 (recognition of rods related by compensation) except that a different pair of rods are to be compared. The two rods of Item 6 also can be related by the compensation schema. Respondent P3-IIIA's answer to Item 6 will be substituted for that of Item 5 because his response to Item 6 more clearly explains the assumptions he made about the influence of the factors on a rod's flexibility.

On Item 1 (selection of an equally flexible rod), Respondent P3-IIIA selected a medium length, thick, round,
brass rod. His reason for this selection was:

"The brass rod is the same length, although brass will flex easier, the brass rod is round and not square so it makes up for the difference."

Respondent P3-IIIA's response to Item 6 (recognition of rods related by compensation) was:

"Yes, because even though it's shorter, you think it is going to be stronger, but since it is smaller around (thickness) it makes up the difference."

Respondent P3-IIIA's responses to Item 8 (selection of rods that test the effect of shape on flexibility) are completely correct and do not serve in this example to add understanding to the interpretation of the subject's reasoning level than can be derived from examination of Item 1 and 6.

Response to Item 1 (selection of an equally flexible rod) is representative of those Respondent P3-IIIA made on Items 1 through 4. They were quite similar to the response given by Respondent P3. The focus rod, rod 5, is a medium length, thin, square, steel rod while the rod Respondent P3-IIIA chose was a medium length, thick, round, brass rod. He related the two rods in a way equivalent to the phrase "thin and square bends like round and thick."  

Note that the subject in the example above assumes that round rods are less flexible than square rods. If the impression that there were absolute correct answers for each item in this task was given by reports of the previous examples, let this example serve to dispel this impression. Subjects' responses were evaluated from the viewpoint that what would be considered a logical response depended on assumptions the subject entertained about rods' flexibility.
His reasoning was correct except that, like Respondent P3-IIB, he ignored the fact that the two rods also differed by one other factor, and that factor would negate the equality in his equation. The rod he chose was thicker than the focus rod; he apparently ignored this factor.

On Item 6 (recognition of rods related by compensation) he demonstrated that he was aware of how the different thicknesses of rods could affect flexibility. He said, "Since it is smaller around (a reference to thickness), it makes up the difference." He did answer this item and Item 5 correctly, which separated his behavior from that of late concrete Respondent P3-IIB. He also was more consistent in his application of the compensation schema throughout his responses to Items 1 through 6. In no case did he make judgments about rods based on a comparison of their physical properties without the consideration of how the properties influenced the rod's flexibility. This feature of his responses then separates his behavior on this task from that observed in the review of Respondent P3-IIB's behaviors.

Respondent P3-IIIA's lack of attention to the one difference in the rods he compared on Item 1 that would render the compensation phrase false was characteristic of subjects labeled lower formal (IIIA) on this task. Piaget points this out as one feature identifying early formal reasoners, and it was used here as the primary factor to separate Respondent P3-IIIA's behavior and that of the others judged to be at the early formal stage from those in the upper
formal stage. 

Level IIIB, Late Formal - Examples of a Subject's Responses on the Flexible Rods Task

To aid the discussion in presenting the example of a late formal (IIIB) reasoner on the Flexible Rods Task (P3), this subject will be referred to as Respondent P3-IIIB.

On Item 1 (selection of an equally flexible rod), Respondent P3-IIIB chose the correct rod. Recall that the focus rod was a medium length, thin, square, steel rod. His reasoning was as follows:

"Brass is softer than steel but I'm not sure if that means it's more flexible. However, if it is, No. 4 would bend about the same amount because, although it is softer--more flexible--it is also thicker, which should compensate."

Respondent P3-IIIB's responses on Items 5 and 8 were also correct. They are omitted only because they would not add to an understanding of this subject's behavior and that of subjects who reasoned at a IIIB level. Notice that Respondent P3-IIIB's response is not only correct, but it has a hypothetical character. Respondent P3-IIIB says, "... if that means it's more flexible ..." and he goes on to explain himself under an assumption he consciously realizes he has made. It seems that for this subject reality consisted of those cases that prove to be true in experimentation with those cases that are possible. Piaget

---

lists the ability to hold such a point of view as a mark of formal reasoning.\textsuperscript{12}

The Proportional Reasoning Task (P4)

The distribution of subjects by substage level is given in Table 6 for the Proportional Reasoning Task. These figures were computed from the actual responses of 231 entering freshmen for the school year 1977-78 at The Ohio State University - Lima Branch.

<table>
<thead>
<tr>
<th>Stage Level</th>
<th>Early Concrete</th>
<th>Late Concrete</th>
<th>Early Formal</th>
<th>Late Formal</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>53</td>
<td>38</td>
<td>54</td>
<td>86</td>
</tr>
<tr>
<td>Percent</td>
<td>22.9</td>
<td>16.5</td>
<td>23.4</td>
<td>37.2</td>
</tr>
</tbody>
</table>

A sizable number of subjects were judged to be in each substage. Thus, on this task as with the other tasks, behaviors characteristic of each substage were well represented.

The distribution on the Proportionality Task seems to be shifted more towards advanced substages than are the distributions for the other tasks. This may be due to the fact that mathematics problems which require proportional reasoning for a solution are introduced quite early in the public

\textsuperscript{12}Ibid., p. 256.
school mathematics curriculum. Although Piaget has claimed many times that his evaluation of subjects' reasoning levels is content free, recently he has admitted that a subject's familiarity with the content of any of his tasks may in some degree influence the subject's performance on the task, especially in older subjects.\(^{13}\) This may be the case for the subjects in this study who are college freshmen. Whether it is or is not does not directly affect the inquiries of this study. The important thing is the investigation of the subjects' different thought processes.

**Items for the Proportional Reasoning Task**

Listed below are the series of questions presented to subjects on the Proportional Reasoning Task in the P-test.

**Item 1:** (the enlarging machine problem)
If you put an 8 inch page into an enlarging machine, it enlarges to 32 inches. What will be the final size of a 16 inch page?

Please explain or show how you arrived at your answer.

**Item 2:** (the auto salesman's commission problem)
An auto salesman receives a commission of $200 on a $2800 car. How big would you expect his commission to be on the $3000 model?

Please explain or show how you arrived at your answer.

Item 3: (the sales tax problem)
The sales tax in one state is $600 on a purchase of $3000. What is the tax on $18,000?

Please explain or show how you arrived at your answer.

Item 4: (the magnifying glass problem)
Letters 2 mm high appear to be 7 mm high under a magnifying glass. How high would 3 mm letters appear under the same magnifying glass?

Please explain or show how you arrived at your answer.

Level IIA, Early Concrete - Example of a Subject's Response on the Proportional Reasoning Task

The responses below are from a subject judged to be at the early concrete stage, IIA, on the Proportional Reasoning Task, P4. This respondent will be referred to as Respondent P4-IIA.

On Item 1 (the enlarging machine problem), Respondent P4-IIA gave 40 inches as his answer. He reasoned incorrectly in the arithmetic he presented as an explanation:

\[
\begin{array}{c}
32 \\
-8 \\
24 \\
\end{array}
\quad \text{and add} \quad \begin{array}{c}
16 \\
24 \\
\end{array}
\quad \begin{array}{c}
40 \\
\end{array}
\]

On Item 2 (the auto salesman's commission problem), Respondent P4-IIA suggested $50 as the answer with the following explanation, "... because it wouldn't be too much more."
On Item 3 (the sales tax problem), Respondent P4-IIIA admits that he does not know. In the space provided on this item he wrote these arithmetic problems:

\[
\begin{array}{c c c c}
1000 & 18,000 \\
3,000 & 3000/18,000 \\
15,000 & 18,000 \\
\end{array}
\]

On Item 4 (the magnifying glass problem), Respondent P4-IIIA gave 8 mm as his answer. Again he reasoned incorrectly and explained, "Add 1."

Respondent P4-IIIA demonstrated no proportional reasoning on any of the items in this task. On each item he used a classical concrete response, an addition strategy.\(^{14}\) For example, on Item 1 (the enlarging machine problem) he reasoned that the 8-inch page increased 24 inches in absolute terms and, thus, he concluded that the 16-inch page would also increase by 24 inches. Some subjects judged to be at the early concrete stage were able to use the fact that the 8-inch page increased by a factor of four and they reasoned that a 16-inch page would increase by four times also. But, like Respondent P4-IIIA, their responses on the other tasks reflected an addition procedure.

Level IIB, Late Concrete - Example of a Subject's Response to the Proportional Reasoning Task

The subject whose responses are below was judged to be a late concrete, IIB, reasoner on the Proportional Reasoning Task, P4. He will be referred to as Respondent P4-IIB.

On Item 1 (the enlarging machine problem), Respondent P4-IIB gave 64 inches as his answer, a correct answer. He reasoned as follows:

"If 8" paper goes to 32 it has increased 4 times itself so I use this to get 64" take 16 x 4 = 64"

To Item 2 (the auto salesman's commission problem), Respondent P4-IIB answered:

"I divided this (an arrow points to the division problem below) gave me how much he made on $200 and that is how much extra he made so I added 14."

\[
\begin{array}{c}
200 \\
14 \\
\hline
200 / 2800 \\
200
\end{array}
\]

(The placement of these numbers are exactly as Respondent P4-IIB put them)

For Item 3 (the sales tax problem), Respondent P4-IIB gave the answer 3600 which was correct. He reasoned:

"18,000 is 6 times more than 3000 so I took 6 x 600 to arrive at an answer."

For Item 4 (the magnifying glass problem), Respondent P4-IIB answered 8 mm which was incorrect. He reasoned:

"Since you increase 5 once I did it again."

The behavior of Respondent P4-IIB is classified as late concrete rather than early concrete because of his responses to Item 1 (the enlarging machine problem) and Item 3 (the sales tax problem). In both situations he found a scale
factor, the ratio, which relates the page size to the copy size on Item 1 and the first purchase price to the second purchase price in Item 3. He then applied the factors he found to compute the missing quantity in the proportional ratio in each problem. This method does require the subject to have some understanding of proportionality, but it was a primitive and intuitive one in Respondent P4-IIB's case. He was unable to apply it to Item 2 (the auto salesman's commission problem) and Item 4 (the magnifying glass problem).

The beginning of subjects' ability to relate quantities spontaneously by a ratio identifies their reasoning at a higher level than stage IIA, early concrete, but their inability to consistently relate to ratios when required to form a proportion places them at a point below the early formal, IIIA, stage. Piaget places these subjects in the late concrete, IIB, stage.15

Respondent P4-IIB used the scaling strategy on Items 1 and 3 but reverted to an addition strategy on item 2 and 4. Thus he demonstrated that he is at a stage where he could begin to use ratios between variables as part of his thought processes.

Level IIIA, Early Formal - Example of a Subject's Responses to the Proportional Reasoning Task

The subject whose responses are listed was judged to be an early formal reasoner, IIIA, on the Proportional Reasoning

15Ibid., pp. 172-173.
Task, P4. Therefore, he will be referred to as Respondent P4-IIIA.

On Item 1 (the enlarging machine problem), Respondent P4-IIIA correctly answered 64 inches. His reason was:

"The enlarging machine increased the page 4 times its original size. If it increases the 16 in. page the same amount it becomes 64 (16 x 4)."

On Item 2 (the auto salesman's commission problem), Respondent P3-IIIA has some difficulty. He gave an answer that was within pennies of the correct answer, but his reasoning was as follows:

\[
\frac{28}{200} \quad \text{"app. $7 for every $100."}\\
\frac{196}{4}
\]

On Item 3 (the sales tax problem), Respondent P4-IIIA correctly answered $3600 and gave as his reason the arithmetic example below.

\[
\frac{6}{3000/18000} \quad \frac{600}{6} \quad \frac{3600}{3600}
\]

On Item 4 (the magnifying glass problem), Respondent P4-IIIA again responded correctly. He presented the arithmetic example below as an explanation of how he arrived at his answer.

\[
\frac{3.5}{2/7} \quad 3.5 \cdot 3 = 10.5
\]

Respondent P4-IIIA demonstrated his ability to use the scaling strategy described in the last example, thus he was
judged to be at a higher stage level than was Respondent P4-IIB. Because his answers did not demonstrate any formal proportional reasoning strategies such as the formation of an equation with two equal ratios, and because he seemed not to be able to apply his scaling method to Item 2 (the auto salesman's commission problem), he was not judged to be a late formal reasoner on this task.

Piaget is not explicit in his description of the difference between an early formal reasoner and a late formal reasoner when he discusses proportionality. However, Karplus, in an extension of Piaget's work, describes the scaling strategy he claimed lower formal subjects use and this strategy was similar to the one used by Respondent P4-IIIA.

Level IIIB, Late Formal - Example of a Subject's Responses on the Proportional Reasoning Problem

The subject whose responses are listed below was judged to be a late formal, IIIB, reasoner on the Proportional Reasoning Task, P4. He will be referred to as Respondent P4-IIIB; his solutions to each of the items of this task consisted of an equation that gives the equivalence of two ratios. To be able to find equivalent ratios consistently was taken as strong evidence that the subject possessed the

---

16 Ibid., pp. 174-181.
formal schema Piaget used to describe subject's abilities to reason proportionally. Piaget stated,

> Mathematical proportions consist simply of the equality of two ratios \( \frac{x}{y} = \frac{x'}{y'} \). Their formation raises a psychological problem only because it does not take place during the concrete operational stage.\(^{18}\)

Piaget continued by describing how and why the equation is constructed at the formal stage. He also indicated that the use of such a mathematical relationship between variables is not overtly expressed by subjects until the late formal stage.\(^{19}\)

On Item 1 (the enlarging machine), Respondent P4-IIIB gives a correct equivalence of two ratios present, but not explicitly given, in the problem. He wrote and solved the equation:

\[
\begin{align*}
\frac{8}{32} &= \frac{16}{x} \\
8x &= 16 \cdot 32 \\
x &= 64
\end{align*}
\]

On Item 2 (the auto salesman's commission problem), Respondent P4-IIIB gave the second of four responses that demonstrated his consistent and correct application of the proportionality schema. He wrote:

\[^{18}\text{Ibid., p. 314.}\]

\[^{19}\text{Ibid., pp. 173-181.}\]
\[
\frac{200}{2800} = \frac{x}{3000}
\]
\[
200 \cdot 3000 = 2800x
\]
\[
\$214.28 = x
\]

Note that the missing variable value, \(x\), was the smaller quantity in the ratio \(x/3000\) for this problem; in Item 1, it was the larger value in the ratio \(16/x\). This is taken as evidence that Respondent P4-IIIB understood the relationships among the variables as Piaget has outlined in his description of how the INRC group of transformations apply to proportional reasoning.\(^{20}\)

On Item 3 (the sales tax problem) and Item 4 (the magnifying glass problem), Respondent P4-IIIB continued to give the same type solutions presented in answer to the first two items.

Not all subjects judged to be late formal on this task gave the mathematical solutions given by Respondent P4-IIIB on each of the Proportional Reasoning Task. Some subjects used the scaling on one or more of the items. This was found primarily on Items 1 and 3 where the scaling factors were easily derived small whole numbers, 4 on Item 1 and 6 on Item 3. Karplus also found that formal reasoners on a task of proportional reasoning used both formal solutions and scaling strategies.\(^{21}\)

\(^{20}\) Ibid., pp. 173-181, 194-198 and 206-209.

All subjects listed at the late formal substage for this task demonstrated ability to form equations with equivalent ratios and solved all problems correctly in the Proportional Reasoning Task.

Conclusions—Description of the P-test Tasks

The twelve examples of subjects' responses to the four P-test tasks, the Pendulum Task (P1), the Triangle Task (P2), the Flexible Rods Task (P3) and the Proportional Reasoning Task (P4) are to be taken as representative of those behaviors that define a given substage level for the task reviewed. Obviously not all subjects gave responses that were as clearly categorized as those who served as examplars. Some subjects' responses within a task were strongly suggestive of one level on one item and another level on a second item. For example, there were subjects who answered the first three items of the Proportional Reasoning Task, the enlarging machine problem, the auto salesman's commission problem, and the sales tax problem, with a very formal proportions equation and, perplexingly, used an addition strategy on Item 4, the magnifying glass problem; the addition strategy was described as the typical early concrete (IIA) level response while the equation with equivalent ratios was given as the typical late formal (IIIB) level response. Such cases were also present in response patterns found on the other tasks, but cases as extreme as the one given here were, fortunately, the exception.
The empirical results obtained give evidence in support of Piaget's developmental stage theory although they seem to contradict the ages at which Piaget claimed that adolescents acquire formal thought schemas; the subjects in this study were college age young adults and Piaget gave the age 15 as that by which a person should be expected to acquire formal operations. 22

Thus, it seems that the P-test development has been successful from the point of view that it has elicited the same cognitive behaviors from subjects with written tasks that Piaget obtained from other subjects with individual interviews and described in The Growth of Logical Thinking from Childhood to Adolescence and in The Child's Concept of Geometry. How consistently the P-test measured the Piagetian behaviors and the claim that the P-test did in fact measure the same cognitive behaviors described by Piaget will be tested by statistical procedures in the next part of this chapter. Specifically, the P-test responses from a sample of 27 subjects who volunteered to be tested individually in a Piagetian-style interview will be compared to the results obtained from the interview sessions; the P-test's internal consistency will also be measured.

P-test Tasks Intercorrelations
and the P-test Reliability

As a first measure of the reliability of the P-test, the correlations among the four P-test tasks, the pendulum Task (P1), the Triangle Task (P2), the Flexible Rods Task (P3), and the Proportional Reasoning Task (P4), are given in Table 7 for the 231 subjects in the study population. All correlations are significantly different from zero at $p < .001$ level.

Table 7
Correlations Among the Four P-test Tasks

<table>
<thead>
<tr>
<th></th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.3837</td>
<td>0.4250</td>
<td>0.4442</td>
</tr>
<tr>
<td>P2</td>
<td></td>
<td>0.3225</td>
<td>0.4239</td>
</tr>
<tr>
<td>P3</td>
<td></td>
<td></td>
<td>0.3869</td>
</tr>
</tbody>
</table>

All correlations significant, $(p < .001, \ N=231)$

A measure of the internal consistency or reliability of the P-test was computed using the Statistical Packages for the Social Sciences (SPSS, 1970) sub-program Reliability with an alpha model. The alpha level was 0.72. Thus, the measure of reliability of the P-test gives evidence of the internal consistency.

Total Score Patterns on the P-test

The subjects in this study were college freshmen. Their mean age was 19.4 years at the time they took the P-test; a
distribution of their ages by percentage is given in Table 8.

<table>
<thead>
<tr>
<th>Age</th>
<th>17-18</th>
<th>18-19</th>
<th>19-20</th>
<th>20 or older</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
<td>11.2</td>
<td>71.7</td>
<td>8.3</td>
<td>8.9</td>
</tr>
</tbody>
</table>

N = 205 (no age data available for 26 subjects)

Each of the 231 subjects in the study population took the P-test and received an overall P-test score. Table 2 in Chapter III lists the ranges used to place subjects into one of the five categories, early concrete (IIA), late concrete (IIB), transitional (t), early formal (IIIA), and late formal (IIIB); these are called the P-test levels. With the exception of the middle category, transitional, they are intended for comparison with the substages Piaget identified for concrete and formal operational children. A summary of the results of subjects classified by P-test levels is given in Table 9. From this table it is evident that the P-test evaluations resulted in over 50 percent of the subjects being placed below the formal operational level.

<table>
<thead>
<tr>
<th>Level</th>
<th>IIA</th>
<th>IIB</th>
<th>T</th>
<th>IIIA</th>
<th>IIIB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
<td>20.8</td>
<td>30.0</td>
<td>10.4</td>
<td>29.1</td>
<td>10.0</td>
</tr>
</tbody>
</table>
If formal thought were to develop for the subjects in this study at the time Piaget predicted (11 to 15 years), then it would be expected that all of them would have been judged in the early or late formal stages as measured by the P-test. If this had happened, there would be little reason to pursue the research outlined in Chapter III. However, because substantial numbers of subjects were sorted into each of the P-test categories, further analysis is necessary. To obtain a complete view of the P-test distribution, Table 10 lists the frequency of each P-test score.

Table 10
P-test Score Distribution

<table>
<thead>
<tr>
<th>Score</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>5</td>
<td>19</td>
<td>23</td>
<td>23</td>
<td>17</td>
<td>29</td>
<td>24</td>
</tr>
<tr>
<td>Percent</td>
<td>2.2</td>
<td>8.3</td>
<td>10.0</td>
<td>10.0</td>
<td>7.4</td>
<td>12.4</td>
<td>10.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>30</td>
<td>23</td>
<td>14</td>
<td>14</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Percent</td>
<td>13.0</td>
<td>10.0</td>
<td>6.1</td>
<td>6.1</td>
<td>3.0</td>
<td>0.9</td>
</tr>
</tbody>
</table>

In addition to the frequency distribution given in Table 10, a graphical representation of this distribution is given in Figure 6. A nearly normal distribution slightly skew to the right seems to be pictured in the graph of Figure 6. Although this distribution may be surprising for

---

a population with a mean age of 19.4 years, it is consistent with other researchers' findings cited in Chapter II.
Figure 6

P-test Score Distribution
Comparison of the P-test to Piagetian Interview Approach

To evaluate the validity of the P-test as a Piagetian measure, 27 subjects who volunteered from the study population were interviewed on two Piagetian tasks. These two tasks, the Colored Chemical Task and the Pendulum Task, are described in *The Growth of Logical Thinking from Childhood to Adolescence*, Chapters 4 and 7 (Inhelder and Piaget, 1958). The interviews were conducted before the P-test tasks were evaluated so there would be no chance the experimenter would be influenced by the knowledge of an interviewed subject's P-test results. These interviews were done from two to four weeks after the administration of the P-test.

The scoring procedure for the Piagetian-interview tasks was similar to that used for the written P-test tasks. On the basis of subjects' responses and actions, they were judged to be operating at one of the four substages—early concrete, late concrete, early formal, or late formal. In the order listed, the substages were equated with a value from 1 to 4. The sums of subjects' two interview task scores was defined as their "interview score."

Hypothesis $H_1$ is "There will be a significant correlation between subjects' P-test scores and their 'interview scores'." The significance level for the test of this hypothesis was set at .05. The correlation between the 27 subjects' interview scores and their P-test scores was computed with an $r=0.72$ significant at $p<.001$. Thus,
hypothesis $H_1$ is accepted.

Further analysis of the validity of the P-test as a measure of Piagetian levels of reasoning was obtained by entering each P-test task as a variable into a multiple regression analysis to predict the interview scores. The order of entry was $P_4$ (Proportionality), $P_1$ (Pendulum), $P_3$ (Rods), and, last, $P_2$ (Triangles), yielding a multiple $R = .7914$ with an $F = 12.850$ ($p < .001$. $df = 3.23$).

In addition to the correlations given above, a factor analysis using the principal components method of the SPSS sub-program Factor was performed using the interview scores and the scores on each of the four P-test tasks as variables. Only one factor was extracted, with each variable loading substantially on that single factor. Factor loading and communalities are given in Table 11.

Table 11

Results of a Factor Analysis on the P-test Tasks and the Interview Tasks

<table>
<thead>
<tr>
<th>Variables</th>
<th>Factor Loadings</th>
<th>Commonality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pendulum (Inter.)</td>
<td>0.75966</td>
<td>0.57708</td>
</tr>
<tr>
<td>Colored Chem. (Inter.)</td>
<td>0.78671</td>
<td>0.61891</td>
</tr>
<tr>
<td>$P_1$ (Pendulum)</td>
<td>0.79114</td>
<td>0.62591</td>
</tr>
<tr>
<td>$P_2$ (Triangle)</td>
<td>0.47008</td>
<td>0.22098</td>
</tr>
<tr>
<td>$P_3$ (Rods)</td>
<td>0.62754</td>
<td>0.39381</td>
</tr>
<tr>
<td>$P_4$ (Proportionality)</td>
<td>0.72336</td>
<td>0.52325</td>
</tr>
</tbody>
</table>

It may be noted that task $P_2$ of the P-test was the only variable not loading better than .6 on the single factor. The task required a congruent copy of a given triangle to be
constructed using only an index card. The procedure is somewhat different than that originally used by Piaget (1960) and the scoring procedure of the task also deviated from Piaget's. On this task, Piaget classified subjects into a five stage range from earlier concrete to late formal while this researcher used only four classifications. A correlation of $r = .78 \ (p < .001)$, which is considerably higher than $r = .72$, was calculated between the interview scores and scores derived from P-test excluding P2.

Even though the Triangle task (p2) appeared to be measuring something in addition to what the other tasks were measuring, in the correlations among the tasks given in Table 7 for the entire sample of 231 subjects, the Triangle Task did not appear to be different from the other tasks.

Comparison of Independent Raters

In addition to verification of the P-test as a reliable measure of a subject's cognitive stage of reasoning, it is important that the results do not depend upon the person rating the subjects' responses. This issue arises because of the manner in which the test is to be analyzed. In some parts of the P-test there is not what could be called a single correct answer. In these instances, general response behaviors are described rather than specific correct answers. In fact, a subject's answer on the pendulum problem may reflect that he believes that weight is the operant variable, when this is actually not the case, but he still could
demonstrate formal operational behavior on this task. It is this characteristic of the P-test which was intended to make it as true to the Piagetian interview methods as possible.

In order to test whether or not the P-test can be objectively evaluated from a set of protocols given in Chapter III, two other raters besides the experimenter independently evaluated a random sample of 30 subjects' P-test responses. One rater (rater 1) is a developmental psychology professor whose area of study derives from Piaget's theories. This rater has a good understanding of Piaget's developmental stages and, as such, served as a counter-check on the experiment's interpretation; his evaluation also adds content validity to the test and its interpretation. The second rater (rater 2) is a philosophy professor with little formal knowledge of Piaget's work. This rater served to test whether the guidelines for evaluation produced an objective scoring of the P-test.

Hypothesis $H_2$ is, "There will be a significant inter-rater reliability coefficient derived for the independent evaluation of 30 subjects' P-test scores by the three raters." An interrater reliability coefficient was computed for three independent scorings of the 30 P-test results to test hypothesis $H_2$ at the .05 level. The sub-program, Rel, of the Ohio State University Statistical Package was used to compute this statistic. The interrater reliability adjusted for the mean of three judges was $r = 0.972$, significant well beyond the .01 level. Thus, hypothesis $H_2$ was accepted.
Table 12
Correlations Among Experimenter (E) and Additional Raters

<table>
<thead>
<tr>
<th></th>
<th>Rater 1</th>
<th>Rater 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>0.915</td>
<td>0.933</td>
</tr>
<tr>
<td>Rater 1</td>
<td>0.914</td>
<td></td>
</tr>
</tbody>
</table>

From these statistics, it seems safe to say that the test can be evaluated objectively, even though the evaluation protocols for the P-test require some responses to be judged by the content rather than by a predetermined fixed answer.

Summary

In component I, the development of the P-test, three hypotheses were stated. All were accepted. In addition, the test did separate college freshmen into different reasoning stages. Surprisingly, but consistent with other research, fifty percent of these freshmen were judged to be at or below the late concrete stage, level IIB.

Based on the review of the sample responses, the correlation between the interview scores, P-test scores, and the reliability coefficient computed for the P-test, it was argued that the P-test is a valid and reliable measure of Piagetian stages of reasoning for the freshman level college students in this study. In addition, it was argued that the guidelines written for the evaluation of the P-test could be used by other researchers to obtain similar results.
In Chapter V, relationships between the stages of reasoning, other mathematics tests, and achievement in mathematics courses will be analyzed.
CHAPTER V
RESULTS FOR COMPONENTS II AND III

Introduction
The results for Component II, an investigation of the relationship between the P-test and a standardized mathematics achievement test (Math ACT), and Component III, an investigation of the power of the P-test to predict mathematics achievement in six freshman mathematics courses, are presented in terms of the hypotheses and statistical analyses listed in Chapter III. For Component III the final determination of the mathematics class groupings will be made based on the P-test results as was planned in Chapter III. Hypotheses $H_6$ and $H_7$ will be rewritten to reflect the partitions among the six mathematics classes tested.

Results for Component II
This section examines the results of the prediction of subjects' Math ACT standard scores from the P-test tasks, the Pendulum Task (P1), the Triangle Task (P2), the Flexible Rods Task (P3), and the Proportional Reasoning Task (P4). In connection with this examination of the results, two hypotheses, $H_4$ and $H_5$, are tested. Hypothesis $H_4$ is "There will be a significant multiple $R$ derived from a regression
analysis with the P-test tasks predicting Math ACT standard scores;" and hypothesis $H_5$ is "The regression coefficients for each P-test task entered into the multiple regression analysis to predict Math ACT standard scores will be significant." To test these hypotheses a stepwise regression analysis will be used, contained in the SPSS (1975) sub-program package called Regression. As a first step in the development of the prediction equation derived from the regression analysis, a correlation matrix for the regression variables is presented in Table 13.

Each of the P-test tasks entered the regression equation with a significant regression coefficient at a level greater than .05. Thus, hypothesis $H_5$ stated in Chapter III is accepted. Also, the overall multiple $R$ for the regression analysis was significant beyond the .001 level and, therefore, hypothesis $H_4$ is also accepted. The implications and limitations of these results will be discussed in Chapter VI.
Table 13
Intercorrelations of Variables Employed in the Regression Analysis for Component II

<table>
<thead>
<tr>
<th>Pendulum Task</th>
<th>Triangle Task</th>
<th>Flexible Rods Task</th>
<th>Proportional Reasoning Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(1)</td>
<td>P(2)</td>
<td>P(3)</td>
<td>P(4)</td>
</tr>
<tr>
<td>Math ACT</td>
<td>0.525</td>
<td>0.456</td>
<td>0.445</td>
</tr>
<tr>
<td>Pendulum Task (P1)</td>
<td>0.397</td>
<td>0.445</td>
<td></td>
</tr>
<tr>
<td>Triangle Task (P2)</td>
<td>0.316</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexible Rods Task (P3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Correlations are significant at 0.001 level if r ≥ 0.241

From Table 14, an equation for the maximum prediction of Math ACT standard scores may be obtained from the predictor variables, the P-test tasks.

Table 14
Maximum Prediction Equation for Math ACT Standard Scores from the P-test Tasks

<table>
<thead>
<tr>
<th>Variable</th>
<th>Regression Coefficient</th>
<th>Standard Deviation of b</th>
<th>F ratio of b</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4</td>
<td>2.82836</td>
<td>0.38874</td>
<td>52.936</td>
</tr>
<tr>
<td>P1</td>
<td>1.83931</td>
<td>0.51488</td>
<td>12.761</td>
</tr>
<tr>
<td>P2</td>
<td>1.30889</td>
<td>0.46980</td>
<td>7.762</td>
</tr>
<tr>
<td>P3</td>
<td>1.18165</td>
<td>0.49400</td>
<td>5.722</td>
</tr>
<tr>
<td>Constant</td>
<td>2.26872</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard Deviation of Residuals = 5.50591
Multiple Correlation, R = 0.72039 \( R^2 = 0.51896 \)
F test of R = 51.51366 Degrees of Freedom, (4,191)

Notes: For a test of the F of b a value F = 6.81 is significant at the .01 level and an F = 3.91 is significant at the .05 level both with 1,191 degrees of freedom.
Because both hypothesis \( H_4 \) and \( H_5 \) were accepted, a post-hoc analysis was carried out as was planned in Chapter III. This analysis consisted of the computation of simple correlations between the P-test and the variables, Math ACT test, English ACT test, Social Science ACT test, Natural Science ACT test, Composite ACT level, the D-placement test, the B-placement test and H.S. Math Points. The resulting correlations are presented in Table 15. Since hypotheses \( H_4 \) and \( H_5 \) give evidence that the P-test is related to mathematics achievement as measured by Math ACT scores, these correlations serve to give additional information about the P-test's relationship to standardized tests of academic achievement.

Table 15

Correlations Between the P-test and Select Measures of Academic Achievement

<table>
<thead>
<tr>
<th></th>
<th>Math ACT</th>
<th>English ACT</th>
<th>Social Science ACT</th>
<th>Natural Science ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r )</td>
<td>0.707</td>
<td>0.353</td>
<td>0.469</td>
<td>0.587</td>
</tr>
<tr>
<td>( N )</td>
<td>196</td>
<td>196</td>
<td>196</td>
<td>196</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Composite Level ACT</th>
<th>D-placement Test</th>
<th>B-placement Test</th>
<th>H.S. Math Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r )</td>
<td>0.656</td>
<td>0.393</td>
<td>0.394</td>
<td>0.522</td>
</tr>
<tr>
<td>( N )</td>
<td>196</td>
<td>80</td>
<td>116</td>
<td>192</td>
</tr>
</tbody>
</table>

Notes: All correlation coefficients are significant at \( p \) .001.
Results for Component III

The Final Mathematics Class Groups and a Restatement of the Hypotheses for Component III

The first analysis for Component III must be made on the distributions of the P-test Scale within the six mathematics classes, Math 151, 148, 103, 102, and 100. The plan outlined in Chapter III called for the combination of Math 151, 150, and 148 as one sub-group population and Math 103, 102, and 100 as a second sub-group population. But whether or not it makes sense to combine classes within these two initial divisions depends on the P-test distributions within these classes. The distributions are listed in Table 16.

Table 16

Distributions of the P-test Scale Within the Six Mathematics Classes

<table>
<thead>
<tr>
<th>P-TEST LEVEL</th>
<th>100</th>
<th>102</th>
<th>103</th>
<th>148</th>
<th>150</th>
<th>151</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>44</td>
<td>12</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>63</td>
</tr>
<tr>
<td>Row Pct</td>
<td>69.8</td>
<td>19.0</td>
<td>9.5</td>
<td>0.0</td>
<td>1.6</td>
<td>0.0</td>
<td>36.8</td>
</tr>
<tr>
<td>Col Pct</td>
<td>80.0</td>
<td>48.0</td>
<td>20.7</td>
<td>0.0</td>
<td>3.4</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Tot Pct</td>
<td>25.7</td>
<td>7.0</td>
<td>3.5</td>
<td>0.0</td>
<td>0.6</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

| Count        | 9   | 9   | 17  | 11  | 13  | .3  | 62    |
| Row Pct      | 14.5| 14.5| 27.4| 17.7| 21.0| 4.8 | 36.3  |
| Col Pct      | 16.4| 36.0| 58.6| 52.4| 44.8| 25.0|       |
| Tot Pct      | 5.3 | 5.3 | 9.9 | 6.4 | 7.6 | 1.8 |       |

| Count        | 2   | 4   | 6   | 10  | 15  | 9   | 46    |
| Row Pct      | 4.3 | 8.7 | 13.0| 21.7| 32.6| 19.6| 26.9  |
| Col Pct      | 3.6 | 16.0| 20.7| 47.7| 51.7| 75.0|       |
| Tot Pct      | 1.2 | 2.3 | 3.5 | 5.8 | 8.8 | 5.3 |       |

| Column       | 55  | 25  | 29  | 21  | 29  | 12  | 171   |
| Total        | 32.2| 14.6| 17.0| 12.3| 17.0| 7.0 | 100.0 |

\[ \chi^2 = 97.05698 \text{ with } df = 10, \quad p = .0001 \]
Contingency Coefficient = 0.60173
Examination of Table 16 indicates that the three mathematics classes, Math 103, 102, and 100, have markedly different distributions for their P-test Scale assessment. Figure 7 pictures graphically the distribution differences for these three classes. If they were combined, standard scores derived from each class's different final examination would have to be used for any analysis carried out, and a problem would arise when P-test scores are compared to derived standard scores. This would be the certain occurrence of subjects in Math 100 with low P-test scores (relative to the two mathematics classes) and high standard scores derived from the Math 100 final examinations together with subjects in Math 103 with high P-test scores (relative to the two classes) and low standard scores derived from the Math 103 final examination. What relationship the P-test scores have to mathematics achievement as measured by the standard scores would be uninterpretable with these conditions. A similar argument can be made for not grouping Math 100 with Math 102 and for not grouping Math 102 with 103.
Further examination of Table 16 reveals that the P-test Scale distributions on Math 148 and 150 are very similar but that they both are different than that of Math 151. The argument made for not combining Math 100, 102 and 103 does not apply to Math 148 and 150 because of the similar P-test Scale distributions. There are subjects in Math 148 who can be matched with subjects in Math 150 based on their P-test scores. Thus, if there are subjects in one class who are low on one measure, the P-test, and high on the other
measure, standardized final exam scores, along with subjects in the other class with high P-test scores and low standardized final exam scores, it will not be an artifact of unequal distributions on the P-test. As for Math 151 compared with either Math 148 or Math 150, the argument for not combining classes with different P-test distributions applies. To be able to see more easily the similarities and the differences among the P-test Scale distributions for Math 148, 150, and 151, these distributions are graphed in Figure 8.

Figure 8
P-test Scale Level Distribution for Math 148, 150 and 151
To test statistically the assertion that the Math 148 and 150 classes do not have different P-test distributions, a contingency table for these two classes and their P-test Scale distributions was constructed; Table 17 contains the contingency table for Math 148 and 150.

Table 17
Contingency Table for Comparisons of Math 148 with Math 150 on the P-test Scale

<table>
<thead>
<tr>
<th>P-test Scale</th>
<th>Math Class</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>148</td>
<td>150</td>
</tr>
<tr>
<td>Count</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Row Pct</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Col Pct</td>
<td>0.0</td>
<td>3.4</td>
</tr>
<tr>
<td>Tot Pct</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Count</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Row Pct</td>
<td>45.8</td>
<td>54.2</td>
</tr>
<tr>
<td>Col Pct</td>
<td>52.4</td>
<td>44.8</td>
</tr>
<tr>
<td>Tot Pct</td>
<td>22.0</td>
<td>26.0</td>
</tr>
<tr>
<td>Count</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Row Pct</td>
<td>40.0</td>
<td>60.0</td>
</tr>
<tr>
<td>Col Pct</td>
<td>47.6</td>
<td>51.7</td>
</tr>
<tr>
<td>Tot Pct</td>
<td>20.0</td>
<td>30.0</td>
</tr>
<tr>
<td>Column</td>
<td>21</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td>42.0</td>
<td>58.0</td>
</tr>
</tbody>
</table>

$\chi^2 = 0.90996$ with df = 2 and $p = 0.6345$

The $\chi^2$ value computed by the SPSS Crosstabs sub-program was 0.09 with two degrees of freedom and the significance level was $p = 0.63$. A hypothesis that the P-test level distributions are different for Math 148 and 150 could be rejected.
Based on the results of the examination of the P-test Scale distributions among the mathematics classes, analyses outlined for Component III will be done with five different populations. These populations will consist of Group a, the Math 100 class; Group b, the Math 102 class; Group c, the Math 103 class; Group d, the Math 148 and 150 classes combined; and Group e, the Math 151 class.

Because of the nature of final mathematics class divisions, hypotheses $H_6$ and $H_7$ will be restated for each of the appropriate groups.

$H_{6a}$: The increase in $R^2$ due to the P-test will be significant after the B-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in Group a (Math 100).

$H_{6b}$: The increase in $R^2$ due to the P-test will be significant after the B-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in Group b (Math 102).

$H_{6c}$: The increase in $R^2$ due to the P-test will be significant after the B-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in Group c (Math 103).

$H_{6d}$: The increase in $R^2$ due to the P-test will be significant after the D-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in Group d (Math 148 and 150 combined).

$H_{6e}$: The increase in $R^2$ due to the P-test will be significant after the D-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in Group e (Math 151).
In this section the power of the P-test to predict mathematics achievement as measured by the Math 100 final examination scores will be examined after the variance in the examination scores accounted for by the variables B-placement test and H.S. Math Points has been removed. To accomplish this investigation, the B-placement test and H.S. Math Points, in that order, were entered into a hierarchical decomposition regression method followed by the P-test. This method gives the increment in $R^2$ at each step to be taken as part of variation in the dependent variable attributable to the particular variable added at that step of the regression. Thus, this method will be used to test hypothesis $H_{6a}$ for Group a, which is "The increase in $R^2$ due to the P-test will be significant after the B-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in Group a (Math 100)."

As a first step in the examination of the independent regression variables relationship to the variance in the Math 100 examination are presented in Table 18.
Table 18

Intercorrelations of Variables Employed in the Regression Analysis for Group a (Math 100) of Component III - the P-test Investigation

<table>
<thead>
<tr>
<th>B-placement Test</th>
<th>H.S. Math Points</th>
<th>P-test Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math 100 Final Exam</td>
<td>0.385</td>
<td>0.315</td>
</tr>
<tr>
<td>B-placement Test</td>
<td>0.081</td>
<td>-0.164</td>
</tr>
<tr>
<td>H.S. Math Points</td>
<td>0.112</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Correlations are significant at the .05 level if $r > 0.332$ (N=35)

The statistics required to test hypothesis $H_6$ for Math 100 will be given in Table 19.

Table 19

Summary Regression Table Used to Test Hypothesis $H_{6a}$: Math 100 Final Exam Variance Accounted for by P-test with Variance Accounted for by other Variables Removed

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables</th>
<th>Multiple R</th>
<th>R Square</th>
<th>Change in R Square</th>
<th>Simple R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B-placement Test</td>
<td>0.38526</td>
<td>0.14843</td>
<td>0.14843</td>
<td>0.38526</td>
</tr>
<tr>
<td>2</td>
<td>H.S. Math Points</td>
<td>0.47938</td>
<td>0.22980</td>
<td>0.08137</td>
<td>0.31537</td>
</tr>
<tr>
<td>3</td>
<td>P-test</td>
<td>0.56911</td>
<td>0.32389</td>
<td>0.09409</td>
<td>0.27284</td>
</tr>
</tbody>
</table>

$F$ for change in $R^2$ from step 2 to step 3 is $F=4.3141$ with degrees of freedom $(1,31)$ and significance $P < .05$.

The results given in Table 19 list the $F$ ratio for the test of significance of change in $R^2$ attributable to the $P$-test above that accounted for by the other variables as significant at the .05 level. Thus, hypothesis $H_{6a}$ is accepted
for Group a.

Since hypothesis $H_{6a}$ was accepted, a post-hoc analysis involving the individual P-test tasks will be given as planned in Chapter III. This analysis consisted of a second regression procedure similar to the first used to test hypothesis $H_{6a}$. The difference between the two regressions was that P-test tasks were allowed to enter as separate variables rather than as a sum, the P-test score. In this way, an investigation may be made of what contribution the individual tasks made in the results obtained from the first regression analysis. Table 20 gives the correlations among the variables in this second regression analysis.

<table>
<thead>
<tr>
<th>B-placement Test</th>
<th>H.S. Math Points</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math 100</td>
<td>0.385</td>
<td>0.315</td>
<td>0.084</td>
<td>0.029</td>
<td>0.187</td>
</tr>
<tr>
<td>Final Exam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-placement Test</td>
<td>0.081</td>
<td>-0.051</td>
<td>-0.067</td>
<td>-0.184</td>
<td>-0.106</td>
</tr>
<tr>
<td>H.S. Math Points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Points</td>
<td>0.002</td>
<td>-0.026</td>
<td>0.239</td>
<td>0.081</td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Correlations are significant at the .05 level if $r > 0.332$ ($N=35$).
Results are given in Table 21 obtained from the hierarchical decomposition of the Math 100 final examinations variance which uses the six independent variables, B-placement test, H.S. Math Points, the Pendulum Task (P1), the Triangle Task (P2), the Flexible Rods Task (p3), and the Proportional Reasoning Task (P4). Again, the B-placement test and H.S. Math Points are entered in the regression equation first in the order listed and then the P-test Tasks entered from the third regression step in a usual stepwise regression procedure to predict the remaining variance in the dependent variable.

Table 21

Summary Regression Table of Math 100 Final Exam Variance
Accounted for by P-test Tasks with Variance
Accounted for by other Variables Removed

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>Multiple R</th>
<th>R Square</th>
<th>Change</th>
<th>R Square</th>
<th>Simple R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B-placement</td>
<td>0.38526</td>
<td>0.14843</td>
<td>0.14843</td>
<td>0.38526</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>H.S. Math Points</td>
<td>0.47938</td>
<td>0.22980</td>
<td>0.08137</td>
<td>0.31537</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>P4</td>
<td>0.58374</td>
<td>0.34076</td>
<td>0.11096</td>
<td>0.31484</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>P3</td>
<td>0.61067</td>
<td>0.37292</td>
<td>0.03216</td>
<td>0.18723</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>P1</td>
<td>0.61313</td>
<td>0.37593</td>
<td>0.00301</td>
<td>0.08491</td>
<td></td>
</tr>
</tbody>
</table>

F ratio for change $R^2$ from Step 2 to Step 3 $5.218^{a}$ 1,31
F ratio for change $R^2$ from Step 2 to Step 4 $3.423^{a}$ 2,30
F ratio for Change $R^2$ from Step 2 to Step 5 $2.264$ 3,29

$^a$Significant, p .05.

While the sum of the P-test tasks, the P-test Score, did add a significant increase to the explained variance in the
dependent variable (9.4 percent), two of the tasks taken together add more. After the variance accounted for by the B-placement test and H.S. Math Points was removed, the two variables, the Proportional Reasoning Task and the Flexible Rods Task, add more than 14 percent to the amount of explained variance. Both these P-test tasks require proportional reasoning skills. These results could have implications for the mathematics placement procedures at The Ohio State University; these implications will be discussed in Chapter VI.

The Group b (Math 102) Investigation for Component III

In this section the P-test's power to predict mathematics achievement as measured by Math 102 final examination scores will be examined after the variance in the examination scores accounted for by the variables the B-placement test and H.S. Math Points has been removed. The same hierarchical decomposition regression procedure described for Group a will be used in the analysis to test hypothesis $H_{6b}$ for Group b. Hypothesis $H_{6b}$ is "The increase in $R^2$ due to the P-test will be significant after the B-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in Group b (Math 103)."

The first step of the analysis of the regression procedure required an examination of the correlations among the variables included. Table 22 lists these correlations.
Table 22

Intercorrelations of Variables Employed in the Regression Analysis for Group b (Math 102) of Component III

<table>
<thead>
<tr>
<th></th>
<th>B-placement Test</th>
<th>H.S. Math Points</th>
<th>P-test Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math 102 Final Exam</td>
<td>0.712</td>
<td>0.380</td>
<td>-0.436</td>
</tr>
<tr>
<td>B-placement Test</td>
<td></td>
<td>0.396</td>
<td>-0.238</td>
</tr>
<tr>
<td>H.S. Math Points</td>
<td></td>
<td></td>
<td>-0.052</td>
</tr>
</tbody>
</table>

Notes: Correlations are significant at the .05 levels if $r > 0.468$ (N=18).

Only for the B-placement test, the present main variable for predictions of success in Math 102, was the correlation coefficient with the mathematics test significant. The statistics required to test hypothesis $H_{6b}$ for Group b are given in Table 23.

Table 23

Summary Regression Table Used to Test Hypothesis $H_{6b}$: Math 102 Final Exam Variance Accounted for by P-test with Variance Accounted for by other Variables Removed

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>Multiple R</th>
<th>R Square</th>
<th>Change R Square</th>
<th>Simple R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B-placement</td>
<td>0.71210</td>
<td>0.50709</td>
<td>0.50709</td>
<td>0.71210</td>
</tr>
<tr>
<td>2</td>
<td>H.S. Math Points</td>
<td>0.72007</td>
<td>0.51850</td>
<td>0.01141</td>
<td>0.38025</td>
</tr>
<tr>
<td>3</td>
<td>P-test</td>
<td>0.77258</td>
<td>0.59688</td>
<td>0.07838</td>
<td>-0.43619</td>
</tr>
</tbody>
</table>

$F$ for change in R Square from Step 2 to Step 3 is, $F=2.7221$ with degrees of freedom $(1, 14)$ and significance $p < .05$. 
The significance level for acceptance or rejection of hypothesis $H_{6b}$ for Group b was .05; thus, this hypothesis is rejected for Group b.

Because hypothesis $H_{6b}$ was rejected, there was no post-hoc regression analysis for Group b. The interpretations of these results will be discussed in Chapter VI.

**The Group c (Math 103) Investigation for Component III**

In this section the power of the P-test to predict mathematics achievement as measured by Math 103 final examination scores will be examined. The same procedures used in the analysis for Groups a and b will be used. The B-placement test and H.S. Math Points are to be entered first and second into a hierarchical decomposition regression analysis followed by the P-test. With this procedure a test of hypothesis $H_{6c}$ for Group c is to be made. Hypothesis $H_{6c}$ states: "The increase in $R^2$ due to the P-test will be significant after the B-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in Group c (Math 103)."

The first step in analyzing the regressions procedure will be to report the correlations among the regression variables. Table 24 lists these correlations.
Table 24

Intercorrelations of Variables Employed in the Regression Analysis for Group c (Math 103) of Component III

<table>
<thead>
<tr>
<th></th>
<th>B-placement Test</th>
<th>H.S. Math Points</th>
<th>P-test Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math 103 Final Exam</td>
<td>0.217</td>
<td>0.625</td>
<td>0.063</td>
</tr>
<tr>
<td>B-Placement Test</td>
<td>0.326</td>
<td></td>
<td>0.214</td>
</tr>
<tr>
<td>H.S. Math Points</td>
<td></td>
<td></td>
<td>0.538</td>
</tr>
</tbody>
</table>

Notes: Correlations are significant at the .05 level if $r \geq 0.468$ (N=18).

For the Math 103 population the principal predictor in the O.S.U. placement procedure, the B-placement test, fails to correlate significantly with the criterion for mathematics achievement.

The statistics required to test hypothesis $H_6$ for Math 103 are presented in Table 25.
Table 25
Summary Regression Table Used to Test Hypothesis $H_{6c}$:
Math 103 Final Exam Variance Accounted for by P-test with Variance Accounted for by other Variables Removed

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>Multiple R</th>
<th>R Square</th>
<th>Change in R Square</th>
<th>Simple R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B-placement Test</td>
<td>0.21715</td>
<td>0.04716</td>
<td>0.04716</td>
<td>0.21715</td>
</tr>
<tr>
<td>2</td>
<td>H.S. Math Points</td>
<td>0.62498</td>
<td>0.34060</td>
<td>0.34345</td>
<td>0.62482</td>
</tr>
<tr>
<td>3</td>
<td>P-test</td>
<td>0.70422</td>
<td>0.49592</td>
<td>0.10532</td>
<td>0.06329</td>
</tr>
</tbody>
</table>

F for change in R Square from Step 2 to Step 3 is $F=2.9251$ with degrees of freedom $(1,14)$ and significance $p = .05$.

From the examination of Table 24 and 25 it may be seen that the only significant predictor of mathematics achievement for Group c is the measure of the subject's mathematics performance in high school, the variable H.S. Math Points. Based on these data $H_{6c}$ for Group c must be rejected.

Because hypothesis $H_{6c}$ was rejected, no post-hoc regression analysis was made for Group c. Interpretation and implications of these results will be discussed in Chapter VI.

The Group d (Math 148 and 150) Investigation for Component III

Math 148 and 150 were combined to form one mathematics class group, Group d, upon which the analyses outlined for Component III will be done. The argument to combine these two classes was given in Chapter III. Also, the D-placement test is the primary traditional predictor for success in
Math 148 and 150. As such, it is used where the B-placement test was used in the analyses and reports made for Groups a, b, and c. Except for these changes, the same statistical procedures were used and will be reported in this section for Group d.

The test of hypothesis $H_{7d}$ is quite similar to the group of $H_6$ hypotheses tests for Groups a, b, and c; the difference is that the D-placement test is one of the dependent variables named in $H_{7d}$ and the B-placement test is named in the same role for the whole group of $H_6$ hypotheses. Hypothesis $H_{7d}$ is "The increase in $R^2$ due to the P-test will be significant after the D-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in Group d (Math 148 and 150 combined)." The test of this hypothesis for Group d is reported in this section.

The same procedures used to test the group of $H_6$ hypotheses will be used to test $H_{7d}$ for Group d. This procedure is a hierarchical decomposition of the dependent variable's variance by a regression analysis. The order of the independent variables required to enter the regression equation in order to test $H_{7d}$ is the D-placement test, H.S. Math Points, followed by the P-test. In this manner, the third step allows for a measure of the variance in the dependent variable that can be accounted for by variance in the P-test after that variance in the dependent variable which can be accounted for by the D-placement test and H.S. Math Points
is removed.

To begin a review of the results, the correlations between the regression variables is reported in Table 26.

Table 26
Intercorrelations of Variables Employed in the Regressions Analysis for Group d (Math 148 and 150) of Component III

<table>
<thead>
<tr>
<th></th>
<th>D-placement Test</th>
<th>H.S. Math Points</th>
<th>P-test Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math 148 and 150</td>
<td>-0.032</td>
<td>0.143</td>
<td>-0.002</td>
</tr>
<tr>
<td>Exam Std Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-placement Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H.S. Math Points</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Correlations are significant at the .05 level if \( r \geq 0.322 \) (N=37)

Another puzzling situation presents itself in these results. All the predictor variables have a near zero correlation with the dependent variable. Since the P-test's correlation with mathematics achievement in this group is statistically not different from zero and nearly zero in absolute terms, hypothesis \( H_{7d} \) must be rejected. The summary of a regression analysis would not serve to add any information that cannot easily be deducted from Table 26.

Again, because the hypothesis for this group was rejected, no post-hoc analysis was warranted. The interpretations and implications of these results will be given in Chapter VI.
The last and fifth mathematics class group to be investigated for Component III is Group e (Math 151). Two limitations are to be noted and kept in mind as the results are reported. First, the population size for Math 151 is ten. There were only twelve freshmen registered for this course and two withdrew from school before the end of the quarter. Second, each of the subjects in this group scored a perfect 4 on the Proportional Reasoning Task (P4). Thus, this task does not discriminate between subjects in this group. The effect of this on the statistics is obvious. The P-test score distribution for this group would be the same with or without the scores from Task P4. Also correlations between the Proportional Reasoning Task (P4) and other variables obviously can not be computed, and, as such, this variable cannot be used in the analyses required in this section.

Other than the limitations mentioned above, the same regression procedures outlined for Group d were applied to the data recorded for Group e. The hierarchical variance decomposition of the dependent variable, mathematics achievement, measured by the Math 151 final exams is reported. This procedure allows for a test of hypothesis $H_{7e}$ to be made for Group e. Hypothesis $H_{7e}$ is, "The increase in $R^2$ due to the P-test will be significant after the D-placement test and H.S. Math Points are entered into the regression equation to predict mathematics achievement in Group
Thus, the entry order for the independent variables required by this regression analysis to test hypothesis $H_7e$ is the D-placement test, H.S. Math Points, and the P-test.

The first step in a review of the results is the presentation of the correlations between the variables in the regression analysis. These correlations are given in Table 27.

### Table 27

<table>
<thead>
<tr>
<th></th>
<th>D-placement Test</th>
<th>H.S. Math Points</th>
<th>P-test Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math 151 Final Exam</td>
<td>0.632</td>
<td>0.493</td>
<td>0.311</td>
</tr>
<tr>
<td>D-placement Test</td>
<td>-0.180</td>
<td>-0.180</td>
<td>0.387</td>
</tr>
<tr>
<td>H.S. Math Points</td>
<td>0.080</td>
<td>0.080</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Correlations are significant at .05 level if $r < 0.632$ (N=10).

It seems that for Group e (Math 151), the only significant predictor of mathematics achievement is the D-placement examination. The results of ten regression analysis are reported in Table 28.
Table 28
Summary Regression Table Used to Test Hypothesis \( H_7e \):
Math 151 Final Exam Variance Accounted for by P-test with Variance Accounted for by other Variables Removed

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>Multiple R</th>
<th>Multiple R Square</th>
<th>Change in R Square</th>
<th>Change in Simple R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>D-placement</td>
<td>0.63242</td>
<td>0.39995</td>
<td>0.39995</td>
<td>0.63242</td>
</tr>
<tr>
<td>2</td>
<td>H.S. Math Points</td>
<td>0.88356</td>
<td>0.78068</td>
<td>0.38072</td>
<td>0.49336</td>
</tr>
<tr>
<td>3</td>
<td>P-test</td>
<td>0.88407</td>
<td>0.78158</td>
<td>0.00090</td>
<td>0.31116</td>
</tr>
</tbody>
</table>

F for change in R Square from Step 2 to Step 3 is, \( F = 0.0147 \) with degrees of freedom \( (1,6) \) and significance \( p .25 \).

From the data reported in Table 28 it is seen that only 21.9 percent of variance in Math 151 final examination scores remains unexplained after the D-placement test and H.S. Math Points are entered into the regression equation. Of this remaining variance, the P-test explains very little of it. Thus, hypothesis \( H_7e \) for Group e must be rejected, and no post-hoc analyses are warranted.

Comparisons of Correlations Between P-test and the Formal and Concrete Portions of the Mathematics Class Groups Final Examinations

Described in Chapter III was the construction of the mathematics final examinations for the mathematics classes included as part of this study. Briefly, instructors in each class constructed their final examinations as usual except that they were asked by the experimenter to include four problems which were not presented as examples in the text or
classroom. This part of the examination was designated the "formal portion," and the rest of the examination was called the "concrete portion." The instructors assured the experimenter that solutions for the problems included in the concrete portion had been demonstrated by examples in their classrooms and in homework for the course.

The general assumption behind this aspect of the investigation was that those capable of formal reasoning as judged by the P-test would do much better on the formal portion than those judged not to be capable of formal reasoning, and that difference between formal and non-formal reasoners on the concrete portion would not be large. This situation, if it holds true, should result in a significantly larger correlation between the P-test and the formal portion of the examination than between the P-test and the concrete portion. Thus, hypothesis $H_8$ was stated as "The correlation between the P-test and the formal part of the mathematics examinations will be significantly larger than the correlation between the P-test and the concrete part of the mathematics examinations within the defined mathematics class groups."

Because the results for each math class group were uniformly not significant, hypothesis $H_8$ was not restated for each group. A t-score for the difference between correlations was used to test hypothesis $H_8$ t-scores and the resulting t-scores with the level of statistical significance are given in Table 29.
Table 29
Correlations Used for the Test of Hypothesis H₈ for each Mathematics Class Group

<table>
<thead>
<tr>
<th></th>
<th>Group a (Math 100)</th>
<th>Group b (Math 102)</th>
<th>Group c (Math 103)</th>
<th>Group d (Math 148 &amp; 150)</th>
<th>Group e (Math 151)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-test with Formal Portion</td>
<td>0.3551</td>
<td>-0.0418</td>
<td>-0.0380</td>
<td>0.1016</td>
<td>0.2610</td>
</tr>
<tr>
<td>P-test with Concrete Portion</td>
<td>0.3237</td>
<td>-0.3230</td>
<td>0.0983</td>
<td>-0.1469</td>
<td>0.2925</td>
</tr>
<tr>
<td>Formal Portion with Concrete Portion</td>
<td>0.4904</td>
<td>0.5553</td>
<td>0.3530</td>
<td>0.1966</td>
<td>0.5708</td>
</tr>
<tr>
<td>Sample Size N</td>
<td>44</td>
<td>22</td>
<td>19</td>
<td>39</td>
<td>10</td>
</tr>
<tr>
<td>t-score</td>
<td>0.217</td>
<td>0.368</td>
<td>0.483</td>
<td>0.5760</td>
<td>0.0950</td>
</tr>
<tr>
<td>Significance p</td>
<td>0.25 p</td>
<td>0.25 p</td>
<td>0.25 p</td>
<td>0.25 p</td>
<td>0.25 p</td>
</tr>
</tbody>
</table>

On the basis of the t-scores presented in Table 29, hypothesis H₈ is rejected. The patterns that may be discerned among the correlations listed in this table will be discussed in Chapter VI.

**A Check for Sex Differences In Mathematics Class Groups**

As was noted in the literature review of Chapter II, there is still an unresolved question concerning sex differences on Piagetian measures. Thus, this section will examine differences which exist between the sexes on the P-test for the mathematics class groups. For this examination one
hypothesis has been stated to be tested for each mathematics class group. This hypothesis was listed as $H_g$: "The mean for male subjects is significantly different from the mean for female subjects on the P-test within the defined mathematics class groups." Again, as was the case with hypothesis $H_g$, the results are uniformly not significant; and, as such, hypothesis $H_g$ is not restated for each mathematics class group.

Since there is only one female subject in the Group e (Math 151), this group will not be analyzed for sex differences. Table 30 lists the data for testing the differences between male and female P-test means within mathematics class groups. A t-score and significance level is reported for each group.

Table 30

Comparison of P-test Means by Sex for Math 100, 102, 103, and Math 148 Combined with 150

<table>
<thead>
<tr>
<th>Sex</th>
<th>100 Group a</th>
<th>102 Group b</th>
<th>103 Group c</th>
<th>148 &amp; 150 Group d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>SD</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>7.09</td>
<td>2.07</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.84</td>
<td>2.47</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.80</td>
<td>2.26</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.87</td>
<td>2.08</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>6.81</td>
<td>2.02</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.08</td>
<td>2.50</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.88</td>
<td>1.83</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.88</td>
<td>2.10</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>t-score</td>
<td>0.486</td>
<td>0.231</td>
<td>0.947</td>
<td>0.008</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>52</td>
<td>23</td>
<td>21</td>
<td>45</td>
</tr>
<tr>
<td>Significance p</td>
<td>0.3</td>
<td>0.4</td>
<td>0.17</td>
<td>0.48</td>
</tr>
</tbody>
</table>
As can be seen from Table 30, none of the mean differences between sexes on the P-test is significant at $p = .05$ level. Differences found within the mathematics class groups would have required additional regression analyses for each sex. Because of these results, further analysis will not be required.

**Summary**

In Component II, an investigation of the relationship of the P-test with the mathematics achievement on standardized tests, two hypotheses were stated. Both were accepted. A strong relationship between the Mathematics ACT standard scores and the P-test was demonstrated. Evidence that the P-test is also related to standardized measures of academic achievement was presented as part of a post-hoc analysis. The primary results were the derivation of a maximum prediction equation for the Mathematics ACT standard scores from the P-test tasks.

In Component III, an investigation of the power of the P-test to predict mathematics achievement in six freshman mathematics courses, four hypotheses were stated. Hypothesis $H_6^*$, the first of the four, was restated in the form of three hypotheses, $H_{6a}^*$, $H_{6b}^*$, and $H_{6c}^*$. This consisted of replacing the phrase "defined mathematics class groups" with the appropriate mathematics group for each hypothesis. Hypothesis $H_{6a}^*$ was accepted and a post-hoc analysis was done to gain additional information for interpretation of the
result. Hypotheses \( H_{6b} \) and \( H_{6c} \) were rejected. The second hypothesis, \( H_7 \), was also split into two hypotheses, \( H_{7d} \) and \( H_{7e} \), for the same reason \( H_6 \) was split. Both hypothesis \( H_{7d} \) and \( H_{7e} \) were rejected.

An examination of the relationship of the P-test to two designated portions of the mathematics final examinations was investigated. The investigation led to the rejection of hypothesis \( H_8 \) for each mathematics class group.

Last, an examination was made of whether there were sex differences on the P-test with the mathematics class groups. No statistically significant sex differences were found. The stated hypothesis, \( H_9 \), was rejected.

The interpretations and implications of the results reported in this chapter will be discussed in Chapter VI; there will also be suggestions for further research that should follow from the results of this study.
CHAPTER VI
CONCLUSIONS

Summary

The primary objective of this study was to develop a version of a set of Piagetian tasks in the form of a group administered paper-and-pencil test. This test was designed to measure a subject's level of concrete or formal reasoning; and it was also designed to obtain results which could be claimed to be equivalent to results from administration of the tasks in the usual Piagetian clinical interview procedure. Thus, investigations based on a subject's level of cognitive development consistent with Piaget's theory could be carried out efficiently and systematically for large groups of subjects.

As stated in Chapter I, the impetus for the development of the P-test was a desire to explore mathematics learning. An underlying assumption for the study was that an extension of Piaget's work would be a useful avenue for exploration of mathematics learning. The reason for making such an assumption was derived from the nature of Piaget's theory. Piaget attempted to explain learning through the development of cognitive structures. These structures were used to explain how children learn and the theory explained how structures
are developed. Thus, by examination of the structures a child possesses, Piaget attempted to explain what concepts a child can learn and which he or she can not learn.

This study did not pretend to attack directly questions about which mathematics concepts a subject is capable of learning nor to explain which he or she is not capable of learning. This is the long term objective of the experimenter for which this dissertation was a first step. Piaget provided the theory and this study attempted to provide a tool to explore and apply that theory to mathematics learning for young adults.

It seemed that if the P-test were a usable tool with which to explore mathematical learning, then it would be highly related to mathematics achievement. Thus as a logical first application of the P-test this study measured the relationship between the P-test and mathematics achievement. Two different types of mathematics achievement were explored in the investigation of the relationship between the P-test and mathematics achievement. The first type used subjects' scores on the Math ACT examination as a measure of subjects' level of mathematics achievement. This provided a broad definition of mathematics achievement that was common to a large number of subjects (N=196) in the study.

The second type of mathematics achievement was tied to the specific college freshman mathematics course in which subjects were enrolled. The subjects' final examination scores in their mathematics class were used as a measure of
their mathematics achievement.

Thus, because of the logic and the opportunity to make the comparisons outlined, this study also included investigations of the relationship between mathematical achievement and the P-test. As a result, three separate yet related investigations formed a natural organization for the study. The study was divided into three components. Component I served to meet the primary objective of the study, the development and validation of the P-test. Components II and III centered on the relationship of the P-test to mathematical achievement. Component II was a global investigation in which Math ACT scores were used as a common measure of mathematical achievement for the largest possible population of subjects in the study. Component III was an investigation in which the relationship of the P-test to mathematical achievement was explored within specific freshman mathematics classes taught at OSU-Lima, Fall quarter, 1977.

Interpretation of the Results

Results for Component I

The P-test was used successfully to classify college freshmen into distinctive levels of concrete and formal reasoning. The distribution for 231 subjects' levels of reasoning as assessed by the P-test was: 20.8 percent at the early concrete level (IIA), 30 percent at the late concrete level (IIB), 10.4 percent at the transitional level (T), 29.1 percent at the early formal level (IIIA), and 10 percent
at the late formal level (IIIB). Surprisingly, but consistent with the research cited in Chapter II, slightly more than 50 percent of the subjects in this study were judged to be at the concrete level.

The sample for this population can not be claimed to be a random sample of the general young adult population nor even a typical college freshman population. The subjects are those who chose to attend a two year branch campus of a large state university. It is even difficult to compare the study population to other two-year institutions of higher learning in other states because the students at the Lima Campus are all enrolled in four-year programs; OSU does not offer an Associate Arts Degree at any of its campuses. Thus, the results of this study should be applied with caution when making generalizations about the total population of young adults.

Despite the above precautions, the results are in agreement with those cited by McKinnon (1976). His survey of college freshmen's reasoning abilities included seven different institutions which ranged from private to public and from large universities to two-year community colleges.

A comparison of the results in this study with those obtained by Lawson and Renner (1974) give evidence that they do fit into a developmental sequence and thus fit Piaget's description of cognitive development except for the ages at which different stages are acquired. Lawson and Renner evaluated students level of Piagetian reasoning in
grades seven through twelve. The results obtained from the present study fit just as might be expected for a measure that is tied to developmental growth. Table 31 list the data for comparison.

Table 31
Comparison of P-test Results to Those Derived by Other Researchers

<table>
<thead>
<tr>
<th>Grade</th>
<th>Sample Size</th>
<th>Percent by Stage</th>
<th>Percent Concrete</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Concrete</td>
<td>Trans.</td>
</tr>
<tr>
<td>7</td>
<td>96</td>
<td>83.3</td>
<td>15.6</td>
</tr>
<tr>
<td>8</td>
<td>108</td>
<td>76.9</td>
<td>19.4</td>
</tr>
<tr>
<td>9</td>
<td>94</td>
<td>81.9</td>
<td>13.8</td>
</tr>
<tr>
<td>10</td>
<td>94</td>
<td>72.3</td>
<td>21.3</td>
</tr>
<tr>
<td>11</td>
<td>99</td>
<td>70.7</td>
<td>23.2</td>
</tr>
<tr>
<td>12</td>
<td>97</td>
<td>66.0</td>
<td>21.6</td>
</tr>
<tr>
<td>Present Study</td>
<td>231</td>
<td>37.8</td>
<td>36.1</td>
</tr>
</tbody>
</table>

Comparison of the results of this study with different research reports only provides circumstantial evidence that the P-test is a valid Piagetian measure of a subject's cognitive level. In Component I, direct comparisons were made of subjects' performance on the P-test and their performance

---

1 The data for subjects in grades 7 through 12 were adopted from those given by Anton E. Lawson and John W. Renner, "A Quantitative Analysis of Responses to Piagetian Tasks and its Implications for Curriculum," Science Education, 58 (1974), p. 549.
on two Piagetian tasks administered as closely as possible to Piaget's clinical interview procedure. The correlation between 27 volunteer subjects' P-test score and their interview score was $r = 0.72$. Also, a multiple $R = 0.79$ was obtained from using the individual P-test tasks to predict the "interview scores." The only other researcher to make a similar comparison of paper-and-pencil tasks with clinically administered Piagetian tasks was Renner (1977). The best he could do with different combinations of a set of written tasks was to obtain an $R = 0.62$.

In addition to the statistical procedures reported for Component I, typical responses elicited from the P-test tasks were listed. These responses were compared to those Piaget had described. The comparisons demonstrated that the written tasks do elicit similar responses to those Piaget obtained through clinical interviews. Thus, it may be concluded that, for the population as a whole, the P-test's written versions of Piagetian tasks did not significantly alter the type of responses that were expected. This does not preclude the fact that a written test may have affected certain individuals differently than a test conducted in the clinical interview mode. But, the evidence here is that, on the whole, the P-test did compare favorably with the clinical approach with respect to the evaluation of a subject's Piagetian level of

---

2 John W. Renner, "Evaluating Intellectual Development Using Written Responses to Selected Science Problems," A Report to the National Science Foundation, 1977 (Number EPP75-19596), p. 120.
cognition.

The internal consistency of the P-test was also respectable, with an \( r = 0.72 \) and the interrater reliability for three independent rater's evaluations of 30 subjects' P-test responses was 0.97. Thus, the P-test is reliable and the evidence supports the claim that the experimenter's evaluation of subjects' cognitive level was assessed objectively.

Only one discrepancy seemed to stand out in the data gathered for Component I. The distribution of subjects by reasoning level on the Triangle Task was much different than that obtained from the other tasks. Over 54 percent were categorized as late concrete (IIB) on this task. The Triangle Task (P2), as the name suggests, is presented in a geometrical context. Solution for this problem requires the subject to be able to organize space by a coordinate system consciously or subconsciously and, thus, to locate points in space by forming perpendicular lines to imagined coordinate axes. The other tasks require either proportional reasoning or the ability to sort out the effects of different variables in an experimental situation. Even though the content of the questions is different, they all reportedly require formal reasoning.

Several possibilities may have caused this discrepancy. It could be that this situation was caused by what Piaget
calls a "horizontal decalage." The term applies to a situation in which subjects judged to be at one level on a given task are also judged to be at a lower or higher level of reasoning on a second task. Both tasks in this situation are supposed to require the same schemata for solution. In addition to Piaget's observation of this type of behavior, Elkind (1962) reported horizontal decalages on a set of three conservation tasks about mass, weight and volume respectively, for college freshmen. Each task involves the transformation of certain properties of an object (the object is usually a ball of clay or beaker of water), but the transformation does not alter the mass, weight or volume of the object. The same schema is required for an understanding that the mass, the weight or the volume has not changed. Subjects demonstrate the ability to conserve mass first. In time many subjects also conserve weight, but many more never seem to conserve volume.

Unlike the set of conservation tasks, the Triangle Task (P2) and the other P-test tasks require different specific schema. What the P-test tasks all have in common is that a subject must be capable of hypothetical reasoning; that is, the ability to formulate possibilities that are not readily

---


available to the subject through the concrete presentation of the problems within the tasks. Because of this, the horizontal decalage explanation does not completely fit this situation. But since there are no investigations exploring the exact relationship of geometrical tasks and non-geometrical tasks at the formal level, the possibility of a horizontal decalage cannot be dismissed.

A second possibility that should be explored as the cause of the distribution discrepancy is that the level of difficulty of the Triangle Task (P2) has been misjudged. Perhaps the task requires more than formal reasoning. Hints of a fifth-stage beyond the formal stage have been reported by Arlin (1976). In response to Arlin's findings, Cowan writes:

> First, we need to investigate formal operations in more detail, to decide which new ideas represent a decalage and which might stem from a more differentiated and integrated structure with the same basic rules.5

In light of the reports of large percentages of young adults not attaining formal reasoning, Cowan's remarks may be equally applicable to old as well as new ideas. Thus, the first and second possibilities may be two sides of one investigation. Assuming that the Triangle Task does measure something different from the other tasks, then one or the other of these two possibilities would form an appropriate explanation for the situation. That is, over 54 percent of

the P-test's subjects were judged to be at the late concrete level on the Triangle Task either because of a possible horizontal decalage in their reasoning abilities or because the Triangle Task may require reasoning beyond the formal stage.

There also remains the possibility that the design of the tasks itself was the cause of the difference. While the first two explanations are the more exciting possibilities, this third factor is equally plausible at this stage of the completed research. Plans are already underway to change the form of the Triangle Task. Specifically, a newer version will ask the subjects to write out an explanation of what they were attempting to accomplish by their constructions.

The factor analysis reported through Table 11 demonstrated that while the Triangle Task was the least related variable among the P-test tasks and the two Piagetian interview tasks, it still shared some measure of commonality (0.22) and had a factor loading of 0.4701 on the single factor produced. This suggests that the Triangle Task does share some common property or properties with the other tasks. Exactly what the common properties are remains a topic for future research.

From the results for Component I, it may be concluded that the P-test does represent a successful conversion of four Piagetian tasks into a paper-and-pencil test. Chapter IV listed examples of those responses elicited by the P-test and demonstrated that the responses compared favorably with those elicited by Piaget and his collaborators. In addition,
the responses on the P-test compared favorably with inter­view responses obtained in this study.

The P-test also sorted a substantial percentage of col­lege freshmen into each substage level of reasoning defined for concrete and formal operations. Thus, the first two general hypotheses stated in Chapter I can be accepted.

Results for Component II

The results of the comparison between P-test tasks scores and Math ACT standard scores demonstrate a strong linear relationship. For this component, a subject's Math ACT standard score was taken as a measure of his prior mathematics achievement. Thus, based on the results reported in Chapter V for Component II, the P-test can be said to be a good measure of a subject's achievement in mathematics. This result is taken as an indication that a measure of a subject's level of concrete and formal reasoning as measured by the P-test should be a fruitful direction in which to explore mathematical learning.

In addition to a strong relationship between the P-test and mathematics achievement, the results of Component II suggest that the P-test is highly related to other areas of academic achievement. In the post-hoc analysis for Component II correlations between the P-test and seven standardized measures of academic achievement were reported in Table 15. All of the correlations were significant beyond the .001 level. Statistically there is less than one
chance in one-thousand that any of these correlations would be as large as they are by chance alone. Thus, in spite of the fact that these results occur as part of a post-hoc analysis, they can be taken as evidence that if a subject has a high level of performance on the P-test, he or she will be likely also to have a high level of performance on standardized measures of academic achievement.

While the results in Component II do not in any way suggest why some subjects did well and others did not do so well on academic achievement test, they do demonstrate that the single variable, the P-test, accounted very well for how subjects performed on the achievement tests. What should be concluded from these results is that investigations to explore learning, especially mathematical learning, should be conducted using a subject's level of concrete or formal reasoning as starting point.

In this component another general hypothesis listed in Chapter I is given support. The P-test was highly related to standard measures of academic achievement.

Results for Component III

A subject's level of concrete or formal reasoning had very little to do with his performance on a final examination in a freshman mathematics course. All the hypotheses except one relative to the relationship between the P-test and mathematics achievement were rejected. Group a, the Math 100 class, was the only class in which the P-test added
to the prediction of mathematics final examination scores
given by the traditional predictors, the placement examina-
tion and high school mathematics points. The only positive
conclusions to draw from the data are speculative directions
for further research.

To partly explain why such different results were ob-
tained for this component from those reported for Component
II, the distribution of the P-test scores within classes was
closely reviewed. The performance on the P-test is highly
correlated with the level of the mathematics courses in the
study. This fact caused a restriction in the P-test scale
within the mathematics class groups. In Group a (Math 100),
80 percent of the subjects scored between 4 and 8; in Group
b (Math 102), 68 percent scored between 7 and 11; in Group c
(Math 103), 70 percent scored between 9 and 12; in Group d
(Math 148 and 150), 70 percent scored between 10 and 13; and
in Group e (Math 151), 67 percent scored between 12 and 14.
In each group there was a measure with at most a five point
range used to predict a measure with a much larger range.
Thus, the results can be partly attributed to what might be
called a deficiency in the P-test scoring. The reason that
the P-test tasks were scored as described in Chapter III
was to make direct comparisons with Piaget's work. Now that
this comparison has proven to be relatively successful, the
development of an alternate scoring scale for the P-test may
prove to remedy the deficiency created by the limited range
of scores.
Since the restricted P-test score ranges within each mathematics class group were nearly equal in length, the affect of these restrictions on the computed statistics should not have been greater or smaller in any one group. This makes the fact that statistically significant results were found only for Group a (Math 100) more impressive.

That the P-test improved the prediction of mathematics achievement in only Group a can be explained in one of two ways. The first way would be to assume that concrete operational thought is the highest level of cognitive functioning required for success in the mathematics courses studied. Crosswhite, in discussing his failure to find variables that measured cognitive reasoning levels and predicted mathematics success in OSU mathematics courses, stated:

Unanswered is the question of whether we are teaching what we should be teaching in the way in which it should be taught.6

He seems to reflect an opinion that is consistent with the assumption upon which the first explanation will be based.

To see how the above assumption could account for the results in Component III, it must be stated that 80 percent of the subjects in Group a had P-test scores distributed in the concrete level range. Thus, a good many of the Group a subjects were not even fully concrete operational. At the same time, over 52 percent of the subjects in the other

mathematics class groups were at or above the upper concrete stage of reasoning.

If concrete operational thought was the highest level a subject needed to be reasoning at in order to be successful in a mathematics course included in this study, then other variables besides the P-test would be more important in the final determination of achievement within the mathematics classes of Group b through Group e. The subjects in these classes would all be capable of high achievement because they were all at least concrete operational. In Group a, on the other hand, the P-test discriminated between subjects capable of concrete operational thought and those who had not fully attained concrete operational reasoning. As such, the possibility remained that the P-test could improve the prediction of mathematics achievement for Group a under the assumption that concrete operational thought was the highest cognitive level necessary for good achievement. Thus, under this assumption the results that did occur would be thought to be very likely.

A second possibility to account for the results in Component III will also be deduced from how the P-test scores were distributed within the mathematical class groups. Group e, besides containing only 10 subjects, contained 10 subjects all judged to be at or above the lower formal, IIIA stage. Large numbers of subjects in Group b, Group c and Group d scored in the middle range of the P-test score range; thirty-six percent in Group b, 58 percent
in Group c, and 48 percent in Group d scored between 9 and 11 on the P-test. As stated earlier, 80 percent of the Group a subjects scored in the lower range (below 8) on the P-test. Thus, subjects in Group a and Group e were judged primarily to be stable with respect to level of reasoning and large numbers of subjects in Group b, Group c, and Group d were judged to be in what could be labeled a transitional stage between concrete and formal operations.

Studies cited in Chapter II reported that it was transitional subjects that seemed to be most affected by attempts to influence their acquisition of formal operational thought patterns. Thus, it would seem that many subjects in Group b, Group c, and Group d might experience cognitive growth because of their ten weeks of college experience.

While the P-test was given in the first week of the Fall quarter, the final examinations were given ten weeks later. If cognitive growth had taken place for subjects in Group b, Group c, and Group d, then their performance on a final mathematics examination would reflect the performance of a subject at higher cognitive level than that at which they were assigned by the P-test ten weeks earlier.

If all these conditions outlined in the second possibility actually existed, then it would be expected that the P-test would be correlated in a positive direction with the mathematics final examination scores in Group a and Group e.

---

7 The P-test score range is from 4 to 16.
Also, near zero correlations would be expected between the P-test and mathematics final examination scores in Group b, Group c and Group d. Examination of the correlations listed in Table 25 shows that the correlations between the P-test and the mathematics final examination in Group a and Group e were in the positive direction while the same type of correlations in Group b, Group c and Group d were near zero or negative.

Thus, there is some evidence that significant prediction results were obtained for Group a and not Groups b, c, and d because of cognitive growth that may have taken place in the ten weeks between the administration of the P-test and the administration of the mathematics final examination.

The first explanation of the results is based on what can be considered a situation that should not exist. Mathematics at the college level, if not at lower levels, should challenge students to use formal reasoning. Thus, it is important that the first possibility be explored with further research. The second explanation presents a more pleasing possibility. It would be nice to think the college studies would cause cognitive growth to take place. This possibility should also be explored further by administering the P-test before and after certain college courses.

For Group a, where statistically significant results were obtained, a post-hoc analysis yielded additional interesting results. While the P-test accounted for an additional 9.4 percent of the variance in Math 100 final examination
scores over that accounted for by the B-placement test and H.S. Math Points, two of the separate P-test tasks together accounted for 14.3 percent of such variance. More significant than the increase is the fact that both P-test tasks, the Flexible Rods Task (P3) and the Proportional Reasoning Task (P4), require proportional reasoning skills for solutions. One implication which can be made from this is that problems requiring proportional reasoning would make good items for mathematics placement examinations.

Suggestions for Further Study

1. The limited results obtained in Component III suggest that this aspect of the study should be redone with a few changes. Since Math 100, 102, and 103 have common topics in their syllabae and, similarly, so do Math 148, 150, and 151, a common examination over common topics should be constructed for each of these two groups. Also, the P-test scale should be expanded. Both the experimenter and the additional P-test raters felt that finer distinctions could be made between subjects within a given category level. An experiment with these changes would be a better test of whether assessment of a subject's cognitive level can be used to predict achievement in a mathematics course.

2. Because the population in this study and those reported in Chapter II consisted of college students, studies to assess the cognitive level of young adults in the general population should be conducted. From this, a more accurate
picture of the extent of young adults' reasoning can be formed.

3. Large numbers of young adult college freshmen are found to be at or below the concrete level of reasoning. What happens to these students in long-term college programs which require a great deal of formal reasoning? This question should be researched using longitudinal methods.

4. Is it too late for subjects who are 19 years or older to acquire formal operations if they have not already done so? The literature seems to suggest that some do go on to obtain it (McKinnon, 1970; Kohlberg and Gilligan, 1971; Tomlinson-Keasey, 1972). But which ones acquire formal skills and why requires study. Administration of the P-test before and after certain college courses or experiences may provide some clues.

5. The results obtained on the four P-test tasks suggest that the one based in a geometrical context, the Triangle Task P(2), was different than the other P-test tasks. Three possibilities were given for the difference. These possibilities included the existence of a horizontal decalage, the existence of a fifth stage of reasoning beyond the formal operational stage, or a deficiency in the manner in which the Triangle Task P(2) was presented. Each of these three possibilities should be given attention in future research.

6. The relationship between subjects' level of reasoning and their mathematics achievement was explored in this
study. Future studies should continue to examine this relationship as well as the relationship between subjects' level of reasoning and the effectiveness of different methods of instruction.
APPENDIX A

THE P-TEST
Instructions

The test you are about to take consists of five problems or situations. Following the description of each problem or situation, you will be asked to respond to statements and questions concerning the problem or situation. Often the way you respond will be more important than whether your answer is correct or not correct; because of this, please be sure to give reasons for your answers when requested. Please be as complete as you can in making your responses.

If you should return to any question to make changes do not erase your first response. Simply cross it out with one single line and write the new response below it.

This test will not be used in any way that will affect your standing in this college or any thing else you pursue. The results will be held in strict confidence, so please relax and do the best you can to answer the questions.

There are 3 things to do before you begin the test:

1. Print your name to the right. Name _____________________________

2. Check the math course you are enrolled in this quarter.

   ___ Math 100
   ___ Math 102
   ___ Math 103
   ___ Math 148
   ___ Math 149
   ___ Math 150
   ___ Math 151
   ___ None of these

3. Print your name on the index card handed you. Put it with your last name first in the upper right corner. You will need this index card to help you answer situation B.
Situation A

A ball suspended from a string and set into a swinging motion is called a pendulum (see figure below). In this situation you will investigate the time it takes a pendulum to make "one swing." One swing of the pendulum will be defined as follows: Draw the ball off center to a point A of your choice (see figure) and release the ball from point A allowing it to swing to point B opposite A and back to point A; the path of the ball from point A to point B and back to A will be called "one swing" of a pendulum.

Figure

(1) What factors do you think could influence the time of one swing of the pendulum?

You will now use the diagrams of pendulums on the next page to help you answer the rest of the questions in this situation. The diagrams picture 9 different pendulums. They differ by the length of their string and/or the weight of their ball. The weight of the ball is given directly below each ball and length of the string can be read from the scale given to the left. Please turn the page and study the pictured pendulums. If you have any questions concerning the way the pendulums are pictured please ask the proctor at this time.
(2) Select three of the pictured pendulums so that your second choice takes a longer time to make one swing than your first choice and that your third choice takes a longer time to make one swing than does your second choice.

Choice 1 is ________
Choice 2 is ________
Choice 3 is ________

(3) Suppose you check the time for one swing of pendulum no. 4 and also for pendulum no. 9 and find that their times are different. Can you conclude from this what caused the time difference?

Yes or No? ________

Explain as best you can the reason for your answer.
Other people have suggested the factors listed below as those which influence the time of one swing of a pendulum.

(1) The distance that the ball is drawn back before it is released.
(2) The length of the string holding the ball.
(3) The weight of the ball.

You may have listed some of these factors or you may have given others.

(4) Tell which pendulums you would select and how you would use them to demonstrate which factors influence and which do not influence the time of one swing of a pendulum. Be sure your demonstration will leave no doubt as to which factors do and which do not have an influence in this situation.
Tell which pendulums you would choose and what observations you would have to make to conclude that the weight of a ball has no influence on the time of one swing of any of the pictured pendulums. It is a fact that weight of a ball does not effect the time, but this question required you to tell only how, through experimentation, you could discover this fact. Please take time to detail your experiment listing which pendulums you would investigate and what observations you would make to discover that weight has no influence and also explain how your experiment would allow you to conclude that weight has no effect on the time of one swing of a pendulum.
Situation B

Aided only by the index card provided you, draw as accurate a copy as you can of the triangle shown above. Draw your copy anywhere below the line of stars. Do not trace the triangle on the index card, but use it in any other manner in which you may choose.
Situation C

In this problem you will be asked questions about the flexibility of different types of rods. The degree of flexibility will be defined as the amount or distance the end of a rod moves downward (bends) when it is extended out over the side of a table with a 3 pound weight attached to it and while the other end is attached firmly to the table top. Naturally you will not be able to observe actual results since your rods are only diagramed. Referring to the diagrams, try to answer the questions that follow to the best of your ability. If you do not understand the diagrams please ask your proctor about them.
To answer question 1 please use diagram number 1 on the next page.

(1) If you tested the flexibility of rod no. 5, which rod do you think would bend about the same amount? Choose the best possible answer and PLEASE explain the reason or reasons for your choice.
Diagram 1
To answer question 2 please use diagram number 2 on the next page.

(2) If you tested the flexibility of rod no. 4, which rod do you think would bend about the same amount? Choose the best possible answer and PLEASE explain the reason or reasons for your choice.
To answer question 3 please use diagram number 3 on the next page.

(3) If you tested the flexibility of rod no. 1, which rod do you think would bend about the same amount? Choose the best possible answer and PLEASE explain your reason or reasons for your choice.
Diagram 3
To answer question 4 please use diagram number 4 on the next page.

(4) If you tested the flexibility of rod no. 5, which rod do you think would bend about the same amount? Choose the best possible answer and PLEASE explain the reason or reasons for your choice.
To answer the two questions on this page please refer to diagram 5 on the next page.

(5) Will rods 3 and 7 bend about the same amount? Why or why not?

(6) Will rods 1 and 5 bend about the same amount? Why or why not?
Diagram 5
(7) List all pairs of rods that could be compared to find out if brass rods bend more than steel rods. Explain how you decide what choices to make.

(8) List all pairs of rods that could be compared to find out if square rods bend more than round rods. Explain how you decided what choices to make.
Situation D

1. If you put an 8 inch page into an enlarging machine, it enlarges to 32 inches. What will be the final size of a 16 inch page?

Please explain or show how you arrived at your answer.

2. An auto salesman receives a commission of $200 on a $2800 car. How big would you expect his commission to be on the $3000 model?

Please explain or show how you arrived at your answer.
3. The sales tax in one state is $600 on a purchase of $3000. What is the tax on $18,000?

Please explain or show how you arrived at your answer.

4. Letters 2mm high appear to be 7mm high under a magnifying glass. How high would 3mm letters appear under the same magnifying glass?

Please explain or show how you arrived at your answer.
APPENDIX B
GUIDELINES FOR EVALUATION OF THE P-TEST
Task A - The Pendulum Situation

The schema required to answer questions in this situation has been described by Piaget as "all things being equal." That is, if all factors which may influence the outcome of an experiment are kept constant in two trials except for a single factor to be investigated, then the results of the two trials may be used to logically judge the effect of the single manipulated factor.

A pendulum is used to provide the setting for this experiment, and the factors under consideration will be limited to three factors. Only the weight of the bob (weight) and the length of the string (length) and the release point are considered.

Question 1 is not used in the evaluation. It is designed to allow the student some time to become familiar with the situation.

Question 2 checks whether or not subjects are capable of ordering the variables in question. It also allows the experimenter to begin to see what assumptions the subject brings to the experiment. Correct ordering relations for the pendulums would consist of a sequence of choices with successively increasing or decreasing values of a single factor while ignoring or controlling the other factor, or increasing decreasing values of both length and weight either in the same or opposite directions.

Examples of some correct ordering are as follows:

<table>
<thead>
<tr>
<th></th>
<th>choice 1</th>
<th>choice 2</th>
<th>choice 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>2</td>
<td>6</td>
<td>9 (length varied-weight controlled)</td>
</tr>
<tr>
<td>(2)</td>
<td>7</td>
<td>8</td>
<td>6 (weight varied-length ignored)</td>
</tr>
<tr>
<td>(3)</td>
<td>7</td>
<td>1</td>
<td>9 (both varied and increasing)</td>
</tr>
<tr>
<td>(4)</td>
<td>5</td>
<td>3</td>
<td>9 (weight varied-length controlled)</td>
</tr>
</tbody>
</table>
Examples of incorrect ordering are as follows:

<table>
<thead>
<tr>
<th>choice 1</th>
<th>choice 2</th>
<th>choice 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 2</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>(2) 2</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>(3) 9</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>(4) 6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>(5) 2</td>
<td>3</td>
<td>9</td>
</tr>
</tbody>
</table>

What the above have in common is the fact that neither weight nor length are ordered. A closer look would indicate that subjects are first using one variable between their first and second choice and then they switch to the other variable (length or weight) to compare their second and third choices. This type of behavior is characteristic of level II-B behavior; but before any subject should be judged to be at this level, his response on the other questions should be considered. For the purpose of scoring, the above response will be taken as incorrect.

Question 3 offers two pendulums for potential experimentation. The pendulums vary on both the weight and length factors. When the periods are compared they will be different, but there is no logical conclusion that may be drawn as to which factor caused the difference. The purpose of this question is to test whether subjects will anticipate the use of the schema "all things being equal." Piaget has claimed that one difference between a level III-B and a level III-A reasoner is that although both are capable of using the schema, a III-B person anticipates its application while a III-A person does not.

Thus question 3 poses a problem in evaluating subjects that already have some knowledge about pendulums. A subject who
knows or believes that a single factor is the only operant factor may misread the question and list the factor as the cause of the difference. For this study such a response will be considered a partial miss. Whether or not it is used to place the subject in level III-B or III-A should depend on the completeness of response to the other questions. This also holds for a judgment between III-A and II-B.

A typical subject in which the judgment should be made in favor of the higher level is as follows:

(1) for III-B over III-A, the subject answers question 3 with a "yes" and says it is because of the longer string. He also answers question 2 holding weight constant and varying string length, and is very complete in his answers to questions 4 and 5.

(2) for III-A over II-B, the subject answers exactly as described in (1) above except for question 4. In the case of question 4, he does not test for different release positions.

Some examples of incorrect responses to question 3 are as follows:

(1) **No**. No definite explanation given, i.e., "not enough information given"

   OR

   An explanation which does not focus on the fact that no conclusion may be drawn.

(2) **Yes**. An explanation as follows: "Pendulum 4 has a shorter string than does 9 so therefore it has less to travel and since pendulum 9 is heavier it also caused a difference.

Some examples of correct responses to question 3 are as follows:

(1) **No**. An explanation as follows: "It's hard to tell because of the different lengths and weights."
Yes. In spite of an answer of "yes" the subject demonstrates that he understands that the differences in both factors makes it impossible to judge and begins to relate what needs to be changed in order to judge what caused the difference. Although this subject misread the question he has demonstrated that he has anticipated the proper schema.

Questions 4 and 5 are such that they require the subject to use the proper schema. Thus a subject who has missed question 3 may demonstrate that he is capable of applying the schema.

In scoring question 4, subjects who correctly test for the effects of weight and length, but miss or ignore release point are to be given a partial miss.

To score this test subjects were given either an H (hit) for a question answered correctly or M (miss) for a question answered incorrectly or a "M?" for a question judged to be partially correct. Thus on a scoring card a subject's record may appear as follows:

<table>
<thead>
<tr>
<th>Questions</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Score</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject n</td>
<td>H</td>
<td>M</td>
<td>M?</td>
<td>H</td>
<td>2 1/2</td>
<td>II-B</td>
</tr>
<tr>
<td>Subject m</td>
<td>H</td>
<td>H</td>
<td>M?</td>
<td>H</td>
<td>3 1/2</td>
<td>III-A</td>
</tr>
<tr>
<td>Subject l</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>4</td>
<td>III-B</td>
</tr>
</tbody>
</table>

Reasoning Level Assignment

III-B Subjects answer questions 2 through 5 successfully as described in the "Nature of the Task" section

III-A Subjects have the equivalent of 3 out of the four questions 2 through 5 correct or 2 1/2 out of four because they missed 2 and partially missed 4.

II-B Subjects have less than the equivalent of 3 out of four questions correct, but demonstrate they have some understanding of the schema "all things being equal." (See examples of level II-B responses).
Subjects answers are quite incomplete and demonstrate that they are confused as how to deal with the situation.

To aid in judging a subject either II-B or II-A the following hypothetical set of responses are given below.

Example 1: The least a subject may score and still be II-B.

Question
(2) Choices are 9, 3 and 5 respectively or some other correct sequence.
(3) Answers "yes" and follows with a response that demonstrates no understanding of the schema necessary for understanding the situation.
(4) Chooses light ball on long string and heavy ball on short string. This seems to be the subject's preconceived notion about how pendulums work. Here no idea that his assumptions may be tested is demonstrated.
(5) Subject gives an indication that string length must be held constant and weight varied.

Since the hypothetical subject answers question 2 correctly and showed an indication in question 5 of understanding the schema, "all things being equal," he is to be judged a level II-B reasoner. If either 2 is incorrect or something equivalent to a correct demonstration in question 5 is missing, such a subject should be judged level II-A.

Example 2: A typical II-A subject.

Question
(2) Choices are 5, 3 and 9 respectively. Here the subject obviously has chosen weight as the operant factor.
(3) Answers "yes" and gives string length as the operant factor and ignores weight. This appears to be in contradiction to the assumptions under which the subject formulated his answer to question 2.
(4) The subject does not set up an experiment but instead tries to explain why a factor is operant or non-operant.
The subject responds in a fashion similar to that stated in (4) above.

Task B - The Triangle Copying Situation:

Subjects are asked to draw an exact copy of a triangle. They are provided only with the triangle, a piece of paper upon which to draw the copy and 3 x 5 index card. They are told to use the index card in any way they choose. The card can be used as both a straight edge and a means of transferring copies of line segments (the sides of the triangle). Also, it may be used as a tool to form perpendicular line segments by using the corner (a right angle) of the card; this application is a critical part of the task.

Piaget describes the acceptance by a subject of the necessities of a situation as a subset of the possible as evidence that the subject is able to reason at the formal stage. In this task the set of "necessities" consist of the three line segments forming the sides and angles of the triangle. In order for a subject to copy the triangle, the subject must be able to frame the triangle in a coordinate plane i.e., as a subset of a two-dimensional space. More specifically in this task, the subject must use a perpendicular from one side of the triangle (or its extension) to the adjacent side to measure angular separation (see illustration below). This act of forming the perpendicular is either an overt or subconscious intuitive superimposition of a coordinate system on the plane containing the triangle.
In the figure above segment D'C' or DC are to be formed using the corner of the index card as copy of a right angle. Length AD' or AD and D'C' or DC are then transferred to the copy using the index card. Once point C' or C is located on the copy, sides AC and BC may be copied forming each angle with exactly the same measure as in the original.

The other correct solution consists of the subject using a measure of AC (or BC) as a radius length with which to draw an arc with center at A (or B). Point C is on the arc and may be found by the intersection of the arc and BC (or AC).

Reasoning Level Assignments:

III-B Subjects use perpendicul ars via the corner of the index card to measure angular separation as well as making linear measurements; or they measure AC (or BC) using this as a radius with which to swing an arc. The point, C, should lie on the arc where segment BC intersects the arc.

III-A Subjects form angular separation by estimation using "trial and error." Note this behavior by the systematic attempt to make the sides meet by altering the angular separation between two given sides. Also count any special attention given to angular separation. The key to evaluation of a III-A response is the attention the subject gives to angular separation in addition to linear measurement.

II-B Subjects use only one-dimensional measurement; the lengths of the sides are reproduced accurately. The most noticeable indication of level II-B reasoning is subjects' lack of attention to angular separation. Often the subjects are unable to locate the measured segments so that they meet to form the triangle, or the subject will draw two sides to the
exact length and settle for an approximation length of the third side so that a triangle is formed.

II-A No attempt is exhibited to accurately copy the triangle. Witness this both from the figure drawn or the lack of measurement marks on the index card.

Task C - The Flexible Rods Situation

Subjects are asked to simulate an investigation of the flexibility of a set of rods. The rods vary by length (3 different lengths), cross-sectional diameter (thick or thin), shape (round or square) and by material type (steel or brass).

Subjects are asked two types of questions, both of which test the subjects' abilities to form relations upon relations. The first type of question asks the subject to select a rod from among 4 choices that will bend about the same amount as a specified rod when a weight is hung from the rods. To be successful, subjects must select a rod that at the same time has one characteristic factor making it tend to be less flexible than the given rod and one factor making it tend to be more flexible than the given rod. For example, two rods of the same material and shape may bend about the same if while one is long and thick, the second is short and thin (long and thick bends the same as short and thin). Thus a sort of multiplication is used as the schema for this part of the task.

The second type of question requires subjects to select pairs of rods that differ in only one factor characteristic. Two such rods may be used to test the effect on flexibility of the one factor which differs in value between the two rods. For example, two rods with all factors identical except the shape factor would allow one to test whether square rods or round rods were more flexible.
The answers to the questions are:

1. Rod 4 (thick brass bends like thin steel).
2. Rod 1 (short and thin bends like long and thick).
3. Rod 4 (thick brass bends like thin steel).
4. Rod 4 (round and thick bends like square and thin).
5. Yes! A reason must accompany the answer which expresses some understanding of the multiplication of "long and thick bends the same as short and thin."
6. Same answer as given in number 5 above.
7. Pairs of rods should be selected with one steel and the other brass but identical in all other respects.
8. Pairs of rods should be selected with one round and the other square but identical in all other respects.

Reasoning Level Assignments:

III-B Subjects answer all questions correctly. At most a subject responding at this level may choose a rod not listed above for questions 1 through 4 but they will use a correct reason, i.e., one involving the multiplication schema. In such a case the subject may have overlooked some factor in which the rods in the comparison differ.

III-A Subjects use the multiplication schema in giving reasons for most of the answers to questions 1 through 4. The rods selected for the four questions would be plausible if not correct and answers questions 5 through 8 with reasonable completeness.

II-B Subjects answer one of the questions 1 through 4 with correct reasoning; but that is all they have complete success on while having partial success on questions 7 and 8 or 5 and 6.

Or
Subjects answer 5 and 6 correctly and have partial success on questions 7 and 8 or questions 1 through 4.

Or
Subjects answer 7 and 8 correctly and have partial success on earlier questions. Correct answers on 7 and 8 will be taken as subjects ability to consistently pick appropriate pairs. The judgment of this should allow for a subject who carries a false assumption into the problem. A good example of this is the subject who assumes that steel and brass rods are equally flexible. Such a subject would choose pairs of rods that are appropriate for testing roundness verses squareness in question 8 except one rod could possibly be steel and the other brass. Such an assumption should be easily recognized from the consistency with which the subject chooses pairs of rods appropriate under the subject's false assumption.

II-A Subjects have little success with questions 1 through 6 and are unable to choose pairs of rods that differ by only the factor to be tested while answering questions 7 and 8.

Task D - The Proportional Reasoning Situation

Subjects are asked to solve four arithmetic problems which require a proportion between two ratios to be formed and solved to attain success on all items.

There are four levels of problems and as such one or more of the problems may be answered successfully with a non-proportional and "incomplete" strategy.

Questions 1 and 3 may be answered successfully with a scaling strategy. A scaling strategy consists of recognizing that the numbers given for one of the ratios are such that one is a rational multiple (usually integral) of the other. For example, in question 1, 32 is 4 times larger than 8 and thus a subject could reason correctly that the
solution is 4 times larger than 16 (see problem 1).

In question 3 the scaling factor is less obvious than in question 1. Unless the subject has internalized the scaling strategy it is unlikely that the subject will be successful on question 3.

Questions 2 and 4 require subjects using a pure scaling strategy to work with a non-integral factor. Question 2 has 1/4 as its factor and question 4 has 2/7 as its factor. Successful application of the scaling strategy on these items would be equivalent to setting up a proportional equation and solving the equation, i.e., it would be a complete strategy.

Reasoning Level Assignments:

**III-B** Subjects use a definite proportional reasoning strategy on questions 2 or 4 and answer the other questions successfully.

**III-A** Subjects use a scaling strategy on questions 1 through 3 or a proportional strategy, but are unable to solve either questions 2 or 4.

**II-B** Subjects use a scaling strategy on questions 1 and 3, but are unable to use an appropriate strategy on questions 2 and 4.

**II-A** Subjects are able at most to solve only question 1.

If an subject has gotten #1 and #4 correct (using scaling/multiplication on #1 and scaling or even proportions on #4) count him as a II-B or III-A if he also used some proportions on #2 and #3's incorrect answers. If someone doesn't fit a given category—use strict count, e.g., 1 correct = II-A, 2 correct = II-B, 3 correct = III-A and 4 correct = III-B.


Keating, Daniel P. and Schaefer, Rosalind A. "Ability and Sex Differences in the Acquisition of Formal Operations." Developmental Psychology, 11 (1975), 531-532,


_______. "Sex Differences in Concrete and Formal Reasoning Ability as Measured by Manipulative Tasks and Written Tasks." Science Education, 59 (1975), 397-405.


Tomlinson-Keasey, C. "Formal Operations in Females from Eleven to Fifty-four Years of Age." Developmental Psychology, 6 (1972), 364.


