WHITTEN, Maurice Mason, 1923-
CHANGES IN UNDERSTANDING OF THE NATURE OF
SCIENCE AND IN CRITICAL THINKING RESULTING
FROM TWO DIFFERENT PHYSICAL SCIENCE COURSES.

The Ohio State University, Ph.D., 1971
Education, scientific

University Microfilms, A XEROX Company, Ann Arbor, Michigan
Please Note:

Some pages have very light type. Filmed as received.

University Microfilms.
ACKNOWLEDGMENTS

The writer wishes to express his appreciation to the individuals who have given of their abilities and time to assist in this investigation. Special thanks are due the following:

To the late Dr. John S. Richardson, Dr. Frederick R. Cyphert, and Dr. Jack G. Calvert for their guidance in planning the doctoral program through completion of the course work.

To Dr. Robert W. Howe for becoming my advisor and to the other members of my reading committee, Dr. Frederick R. Schlessinger and Dr. William R. Riley, for their guidance.

To my colleagues in the Physical Science Department of Gorham State College, Professors James W. Pendleton, Parnell S. Hare, and George H. Ayers, for their constructive suggestions and willingness to participate in the experimental course.

To the students whose participation made this investigation possible.

To Mrs. Doris R. Mann for her assistance in tabula-
ting data and for her secretarial work.

To Mrs. Bette Fitzgerald, Mr. Alfred A. Mann, and Mr. Alan A. Mann for their specific contributions.

To my mother, Mrs. Caroline A. Whitten, for her original encouragement in pursuing a doctoral program.
October 1, 1923 . . Born: Providence, Rhode Island
1945. . . . . A.B., Colby College, Waterville, Maine
1948-1955 . . . . Chemistry and Physics Teacher, Lewiston High School, Lewiston, Maine
1952-1953 . . . . Lecturer in Chemistry, Central Maine General Hospital, School of Nursing, Lewiston, Maine
1955-1959 . . . . Instructor of Physical Science, Gorham State Teachers College, Gorham, Maine
1960-1961 . . . . National Science Foundation Fellowship, Academic Year Institute, The Ohio State University, Columbus, Ohio
1961-1963 . . . . Assistant Professor of Physical Science and Chemistry, Gorham State Teachers College, Gorham, Maine
1963-1964 . . . . National Science Foundation, Science Faculty Fellowship, The Ohio State University, Columbus, Ohio
1964- . . . . . . Associate Professor of Physical Science and Chemistry, Gorham State College (since 1968: of the University of Maine), Gorham, Maine
FIELDS OF STUDY

Major Field:  Science Education

Studies in Science Education. Professors the late John S. Richardson and Frederick R. Schlessinger.

Minor Field:  Secondary Education


Minor Field:  Chemistry


Studies in Recent Advances in Chemistry. Professor Alfred B. Garrett.


Studies in Chemical Literature. Professor Earle R. Caley.
## CONTENTS

<table>
<thead>
<tr>
<th>Acknowledgments</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vita</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
</tbody>
</table>

### Chapter

#### I. Introduction

<table>
<thead>
<tr>
<th>The Problem</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement of the Problem</td>
<td>6</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>9</td>
</tr>
<tr>
<td>Basic Assumptions</td>
<td>12</td>
</tr>
<tr>
<td>Delimitation of the Study</td>
<td>13</td>
</tr>
<tr>
<td>Limitations of the Study</td>
<td>13</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Procedures</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Population</td>
<td>15</td>
</tr>
<tr>
<td>The Course</td>
<td>15</td>
</tr>
<tr>
<td>Sources of Data</td>
<td>17</td>
</tr>
<tr>
<td>Analysis of the Data</td>
<td>17</td>
</tr>
<tr>
<td>Compendium of Succeeding Chapters</td>
<td>18</td>
</tr>
</tbody>
</table>

#### II. Review of Selected Related Literature

<table>
<thead>
<tr>
<th>Critical Thinking</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitions and Significant Characteristics</td>
<td>20</td>
</tr>
<tr>
<td>Critical Thinking as an Educational Objective</td>
<td>27</td>
</tr>
<tr>
<td>Review of Pertinent Research</td>
<td>32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Understanding of the Nature of Science</th>
<th>49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitions and Significant Characteristics</td>
<td>49</td>
</tr>
<tr>
<td>Understanding of the Nature of Science as an Educational Objective</td>
<td>59</td>
</tr>
<tr>
<td>Review of Pertinent Research</td>
<td>71</td>
</tr>
</tbody>
</table>
### III. DESIGN OF THE STUDY

<table>
<thead>
<tr>
<th>Purpose</th>
<th>89</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Population</td>
<td>90</td>
</tr>
<tr>
<td>The Evaluation Instruments</td>
<td>91</td>
</tr>
<tr>
<td>Scholastic Aptitude Test of the College Entrance Examination Board</td>
<td>92</td>
</tr>
<tr>
<td>Test on Understanding Science, Form W</td>
<td>93</td>
</tr>
<tr>
<td>Cornell Critical Thinking Test, Level Z</td>
<td>96</td>
</tr>
<tr>
<td>Wisconsin Inventory of Science Processes</td>
<td>99</td>
</tr>
<tr>
<td>The Experimental Design</td>
<td>102</td>
</tr>
<tr>
<td>Collection of the Data</td>
<td>104</td>
</tr>
<tr>
<td>Procedure for Analysis of the Data</td>
<td>105</td>
</tr>
</tbody>
</table>

### IV. RESULTS AND INTERPRETATION

| Non-Criterion Variables of the Groups | 110 |
| Criterion Variables of the Groups | 113 |
| Testing the Null Hypotheses | 122 |
| Analysis of Gains and Losses on the TOUS for the Experimental Group | 147 |
| Summary | 158 |

### V. CONCLUSIONS AND RECOMMENDATIONS

| Conclusions | 161 |
| Critical Thinking | 161 |
| Understanding of the Nature of Science | 166 |
| Recommendations | 171 |

### APPENDIX A

|  | 173 |

### BIBLIOGRAPHY

|  | 195 |
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Comparison of Two Freshman Groups at Oregon State University, Sub-test and Total-test Means, TOUS</td>
<td>72</td>
</tr>
<tr>
<td>2.</td>
<td>Group Mean Scores on TOUS for Both Experimental Periods at Kansas State Teachers College</td>
<td>73</td>
</tr>
<tr>
<td>3.</td>
<td>Enrollment and Losses in Physical Science</td>
<td>90</td>
</tr>
<tr>
<td>4.</td>
<td>Enrollment Distribution by Sex and Class</td>
<td>91</td>
</tr>
<tr>
<td>5.</td>
<td>TOUS, Form W, Distribution of Questions</td>
<td>95</td>
</tr>
<tr>
<td>6.</td>
<td>TOUS, Form W, Reliability Using K-R Formula 20</td>
<td>96</td>
</tr>
<tr>
<td>7.</td>
<td>TOUS, Form W, Tentative Norms for High School Students, 1960</td>
<td>96</td>
</tr>
<tr>
<td>8.</td>
<td>CCTT, Level Z, Distribution of Questions</td>
<td>97</td>
</tr>
<tr>
<td>9.</td>
<td>CCTT, Level Z, Results of 48 High School Students</td>
<td>98</td>
</tr>
<tr>
<td>10.</td>
<td>CCTT, Level Z, Results of 62 Cornell Undergraduates</td>
<td>99</td>
</tr>
<tr>
<td>11.</td>
<td>Summary of Scores on Scholastic Aptitude Tests and Achievement Tests</td>
<td>111</td>
</tr>
<tr>
<td>12.</td>
<td>Distribution of Scores on Scholastic Aptitude Tests and Achievement Tests</td>
<td>112</td>
</tr>
<tr>
<td>13.</td>
<td>Distribution of Scores on Scholastic Aptitude Tests and Achievement Tests</td>
<td>112</td>
</tr>
<tr>
<td>Table</td>
<td>Test on Understanding Science, Effect of Deleting Questions 18, 20, and 47 on Sub-test and Total-test Means: for the Control Group</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Effect of Deleting Questions 18, 20, and 47 on Sub-test and Total-test Means: for the Experimental Group.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Summary of TOUS Posttest Mean Scores for Gorham State College and Kansas State Teachers College.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficients for Posttest TOUS Total Scores With CEEB Scores and Other Posttest Scores.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cornell Critical Thinking Test, Level Z, Means for the Control Group and the Experimental Group.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficients for Posttest CCTT Scores With CEEB Scores and Other Posttest Scores.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wisconsin Inventory of Science Processes, Means for the Control Group and the Experimental Group.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation Coefficients for Posttest WISP Scores With CEEB Scores and Other Posttest Scores.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analysis of Variance for TOUS-Part I Control and Experimental Groups</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analysis of Variance for TOUS-Part II Control and Experimental Groups</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analysis of Variance for TOUS-Part III Control and Experimental Groups</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analysis of Variance for TOUS-Total Control and Experimental Groups</td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>26. Analysis of Variance for CCTT Control and Experimental Groups</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>27. Analysis of Variance for WISP Experimental Group</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>28. Analysis of Covariance for TOUS-Part I</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>29. Analysis of Covariance for TOUS-Part III</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>30. Analysis of Covariance for TOUS-Total</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>31. Analysis of Covariance for CCTT</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>32. Analysis of Variance for Difference Among Instructors in Producing Changes</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>33. Analysis of TOUS Questions Registering Greatest Gains by the Experimental Group</td>
<td>148</td>
<td></td>
</tr>
<tr>
<td>34. Analysis of TOUS Questions Registering Greatest Losses by the Experimental Group</td>
<td>151</td>
<td></td>
</tr>
<tr>
<td>35. Analysis of WISP Questions Registering Greatest Gains by the Experimental Group</td>
<td>153</td>
<td></td>
</tr>
<tr>
<td>36. Summary of Scores on Scholastic Aptitude Tests</td>
<td>162</td>
<td></td>
</tr>
<tr>
<td>37. Percentage Distribution of High and Low Scores</td>
<td>162</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

For many years there has been general agreement by scientists and science educators that science teaching at all levels should have broader objectives than memorization of scientific facts and principles. Objectives mentioned with increasing frequency include: (1) improvement of critical thinking skills or abilities and (2) understanding of the nature of science and its processes.

Noll (80, p. 2), in 1933, wrote:

It is often said that this is an age of science. Perhaps it is for a few scientists; but the habit of scientific thought has scarcely touched the vast majority of us. . . . We teach science in our schools so that pupils may pass examinations, but it is highly improbable that many of them form habits of thinking that will be of value later.

Almost a decade later, Tyler (103, p. 1) stated:

The first step in this rebuilding of science education is to recognize scientific method or critical thinking as a major objective of science instruction. It is true that lip-service has been given to the importance of scientific method as an objective of education, but rarely have courses of study actually been developed by considering the kinds of educational experiences and the teaching methods most likely to be effective in teaching scientific method.
Referring to college students, Dressel and Mayhew (32, p. 153), in 1954, stated:

As a starting point in consideration of critical thinking, it perhaps can be accepted that students, as a result of their education experience, should be able to carry on types of mental activity more complicated than simple recall and re-statement of ideas, facts, principles, etc., given in the textbook or presented by the instructor in his lectures. . . .

A major aim of general education is for the student to acquire and use the skills and habits involved in critical and constructive thinking.

The Educational Policies Commission (35, p. 4) in its 1961 pronouncement on the purposes of American education, stated:

To be free, a man must be capable of basing his choices and actions on understandings which he himself achieves and on values which he examines for himself. He must be aware of the bases on which he accepts propositions as true. He must understand the values by which he lives, the assumptions on which they rest, and the consequences to which they lead. . . . must be capable of analyzing the situation in which he finds himself and of developing solutions to the problems before him. He must be able to perceive and understand the events of his life and time and the forces that influence and shape those events.

At the college level, general education physical science courses came into existence in the 1930's and gained widespread acceptance in the 1940's. There was a growing belief that survey courses which cut across traditional science course boundaries were more relevant in promoting understanding of science and a more comprehen-
sive view of the universe than one or two courses in specialities such as chemistry or physics. The demand for the integrated course increased rapidly; in some institutions it was an elective, and in others it was a general education "core" requirement. Some of the early textbooks written for these courses included: Man and the Nature of His Physical Universe by Jean et al. (1934); College Physical Science by McCorkle and Lewis (1934); Man's Physical Universe by Bawden (1937); and The Universe Surveyed by Richards et al. (1937).

The President's Commission on Higher Education (89, p. 49) reported in 1947:

"General education" is the term that has come to be accepted for those phases of nonspecialized and nonvocational learning which should be the common experience of all educated men and women. . . . It should enable him to identify, interpret, select and build into his own life those components of his cultural heritage that contribute richly to understanding and appreciation of the world in which he lives.

The Commission (89, pp. 50-58) then enunciated eleven objectives of general education which should be understood in terms of performance and behavior rather than in terms of mastering certain bodies of knowledge. Four of the objectives, selected because of their pertinence to science courses, include:

2. To participate actively as an informed and responsible citizen in solving the social, eco-
nomic, and political problems of one's community, State and Nation.

4. To understand the common phenomena in one's physical environment, to apply habits of scientific thought to both personal and civic problems, and to appreciate the implications of scientific discoveries for human welfare.

5. To understand the ideas of others and to express one's own effectively.

11. To acquire and use the skills and habits involved in critical and constructive thinking.

In the 1960's there was increasing concern for the scientific literacy of those students who will not become scientists, engineers, or technicians. It is especially important that college students preparing to be elementary school teachers be exposed to courses which provide opportunities to increase their understanding of the nature of science and to improve their critical thinking abilities which will enable them to make wise decisions as citizens and teachers.

Description of the Physical Science Course at Gorham State College

Physical science became a required course for all sophomores at Gorham State (Teachers) College at the close of World War II. It was a non-laboratory course because of a lack of laboratory facilities until 1958-59. Since then, the course has had a two-hour per week laboratory period associated with it. In 1964-65, physical
science (Science 100-101) became a freshman course required of all majors in elementary education. After 1964-65, the physical science course was essentially the control course of 1968-69.

The control course taught in 1968-69 had as its primary objective, and in practice almost its exclusive objective, that students should acquire the basic concepts, principles, and facts of physical science to improve their understanding of the universe. There were no specific opportunities provided to promote critical thinking or to develop understandings of the nature of science and its processes.

The experimental course taught in 1969-70 had, in addition to the subject matter objective, two other objectives: (1) improvement of the students' understandings of the nature of science and its processes, and (2) improvement of the students' critical thinking skills. The course was carefully planned to provide opportunities, both in the classes and in the laboratories, to develop these objectives. Further description of the courses is found in the latter part of this chapter under Procedures and also in Appendix A.

There has been a lack of adequate research attempting to measure the attainment of the objectives of general education physical science courses at the college lev-
el and none at Gorham State College.

THE PROBLEM

Statement of the Problem

Do changes in college students' understanding of the nature of science and in critical thinking result from modifying a "traditional," general education, laboratory, physical science course?

This problem was divided into three sub-problems; each has six null hypotheses to be tested.

I. Do students, after taking a year of physical science, make significant gains as shown by their pretest and posttest scores on the criterion instruments?

The following null hypotheses apply to each group separately, the control group (1968-69) and the experimental group (1969-70).

A. There is no significant gain pre- to posttest in understanding the scientific enterprise as measured by Part I of the TOUS.

B. There is no significant gain pre- to posttest in understanding of scientists as measured by Part II of the TOUS.

C. There is no significant gain pre- to posttest in understanding the methods and aims of
D. There is no significant gain pre- to posttest in understanding of the nature of science as measured by the total score of the TOUS.

E. There is no significant gain pre- to posttest in critical thinking ability as measured by the total score of the CCTT.

F. There is no significant gain pre- to posttest in knowledge of the processes of science as measured by the total score of the WISP.

II. When adjusted for differences on pretest scores, scholastic aptitude, and achievement in English if significant, were there significant differences in learning outcomes as measured by the criterion instruments between the control physical science group and the experimental physical science group?

A. There is no significant difference in outcomes in understanding the scientific enterprise as measured by Part I of the TOUS between the experimental group and the control group.

B. There is no significant difference in outcomes in understanding of scientists as measured by Part II of the TOUS between the ex-
periental group and the control group.

C. There is no significant difference in outcomes in understanding the methods and aims of science as measured by Part III of the TOUS between the experimental group and the control group.

D. There is no significant difference in outcomes in understanding the nature of science as measured by the total score of the TOUS between the experimental group and the control group.

E. There is no significant difference in outcomes in critical thinking ability as measured by the total score of the CCTT between the experimental group and the control group.

F. There is no significant difference in outcomes in knowledge of the processes of science as measured by the total score of the WISP between the experimental group and the control group.

III. Do the teachers influence significantly the changes in learning outcomes as measured by the criterion instruments?

A. There is no significant difference between teachers in producing changes in understand-
ing the scientific enterprise as measured by Part I of the TOUS.

B. There is no significant difference between teachers in producing changes in understanding of scientists as measured by Part II of the TOUS.

C. There is no significant difference between teachers in producing changes in understanding the methods and aims of science as measured by Part III of the TOUS.

D. There is no significant difference between teachers in producing changes in understanding the nature of science as measured by the total score of the TOUS.

E. There is no significant difference between teachers in producing changes in critical thinking ability as measured by the total score of the CCTT.

F. There is no significant difference between teachers in producing changes in knowledge of the processes of science as measured by the total score of the WISP.

Definition of Terms

For the purpose of this study, the following terms
are defined:

I. Critical Thinking

Ennis (42, p. 599), whose test is used in this investigation, stated: "As a root notion, critical thinking is here taken to mean the correct assessing of statements."

Critical thinking, as used in this study, refers specifically to Ennis's list of seven skills which a student demonstrates in:

1. Drawing conclusions
2. Identifying faulty thinking
3. Identifying reliable statements
4. Deciding on the certainty of conclusions
5. Selecting the best prediction
6. Selecting the best definition
7. Identifying an unstated assumption

II. The Nature of Science

Welch (111), in his study, described the process used to identify elements of the scientific process. His work was the basis for the instrument, the Wisconsin Inventory of Science Processes, used by the present investigator. Cooley and Klopfer (23) developed the Test on Understanding Science, also used by the present investigator. From an examination of these two instruments one may arrive at a descriptive definition of the "nature of science."

The nature of science, as the term is used in this
study, means a human, intellectual, dynamic, knowledge-developing activity which involves assumptions, observing, hypothesizing, experimenting, measuring, classifying or organizing, evaluating, and reporting or communicating. Scientific activity makes extensive use of ideal concepts or models, hypotheses, theories which are regarded as tentative, laws, and probability type explanations rather than absolute statements. The activity, international in character, affects society and is affected by society.

III. Scholastic Aptitude

Scholastic aptitude is the capacity to engage successfully in academic work. For the purpose of this study, scholastic aptitude is defined as the abilities measured by the scores on the verbal and mathematical sub-tests of the Scholastic Aptitude Test (SAT) of the College Entrance Examination Board.

IV. Criterion Instruments

Criterion instruments are defined as those evaluation tests which are used in a study to test the hypotheses under investigation. The criterion instruments used in this investigation were the Test on Understanding Science, Form W; the Cornell Critical Thinking Test, Level Z; and the Wisconsin Inventory of Science Processes.
Basic Assumptions

In this study, it is assumed that:

1. The physical science course will produce changes in the student's understanding of the nature of science and its processes.

2. The physical science course will produce changes in the student's critical thinking skills.

3. The changes in the students will result primarily from taking the physical science course; not from other courses taken concurrently.

4. The College Entrance Examination Board's Scholastic Aptitude Test scores will provide a valid and reliable measure of the student's general scholastic ability.

5. The Test on Understanding Science, the Wisconsin Inventory of Science Processes, and the Cornell Critical Thinking Test will provide valid and reliable measures of the qualities claimed by their respective constructors.

6. Maturation and history between pretesting in September or October and posttesting in May will not significantly affect the results.
Delimitation of the Study

1. This study will be limited to students enrolled in the physical science course (Science 100-101) at Gorham State College of the University of Maine during 1968-70.

2. The control group will be limited to students who completed the full year course in 1968-69. The experimental group will be limited to students who completed the full year course in 1969-70.

3. Pretesting and posttesting will be limited to measuring understanding of the nature of science and its processes, and critical thinking skills.

Limitations of the Study

1. The Wisconsin Inventory of Science Processes was given only at the end of the year (May, 1969) to the control group. Therefore, no pretest data are available for this instrument for 1968-69.

2. The students enrolled in physical science (Science 100-101) are prospective elementary (K-6 inclusive) education majors. This group is not necessarily representative of all of the students at Gorham State College nor of college students universally.
Significance of the Study

It is often claimed that general education courses in science will help to produce citizens who are scientifically literate. Pella, O'Hearn, and Gale (84, p. 44) found that understanding of the nature of science was among the three most frequently used referents by authors writing about scientific literacy. It is generally agreed that critical thinking is a requisite for intelligent citizenship, and that science courses for prospective teachers should include improving understanding of the nature of science and its processes and improving critical thinking skills as two of their important objectives.

Tyler (105, pp. 44-45), in discussing the importance of research in science, has indicated that one of the areas in need of good research is that of the objectives of science instruction.

Despite the large number of colleges and universities offering physical science and the large number of students enrolled, there has been relatively little research conducted to ascertain the effectiveness of the courses in contributing to the objectives. Some investigations have given negative results, others have been inconclusive, and some have given positive results. These investigations will be considered in Chapter II.

The results of the present study should provide a
basis for an improved general education physical science program. It should also provide important information for institutions that are concerned with the preparation of teachers, especially those for the elementary schools. The results should serve to stimulate further research.

PROCEDURES

The Population

The control group consisted of all students (144) enrolled who completed physical science (Science 100-101) for 1968-69 at Gorham State College of the University of Maine.

The experimental group consisted of all students (190) who completed the course in 1969-70.

The Course

The physical science course, Science 100-101, at Gorham State College is a general education survey course for non-science majors. Course content is selected from astronomy, chemistry, geology, and physics. The course is a two-semester course which meets for two one-hour class (lecture) periods per week and one two-hour laboratory period per week. Each laboratory section is limited to 24 students.

It was assumed that the academic rigor of the course
for the control group and the experimental group would be comparable. Descriptions of the control and experimental courses may be found in Appendix A, pages 173-194.

Major differences, in addition to the objectives mentioned earlier in this chapter, between the control course and the experimental course include:

1. The historical approach was used much more extensively in the experimental course than in the control course. One of the Harvard Case Histories, Robert Boyle's *Experiments in Pneumatics*, was used.

2. The textbook used with the control group was *Fundamentals of Physical Science* (5th edition) by Krauskopf and Beiser, 1966. The experimental group used *College Physical Science* (2nd edition) by Miles, Sherwood, and Parsons, 1969. The latter book was more in harmony with the objectives of the course; made more use of the historical approach, and included material dealing with procedures and methods in science.

3. The control group used a laboratory manual and most of the exercises were of the "cook-book" type. The experimental group used laboratory exercises prepared by the physical science staff to promote the objectives of the course.
Sources of Data

Scores on the College Entrance Examination Board's Scholastic Aptitude Tests and Achievement Tests were obtained from the registrar's office at Gorham State College.

Pretest and posttest scores were obtained on the following criterion instruments which were administered by the investigator and two other members of the physical science staff:

1. Test on Understanding Science, Form W, by Cooley and Klopfer.
2. The Wisconsin Inventory of Science Processes, Test 671, in the Scientific Literacy Series.
3. The Cornell Critical Thinking Test, Level Z, by Ennis and Millman.

Analysis of the Data

Data were analyzed with the use of the facilities of the Numerical Computation Center at The Ohio State University.

A BMD02D Correlation with Transgeneration Analysis was used to determine the coefficients of correlation among all of the variable factors involved.

An Analysis of Variance for Factorial Design with Unequal Cell Frequency, Least Square Solution was used to determine variance attributable to the course and to the
different instructors.

A BMD04V Covariance Analysis was used to control the variables of scholastic aptitude and prior understanding or ability as determined by the particular instance.

These tables of data were analyzed to obtain the information to make the decisions whether to accept or reject the null hypotheses.

Compendium of Succeeding Chapters

The chapter organization of the remainder of this study is as follows:

Chapter II is a review of selected related literature;
Chapter III deals with the design of the study;
Chapter IV gives the results and interpretation; and
Chapter V lists the conclusions and recommendations.
CHAPTER II

REVIEW OF SELECTED RELATED LITERATURE

A review of the literature on education, especially science education, revealed that considerable research had been conducted in the last fifty years on one or more phases of the problem studied in this investigation. There have been, however, comparatively few studies in physical science at the college level dealing with critical thinking and understanding of the nature of science.

This chapter will be divided into two major sections:

1. Critical thinking and its development.
2. Understanding of the nature of science.

Each section will follow a similar pattern of development, reviewing:

1. Definitions and significant characteristics.
2. Objectives of general education and science education.
3. Pertinent research studies.
Definitions and Significant Characteristics

There is no universally accepted definition of critical thinking. Investigators and authors differ in their interpretation and meaning of the term. The nature of critical thinking will be examined in greater detail than given in the definitions of Chapter I.

Warren, in 1934, defined thinking (109, p. 277) as "a determined course of ideas, symbolic in character, initiated by a problem or task, and leading to a conclusion."

Warren's definition has been accepted by Russell and others as generally satisfactory. Russell (94, p. 283) noted that a definition of critical thinking "must be sharpened beyond its relationship to problem solving and creative thinking. . . . knowledge of a particular field is prerequisite to critical thinking." One cannot be critical about something of which he knows nothing.

Russell (94, p. 82) proposed that critical thinking ordinarily involves four conditions:

1. A knowledge of the field or fields in which the thinking is being done

2. A general attitude of questioning and suspended judgment; a habit of examining before accepting

3. Some application of methods of logical analysis or scientific inquiry
4. Taking action in light of this analysis or reasoning.

A wide range of abilities and attitudes has been included under critical thinking by different writers. Russell (94, pp. 283-284) has summarized:

A number of writers have suggested that it involves not only inspection, comparison, and appraisal but some measure of belief and other affective factors. Anderson and others (1944) have included in the term such items as selecting and organizing relevant facts, making inferences, distinguishing fact from opinion, and recognizing insufficient evidence. In one of the best studies concerning the possibility of teaching critical thinking in the high school, Glaser (1941) has included in the term '1) an attitude of being disposed to consider in a thoughtful way the problems and subjects that come within the range of one's experience, (2) knowledge of the methods of logical inquiry and reasoning, (3) some skill in applying these methods.' DeBoer (1946) has emphasized that the establishment of criteria for judging a statement or product is one of the most important parts of critical thinking. In relation to printed materials he believed that criteria can be established at three levels: (1) the question of fact and of relevance, (2) the problem of accuracy of the item and of the reliability of its source, and (3) the ability to appraise the validity of the author's conclusions.

McBurney and Hance (73, pp. 204-209) have divided the causes of uncritical thinking into two major categories, personal and social. Personal has been subdivided to include unintentional and intentional causes. The first included inability to observe, to remember, to organize, or to test hypotheses. Intentional causes included the desire to imitate, to bolster self-esteem, to convert;
along with the operation of prejudice and the tendency to rationalize.

They listed under social causes: (1) "group pressures" which lead to censorship, the force of tradition and (2) Francis Bacon's "idols of the forum," meaning creeds, dogmas and parties.

Ennis, whose test is used in the present study, enumerated (41, p. 84) twelve aspects of critical thinking as follows:

1. Grasping the meaning of a statement.
2. Judging whether there is ambiguity in a line of reasoning.
3. Judging whether certain statements contradict each other.
4. Judging whether a conclusion follows necessarily.
5. Judging whether a statement is specific enough.
6. Judging whether a statement is actually the application of a certain principle.
7. Judging whether an observation statement is reliable.
8. Judging whether an inductive conclusion is warranted.
9. Judging whether the problem has been identified.
10. Judging whether something is an assumption.
11. Judging whether a definition is adequate.
12. Judging whether a statement made by an al-
leged authority is acceptable.

Ennis (41, pp. 84-86) further stated, "There are three basic analytically distinguishable dimensions of the proposed concept of critical thinking: a logical dimension, a criterial dimension, and a pragmatic dimension."

The logical dimension, in general, involves judging "alleged relationships between meanings of words and statements." Anyone competent in this dimension knows what follows from one or more statements because of their meaning. He knows the use of the logical operators, "all," "some," "none," "and," "if...then," "unless," etc. One also knows what it is for something to belong to a class of things.

The criterial dimension involves knowledge of criter-
ia for judging statements excluding the logical criteria which are included in the previous dimension.

The pragmatic dimension "covers the impression of the background purpose on the judgment, and it covers the decision as to whether the statement is good enough for the purpose." It recognizes that various factors must be balanced before making the judgment, "This is enough evi-
dence." The pragmatic dimension also requires an "admis-
sion that complete criteria can not be established for critical thinking. An element of intellectual judgment is usually required in addition to applying criteria and
knowing the meaning."

Ennis, Russell, and others have indicated that critical thinking involves a considerable amount of judgment which is not necessarily a purely intellectual process but may involve emotional factors. Weber (110, p. 266), after questioning various psychologists, concluded that judgment is "a cognitive reaction initiated by a query, and involving inspection, discrimination, comparison, appraisal, and a degree of belief."

Therefore, the problem in critical thinking may not be the elimination of emotional factors but giving them a suitable place in the total process. To make definitions and descriptions more meaningful, Postman and Egan (87, pp. 217-238) have summarized some research on judgment which is pertinent:

1. There is a central tendency in judgment--subjects tend to avoid extremes.

2. Judgments around a center are more difficult and more variable than those at the extremes.

3. Some judgments are made on the basis of the 'halo effect'; the subject uses irrelevant information.

4. The atmosphere effect may be important--the judgment is made on general impression rather than exact discrimination.

5. The speed of the judgment depends upon the difficulty of discrimination and the subject's confidence. Correct judgments usually take a shorter time than incorrect.
Black (9, p. 7) also emphasized the role of judgment and the making of comparisons. He used the adjective critical in connection with thinking to mean critical of thinking. He stated:

The logician searches for standards of reasoning. To be in a position to improve reasoning means to be in a position to distinguish good from bad. . . . a thinker who tries to improve his thinking must have in mind some standard for discriminating better thinking from worse.

Burke (14, p. 527) wrote an article in 1949, listing fifteen behaviors which he believed exemplified critical thinking. He used the verbs "criticizes" (7 times), "differentiates" (2 times), and "selects" (2 times) which necessitate the use of judgment, thus agreeing with the ideas of Ennis, Russell, Weber, and Black.

A comparison of Burke's list of behaviors with Ennis's list of aspects of critical thinking revealed many ideas common to both. Ennis placed more direct emphasis on "judging" than Burke, having used the term in eleven out of twelve statements. The latter described behaviors with more specificity and included the following, not listed by Ennis: recognizes evidence of personal bias, draws valid inferences from graphs, estimates probability of an inference, recognizes tentative nature of hypotheses, criticizes hypotheses, criticizes experimental procedures, recognizes errors of measurement, and what
assumptions must be maintained in generalizations from experimental results.

Burmester (15, pp. 259-263) developed a comprehensive outline based on an analysis of the types of behaviors involved in scientific thinking which, it was believed, could be measured by objective tests. Burmester emphasized organizing behaviors around major "abilities" with her unique contribution being the ability to delimit.

Burton, Kimball, and Wing (17, pp. 38-39), in a review of research, discussed critical attitudes which are necessary if one is to engage successfully in critical thinking. These attitudes included objectivity, intellectual curiosity, intellectual honesty, open-mindedness, and intellectual skepticism or suspension of judgment. Additional attitudes which they regarded as prerequisites for critical thinking included conviction of universal cause-and-effect relationships, a disposition to be systematic, flexibility in being able to give up a previous attractive conclusion, and decisiveness in reaching conclusions.

The definition of critical thinking used in the present investigation is Ennis's (42, p. 599) statement, "Critical thinking is here taken to mean the correct assessing of statements." This definition refers specifically to Ennis's list of seven behavioral skills referred
Summary. There is general agreement that if one is to carry on effective critical thinking, it is essential to have some intelligence; knowledge; abilities to compare, select, organize; and, finally, to come to some warranted conclusion. Russell, Ennis, Black, and others have emphasized the role of judgment in critical thinking. Ennis pointed out that critical thinking may be considered as having three dimensions—logical, criterial, and pragmatic.

A comparison of Burmester's and of Burke's lists of behaviors exemplifying scientific thinking with the list of Ennis showed considerable overlapping of ideas. The latter emphasized judging more than the other two. Burton, Kimball, and Wing were in general agreement with the others cited but devoted special attention to the necessity of critical attitudes.

Critical Thinking as an Educational Objective

The development of critical thinking has been recommended as an important objective of science education for approximately fifty years. Some writers have used such terms as scientific thinking, reflective thinking, or scientific attitude to describe basically the same type of intellectual activity or thought processes.
Curtis (28, p. 43), in 1924, developed a descriptive outline of scientific attitudes. He included the habit of delayed response, that is, holding views tentatively for proper reflection to permit adequate consideration of various options; and the habit of weighing evidence with respect to its pertinence, soundness, and adequacy. He also included respect for another's viewpoint, open-mindedness and willingness to change when presented with adequate evidence.

After Curtis's outline of scientific attitudes appeared in 1924, there was a gradual acceptance by science educators of objectives, such as: (1) development of scientific thinking, (2) improving problem-solving ability, and (3) developing scientific attitudes. Caldwell and Curtis (18), in 1929, listed 14 scientific attitudes in their general science textbook. Hunter (54, p. 205), five years later in 1934, devoted a chapter to "The Method of Science" in his book on science teaching. He stated:

Any work on objectives in science rates high the ability to use the method of science. . . . The development of a scientific attitude toward the problems of life and the establishment of habits of scientific thinking which will be used in daily life are outcomes which educators believe should come directly from the study of science. . . . Snap judgments rather than reflective thinking lead the average person to his daily conclusions. Dogma still holds sway rather than evidence. Superstitions, hunches, and the rabbit's foot all play their part in the lives of the average person. We are swayed by propaganda . . .
Tyler (103, p. 1), in 1942, emphasized the importance of critical thinking becoming a main objective of science instruction. In 1960, in "The Behavioral Scientist Looks at the Purposes of Science-Teaching," he (104, p. 32) wrote:

A major objective . . . in science-teaching is to help students develop the ability to carry on the whole process of scientific inquiry, including raising questions, identifying particular problems growing out of these questions, suggesting possible explanations, devising ways of testing these explanations, making relevant observations and collecting relevant data, interpreting the data, and restating the explanations and the new questions and problems that result from this cycle of inquiry.

Russell (94, p. 281) emphasized the importance of critical thinking in a democratic society. Children, youth, and adults in a democracy will always be confronted with controversial issues. They will often have to make a choice; therefore, they should become increasingly skilled in distinguishing, criticizing, and evaluating the various possibilities. "Critical thinking would seem to be one of the first requirements of a full-fledged citizen in a democracy." Similarly, Richardson (91, pp. 1, 8) thought one of the greatest needs of our society was for leadership inculcated with the habits of critical thinking. His first objective for teaching science was to develop the ability to think critically.

Cohen and Watson (20, p. 25), in 1952, in discussing
the role of science in general education, stated:

Through the use of selected science materials and techniques we have a chance to find out whether students can learn to think—provided we focus sharply on this task. If they can be helped materially, then science will have something unique to offer non-science students. For that reason I would put first in any General Education program in science the objective of learning to reason critically, imaginatively, and constructively about problems in science.

A decade later, Lehmann and Dressel (69, pp. 1-2) included in the objectives of a college education, "the development of skill in critical thinking and problem solving."

Keeslar (60, pp. 212-213), after describing the procedures of the scientific method, indicated that many people believe the most valuable product of scientific effort has been "the method of inquiry developed" not the quantity of knowledge accumulated. The "scientific method" of inquiry has been emphasized in various forms as an educational objective by several national committees including the committee for the 46th Yearbook of the National Society for the Study of Education. They (78, p. 62) noted there have been few points in educational discussions on which there has been greater agreement than the desirability of teaching the "scientific method." They believed that the elements of the scientific method should make an important contribution to developing an
attitude of critical mindedness. Concurrently, there should be: growth in willingness to consider new information, decreasing gullibility, willingness to change attitudes, distinctions made between facts and fiction, and willingness to seek reliable information as a basis for action.

A committee of the National Science Teachers Association, in 1961, formulated some long range objectives (99, p. 28) for science teaching, including the acquisition by students of the habitual and skillful use of sound thinking in coping with problem situations. To do this, students must have "understanding of, faith in, and direct practice with sound methods and attitudes of thought."

The Commission on Science Education of the American Association for the Advancement of Science formulated objectives for modern science teaching which were summarized by Kessen (61, pp. 4-6) in 1964. The pertinent points included: (1) "science is best taught as a procedure of enquiry," (2) the scientific attitude requires questioning the claims of authority while respecting the work of the past, and (3) the procedures (methods) of science should be developed by teachers.

Critical thinking, as an educational objective, is not limited to the sciences. It is often mentioned as an objective in the social studies, Aldrich's (1, pp.
description of the three year effort of New York University's School of Education; and in mathematics, accounts by Pingry (86, pp. 466-470) and Brown (12, p. 104).

Summary. From a review of the literature, it is concluded that increasing the critical thinking skills of students has been a concern and an objective of science educators for the last half century. This objective has been stated in various forms by authors of books on science teaching, individuals concerned with the objectives of general education, those concerned with the objectives of science instruction, individuals in other academic disciplines, and committees of national professional organizations (AAAS, NSSE, NSTA). Representatives cited, of the concerned groups, included Hunter, Richardson, Russell, Lehmann and Dressel, Caldwell, Curtis, Tyler, Cohen and Watson, Aldrich, Brown, and Pingry. Improving critical thinking skills has been regarded as a major objective of science instruction at both the college and the secondary school levels.

Review of Pertinent Research

The related research literature has been divided into three major categories: (1) those studies which have shown no significant gains in attempts to improve critical thinking, (2) those studies which have shown significant
gains in attempts to improve critical thinking, and (3) studies which show relationship of critical thinking to other factors.

Studies Showing No Significant Gains

Craven (25, pp. 130-131), in 1965, using the Cornell Critical Thinking Test, Form X, concluded that critical thinking ability "was not a major learning outcome of the study of college science." He also concluded that the critical thinking abilities of science teacher-candidates were significantly greater than those of freshmen in social science education, freshmen in science education, elementary teacher-candidates, and in-service teachers.

Novak (81, p. 229) compared a conventional and a project centered method of teaching college general botany at the University of Minnesota in 1956-57. It was concluded that the two groups did not differ significantly in mean score on the problem solving posttest. It was also concluded that neither group "made a significant mean gain in scientific attitude over the two month interval, as measured by the tests used."

Yudin (120, p. 935), in 1957, compared two methods of instruction in a program designed to improve the critical thinking ability of college freshmen. None of the groups used in the study showed significant gains on the Watson-Glaser Critical Thinking Appraisal.
Lyle (71, pp. 129-133) investigated the teaching of critical thinking in college general psychology in 1958. His emphasis was on the problem solving method as opposed to the teacher centered method. His conclusion, based on the results of *A Test of Critical Thinking, Form G*, by the American Council on Education, was that both methods produced about the same small gain. There was no significant difference between experimental and control groups. He stated, "It seems improbable that instruction in one three-hour course (one semester) could be expected to produce any great degree of improvement which would be reflected" in such a test.

In addition to the studies at the college level just cited, there have been a number of studies conducted at the secondary school level which did not show significant gains. Only three are cited here.

Brown (13, p. 1611-A) compared 153 chemistry students with 149 non-chemistry students in high school during 1966-67. Using the Cornell Critical Thinking Test, Form X, he concluded there was "no significant growth in critical thinking abilities between any of the groups of chemistry and non-chemistry students." However, the chemistry students did make greater gains in critical thinking abilities.

Downing (30, pp. 87-89) made a study of 2500 pupils
in grades eight to twelve. Some of them studied science and some did not. He found there was no evidence that high school students acquired skills in scientific thinking as a necessary by-product of the study of scientific subjects as they were taught at the time of his study in 1932.

Alpern (3, p. 223) studied the ability to test hypotheses of approximately 600 students in the science classes of a New York City high school in 1946. He concluded:

Considering the only factors selected for study, the net relationships, as indicated by the coefficients of partial correlation, between the ability to select procedures to test hypotheses and chronological age, intelligence, reading grade and previous terms of high-school science, respectively, were not significant.

Summary. Among the college level studies cited which reported no significant gains, only Craven used the Cornell Critical Thinking Test and that was Form X which differs considerably from Level Z. Some of the studies reporting no significant gains were evaluating conventional courses which did not make any special effort to promote critical thinking; other studies were one semester or less in duration. Novak reported the shortest time interval between pretesting and posttesting—two months. Lyle was skeptical of a one-semester course being able to make any
great improvement in critical thinking skills.

Among the secondary school science investigations cited which reported no significant gains, only Brown used the Cornell Critical Thinking Test, Form X. Downing believed his instrument to be in need of considerable revision. Both the validity and the reliability of the early instruments, especially those of limited local use, can be questioned.

Studies Showing Positive Gains

Tolman (101, p. 2432-A) used the Cornell Critical Thinking Test, Form Z, in a study of student performance in general biology in three community colleges (261 students) and in two four-year institutions (465 students). He found there was a significant difference between students in community colleges and four-year institutions in terms of critical thinking ability. It was concluded that the four-year institutions provided a superior educational experience in terms of developing critical thinking skills.

Rickert (92, pp. 4226-27) studied improvement in critical thinking ability. The American Council on Education Test of Critical Thinking, Form G was used to assess learning with college freshmen in three different physical science courses. He concluded (92, p. 4227) that
ability to think critically was improved significantly by an experimental physical science course that gave students "opportunities to analyze problems, examine assumptions, collect and organize data, and test hypotheses." It was found that "there was no significant correlation between the gains on the test of critical thinking and academic ability" although there was a "significant correlation between scores made on the School and College Ability Test" and the preliminary critical thinking test scores.

Yoesting (119, pp. 40-41) conducted a study in 1964 to determine the effectiveness of a college physical science course in developing critical thinking skills. He used two groups of first semester freshmen, matched with respect to age, sex, American College Testing Program scores, and high school mathematics-science background. There were 82 in each group; one group was enrolled in a general physical science course, the other group was not. The results, using the Watson-Glaser Critical Thinking Appraisal showed that the science group made significant improvement over the non-science group at the end of one semester.

Jones (58, pp. 112-114) made a study of a sample of 148 freshmen in twelve classes of physical science involving seven different instructors. He found that stu-
dents of the "second, third, and lower quartiles attained positive gains in critical thinking skills as measured by the WGCTA." The gain of the second quartile was not significant at the .1 level of confidence. The gain of the third quartile was significant at the .1 level of confidence while the gain of the fourth quartile was significant at the .05 level. The upper quartile students of four instructors made slight gains in critical thinking but none were significant at the .1 level. The upper quartile students of three instructors sustained a loss in their scores on WGCTA.

Montague (75, pp. 84-85) investigated the contributions of the college general chemistry laboratory to promote critical thinking. He designed experiments which emphasized problem solving for his experimental group. He found a significant increase in ability to think scientifically using Burmester's A Test of Aspects of Scientific Thinking for both the experimental and control groups based on pre- and posttest scores. There was no significant difference between the two groups on this test. The experimental group did do significantly better than the control group on the Watson-Glaser Critical Thinking Appraisal.

Olson's (82, p. 230) study compared a student-centered method with a teacher-centered method in teaching
college general education biological science to non-science majors. He used the Burmester Ability to Think Scientifically Test and concluded "the mean scientific thinking ability performance was increased significantly in both treatments." There was no significant difference in performance of scientific thinking between the two groups.

In addition to the studies at the college level which have been cited, there have been various studies conducted at the secondary school level which have shown positive gains.

One of the pioneering studies on scientific attitudes was reported by Curtis (28) in 1924. This involved an experimental group and control groups in grades seven, eight, and nine. The group receiving special instruction in scientific attitudes scored significantly higher on the test used, even after four to six months had elapsed. He further concluded:

1. There is some evidence of a positive correlation between brightness quotients and scientific attitudes.

2. Scientific attitudes do not seem to be developed by regular classroom instruction in general science to an extent comparable with that secured by definite instruction in scientific attitudes.

Teichman (100, p. 272) demonstrated that a teaching technique which "emphasizes the ability to make conclusions, and to state reasons why some conclusions were
faulty" resulted in significantly higher scores on tests designed to measure such abilities. The investigation was made with ninth grade general science students. He also found that the ability to state a conclusion from several offered, and the ability to give the best reason for a faulty conclusion, are not identical abilities.

Kastrinos (59, pp. 86-88) concluded that high school biology classes taught by a principle-critical thinking method made greater gains in critical thinking than did classes taught by the same instructors using a textbook recitation method. He developed a critical thinking test and found that it correlated favorably with the other critical thinking tests (Watson-Glaser Critical Thinking Appraisal and University of Illinois Test of Ability to Judge Interpretation of Data) used in his investigation.

Howe (53, pp. 202-203), in a study of tenth grade biology classes in Oregon public schools during 1962-63, found 44 classes out of 51 made positive gains in critical thinking skills. The instrument used was the Watson-Glaser Critical Thinking Appraisal. His analysis indicated that "most of the high gains were associated with classes utilizing problem-solving techniques with direct instruction and practice in critical thinking."
Summary. Among the college level studies cited which reported significant gains, only Tolman used the Cornell Critical Thinking Test, Level Z. He concluded that four-year institutions provided a superior educational experience in improving critical thinking skills compared with two-year colleges. Rickert and Yoesting both reported significant gains for physical science students while Jones noted the more significant gains being made by the lower quartiles. Rickert, in physical science, and Montague, in chemistry, had designed courses or experiments which emphasized problem solving and scientific procedures. Olson reported significant gains in biological science classes for non-science majors.

In general, one may conclude that, at the college level, those studies which reported the most decisive gains involved courses in which a special effort was made to improve critical thinking skills. Yoesting (students in physical science vs. students not taking any science) and Olson (students in biology with two different approaches) were exceptions to the general rule; both reported significant gains.

All of the secondary school science studies cited which reported positive gains in critical thinking, occurred when a teacher made a direct effort and employed procedures to promote improvement of such skills.
Studies Showing Relationship of Critical Thinking to Other Factors

A considerable amount of experimental work has been done in the development of tests of critical thinking. Some tests have been developed to evaluate certain aspects of critical thinking; others to measure the relationship of critical thinking ability to other factors. The instruments already mentioned in this chapter include the Watson-Glaser Critical Thinking Appraisal used by Yudin (120), Yoesting (119), Jones (58), Montague (75), and Kastrinos (59); Cornell Critical Thinking Test, Forms X and Z used by Craven (25), Brown (13), Tolman (101); ACE Test of Critical Thinking used by Lyle (71), and Rickert (92); Burmester's A Test of Aspects of Scientific Thinking used by Montague (75), Olson (82); and Kastrinos' Critical Thinking Test used by Kastrinos (59). A brief description of several additional studies involving various tests will be given.

Downing (31, pp. 121-128) developed an instrument to measure skill in using some elements of scientific thinking. He defined fifteen elements of scientific thinking which included defining problems, accurate observation, judging accuracy of data, ability to synthesize, and testing hypotheses. Results obtained with the instrument indicated no increase in scientific thinking ability re-
sulting from science instruction. He suggested the need of a supplementary test for measuring those aspects of scientific thinking that result from instruction and/or experience. His suggestion implied that instruction and experience are two factors related to scientific or critical thinking.

Furst (46, pp. 621-624) found that correlations between tests of critical thinking involving reading and analysis, judging relations, judging validity of evidence, interpretation of data, summarization of trends, and application of principles, with tests of scholastic aptitude as measured by the Quantitative and Linguistic scores of the A.C.E. Psychological Examination changed very slightly in a two-year period. He noted that, of a total of 30 initial coefficients of correlation, zero per cent were above .61; 33% were from .41 to .60; 47% ranged from .21 to .40; and 20% were .20 or below. He further observed that none of the changes in \( r \) (coefficient of correlation) from initial to final testing were statistically significant. He concluded that correlations between critical thinking scores and scholastic aptitude scores "were rather low to begin with and hence could not have decreased much more after the two year period and still have maintained some relationship, which they generally do."

Fogg and Calia (44, pp. 1-14) reported on the rela-
tionship of critical thinking ability to the format of science tests. They studied the effects of two testing techniques on achievement in science and critical thinking ability using 551 college freshmen at Boston University College of Basic Studies. The standard test format was to select "one best answer" from five alternatives; the experimental format was for the student to select from five choices those which he was sure were wrong. Both groups improved significantly beyond the .01 level of confidence in critical thinking ability as measured by the WGCTA. The investigators also concluded that "improvement in critical thinking ability tenuously associated with the negative form" of the tests suggests that "fact-permeated subject matter may be less amenable to test via the experimental format than ideational-oriented courses."

Burton, Kimball, and Wing (17, pp. 240-265) have summarized a considerable amount of research up to 1960 on various factors affecting critical thinking. Their survey of the research caused them to group the factors into four major categories, as follows: (1) intellectual, (2) personal-emotional, (3) experiential, and (4) procedural.

There is quite general agreement that intelligence definitely is an important factor in determining ability to think critically. Burton, Kimball, and Wing cited
studies by Glaser (50), Furst (46), Alpern (3), and Teichman (100) which showed positive correlations between tests measuring intelligence and tests measuring critical thinking skills. Burton, Kimball, and Wing concluded that ability to do inductive and deductive reasoning and dexterity with abstract symbols seem to be two intellectual traits which relate intelligence and critical thinking most closely. They also concluded that correlations reported from various studies indicate that less than half of the variability in problem-solving skills may be accounted for by scores on intelligence tests.

In addition to intelligence, certain academic skills and subject matter knowledge are necessary to solve problems competently in various subject matter areas. Burton, Kimball, and Wing (17, p. 243) emphasized that though "academic skills are necessary, they are not sufficient" to insure critical thinking. The studies they cited in support of this statement included, "Burack, Burack and Moos, Bloom and Broder, Duncker, Horrocks, Maier, and Maltzman and others."

Burton, Kimball, and Wing (17, pp. 245-251) discussed various personal-emotional factors including rigidity or "set," emotional stress, biases, and attitude toward the problem. Most of the theoretical analyses of scientific thinking have emphasized the necessity of flexibility in
thinking and facility in generating hypotheses. The antithesis of this characteristic is rigidity or set. The latter term is sometimes regarded as the tendency to continue using a particular mode of thinking when it is no longer fruitful. Among the studies cited which indicate that thinking ability is handicapped in relation to one's susceptibility to set are those by Luchins (70), Maier (72), Solomon (97), and Rokeach (93). Christie (19) found that frustrating conditions promote rigidity, and Beir (8) observed that a state of anxiety tended to cause "a loss of flexibility of intellectual function." The role of biases, as barriers to clear thinking, have been investigated by Duncker (33), Lefford (68), Morgan and Morton (76) and others. The pressure of majority opinion is very significant in contributing to bias. The effect of attitude toward the problem was investigated by Bloom and Broder (11). They found that good problem solvers were more aggressive, confident, tenacious, and attentive to detail, and placed more faith in reasoning out the solution than did poor problem solvers.

Burton, Kimball, and Wing (17, pp. 251-256), under experiential factors, summarized research relevant to functional fixedness, age trends in thinking abilities, the difficulty reduction tendency, and the role of mistakes. Functional fixedness seems to be another facet
of psychological rigidity. It tends to impede effective thinking by decreasing "the versatility of the thinker in his manipulation of both thoughts and materials." Diversity of experience and assistance in generalizing the significance of specific experiences appear to be helpful counter-measures to functional fixedness. Abstract and inductive reasoning abilities apparently develop most rapidly during adolescence. Mistakes can be used advantageously in improving critical thinking.

Under procedural factors, Burton, Kimball, and Wing (17, pp. 256-261) summarized research dealing with insight, trial-and-error, conceptual and perceptual approaches, influence of the problem setting, and methodological considerations. They concluded that "trial and error" and "insight" may be regarded as representing different sections of the spectrum of problem-solving behavior and that, in most problem situations, the thinker makes use of both types of behavior. Individuals whose mode of thinking can be classified as "abstract, conceptual, general, oriented toward the problem goal" appear to be more effective in most problem situations than individuals whose mode of thinking can be classified as "concrete, perceptual, specific, oriented toward the problem materials." Each problem situation presents itself in a particular context which is a definite factor in solving
the problem. Sometimes the context aids objectivity, generalization, and direction; sometimes it inhibits. To the time (1960) of their review, Burton, Kimball, and Wing found that attempts to isolate and evaluate specific problem-solving techniques had not been successful. They concluded that "hope for a neat set of 'rules for thinking' is presently a forlorn one."

**Summary.** Skill in critical thinking is dependent on and is influenced by many factors as revealed by the research literature. Intelligence is an important factor in determining ability to think critically but factors not measured by intelligence tests are important in problem solving.

Certain academic skills and subject matter knowledge are essential for critical thinking but do not guarantee that it will take place.

Personal-emotional factors affect ability to think critically. Minimizing rigidity or "set"; freedom from anxiety; recognition of bias; and an aggressive, confident, and tenacious attitude toward problems are characteristics or conditions that favor critical thinking.

Experiential factors also influence critical thinking. Diversity of experience plus assistance in generalizing specific experiences aid the ability to think cri-
tically. Abstract and inductive reasoning abilities appear to develop most rapidly during adolescence.

Procedural factors also have a bearing on critical thinking skills. The most effective thinkers use trial-and-error as well as insight, and have thought patterns characterized as abstract, conceptual, general, oriented toward the problem goal. The context of each problem is important. Finally, there is no neat, pat, universal set of rules that insures critical thinking.

UNDERSTANDING OF THE NATURE OF SCIENCE

Definitions and Significant Characteristics

There is no universally accepted, concise definition of the term, nature of science. Both the "understanding of" and "the nature of science" have changed greatly over the years. Two hundred years ago, the terms natural history and natural philosophy were used instead of science. Schwab (95, pp. 9-11) has indicated that until the latter part of the nineteenth century, science was concerned primarily with "seeking the facts of nature and reporting them." The discovery of radioactivity and the introduction of the theory of relativity brought about a revolution in thinking about the nature of science. The gradual recognition of the role of conceptual schemes became pre-eminent.
Schwab (95, p. 12) stated:

A fresh line of scientific research has its origin, not in objective facts alone, but in a conception, a construction of the mind. And on this conception, all else depends. It tells us what facts to look for in the research. It tells us what meaning to assign these facts.

He later pointed out that the conceptions of science are the guiding principles of enquiry and "not its immediate fruits." The scientific knowledge of a particular time rests on "selected facts"; the selectivity is dependent on the conceptual principle of the enquiry. One of the distinguishing features of the nature of science now is the "revisionary character of scientific knowledge."

Scientific conceptions are products of the human mind and all science is a human enterprise. Wolf (117, p. 17) stated, in 1925:

Nature, with all her regularities and irregularities, might have been just as real even if there were no men to observe and to study her. But there could have been no science without human beings, or beings like them. It is the spirit of man brooding over the stream of natural events that has given birth to science.

Selections from the writings of science educators and others are cited to give additional meaning to the significant characteristics of "understanding of the nature of science."

Van Deventer (106, pp. 391-393), in 1955, summarized basic principles which are characteristics of the modus
operandi of scientists. Many of these principles have been incorporated into examination items in evaluative instruments dealing with understanding of the nature of science and its processes. The principles listed by Van Deventer were:

1. Principle of Objectivity
2. Principle of Tentativeness
3. Principle of Consistency
   A scientist assumes that the behavior of the universe is not capricious, but is describable in terms of consistent laws.
4. Principle of Uniformity
   A scientist believes that the forces which are now operating in the world are those which have always operated.
5. Principle of Causality
6. Principle of Parsimony
   A scientist prefers simple and widely applicable explanations of phenomena.
7. Principle of Materiality
   A scientist prefers material and mechanical explanations of phenomena, rather than those which depend on supernatural factors.
8. Principle of Dynamism
9. Principle of Relativeness
10. Principle of Intergradation
    A scientist thinks in terms of continua; he distrusts sharp boundary lines.
11. Principle of Practicality
    A scientist expects that in any situation involving competition among units of varying potentialities, those which work best under existing circumstances will tend to survive and be perpetuated.
12. Principle of Continuous Discovery

13. Principle of Social Limitation

14. Principle of Complementarity
   A scientist attempts to incorporate all phenomena into a single consistent, natural scheme, but he recognizes that contradictory generalizations may be necessary to describe different aspects of certain things as they appear to us.

Lawson (67, pp. 86-87) contacted a number of people in a wide variety of fields of the physical, biological, and social sciences to obtain their opinions as to the most important characteristic of the "scientific attitude." He combined their responses with the writings of other scientists to formulate the following summary statements (some in abbreviated form) as a general consensus:

a. It is free of dogmatism, a stranger to doctrinaire assertion;

b. Its conclusions are tentative;

c. It is cautious and involves a withholding of judgment;

d. It is not absolutistic, preferring to state facts as being apparent facts within a presently considered frame of reference;

e. It is objective and impersonal in the major task of seeking for answers and solutions;

f. It is motivated by intellectual curiosity—the desire to know, to explore, to experiment;

g. It involves a willingness to accept a rigorous course and follow it with single-mindedness;

h. It constantly is on guard against the danger of mistaking subjective inference for objective
evidence;

i. It shows a basic skepticism and a self-critical taste for self-evaluation;

j. It leans heavily upon controlled experiment and observations in the use of materials capable of providing sensori-perceptual data;

k. It seeks for unification of principles but does not insist on inferring from such unification any discovery of final or ultimate truth.

Vitrogan (107, pp. 175-186) sought to develop an instrument to measure a generalized attitude toward science. For this purpose, he formulated certain hypothetical criteria which were predicated to relate to a generalized attitude toward science. Instead of contacting people as Lawson did, Vitrogan derived his criteria primarily by analyzing the writings of the philosophers and scientists: Dewey, Pearson, Wendell Johnson, Cohen, Nagel, Russell, Kahn, Schwab, L. K. Frank, Conant, Bronowski, Poincaré, Shapley, Eddington, Reichenbach, Peirce, Douglas Johnson, Beveridge, Slichter, and Cannon. "A positive generalized attitude toward science was characterized" by the following statements, some in modified form:

1. a predisposition to discern the degree in which one person or thing differs from another

2. a tendency to challenge authority, to test traditional beliefs and customs

3. a readiness to change as changing conditions require
4. an ability to differentiate between controlled and reliable observation as opposed to casual observation

5. a basic notion that reality is to be regarded as a process implying continuous change

6. structure in the form of relations and equations will be stressed over functions

7. greater concern for research rather than findings

8. an emphasis on probability type explanations rather than absolute solutions.

Hurd (56, pp. 14-15), in 1970, wrote that a scientifically enlightened person has an operational understanding of certain knowledges and attitudes. Eight of his twelve included:

1. (He) understands the purposes of the scientific endeavor to be the establishment of general laws and the conceptualization of knowledge about the natural environment.

2. He recognizes that scientific knowledge grows, possibly without limit, where each new step depends upon and engulfs all that was known before.

4. He appreciates the worthiness of systematic investigation in the sciences.

5. He recognizes the interdependency of inquiry processes and the derived concepts, laws or theories.

7. He sees the need to view the scientific enterprise within the broad perspectives of culture, society, and history.

10. He views science and technology as interrelated and dependent upon each other; however, he is also aware that they are not synonymous and that their goals are different.
11. He appreciates the universality of scientific endeavors, their lack of national, cultural, or ethnic boundaries.

12. He has some awareness of the need to generate a system of concepts within which science, society, and the humanities can fit.

Others have attempted to contribute to a better understanding of the nature of science by describing some of the misunderstandings and fallacies. Gordon (52, pp. 1-8) described several misconceptions and erroneous attitudes regarding science which are held by many college freshmen and other students. Many believe that science is primarily techniques and gadgetry; that the measure of science's success is its contribution to the production of more gadgets. Some have an idea that any scientific law must be an algebraic equation. Another misconception is, "it must be infallible or it isn't science."

Nagel (77, p. 204) also believed that students had many erroneous ideas about the nature of science. He thought students should realize that collecting facts is neither the beginning nor the goal of scientific inquiry. In discussing the importance of method in science, in a liberal education, he stated:

Emphasis must be placed on the theoretical motivations that underlie the gathering of data, upon the selective character of observation and experiment, and upon the need to analyze and interpret the primary data of observation before they can be admitted as significant fact. Moreover, the student should be made to rec-
ognize that the concepts to which he is introduced . . . are intellectual creations, often suggested by the data, and are the production of a constructive imagination.

Fox (45, p. 58) pointed out understanding the effects of science, held by many high school teachers, is very limited. When asked, "what impact science has had on mankind, the responses usually include technological developments such as transistors, drugs, plastics, fertilizers, . . ." He observed that teachers rarely comprehend the "conceptual consequences of removing man from the center of the universe, putting him into the mainstream of nature . . ." Likewise, the significance of recognizing the physical autonomy of the universe, as contrasted with belief in a discernible purpose for man's benefit, is seldom comprehended by teachers.

Tyler (104, pp. 31-32), noting the rapid development of science and technology in contemporary life, stated that the individual, "... must understand science as a continuing process of inquiry, not as a set of firm answers to particular questions." He observed that if science is taught as a set of final answers, the students would be no better prepared for intelligent, responsible behavior in the changing, scientific, and technological world than uneducated savages. Tyler then described the
process of scientific inquiry (page 29 of this chapter).

For the purpose of this investigation, the descriptive definition (Chapter I, p. 11) of the nature of science derived from the Test on Understanding Science and the Wisconsin Inventory of Science Processes will be used. The nature of science means a human, intellectual, dynamic, objective, knowledge-developing activity which involves assumptions, observing, hypothesizing, experimenting, measuring, classifying or organizing, evaluating, and reporting or communicating. Scientific activity makes extensive use of ideal concepts or models, hypotheses, theories which are regarded as tentative, laws, and probability type explanations rather than absolute statements. The activity, international in character, affects society and is affected by society.

Summary. The definition of understanding of the nature of science used in this study is given in the preceding paragraph.

A comparison of the ideas presented in the writings of the four authors cited who itemized their ideas in considerable detail revealed many traits in common. The various items on the evaluative instruments, TOUS, and WISP, are generally in close agreement with the ideas expressed by one or more of the authors cited in this section. No
points of genuine disagreement were found. The following are believed to be the more important characterizations of the nature of science or of scientists investigating natural phenomena: objectivity (Van Deventer, l; Lawson, e, h), tentativeness (Van Deventer, 2; Lawson, b), free of dogmatism (Lawson, a; Vitrogan, 2), motivated by the desire to know (Lawson, f; Vitrogan, 7), controlled experimentation (Lawson, j; Vitrogan, 4; Hurd, 4), not absolutistic (Lawson, d; Vitrogan, 3, 8), dynamism (Van Deventer, 8; Vitrogan, 5), relativeness (Van Deventer, 9, 10; Vitrogan, 1; Hurd, 5), suspended judgment (Lawson, c, i), materiality (Van Deventer, 7; Hurd, 1), continuous discovery (Van Deventer, 12; Hurd, 2), seeks for unification of principles (Lawson, k; Van Deventer, 3, 14), and social limitations (Van Deventer, 13; Hurd, 7, 12).

Characteristics given by Van Deventer and not by the others were causality, uniformity, parsimony, and practicality. Vitrogan's unique contribution was that structure and relations should be emphasized over function. Hurd's contributions, not repeated by the others, were an appreciation of the universality of scientific endeavors and a view of science and technology as interrelated and dependent on each other.

Careful examination of the TOUS revealed the following topics were not mentioned or strongly alluded to in
any of the cited writings: scientific societies, instruments, money, institutional pressures on scientists, and controversies in science.

Similarly, examination of the WISP revealed the only topics not mentioned or strongly alluded to in any of the writings cited were: the results of science are amoral, measurement, and prediction.

**Understanding of the Nature of Science as an Educational Objective**

Many committees and individuals, concerned with the objectives of science teaching, have made recommendations that some understanding of the nature of science, including its methods and processes, be among the major objectives. The pronouncements of several committees will be considered first.

In 1927, the American Association for the Advancement of Science had a special committee study and report on science in education. This report (4, p. 664) emphasized the increasing importance of scientific method as a major objective of science instruction in a scientifically oriented society.

Scientific method, the development of science, and the societal implications of science as objectives in the AAAS report were stated more clearly by a committee of the National Science Teachers Association (99, p. 28) in
1961 when it formulated long-range objectives of science education. Two of the objectives pertinent to this part of the present study were:

1. Students should know something of the development of science and of the people who have contributed toward it.

2. Students must acquire a working concept of the relations between science and society, science and individuals, and science and technology.

One of the more recent (1966) sets of recommendations regarding objectives of science instruction was made by the Educational Policies Commission in *Education and the Spirit of Science* (36, pp. 15-27). They believed that to communicate and develop the spirit of science should be among the principal goals of education throughout the world. The values which they thought characterized the scientific enterprise as a whole, and that underlie all science, are:

1. Longing to know and to understand
2. Questioning of all things
3. Search for data and their meaning
4. Demand for verification
5. Respect for logic
6. Consideration of premises
7. Consideration of consequences

Two different groups were concerned with objectives specifically at the college level. The first was the
American Council on Education which sponsored the Cooperative Study of Evaluation in General Education (32, pp. 102-105)—one of the most extensive evaluations of college educational objectives that has been made. The three and one-half year study was started in 1950 by about 50 people who represented 19 colleges and universities. All participating institutions accepted the objective, "To understand the common phenomena in one's physical environment, to apply habits of scientific thought to both personal and civic problems, and to appreciate the implications of scientific discoveries for human welfare," as expressive of their desires for general education science courses. The science committee enunciated this general objective more specifically as consisting of developing the student's:

1. Ability to apply science knowledge to new problems and situations.

2. Ability to read and evaluate news articles and popular writing on scientific developments.

3. Understanding of the point of view with which a scientist approaches his problems and of the kinds of things that he does.

These three objectives were regarded as most important. One other, pertinent to this part of the study, was the importance and limitations of science in the modern world.

The second group concerned with college level objectives was the American Association of Physics Teachers.
which sponsored a conference, in 1956, to consider the quality and effectiveness of introductory physics courses. The Conference (6, pp. 419-420) adopted as objectives for any introductory physics course, regardless of level of preparation, these goals outlined by the American Institute of Physics in 1955. These were originally formulated for engineering students but were considered to be so basic and comprehensive as to be worthy objectives for any course. Six of the objectives having a bearing on this part of the present study are listed. The course should:

1. Impart some of the physicist's curiosity about the physical world.

2. Indicate the scope and limitations of descriptions and interpretations.

3. Show that underlying, unifying principles can be expressed in mathematical terms.

4. Emphasize the historical development of physics and the struggle to get new ideas accepted.

5. Direct attention to the development of precise methods of measurement and their importance.

6. Through an understanding of the laws of nature, promote more objective judgments, freeing one from superstitions and fears.

Two comprehensive surveys have been made of the objectives of science teaching. The first was by Blanc (10, pp. 47-52) who made a review of the general objectives in science teaching covering a period of approximately 30 years (1920-1950). He attempted to reduce all
of the proposed objectives to five major objectives with several sub-divisions. Those most closely related to this part of the present study include: (1) developing interests and appreciations in the benefits of science, especially gaining new meanings in the contributions of science, and (2) developing democratic social attitudes toward the resources of science. The latter included understanding the impact of technology on society and seeing scientific developments as a solution to some social problems.

The second survey was by Hurd (55, pp. 89-96) who made an analysis of 1373 opinionated articles, written by secondary school science teachers, which were published in School Science and Mathematics (1901 to 1951) and in Science Education (1931 to 1951). There were more articles about objectives than any other topic in every decade—a total of 791 articles. He found the major objectives of science teaching supported by teachers during the half-century, were:

1. To train students in the scientific method of thinking (191)
2. To develop certain scientific attitudes within students (105)
3. To acquire a fund of useful information (60)
4. To develop an understanding of the major principles of science (59)
He noted the last two objectives had decreased in frequency considerably since 1930 and that the following two had increased.

To appreciate the contribution of science to society (42)

To aid the student to discover a vocational interest (21)

This shift in objectives was evidence of increasing concern that attention should be given to the relation of science to society and that science education should have some utilitarian objective.

During and after World War II, many individuals and some committees expressed their concern that science teachers should make their students aware of the interaction of science and society. Expressions of this concern which follow are in chronological order. Powers (88, pp. 136-141), in 1944, proposed five major outcomes of science education to meet the needs of youth in the years following World War II. Among the outcomes were "comprehension of the impact of science on our society" and the "ability to select and use materials made available by science in solving social problems."

Richardson (90, pp. 250-251) was concerned with the social implications of science in the education of science teachers. In 1945, he stated as one problem: "Prospective science teachers exhibit a lack of understanding of
the social function of science." He indicated that their idea of the contribution of science to society was in terms of gadgets or a research laboratory. He deplored the limited objective of prospective science teachers who appear "to believe that the only goal of science teaching is the imparting of information."

Also concerned with the social function of science was the Cooperative Committee on the Teaching of Science of the AAAS (5, p. 112). In 1946, this Committee recommended that prospective science teachers study "the social consequences and historical background of science."

The Committee responsible for the 46th Yearbook of the National Society for the Study of Education emphasized the process outcomes of science teaching which have been referred to earlier in this chapter. The Committee also recognized (78, p. 141) the causal role of science as a factor in many social and economic problems. These problems demand that "the science teacher set the goals of instruction well beyond the narrow confines of the specialized sciences such as biology, physics, and chemistry."

Garrett (49, pp. 216-217), in 1949, was concerned with the social objective and other broad objectives of science instruction in a liberal arts education. Among his objectives to produce better informed citizens were: (1) awareness of the impact of "scientific developments
upon social, economic, political, and religious problems," (2) some "perspective of what can be expected of science in the future," and (3) "conviction that science is power."

Krauskopf (65, p. 62), in 1951, described the efforts of teachers to point out the importance of science in daily life to college students who question science requirements in general-education programs. He stated:

What we do mean is that throughout his life he will be hearing about new inventions and new theories; that as a voter and taxpayer he will help to decide such currently important questions as whether research should be government-supported and whether some kinds of research should be shrouded in secrecy; that as a business man or public official he may be called upon to allocate funds for research, to recommend for or against hiring a scientist; that he will find his thinking in economics, politics, philosophy, and religion subtly but powerfully influenced by science; and that he will hear much about possible applications of the scientific method to economic and social questions. Because science will influence his life in so many diverse ways, he needs to know something of its methods, its goals, its capabilities, and its limitations.

Wilson (114, p. 164), like Garrett and Krauskopf, was concerned with the objectives of science courses at the college level. In 1954, referring to the great majority of college students who will not enter scientific occupations, he stressed attitudes toward science as being more important than detailed knowledge of scientific laws. He stated:

Certainly an understanding of science in the
broader sense is dependent upon some knowledge of scientific laws and principles. However, the student in the college science course may fail to see many of the implications of science and its relationships to other fields of knowledge if no specific attention is given to this phase of science teaching. Our science courses should be giving definite attention to the purposes, methods, possibilities, and limitations of science and its relationships to other activities of men.

Wolfle (118, p. 181), in 1957, discussed science and public understanding generally. He pointed out that the time is past when educated people can get along successfully without an understanding of science. He thought it as important that a citizen understand the influence of science and technology on the economy of his country as to understand the role of historical forces. He further stated: "Science has become an instrument of such power in changing society--whether that be good or bad--that no nation that pretends to have an educated citizenry can neglect it."

Cohen (21, pp. 32-33), in 1964, proposed eleven values and goals for present day science teaching. Four which are relevant to this part of the present study are:

1. To provide enough understanding to enable the educated citizens to collaborate intelligently with those who are actively engaged in scientific pursuits.

2. To enable the citizen both to criticize and to appreciate the effects of the sciences on his society. . . .
3. To give a practical grasp of scientific methods of grappling with problems, at least sufficient for problems which the student will face in his individual and social life.

6. To understand the place of science among other intellectual and esthetic pursuits; briefly, to see the sciences themselves a humanistic enterprise.

In addition to knowledge of the interaction of science and society, various individuals have emphasized the need for understanding the processes of science, its historical and philosophical development.

Conant (22, p. 4), in stressing the need for developing understanding of scientific processes, stated:

The remedy does not lie in greater dissemination of scientific information among non-scientists. Being well-informed about science is not the same thing as understanding science, though the two propositions are not antithetical. What is needed is methods for imparting some knowledge of the tactics and strategy of science.

Allen (2, p. 38) believed one of the most important areas of need, revealed by his study, was the "evident misunderstanding and ignorance of the nature of science." To cope with this problem, he suggested that serious attention be given to the history and philosophy of science. He believed that systematic attempts should be made at every grade level to help students comprehend the meaning of science. He stated: "An understanding of the evolution of scientific thought, its impact on other modes of think-
ing, and its relation to the success of scientific endeavor are in considerable measure essential to a comprehension of the nature of science."

In the same vein, Fitzpatrick (43, pp. 188-189) discussed some of the deficiencies that existed in many professional programs for educating science teachers. He observed that most science teachers appeared naive about the historical and philosophical aspects of science. This resulted from their college course work avoiding these aspects of science; consequently, it is difficult for these teachers to effectively enlighten their students as to how scientific conclusions are established and used.

Finally, Goodlad (51, p. 54), in a comparison of curricula in 1964, observed that there is a high degree of similarity in the objectives of the various "new" curricula in physics (PSSC), chemistry (CBA and CHEM), biology (BSCS), mathematics, and social studies. He commented: "Objectives . . . stress the importance of understanding the structure of the discipline, the purposes and methods of the field, and the part that creative men and women played in the development of the field."

Summary. Some concern was expressed over forty years ago by the American Association for the Advancement of Science that scientific method should receive more atten-
tion as an objective of science education. Since the mid-1940's, there has been an increasing number of scientists, science educators, representatives of professional organizations, and others who have recommended that understanding of the nature of science and its processes should be a major educational objective.

Statements of support for the understanding of the nature of science as an objective at the college level have been made by Garrett (49), Krauskopf (65), Wilson (114), the American Council on Education's Committee on Cooperative Study of Evaluation in General Education (32), and the American Association of Physics Teachers (6). Support of the objective has come from those—Richardson (90), Fitzpatrick (43), and the Cooperative Committee on the Teaching of Science of the AAAS (5)—concerned with the preparation of teachers.

Statements for support for the understanding of the nature of science as an objective, primarily at the secondary level, have been made by Powers (88), Allen (2), Cohen (21), and committees of the National Science Teachers Association (99). Statements of support as an educational objective without respect to level, in order to have an enlightened citizenry, have been made by Conant (22), Wolfle (118), committees of the National Society for the Study of Education (78) and the Educational Poli-
cies Commission (35).

**Review of Pertinent Research**

The related research literature for this section of the study has been divided into three major categories, depending on the evaluation instruments used: (1) those studies which used the Test on Understanding Science, (2) those studies which used either the Wisconsin or the Welch Science Process Inventory, and (3) those studies which used some other instrument.

**Studies Using the Test on Understanding Science**

Craven (25, pp. 133-134) concluded that understanding of the nature of science, as measured on the criterion instrument, TOUS, Form W, was not a major learning outcome of the study of college science. Evidence for this included:

a) With only one exception, TOUS scores were not significantly correlated to either the total number of college science credits completed or to the total number of science grade points earned by members of any of the participating groups. The one exception was a significant negative correlation between TOUS scores and the total number of science grade points earned by the science teacher-candidates.

b) Failure to find a significant difference in understanding of science between science teacher-candidates and social science teacher-candidates, the former groups having completed twice as many credits as had the latter.
Craven (25, p. 88) found at Oregon State University that freshmen social science education majors differed from freshmen science education majors in their mean scores on the TOUS as shown in Table 1.

**TABLE 1**

<table>
<thead>
<tr>
<th>Major</th>
<th>N</th>
<th>Part I</th>
<th>Part II</th>
<th>Part III</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresh.-Soc. Sci. Ed.</td>
<td>19</td>
<td>11.21</td>
<td>12.58</td>
<td>12.21</td>
<td>36.00</td>
</tr>
<tr>
<td>Fresh.-Sci. Ed.</td>
<td>41</td>
<td>11.76</td>
<td>12.17</td>
<td>13.10</td>
<td>37.02</td>
</tr>
</tbody>
</table>

* Craven (25, p. 88)

Crawford and Backhus (26, pp. 1-7, 18-20) investigated the effectiveness of three different laboratory approaches — (1) scheduled or structured, (2) free or audio-tutorial, and (3) home or loosely structured — in physical science at Kansas State Teachers College in 1969. They sought to determine the effectiveness of each of these types on the development of: (1) specified cognitive skills, (2) favorable attitudes toward science, and (3) an understanding of the scientific enterprise as measured by the TOUS, Form W. A total of 333 students were involved. No significant differences were found among the laboratory groups on the TOUS for the first experimental period (beginning January, 1969). They did find significant differences
among the laboratory groups on the TOUS for the second experimental period (beginning September, 1969). The group mean scores for the three different types of laboratory experience for both experimental periods is summarized (26, p. 9) in Table 2.

TABLE 2

GROUP MEAN SCORES ON TOUS FOR BOTH EXPERIMENTAL PERIODS AT KANSAS STATE TEACHERS COLLEGE*

<table>
<thead>
<tr>
<th>Experimental Period</th>
<th>N</th>
<th>Part I</th>
<th>Part II</th>
<th>Part III</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scheduled Lab</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First, Jan. '69</td>
<td>47</td>
<td>11.0</td>
<td>11.7</td>
<td>12.7</td>
<td>35.5</td>
</tr>
<tr>
<td>Second, Sep. '69</td>
<td>70</td>
<td>11.0</td>
<td>11.2</td>
<td>12.1</td>
<td>34.3</td>
</tr>
<tr>
<td><strong>Free Lab</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First, Jan. '69</td>
<td>44</td>
<td>11.4</td>
<td>11.9</td>
<td>12.4</td>
<td>35.7</td>
</tr>
<tr>
<td>Second, Sep. '69</td>
<td>66</td>
<td>11.7</td>
<td>12.3</td>
<td>13.7</td>
<td>37.9</td>
</tr>
<tr>
<td><strong>Home Lab</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First, Jan. '69</td>
<td>49</td>
<td>11.2</td>
<td>11.7</td>
<td>12.7</td>
<td>35.5</td>
</tr>
<tr>
<td>Second, Sep. '69</td>
<td>57</td>
<td>11.2</td>
<td>12.3</td>
<td>13.1</td>
<td>36.7</td>
</tr>
<tr>
<td><strong>Total Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First, Jan. '69</td>
<td>140</td>
<td>11.2</td>
<td>11.8</td>
<td>12.6</td>
<td>35.6</td>
</tr>
<tr>
<td>Second, Sep. '69</td>
<td>193</td>
<td>11.3</td>
<td>11.9</td>
<td>13.0</td>
<td>36.2</td>
</tr>
</tbody>
</table>

* Crawford and Backhus (26, p. 9)
Crawford and Backhus (26, p. 18) concluded that, for the second experimental group, free laboratory and home laboratory groups "scored significantly higher than the scheduled lab groups in terms of their understanding of science as measured by TOUS."

Durst (34, p. 2438-A) investigated the relative effectiveness of a college general biology course which utilized an audio-tutorial laboratory as contrasted with a conventional lecture course. The study involved 720 students in five biology classes at Kansas State Teachers College, Emporia, Kansas. Three of the classes were audio-tutorial, while the other two were traditional biology. Seventy students, thirty-five males and thirty-five females, were selected from each of the two groups. Among the evaluative instruments used was the TOUS, Form W. Results indicated that the students in the audio-tutorial classes achieved higher scores on the evaluative tests than the students in the conventional classes did. It was noted that "none of the differences were significant."

There have also been a number of studies conducted at the secondary school level. Several will be reviewed here.

Klopffer and Cooley (64, pp. 33-47) investigated the effectiveness of using the History of Science Cases Instruction Method in high schools in 1960-61. There were
approximately 2800 students involved in large, medium, small, urban, suburban, rural, public, and private schools throughout the nation. Each participating teacher taught two units of his choice by the HOSC Instruction Method during the year to his experimental classes. In general, each unit took 10-11 teaching days. The TOUS, Form W, was the criterion instrument. They concluded the HOSC Instruction Method "is definitely effective in increasing student understanding of science and scientists when used in biology, chemistry and physics classes in high schools." It was further noted that students made these significant gains in understanding of science and scientists "with little or no concomittant loss of achievement in the usual content of high school science courses."

Klepinger (63, p. 5789) found that teaching a unit on atomic structure using the historical approach resulted in high school physics students of superior ability making significant gains on the TOUS.

Trent (102, pp. 224-229) investigated the relative effectiveness of the traditional high school physics course and the PSSC physics course in obtaining the objective "understanding of science" as measured by the TOUS. The experimental group consisted of 26 randomly selected California high schools that were teaching PSSC physics. The control group consisted of 26 randomly selected Cal-
ifornia high schools that were teaching traditional college preparatory physics in 1963-64. He concluded that when "the variables of prior science understanding and mental ability are statistically controlled, there is no difference between the control and the experimental group" in mean scores on the TOUS. Without statistical control of mental ability and prior science understanding, students studying the PSSC course did significantly better than those studying traditional physics.

Crumb (27, pp. 246-250) conducted a study to determine if there were a significant difference in understanding science between students who had studied PSSC physics and those who had studied traditional physics. The study population consisted of 1275 physics students from 29 rural and urban high schools in Iowa, Kansas, Missouri and Nebraska in 1963-64. The TOUS was administered at the beginning of the school year and near the end of each semester. He concluded that "when adjustments are made for scholastic aptitude, background in natural science, and prior science understanding, significant differences" are found between the PSSC students' and traditional physics students' understanding, in favor of the PSSC group.

Cossman (24, pp. 69-70) investigated learning outcomes of a special course in "Science and Culture" that he designed for secondary school students. The 21 eleventh
and twelfth grade students in "Science and Culture" showed substantial mean gains in scores on a variety of evaluative instruments. A group of effectively matched subjects who did not take the course "failed to achieve such mean gains." Students in the experimental course evidenced a significantly greater increase in the following than those who did not take the course:

1. Understanding of the scientific process, as measured by the TOUS and the Science Opinion Survey.
2. Understanding of scientists as an occupational group, as measured by the Facts About Science Test and the TOUS.
3. Understanding of science as an institution and its relationship to other institutions in our society, as measured by the FAST and the TOUS.

Summary. Significant gains in understanding of science as measured by the TOUS were reported by most of those who offered some special form of instruction. A special course—Cossman; a special unit as part of a course—Klopfer and Cooley, Kleppinger; a special laboratory (free and home laboratory) program—Crawford and Backhus; and one group taught PSSC Physics—Crumb; all reported significant gains.

Durst reported his experimental college biology group
made higher scores on the TOUS than the conventional group but the difference was not significant. When no special instruction in science was involved, no significant gains at the college level were reported by Craven. Trent's study of PSSC and non-PSSC physics classes reported no significant differences between the two groups.

Studies Using Science Process Inventory Tests

Miller (74, p. 4313-A) administered the Welch Science Process Inventory, Form D, to 161 prospective elementary teachers, 71 future secondary school social studies teachers, and 29 prospective secondary school science teachers—all at The Ohio State University. He found that there was a significant difference in the knowledge of science processes held by the prospective elementary teachers and the prospective high school science teachers in favor of the latter group. There was no significant difference between the prospective elementary teachers and the future high school social studies teachers.

Niman (79, p. 226-A) developed instructional materials to aid college students in setting up mathematical formulations of physical situations. The experiment involved 76 (75 women, 1 man) undergraduate liberal arts students in a one-semester physical science course at Hunter College of the City University of New York in
1968. He used the Welch Science Process Inventory as a pretest and posttest. His results showed that the means of the experimental and the control groups did not differ from each other. However, "supplementary analyses suggested that some differences did exist between the experimental and the control groups on the Welch."

Wittwer (116, pp. 4519-4520-A) reported on evaluation of various objectives for participants in the National Science Foundation Research Participation Program for High School Science Teachers at the University of Wisconsin. One of the objectives was to develop in the participants "understanding of the methods or processes of science." Evaluation of the attainment of this objective was made by means of a specially designed Science Process Inventory administered to the 67 RPP teachers and to a comparable group of science teachers who had not participated in the research program. The RPP science teachers did score significantly higher on the instrument than did the teachers who lacked the research experience.

Larson (66, p. 653-A) investigated the preparation in science of prospective elementary teachers enrolled at the University of Wisconsin. More specifically, he sought to determine the effectiveness of one science methods course in contributing to several factors, among them, understanding of science processes as measured by the Wis-
Inventory of Science Processes which was given as a pretest and posttest to 88 prospective elementary teachers. "The Wilcoxon test data indicated that the science methods course was effective in increasing knowledge of science processes. . . ."

Four studies at the secondary school level will be reviewed.

Welch and Walberg (113, pp. 605-614) reported on an extensive investigation of "pretest effect" and "sensitization" when the same criterion instruments were used in pretest and posttest evaluations. "Pretest effect" is the hypothesis that students who take pretests get higher scores on the posttests because of item-practice effects. "Sensitization" is the hypothesis that students in the experimental group get higher scores on the posttests because they learn what to concentrate on during the pretest. The sample consisted of approximately 2200 physics students from 57 high schools selected at random throughout the country. Among the six evaluative measures Welch and Walberg used were the Welch Science Process Inventory and the Test on Understanding Science. They concluded (113, p. 613) that pretesting effect "was not significant at the .05 level on any of the six criteria, suggesting that the pretest alone does not lead to significantly higher posttests when the treatment duration is as long
as seven months . . ." They also concluded the sensitization effect "was not significant on any of the six criteria, which implies that the test did not sensitize the subjects in the experimental course."

Petit (85, p. 1923-A) studied the teaching behavior of eight PSSC and non-PSSC physics teachers and the effects on their 368 students, of whom 55% were in the non-PSSC courses. She concluded that the PSSC and non-PSSC physics students were similar in their understanding of the nature of science as measured by the Welch Science Process Inventory.

Jaffarian (57, p. 62-A), in a study of twelfth grade students in 36 Wisconsin high schools, administered the Wisconsin Inventory of Science Processes as one of several instruments in the spring of 1967. He concluded that the average twelfth grade student "has a greater knowledge of the assumptions made by the scientist than knowledge of the processes of science (as indicated by the WISP scores) regardless of his academic science background."

Walberg (108, pp. 529-542) reported on a study to predict class learning by analyzing interaction of classroom climate, personality needs, and role expectations. The criterion instruments included the Test on Understanding Science and the Welch Science Process Inventory (Welch and Pella, 1968). These were administered to about 3700
high school physics students in 144 classes. Analysis of data revealed two variates; the first characterized by "many independent variables positively correlated with learning." The second was characterized by "an antithesis of cognitive and non-cognitive learning." The first variate was characterized by classes with high IQ's, high marks, non-authoritarian students who regarded their classes as difficult. These classes gained the most on the Test on Understanding Science, the Welch Science Process Inventory, and in physics achievement.

Summary. The Welch Science Process Inventory has been used both in high school and college studies. Miller found a significant difference in knowledge of science processes held by prospective elementary teachers and prospective high school science teachers in favor of the latter. Niman, with a one-semester college physical science course, found no difference in the means of his experimental and control groups. Larson found that a science methods course was effective in increasing knowledge of science processes, as measured by the Wisconsin Inventory of Science Processes, when used with prospective elementary school teachers. Wittwer reported that a group of high school science teachers who participated in a research program scored significantly higher on a special Science
Process Inventory than teachers without the research experience.

For high school physics students, Welch and Walberg reported that pretest effect and sensitization effect were not significant with the WSPI and the TOUS as criterion instruments. Petit found no difference between PSSC and non-PSSC students as measured by the WSPI. Walberg found that high school physics classes with high IQ's, high marks, and non-authoritarian students gained the most on the WSPI.

Jaffarian, who used the Wisconsin Inventory of Science Processes, concluded the average twelfth grade student had a greater knowledge of assumptions made by scientists than of the processes of science.

Studies Using Other Tests

Wilson (114, pp. 159-164) prepared a set of 26 statements about science. The basic ideas in many of them were suggested by *Science and Common Sense* by Conant, and *The Path of Science* by Mees. The test was given to a wide variety of students totaling 285 in three colleges. Some of the conclusions reported in 1954 were:

1. Failure to distinguish between "pure" and applied science. Fifty-seven per cent agreed that "real advances in science consist in the production of
useful devices such as automobiles and radios."

2. A lack of understanding of the necessity for freedom of investigation in science.

3. A majority considered the primary purpose of science was concern with man's physical comfort. Thirty-one per cent thought science was responsible for much evil in the world.

4. A relatively low percentage thought scientists more intelligent than those in other lines of work. A large percentage thought they "made use of lucky guesses."

5. In general, the response of these students indicated "a considerable lack of understanding of science and its place in our society."

Diehl (29, pp. 874-875-A) investigated the use of two teaching techniques, one rigid (control) and the other non-directive (experimental). Both groups used Physical Science for the Non-Scientist materials. The experiment was conducted for one trimester at Miami University. He concluded there was no significant difference between experimental and control groups "in providing a greater understanding of the role of science in contemporary society through the use of PSNS curriculum." The criterion instrument for this conclusion was the Facts About Science Test.
Gaides (48, p. 4241) investigated learning outcomes of prospective elementary education majors by pre- and posttesting those enrolled in a physical science course. He used the Facts About Science Test to measure attitudes toward science. He found that having had one or more years of science in high school was related to students scoring higher on the test than those who did not take any science. He also found that three courses, physical science survey-I, introductory physical geology, and descriptive astronomy, "all appeared equivalent in providing students with experiences which would improve their attitude toward science."

Behnke (7, pp. 193-207) conducted an investigation involving two groups: (1) a sample of 400 biology and 600 physical science teachers, and (2) a sample of 100 scientists. A questionnaire of 50 items in four categories was developed. The categories were:

A. The nature of science: statements on the character, importance, and limitations of science.

B. Science and society: statements on the relation of science to the social order.

C. The scientist and society: statements on the role and place of scientists in society.

D. The teaching of science.

She concluded that when all data were taken into account, the significant differences among sub-groups "were
very much fewer than the similarities." Some items were surprising. "Over one-half the teachers felt that scientific findings were not tentative," and 20% of the scientists felt the same. Only 20% of the scientists believed that the objective of scientific work was to improve human welfare. Over 50% of the teachers thought it was the objective. She stated: "The modern science teacher needs much more knowledge . . . but he must have more than knowledge of science content. Knowing a great deal about science is no guarantee he will know what science is about."

Several years later, Kimball (62, pp. 1498-1499) constructed a model of the nature of science and a scale to measure departures from the model. He sought to determine, "Do qualified science teachers understand the nature of science in the same way that practicing scientists do?"

Population samples were drawn from 965 science and philosophy majors graduated in five selected years from Stanford University and San Jose State College. He found no difference in understanding by science teachers or scientists "when the variable of undergraduate education was controlled." In all comparisons made, "the teachers scored higher than the scientists, but never significantly." The philosophy majors showed "closer agreement with the model than did science majors or scientists," especially in dealing with methodology in science.
Summary of Pertinent Research Studies. Significant gains in understanding the nature of science and its processes occurred in most cases when some special form of instruction was used. This occurred with the TOUS when used by Cossman, Klopfer and Cooley, Kleppinger, Crawford and Backhus (for one experimental period) and Crumb. It occurred with a Science Processes Inventory used by Wittwer, and with the WISP used by Larson in a science methods course.

No significant gains or no significant differences occurred when a course was taught with no special emphasis on the nature of science and its processes. This was reported by Craven using the TOUS at the college level. Trent, using the same instrument, reported no difference between experimental and control high school physics classes. Petit found no significant differences in high school physics classes using the WSPI. Gaides, using the FAST to compare three college courses--physical science, introductory geology, and descriptive astronomy--found no significant difference among them in improving attitudes toward science.

No significant differences between the control and experimental (a special course) groups toward understanding the nature of science were reported by Niman using the WSPI, Diehl using the FAST, and Durst using the TOUS.
The first two were college physical science courses; the first was a one semester course, and the second was a one trimester course. The third was college general biology.

Behnke and Kimball each sought, by different techniques, to compare science teachers with practicing scientists in their understanding of the nature of science. Behnke reported that significant differences were much fewer than similarities. Kimball found no difference "when the variable of undergraduate education was controlled." Wilson, using his own instrument, found a wide variety of college students were lacking in understanding of the scientific enterprise and its purpose.

Miller, using the WSPI, found significant differences between prospective elementary teachers and prospective high school science teachers in their understanding of science processes. The latter group was favored.

Walberg found that high school physics classes, characterized by high IQ's, high grades, and non-authoritarian students, gained the most on both the TOUS and the WSPI.

Welch and Walberg reported in a study of high school physics students that pretest effect and sensitization effect were not significant with the WSPI and the TOUS.

Jaffarian, using the WISP with twelfth grade students, concluded that they knew more about the assumptions made by scientists than about the science processes.
CHAPTER III

DESIGN OF THE STUDY

This chapter is divided into six sections:

1. Purpose
2. The population
3. The evaluation instruments
4. The experimental design
5. Collection of the data
6. Procedures for analysis of the data

PURPOSE

This study was designed to investigate the problem as stated in Chapter I: Do changes in college students' understanding of the nature of science and in critical thinking result from modifying a "traditional," general education, physical science course?

Answers were sought to the following three sub-problems:

I. Do students, after taking a year of physical science, make significant gains as shown by their pretest and posttest scores on the criterion instruments?

II. Were there significant differences in learning outcomes as measured by the criterion instru-
ments between the control physical science group and the experimental physical science group?

III. Do the teachers influence significantly the changes in learning outcomes as measured by the criterion instruments?

THE POPULATION

The subjects of this experiment were all of the students regularly enrolled in Physical Science (Science 100-101) at Gorham State College of the University of Maine for the academic years 1968-69 and 1969-70. Only those students who completed the full year's work were included in the experiment—144 for the first year and 190 for the second year as shown in Table 3.

TABLE 3
ENROLLMENT AND LOSSES IN PHYSICAL SCIENCE

<table>
<thead>
<tr>
<th>Year</th>
<th>Starting Enrollment</th>
<th>Loss</th>
<th>%Loss</th>
<th>Final Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1968-69</td>
<td>174</td>
<td>30</td>
<td>17.24</td>
<td>144</td>
</tr>
<tr>
<td>1969-70</td>
<td>230</td>
<td>40</td>
<td>17.39</td>
<td>190</td>
</tr>
</tbody>
</table>

The course is required of all students who are prospective elementary (K-6 inclusive) education majors. The majority of the subjects were freshmen women as in-
dicated in Table 4.

### TABLE 4

ENROLLMENT DISTRIBUTION BY SEX AND CLASS

<table>
<thead>
<tr>
<th>Class</th>
<th>Control Group 1968-69</th>
<th>Experimental Group 1969-70</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>Freshmen</td>
<td>114</td>
<td>12</td>
</tr>
<tr>
<td>Sophomores</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Juniors</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Seniors</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Special</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>126</td>
<td>18</td>
</tr>
</tbody>
</table>

Approximately 85% of the subjects were residents of Maine and had graduated from Maine secondary schools. Of the non-resident students, over 50% were from Massachusetts.

Each year's group was similar in terms of their enrollment in secondary school science courses: 95% had taken biology, 85% had chemistry, and 21% had physics.

THE EVALUATION INSTRUMENTS

In order to investigate the problems originally raised in this study, it was necessary to obtain data with valid and reliable instruments.
The Scholastic Aptitude Test was selected to provide a statistical control for the scholastic aptitude of the subjects of this study. There were three reasons for selecting this instrument.

1. It has been used extensively with various studies involving college students for many years and is generally accepted as an excellent instrument. Zimmerman (16, p. 708), in a review in Buros, stated: SAT has a number of major strengths which should be noted. In the first place, the test is very carefully constructed and it is thoroughly analyzed by appropriate methodology before being presented for use in the selection of students. Its range is sufficiently wide to differentiate among both the lowest and highest levels in the college bound or freshman samples. No other competing test publisher prepares so many alternate forms in which items are so carefully matched for content and difficulty and scores are so painstakingly equated. . . . there is probably no better test available to estimate the entering students' college level verbal and mathematical comprehension.

2. The Scholastic Aptitude Tests consistently maintain a high degree of reliability. Bowers (16, pp. 705-706), in a critique in Buros, noted test-retest reliability coefficients of 0.89 for the verbal scale and 0.85 for the mathematical scale for time intervals up to ten months. He also pointed out that reliability coefficients
for both scales for fourteen SAT forms in 1959-1962 consistently approximated 0.90. Standard errors of measurement for both scaled scores of these fourteen forms varied generally between 30 and 35 points. The verbal scale has been found to be a better predictor of freshman grades than the mathematical scale in liberal arts type colleges.

3. The Scholastic Aptitude Test scores, both verbal and mathematical, were readily available in the registrar's office for 98% of the subjects for each year of the study.

Test on Understanding Science, Form W

The Test on Understanding Science, Form W, was developed by Cooley and Klopfer at the Harvard University Graduate School of Education. Form W, published by Educational Testing Service, became available in 1961. Three preliminary forms had been used and revised prior to that time. Form W consists of 60 four-choice items designed to sample student understanding of science in three areas. The areas were established by Cooley and Klopfer in consultation with a number of scientists, science educators, and professors of philosophy of science.

The three areas and the themes associated with each,
for which specifications were developed as a basis for TOUS according to Cooley and Klopfer (23, pp. 3-6), are:

Area I--The Scientific Enterprise, 18 items

1. Human element in science
2. Communication among scientists
3. Scientific societies
4. Instruments
5. Money
6. International character of science
7. Interaction of science and society

Area II--The Scientists, 18 items

1. Generalizations about scientists as people
2. Institutional pressures on scientists
3. Abilities needed by scientists

Area III--The Methods and Aims of Science, 24 items

1. Generalities about scientific methods
2. Tactics and strategy of sciencing
3. Theories and models
4. Aims of science
5. Accumulation and falsification
6. Controversies in science
7. Science and technology
8. Unity and interdependence of the sciences

The distribution of questions by areas on TOUS, Form W, (23, p. 18) is summarized in Table 5, page 95.

In TOUS, Area I--"The Scientific Enterprise," there were three topics not mentioned directly in any of the writings cited in Chapter II: (3) Scientific societies, (4) Instruments, and (5) Money. In TOUS, Area II--"The Scientists," the only topic not mentioned in the cited writings was (2) Institutional pressures on scientists.
In **TOUS**, Area III—"The Methods and Aims of Science," the only topic not mentioned or strongly alluded to in the writings cited was (6) Controversies in science.

**TABLE 5**

**TOUS, FORM W, DISTRIBUTION OF QUESTIONS**

<table>
<thead>
<tr>
<th>Part</th>
<th>Area</th>
<th>Number of Questions</th>
<th>Per Cent of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>The Scientific Enterprise</td>
<td>18</td>
<td>30.0</td>
</tr>
<tr>
<td>II</td>
<td>Scientists</td>
<td>18</td>
<td>30.0</td>
</tr>
<tr>
<td>III</td>
<td>Methods and Aims of Science</td>
<td>24</td>
<td>40.0</td>
</tr>
</tbody>
</table>

The publisher provides keys for scoring so that it is possible to obtain subscores for each area of the test in addition to obtaining a total score.

Although designed to be used with upper grade students in high school, the instrument has been used by various investigators, Craven (25), Durst (34), and Crawford and Backhus (26) for experimental purposes at the college level.

According to Cooley and Klopfer (23, p. 10), the reliability was determined for test **Form X** by applying the Kuder-Richardson Formula 20 to test data from 2535 students. **Form W** is essentially the same as **Form X** with only minor changes according to the authors. Reliability data is given in Table 6, page 96.
TABLE 6

TOUS, FORM X, RELIABILITY USING K-R FORMULA 20*

<table>
<thead>
<tr>
<th>Part</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area I</td>
<td>0.58</td>
</tr>
<tr>
<td>Area II</td>
<td>0.52</td>
</tr>
<tr>
<td>Area III</td>
<td>0.58</td>
</tr>
<tr>
<td>Total</td>
<td>0.76</td>
</tr>
</tbody>
</table>

*Cooley and Klopfer (23, p. 10)

Cooley and Klopfer (23, p. 12) determined means and standard deviations based on 2980 high school students in a nationwide sample tested with Form X in October, 1960. The results for grades 10-12 are given in Table 7.

TABLE 7

TOUS, FORM X, TENTATIVE NORMS FOR HIGH SCHOOL STUDENTS, 1960*

<table>
<thead>
<tr>
<th>Grade 10</th>
<th>Grade 11</th>
<th>Grade 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Students</td>
<td>1064</td>
<td>994</td>
</tr>
<tr>
<td>Mean Score</td>
<td>28.58</td>
<td>31.57</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>7.66</td>
<td>7.02</td>
</tr>
</tbody>
</table>

*Cooley and Klopfer (23, p. 12)

Cornell Critical Thinking Test, Level Z

The Cornell Critical Thinking Test, Level Z, was selected to evaluate changes in student critical thinking
abilities. Level Z, by Ennis and Millman of Cornell University, was copyrighted in 1961 and is labeled an "Experimental Edition." It consists of 52 three-choice items distributed among seven sections which are concerned with various aspects of critical thinking. The seven sections are:

1. Drawing conclusions
2. Identifying faulty thinking
3. Identifying reliable statements
4. Deciding on the certainty of conclusions
5. Selecting the best prediction
6. Selecting the best definition
7. Identifying an unstated assumption

The distribution of questions on CCTT, Level Z is summarized in Table 8.

<table>
<thead>
<tr>
<th>Section</th>
<th>Number of Questions</th>
<th>Per Cent of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>19.23</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>21.15</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>7.69</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>25.00</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>7.69</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>7.69</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>11.54</td>
</tr>
</tbody>
</table>

Examination of the items of the Cornell Critical Thinking Test, Level Z, tends to indicate that it would be less susceptible to other class influences than the Watson-
Glaser Critical Thinking Appraisal. Another factor in the selection of this test was that it also measured abilities which are frequently identified as objectives of science education, often being associated with the "scientific method."

Information on earlier editions is contained in an unpublished Ph.D. thesis by Ennis (38, p. 2876). The earlier editions were given to 2456 students from ten high schools and two colleges.

The results of the CCTT, Level Z, for a small group of high school students are given in Table 9.

**TABLE 9**

<table>
<thead>
<tr>
<th>Grade 10 College Prep.</th>
<th>Grade 11 College Prep.</th>
<th>Grade 12 General</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Highest</td>
<td>33</td>
<td>39</td>
</tr>
<tr>
<td>Median</td>
<td>27.5</td>
<td>31</td>
</tr>
<tr>
<td>Lowest</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Mean</td>
<td>26.75</td>
<td>30.67</td>
</tr>
</tbody>
</table>

*Ennis (40)

In June, 1961, Level Z was administered to 62 Cornell undergraduate students in Social Foundations of Education, a required course for prospective teachers. The results obtained are given in Table 10, page 99.
TABLE 10

CCTT, LEVEL Z, RESULTS OF 62 CORNELL UNDERGRADUATES*

<table>
<thead>
<tr>
<th>Number of Students</th>
<th>62</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (R-W/2 formula)</td>
<td>25.2</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>7.7</td>
</tr>
<tr>
<td>Split-Half Reliability Estimate</td>
<td>0.76</td>
</tr>
<tr>
<td>Median D-index (using total score as criterion)</td>
<td>23.5</td>
</tr>
</tbody>
</table>

*Ennis (39)

Wisconsin Inventory of Science Processes

The Wisconsin Inventory of Science Processes was developed by the Scientific Literacy Research Center at the University of Wisconsin. Its purpose is to inventory knowledge of the processes of science possessed by high school students. The development of another form (Form C) of the instrument was described by Welch and Pella (112, pp. 64-68). A list of elements of scientific processes was obtained from books by Beveridge, Conant, Kemeny, Lachman, Nash and Wilson. Only those elements which appeared in three or more of the six reference books were presented for validity judgment to fourteen research scientists. Revisions based on their suggestions were incorporated into the descriptive outline which became the basis for developing the instrument, Form C.

The mean, based on twelfth grade students only, was
54.2. The mean for teachers was 66.9. The reliability for twelfth grade students was 0.82 and for teachers was 0.827 (84).

Information on Form C is contained in the reference previously cited, Welch and Pella (112, pp. 64-68).

This investigator used Test 671 in the Scientific Literacy Series with permission of the Scientific Literacy Research Center, the University of Wisconsin. This form of the test, copyrighted in 1967, consists of 93 statements for which three choices of answers were available: (A) if the statement is always or nearly always accurate; (I) if the statement is always or nearly always inaccurate; and (D) if the statement is not understood or the student does not know. The 93 statements pertaining to major assumptions, objectives, activities, and products of science are so distributed that each of 31 specific attributes in the following outline (115) is covered by three questions.

I. The Scientist Assumes That:
   A. The Universe and Natural Phenomena Are:
      1. Real 2. Intelligible
      3. Consistent 4. Causal
   B. The Products (Results) of Science Are:
      1. Amoral 2. Repeatable
      3. Parsimonious 4. Tentative
      5. Probabilistic
C. He and His Fellow Scientists Are:

1. Objective
2. Anti-authoritarian
3. Motivated by a desire to understand the physical universe

II. The Actions or Operations of the Scientist Include:

A. Observations Which Are:

1. Selected
2. Influenced by instrumentation and past experience
3. Described and recorded accurately
4. Sometimes unexpected

B. Measurement Which:

1. Has inherent error
2. Is a method of quantitative expression

C. Classification, Which Is:

1. An invention of man used for the organization of data
2. Based on observed relations between variables and hence has inherent weaknesses

D. Experiment, Which Is:

1. Used to test hypotheses and theories and expose new areas to empirical exploration
2. The identification, manipulation and control of variables

E. Communication, Which Is:

1. A method of recording scientific information and adding it to the cumulative fund of "knowledge"
2. An academic obligation which makes scientific information available for independent confirmation and verification
F. Prediction, Which Is Achieved By the Utilization of:
1. Inductive logic
2. Deductive logic
3. A multiplicity of techniques and procedures

G. The Formulation of:
1. Hypotheses  2. Theories
3. Laws        4. Models

In WISP, section I—Assumptions of the Scientist, the only topic not mentioned or strongly alluded to in the writings cited in Chapter II was (B-1) the Results of Science Are Amoral. In WISP, section II—The Actions or Operations of the Scientist, the only topics not mentioned directly in the cited writings were (B) Measurement, and (F) Prediction.

THE EXPERIMENTAL DESIGN

Campbell and Stanley (47, p. 175) listed and described twelve factors which can jeopardize the validity of various experimental designs. They defined internal validity as "the basic minimum without which any experiment is uninterpretable: Did in fact the experimental treatments make a difference in this specific experimental instance?"

External validity raises the question of generalizability: "To what populations, settings, treatment variables, and measurement variables can this effect be gener-
The investigator sought for this study a design that would minimize the threats to internal and external validity while still being practicable for his situation. The nonequivalent control group design, Campbell and Stanley's (47, p. 217) Design 10, was selected. This used a control group, the physical science students of 1968-69, and an experimental group, the physical science students of 1969-70. Both groups were pretested and posttested. The design could be diagrammed thus:

\[
\begin{array}{cccc}
\text{O} & \text{O} & \text{X} & \text{O} \\
\text{O} & \text{X} & \text{O}
\end{array}
\]

where O represents an observation or measurement and X represents exposure of a group to experimental conditions. The Os and Xs in a horizontal row are applied to the same specific persons.

Two other aspects of the study had to be taken into consideration. The first factor was differences that might exist between the control group of 1968-69 and the experimental group of 1969-70. The scores on the College Entrance Examination Board's Scholastic Aptitude Tests (Verbal and Mathematical), their Achievement Tests (English and Mathematics) and the pretest scores on the criterion instruments were used to provide statistical control for these differences. The control group was
taught by the investigator and one other physical science faculty member. The experimental group, because of its larger size, required another physical science instructor in addition to the two that taught the control group. Analysis of variance was used to determine if there were differences due to instructors.

COLLECTION OF THE DATA

Scholastic Aptitude Test scores, both verbal and mathematical for all subjects, were obtained from the registrar's office at Gorham State College. Also, if the student had taken the English and/or Mathematics Achievement Tests of the College Entrance Examination Board, these scores were recorded. All of these tests were administered and scored prior to the students being admitted to the College.

The control group was pretested with the Test on Understanding Science, Form W, and the Cornell Critical Thinking Test, Level Z during the first week of October, 1968. Posttesting with the same instruments took place the third week in May, 1969. The Wisconsin Inventory of Science Processes, Test 671, was administered to this group for the first time in May, 1969.

The experimental group was pretested with TOUS,
Form W; CCTT, Level Z; and WISP, Test 671 in the last week of September, 1969. Posttesting with the same three instruments took place May 11 to May 17, 1970.

The tests were administered by the investigator with the assistance of two other physical science faculty members. All of these tests for both groups were corrected and scored by the same especially instructed secretary and were randomly checked by the investigator.

Revision of the physical science course was based on the analysis of the data for the 1968-69 group.

**PROCEDURE FOR ANALYSIS-OF THE DATA**

The null hypotheses to be tested were grouped into three major categories dealing with: (1) pre- to posttest gains, (2) differences in outcomes between the control and experimental groups, and (3) differences between instructors in producing changes.

The three sub-problems, as stated in Chapter I, are repeated here.

I. Do students, after taking a year of physical science, make significant gains as shown by their pretest and posttest scores on the criterion instruments?

The six null hypotheses (one for each part of TOUS, one for TOUS-total, one for CCTT, and
The six null hypotheses were tested for the control group (1968-69) and the experimental group (1969-70) separately by an Analysis of Variance for Factorial Design with Unequal Cell Frequency, Least Square Solution.

II. Were there significant differences in learning outcomes as measured by the criterion instruments between the control physical science group and the experimental physical science group?

The six null hypotheses (one for each part of TOUS, one for TOUS-total, one for CCTT, and one for WISP) were of the general form, "There is no significant difference in outcomes in (understanding or ability) as measured by (instrument) between the experimental group and the control group."

First, a BMD02D Correlation with Transgeneration Analysis was used to determine the coefficients of correlation among all of the variables involved, in order to identify the factors related to achievement.

Second, EMDO4V Covariance Analysis was used
to control the variables of scholastic aptitude and prior knowledge, understanding, or ability as determined by the particular instance.

III. Do the teachers influence significantly the changes in learning outcomes as measured by the criterion instruments?

The six null hypotheses (one for each part of TOUS, one for TOUS-total, one for CCTT, and one for WISP) were of the general form, "There is no significant difference between teachers in producing changes in (understanding or ability) as measured by (instrument)."

First, a BMD02D Correlation with Transgeneration Analysis was used to determine the coefficients of correlation among all of the variables involved, in order to identify the factors related to different instructors teaching the course.

Second, an Analysis of Variance for Factorial Design with Unequal Cell Frequency, Least Square Solution was used to identify variance due to different instructors.

Third, a BMD04V Covariance Analysis was used to control the variables of scholastic aptitude
and prior knowledge, understanding, or ability in determining the significance of students having different instructors.
CHAPTER IV

RESULTS AND INTERPRETATION

This investigation was designed to determine if significant changes in college students' understanding of the nature of science and in critical thinking skills result from modifying a conventional, general education, laboratory, physical science course. The investigation was conducted at Gorham State College of the University of Maine during the academic years of 1968-69 and 1969-70. The control group consisted of all students who completed the physical science course (Science 100-101) in 1968-69. The experimental group consisted of all students who completed the revised course in 1969-70.

Data for the analyses were obtained from records in the registrar's office and from the criterion instruments administered by members of the physical science faculty. The control group was pretested the first week in October, 1968, to provide statistical controls for pretest knowledge, understanding of the nature of science, and critical thinking skills. Posttesting was conducted the third week in May, 1969, to determine student achievement in these areas. The experimental group was pretested in
September, 1969, and posttested in May, 1970. The same criterion instruments were used with both groups.

The collected data were used in three statistical analyses: (1) a BMD02D Correlation with Transgeneration Analysis determined the coefficients of correlation among all of the variable factors for which data were available and provided a correlation matrix for each year of the study; (2) an Analysis of Variance for Factorial Design with Unequal Cell Frequency determined variance attributable to the course and to the different instructors; and (3) a BMD04V Covariance Analysis was used to control the variables of scholastic aptitude and pretest understanding and ability as measured by the criterion instruments.

Data presented and interpreted in this chapter are discussed in four sections:

1. Non-Criterion Variables of the Groups
2. Criterion Variables of the Groups
3. Testing the Null Hypotheses
4. Analysis of Gains and Losses Made by the Experimental Group

**Non-Criterion Variables of the Groups**

The scores on the College Entrance Examination Board's **Scholastic Aptitude Tests**, Verbal and Mathematical, were used to provide statistical control for scholas-
tic aptitude. The scores on the Board's Achievement Tests in English and Mathematics were used to provide statistical control for knowledge in these areas. These tests were administered prior to the students' admission to the College. The scores were obtained from records in the registrar's office. A summary of information for the control and experimental groups is provided in Table 11.

### TABLE 11

**SUMMARY OF SCORES ON SCHOLASTIC APTITUDE TESTS AND ACHIEVEMENT TESTS**

| Tests       | Control Group | | | | | | Experimental Group | | | | | |
|-------------|---------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|             | *N | H | L | Mean | | | *N | H | L | Mean | | | | | |
| SAT Verbal  | 142 | 669 | 309 | 487.514 | | 186 | 650 | 267 | 473.667 | | | | | | |
| SAT Math    | 142 | 752 | 328 | 496.092 | | 186 | 692 | 271 | 490.145 | | | | | | |
| Eng. Ach.   | 127 | 740 | 355 | 515.260 | | 170 | 703 | 363 | 504.229 | | | | | | |
| Math Ach.   | 88 | 727 | 315 | 484.284 | | 132 | 686 | 326 | 488.417 | | | | | | |

*N=number, H=highest, L=lowest

The distribution of scores on the Scholastic Aptitude Tests and the Achievement Tests by score intervals is given in Table 12, page 112, for the control group and in Table 13, page 112, for the experimental group.
### TABLE 12

**DISTRIBUTION OF SCORES ON SCHOLASTIC APTITUDE TESTS AND ACHIEVEMENT TESTS FOR THE CONTROL GROUP**

<table>
<thead>
<tr>
<th>Score Interval</th>
<th>Verbal</th>
<th>Math</th>
<th>English</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%N</td>
<td>N</td>
<td>%N</td>
</tr>
<tr>
<td>750-800</td>
<td>1</td>
<td>0.70</td>
<td>1</td>
<td>0.79</td>
</tr>
<tr>
<td>700-749</td>
<td>4</td>
<td>2.82</td>
<td>3</td>
<td>2.11</td>
</tr>
<tr>
<td>650-699</td>
<td>5</td>
<td>3.52</td>
<td>8</td>
<td>5.63</td>
</tr>
<tr>
<td>600-649</td>
<td>20</td>
<td>14.08</td>
<td>26</td>
<td>18.31</td>
</tr>
<tr>
<td>550-599</td>
<td>28</td>
<td>19.72</td>
<td>26</td>
<td>18.31</td>
</tr>
<tr>
<td>500-549</td>
<td>42</td>
<td>29.58</td>
<td>36</td>
<td>25.35</td>
</tr>
<tr>
<td>450-499</td>
<td>29</td>
<td>20.42</td>
<td>26</td>
<td>18.31</td>
</tr>
<tr>
<td>400-449</td>
<td>11</td>
<td>7.75</td>
<td>11</td>
<td>7.75</td>
</tr>
<tr>
<td>350-399</td>
<td>3</td>
<td>2.11</td>
<td>5</td>
<td>3.52</td>
</tr>
<tr>
<td>300-349</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>250-299</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

### TABLE 13

**DISTRIBUTION OF SCORES ON SCHOLASTIC APTITUDE TESTS AND ACHIEVEMENT TESTS FOR THE EXPERIMENTAL GROUP**

<table>
<thead>
<tr>
<th>Score Interval</th>
<th>Verbal</th>
<th>Math</th>
<th>English</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%N</td>
<td>N</td>
<td>%N</td>
</tr>
<tr>
<td>750-800</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>700-749</td>
<td>3</td>
<td>1.61</td>
<td>5</td>
<td>2.69</td>
</tr>
<tr>
<td>650-699</td>
<td>10</td>
<td>5.38</td>
<td>8</td>
<td>4.30</td>
</tr>
<tr>
<td>600-649</td>
<td>7</td>
<td>3.76</td>
<td>23</td>
<td>12.37</td>
</tr>
<tr>
<td>550-599</td>
<td>41</td>
<td>22.04</td>
<td>46</td>
<td>24.73</td>
</tr>
<tr>
<td>500-549</td>
<td>61</td>
<td>32.80</td>
<td>53</td>
<td>28.49</td>
</tr>
<tr>
<td>450-499</td>
<td>34</td>
<td>18.28</td>
<td>28</td>
<td>15.05</td>
</tr>
<tr>
<td>400-449</td>
<td>25</td>
<td>13.44</td>
<td>17</td>
<td>9.14</td>
</tr>
<tr>
<td>350-399</td>
<td>3</td>
<td>1.61</td>
<td>4</td>
<td>2.15</td>
</tr>
<tr>
<td>300-349</td>
<td>2</td>
<td>1.08</td>
<td>2</td>
<td>1.08</td>
</tr>
<tr>
<td>250-299</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
Criterion Variables of the Groups

The Test on Understanding Science, Form W, was administered as a pre- and posttest to the control group and to the experimental group. After administering the test to the control group, it was decided to delete question 20 from Part I and questions 18 and 47 from Part II because of their questionable relevance. The effects of the deletion on the sub-test means and the total-test means are given in Table 14 for the control group.

### Table 14

**Test on Understanding Science, Effect of Deleting Questions 18, 20, and 47 on Sub-Test and Total-Test Means for the Control Group**

<table>
<thead>
<tr>
<th>Questions</th>
<th>Part I</th>
<th>Part II</th>
<th>Part III</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18, 20, 47 in</td>
<td>10.930</td>
<td>12.313</td>
<td>11.826</td>
<td>35.062</td>
</tr>
<tr>
<td>18, 20, 47 out</td>
<td>10.201</td>
<td>11.014</td>
<td>11.826</td>
<td>33.042</td>
</tr>
<tr>
<td>Difference</td>
<td>0.729</td>
<td>1.299</td>
<td>.</td>
<td>2.020</td>
</tr>
<tr>
<td><strong>Posttest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18, 20, 47 out</td>
<td>10.778</td>
<td>10.875</td>
<td>12.340</td>
<td>33.993</td>
</tr>
<tr>
<td>Difference</td>
<td>0.778</td>
<td>1.465</td>
<td>.</td>
<td>2.243</td>
</tr>
</tbody>
</table>
The means on the same instrument with questions 18, 20, and 47 deleted are given in Table 15, for the experimental group.

### TABLE 15

**TEST ON UNDERSTANDING SCIENCE, SUB-TEST AND TOTAL-TEST MEANS FOR THE EXPERIMENTAL GROUP**

<table>
<thead>
<tr>
<th>Test</th>
<th>Part I</th>
<th>Part II</th>
<th>Part III</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>9.658</td>
<td>10.632</td>
<td>11.263</td>
<td>31.547</td>
</tr>
<tr>
<td>Posttest</td>
<td>10.900</td>
<td>10.837</td>
<td>12.932</td>
<td>34.626</td>
</tr>
</tbody>
</table>

Cooley and Klopfer (23, p. 12) used the TOUS, Form X, which they claim is similar to Form W, in a nation-wide sample of 2980 high school students in October, 1960. They reported a total-test mean score of 32.25 (Table 7, page 96) based on 753 Grade 12 students.

At Gorham State College, the pretest (October, 1968) mean of the control group was 35.062 and the pretest (September, 1969) mean of the experimental group was 31.547. The latter figure is with questions 18, 20, and 47 deleted which, in effect, lowers the mean score slightly over 2.0 as seen in Table 14, page 113. If deletion is compensated for by adding 2.0 to the 31.547 giving 33.547, then both groups of physical science students at Gorham had somewhat higher means on their pretests than the
twelfth grade students in the Cooley and Klopfer study.

Comparison of the control and experimental pretest means at Gorham with Craven's (25, p. 88) freshmen in social science education and science education at Oregon State University reveals somewhat higher means (Table 1, page 72) for both of the latter groups. The pretest mean for the Gorham control group was 35.062 and for the experimental group 33.547 (compensated). The mean for the Oregon freshmen social science education group was 36.00 and for the science group 37.02.

Comparison of the TOUS posttest mean scores of the Gorham State College control and experimental groups with the Kansas State Teachers College total groups (Table 2, p. 73) revealed comparable scores as shown in Table 16, page 116.

A correlation with Transgeneration Analysis (BMD02D program) was used to determine the coefficients of correlation among all of the variables involved, first for the control group and then for the experimental group at Gorham State College. The posttest TOUS total score was found to have a correlation significant at the .01 level of significance with each part of posttest TOUS as seen in Table 17, page 117.
TABLE 16
SUMMARY OF TOUS POSTTEST MEAN SCORES FOR GORHAM STATE COLLEGE AND KANSAS STATE TEACHERS COLLEGE

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Part I</th>
<th>Part II</th>
<th>Part III</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gorham State College</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>144</td>
<td>11.556</td>
<td>12.340</td>
<td>12.340</td>
<td>36.236</td>
</tr>
<tr>
<td>Experimental (Compensated)</td>
<td>190</td>
<td>11.678</td>
<td>12.302</td>
<td>12.932</td>
<td>36.869</td>
</tr>
<tr>
<td><strong>Kansas State Teachers College</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Exp. Per. (Total)</td>
<td>140</td>
<td>11.2</td>
<td>11.8</td>
<td>12.6</td>
<td>35.6</td>
</tr>
<tr>
<td>2nd Exp. Per. (Total)</td>
<td>193</td>
<td>11.3</td>
<td>11.9</td>
<td>13.0</td>
<td>36.2</td>
</tr>
</tbody>
</table>

*Crawford and Backhus (26, p. 9)

The correlation was most significant for TOUS-Part III and least significant for TOUS-Part II when considering only the TOUS instrument. There was also a correlation significant at the .01 level between posttest TOUS-total and the scores on all of the other evaluation instruments except CEEB Achievement Test in Mathematics. The Mathematics Achievement Test score was significant at the .05 level for the experimental group but was not significant even at the .05 level for the control group.

The Cornell Critical Thinking Test, Level Z, was
administered as a pre- and posttest to the control group and the experimental group. The means for both groups on the pre- and posttests are given in Table 18.

**TABLE 17**

CORRELATION COEFFICIENTS FOR POSTTEST TOUS TOTAL SCORES WITH CEEB SCORES AND OTHER POSTTEST SCORES

<table>
<thead>
<tr>
<th>Test</th>
<th>Control Group</th>
<th>Experimental Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOUS-Part I</td>
<td>0.7684</td>
<td>0.7642</td>
</tr>
<tr>
<td>TOUS-Part II</td>
<td>0.6896</td>
<td>0.5715</td>
</tr>
<tr>
<td>TOUS-Part III</td>
<td>0.7918</td>
<td>0.7898</td>
</tr>
<tr>
<td>SAT-Verbal</td>
<td>0.4936</td>
<td>0.5136</td>
</tr>
<tr>
<td>SAT-Math</td>
<td>0.2505</td>
<td>0.2624</td>
</tr>
<tr>
<td>English-Ach.</td>
<td>0.3801</td>
<td>0.4332</td>
</tr>
<tr>
<td>Math-Ach.</td>
<td>0.1433</td>
<td>0.1729</td>
</tr>
<tr>
<td>CCTT</td>
<td>0.2887</td>
<td>0.3098</td>
</tr>
<tr>
<td>WISP</td>
<td>0.2229</td>
<td>0.3501</td>
</tr>
</tbody>
</table>

Degrees of Freedom 142 188
Value of $r$ for .05 0.163 0.143
Value of $r$ for .01 0.214 0.187

**TABLE 18**

CORNELL CRITICAL THINKING TEST, LEVEL Z, MEANS FOR THE CONTROL GROUP AND THE EXPERIMENTAL GROUP

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td>29.604</td>
<td>30.430</td>
</tr>
<tr>
<td>Experimental Group</td>
<td>27.984</td>
<td>29.326</td>
</tr>
</tbody>
</table>
The Correlation with Transgeneration Analysis for the experimental group showed the posttest CCTT score to have a correlation significant at the .01 level with scores on the other evaluation instruments except TOUS-Part II which was significant at the .05 level as seen in Table 19. The control group's posttest CCTT score had a correlation significant at the .01 level with scores on the other instruments except TOUS-Part II and TOUS-Part III, both of which were significant at the .05 level.

TABLE 19
CORRELATION COEFFICIENTS FOR POSTTEST CCTT SCORES WITH CEEB SCORES AND OTHER POSTTEST SCORES

<table>
<thead>
<tr>
<th>Test</th>
<th>Control Group</th>
<th>Experimental Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT-Verbal</td>
<td>0.3008</td>
<td>0.3689</td>
</tr>
<tr>
<td>SAT-Math</td>
<td>0.4241</td>
<td>0.3072</td>
</tr>
<tr>
<td>English-Ach.</td>
<td>0.3428</td>
<td>0.2694</td>
</tr>
<tr>
<td>Math-Ach.</td>
<td>0.2526</td>
<td>0.1934</td>
</tr>
<tr>
<td>TOUS-Part I</td>
<td>0.2714</td>
<td>0.2512</td>
</tr>
<tr>
<td>TOUS-Part II</td>
<td>0.1860</td>
<td>0.1469</td>
</tr>
<tr>
<td>TOUS-Part III</td>
<td>0.1956</td>
<td>0.2525</td>
</tr>
<tr>
<td>TOUS-Total</td>
<td>0.2887</td>
<td>0.3098</td>
</tr>
<tr>
<td>WISP</td>
<td>0.2195</td>
<td>0.2778</td>
</tr>
</tbody>
</table>

Degrees of Freedom 142 188
Value of r for .05 0.163 0.143
Value of r for .01 0.214 0.187

The correlation was most significant with the SAT-Verbal for the experimental group and with the SAT-Math
for the control group. The correlation was least signifi-
cant with TOUS-Part II for both the control and the exper-
imental groups.

The Wisconsin Inventory of Science Processes was ad-
ministered as a pre- and posttest to the experimental
group and as a posttest only to the control group. The
means for both groups are given in Table 20.

TABLE 20

<table>
<thead>
<tr>
<th>WISCONSIN INVENTORY OF SCIENCE PROCESSES, MEANS FOR THE CONTROL GROUP AND THE EXPERIMENTAL GROUP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Control Group</td>
</tr>
<tr>
<td>Experimental Group</td>
</tr>
</tbody>
</table>

A comparison of the means for the experimental group
showed their pretest mean of 60.247 to be considerably
above the mean of 54.2 for twelfth grade students report-
ed by Pella (83). The experimental group's posttest mean
of 65.453 was somewhat below the mean of 66.9 for teach-
ers reported by Pella (83).

The Correlation with Transgeneration Analysis for the
experimental group showed the posttest WISP score to have
a correlation significant at the .01 level with scores on
the other evaluation instruments except for the SAT-Math
and the Math-Achievement tests. Neither of these was significant at the .05 level although the SAT-Math was approaching the .05 level of significance as seen in Table 21.

<table>
<thead>
<tr>
<th>Test</th>
<th>Control Group</th>
<th>Experimental Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT-Verbal</td>
<td>0.2078</td>
<td>0.3292</td>
</tr>
<tr>
<td>SAT-Math</td>
<td>0.1016</td>
<td>0.1413</td>
</tr>
<tr>
<td>English-Ach.</td>
<td>0.2337</td>
<td>0.2752</td>
</tr>
<tr>
<td>Math-Ach.</td>
<td>0.0532</td>
<td>0.0330</td>
</tr>
<tr>
<td>TOUS-Part I</td>
<td>0.2539</td>
<td>0.2491</td>
</tr>
<tr>
<td>TOUS-Part II</td>
<td>0.0345</td>
<td>0.2176</td>
</tr>
<tr>
<td>TOUS-Part III</td>
<td>0.1922</td>
<td>0.2739</td>
</tr>
<tr>
<td>TOUS-Total</td>
<td>0.2229</td>
<td>0.3501</td>
</tr>
<tr>
<td>CCTT</td>
<td>0.2195</td>
<td>0.2776</td>
</tr>
</tbody>
</table>

Degrees of Freedom 142 188
Value of r for .05 0.163 0.143
Value of r for .01 0.214 0.187

The control group's posttest WISP score had a significant correlation at the .01 level with scores on the English-Achievement, TOUS-Part I, TOUS-Total and the CCTT.

The most significant correlation for the experimental group was between the WISP score and the TOUS-Total score. The least significant correlation for the experimental group was between the WISP score and the Math-Achievement score.
Examination of the correlation coefficients for post-test WISP showed that the value of the coefficients was greater for the experimental group than for the control group with all scores except two, the Math-Achievement and the TOUS-Part I.

Summary. Examination of the SAT-Verbal and Math mean scores, Table 11, page 111, reveals that the control group scored higher on both parts of the aptitude test than the experimental group. The control group also had a higher mean score by 11.03 on the English-Achievement test than the experimental group.

Examination of the distribution of SAT-Verbal scores, Tables 12 and 13, page 112, reveals that 20.42% of the control group scored 550 or above while only 10.75% of the experimental group scored 550 or above. The control group had 9.86% score below 400 on the SAT-Verbal while the experimental group had 16.14% score below 400.

On the criterion instruments, the control group scored higher than the experimental group on the pretest TOUS-Total as well as on each part of TOUS. The control group also scored higher than the experimental on the pretest CCTT. No pretest WISP was given to the control group.

Comparison of the control and experimental groups with twelfth grade students reveals both groups having
higher pretest mean scores on the TOUS than the twelfth grade students. The experimental group also had a higher pretest mean score on the WISP than twelfth grade students. The control group did not take the pretest WISP.

Comparison of the posttest TOUS scores, Table 16, page 116, of the control and experimental groups with physical science students taking the same posttest at Kansas State Teachers College shows comparable mean scores.

Results of the Correlation with Transgeneration Analysis show correlations significant at the .01 level between posttest TOUS-Total score and the scores on all other evaluation instruments except Mathematics-Achievement test. The posttest CCTT score showed correlations significant at the .01 level with the scores on all other evaluation instruments except TOUS-Part II which was significant at the .05 level. The posttest WISP score showed correlations significant at the .01 level with the scores on all other evaluation instruments except for the SAT-Math and the Math-Achievement tests. The correlation coefficients for posttest WISP have higher values for the experimental group than for the control group except for Math-Achievement and TOUS-Part I.

**Testing the Null Hypotheses**

The null hypotheses to be treated were grouped into
three major categories (listed as sub-problems with Roman Numerals) dealing with: (I) pre- to posttest gains after a year’s study of physical science, (II) differences in learning outcomes between the control and experimental groups, and (III) differences among instructors in producing changes.

The three sub-problems, as stated in Chapters I and III, are repeated with appropriate analysis and interpretation of each of the null hypotheses.

Sub-Problem I

Do students, after a year’s study of physical science, make significant gains as shown by their pretest and posttest scores on the criterion instruments?

The following null hypotheses apply to each group separately, the control group (1968-69) and the experimental group (1969-70).

A. There is no significant gain pre- to posttest in understanding the scientific enterprise as measured by TOUS-Part I.

Table 22, page 124, contains the statistical results of the analysis of variance for TOUS-Part I.
TABLE 22

ANALYSIS OF VARIANCE FOR TOUS-PART I
CONTROL AND EXPERIMENTAL GROUPS

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Level of Signif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A, Instr.</td>
<td>1.5607</td>
<td>1</td>
<td>1.5607</td>
<td>0.2627</td>
<td></td>
</tr>
<tr>
<td>B, Course</td>
<td>23.9201</td>
<td>1</td>
<td>23.9201</td>
<td>4.0266</td>
<td>.05</td>
</tr>
<tr>
<td>AB</td>
<td>1.3611</td>
<td>1</td>
<td>1.3611</td>
<td>0.2291</td>
<td></td>
</tr>
<tr>
<td>Error, S_y</td>
<td>1687.1289</td>
<td>284</td>
<td>5.9406</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Experimental Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A, Instr.</td>
<td>10.1608</td>
<td>2</td>
<td>5.0804</td>
<td>1.0362</td>
<td></td>
</tr>
<tr>
<td>B, Course</td>
<td>149.0631</td>
<td>1</td>
<td>149.0631</td>
<td>30.4035</td>
<td>.001</td>
</tr>
<tr>
<td>AB</td>
<td>2.6595</td>
<td>2</td>
<td>1.3298</td>
<td>0.2712</td>
<td></td>
</tr>
<tr>
<td>Error, S_y</td>
<td>1833.6563</td>
<td>374</td>
<td>4.9028</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

1. For the control group.

To be significant at the .05 level of significance, the F-value with 1 and 284 degrees of freedom must equal or exceed 3.88. A value of 4.0266, Table 22, was computed for this part resulting in the null hypothesis being rejected. A value of 4.0266 for F indicated that the control course did make a significant contribution to students' understanding of the scientific enterprise.
2. For the experimental group.

To be significant at the .001 level, the F-value, with 1 and an infinite number of degrees of freedom, must be 10.83 or greater. A value of 30.4035, Table 22, page 124, was computed for this part resulting in the null hypothesis being rejected. This large F-value of 30.4035 clearly indicated that the experimental course made a highly significant contribution to students' understanding of the scientific enterprise.

B. There is no significant gain pre- to posttest in understanding of scientists as measured by TOUS-Part II.

Table 23, page 126, contains the statistical results of the analysis of variance for TOUS-Part II.

1. For the control group.

To be significant at the .10 level of significance, the F-value, with 1 and an infinite number of degrees of freedom, must equal or exceed 2.71. A value of 0.3103, Table 23, page 126, was computed for this part, resulting in the null hypothesis being accepted. A value of 0.3103 for F indicated that the control course did not make a contribution significant at the .10 level to students' understanding of scientists.
TABLE 23
ANALYSIS OF VARIANCE FOR TOUS-PART II
CONTROL AND EXPERIMENTAL GROUPS

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Level of Signif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A, Instr.</td>
<td>1.4592</td>
<td>1</td>
<td>1.4592</td>
<td>0.3260</td>
<td></td>
</tr>
<tr>
<td>B, Course</td>
<td>1.3889</td>
<td>1</td>
<td>1.3889</td>
<td>0.3103</td>
<td>Not .10</td>
</tr>
<tr>
<td>AB</td>
<td>0.9119</td>
<td>1</td>
<td>0.9119</td>
<td>0.2037</td>
<td></td>
</tr>
<tr>
<td>Error, $S_e$</td>
<td>1271.3516</td>
<td>284</td>
<td>4.4766</td>
<td>4.4766</td>
<td></td>
</tr>
</tbody>
</table>

| Experimental Group  |                |    |             |         |                  |
| A, Instr.           | 11.6268        | 2  | 5.8134      | 1.5642  |                  |
| B, Course           | 4.4237         | 1  | 4.4237      | 1.1903  | Not .10         |
| AB                  | 3.0958         | 2  | 1.5479      | 0.4165  |                  |
| Error, $S_y$        | 1389.9414      | 374| 3.7164      | 3.7164  |                  |

2. For the experimental group.

To be significant at the .10 level, the $F$-value, with 1 and an infinite number of degrees of freedom, must equal or exceed 2.71. A value of 1.1903, Table 23, was computed for this part, resulting in the null hypothesis being accepted. A value of 1.1903 for $F$ indicated that while the experimental group did not make a contribution significant at the .10 level, it did make a more important contribution to understanding scien-
tists than the control course did.

C. There is no significant gain pre- to posttest in understanding the methods and aims of science as measured by TOUS-Part III.

Table 24 contains the statistical results of the analysis of variance for TOUS-Part III.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Level of Signif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A, Instr.</td>
<td>0.0178</td>
<td>1</td>
<td>0.0178</td>
<td>0.0023</td>
<td></td>
</tr>
<tr>
<td>B, Course</td>
<td>19.0139</td>
<td>1</td>
<td>19.0139</td>
<td>2.5058</td>
<td>Not .10</td>
</tr>
<tr>
<td>AB</td>
<td>-0.0061</td>
<td>1</td>
<td>-0.0061</td>
<td>-0.0068</td>
<td></td>
</tr>
<tr>
<td>Error, S_y</td>
<td>2154.9766</td>
<td>284</td>
<td>7.5879</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Experimental Group

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Level of Signif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, Instr.</td>
<td>43.4934</td>
<td>2</td>
<td>21.7467</td>
<td>3.1637</td>
<td>.001</td>
</tr>
<tr>
<td>B, Course</td>
<td>262.7786</td>
<td>1</td>
<td>262.7786</td>
<td>38.2292</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>17.5300</td>
<td>2</td>
<td>8.7650</td>
<td>1.2751</td>
<td></td>
</tr>
<tr>
<td>Error, S_y</td>
<td>2570.7891</td>
<td>374</td>
<td>6.8738</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

1. For the control group.

To be significant at the .10 level, the F-value, with 1 and an infinite number of degrees of freedom, must equal or exceed 2.71. A value
of 2.5058, Table 24, page 127, was computed for this part, resulting in the null hypothesis being accepted. A value of 2.5058 for F indicated that the control group did not make a contribution significant at the .10 level to students' understanding of the methods and aims of science.

2. For the experimental group.

To be significant at the .001 level, the F-value, with 1 and an infinite number of degrees of freedom, must be 10.83 or greater. A value of 38.2292, Table 24, page 127, was computed for this part resulting in the null hypothesis being rejected. This large F-value of 38.2292 clearly indicated that the experimental course made an extremely significant contribution to students' understanding of the methods and aims of science. The most significant gain for any of the three parts of TOUS was made on Part III.

D. There is no significant gain pre- to posttest in understanding of the nature of science as measured by TOUS-Total.

Table 25, page 129, contains the statistical results of the analysis of variance for TOUS-Total.
## TABLE 25
ANALYSIS OF VARIANCE FOR TOUS-TOTAL
CONTROL AND EXPERIMENTAL GROUPS

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Level of Signif.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A, Instr.</td>
<td>5.4826</td>
<td>1</td>
<td>5.4826</td>
<td>0.1794</td>
<td></td>
</tr>
<tr>
<td>B, Course</td>
<td>65.1701</td>
<td>1</td>
<td>65.1701</td>
<td>2.1326</td>
<td>Not .10</td>
</tr>
<tr>
<td>AB</td>
<td>4.6424</td>
<td>1</td>
<td>4.6424</td>
<td>0.1519</td>
<td></td>
</tr>
<tr>
<td>Error, S&lt;sub&gt;y&lt;/sub&gt;</td>
<td>8678.6875</td>
<td>284</td>
<td>30.5587</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Experimental Group** | | | | | |
| A, Instr.              | 74.0859       | 2  | 37.0430     | 1.6617|                  |
| B, Course              | 922.2734      | 1  | 922.2734    | 41.3725| .001            |
| AB                     | 32.3516       | 2  | 16.1758     | 0.7256|                  |
| Error, S<sub>y</sub>  | 8337.1875     | 374| 22.2919     |       |                  |

1. For the control group.

To be significant at the .10 level of significance, the F-value, with 1 and an infinite number of degrees of freedom, must equal or exceed 2.71. A value of 2.1326, Table 25, was computed for this part resulting in the null hypothesis being accepted. A value of 2.1326 for F indicated that the control course did not make a contribution significant at the .10 level to the students' understanding of the nature of science.
as measured by TOUS-Total.

2. For the experimental group.

To be significant at the .001 level, the F-value, with 1 and an infinite number of degrees of freedom, must be 10.83 or greater. A value of 41.3725, Table 25, page 129, was computed for this part resulting in the null hypothesis being rejected. This large F-value of 41.3725 clearly indicated that the experimental course made an extremely significant contribution to students’ understanding of the nature of science as measured by TOUS-Total.

E. There is no significant gain pre- to posttest in critical thinking skills as measured by the CCTT.

Table 26, page 131, contains the statistical results of the analysis of variance for CCTT.

1. For the control group.

To be significant at the .10 level of significance, the F-value for 1 and an infinite number of degrees of freedom, must equal or exceed 2.71. A value of 2.3110, Table 26, page 131, was computed for this part resulting in the null hypothesis being accepted. A value of 2.3110 for F indicated that the control course did not make a contribution significant at the .10 level to im-
proving the critical thinking skills of the students.

TABLE 26
ANALYSIS OF VARIANCE FOR CCTT
CONTROL AND EXPERIMENTAL GROUPS

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>( F )</th>
<th>Signif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A, Instr.</td>
<td>16.3576</td>
<td>1</td>
<td>16.3576</td>
<td>0.7688</td>
<td></td>
</tr>
<tr>
<td>B, Course</td>
<td>49.1701</td>
<td>1</td>
<td>49.1701</td>
<td>2.3110</td>
<td>Not .10</td>
</tr>
<tr>
<td>AB</td>
<td>4.8924</td>
<td>1</td>
<td>4.8924</td>
<td>0.2299</td>
<td></td>
</tr>
<tr>
<td>Error, S_y</td>
<td>6042.6250</td>
<td>284</td>
<td>21.2768</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

| Experimental Group  |                |     |             |       |         |
| A, Instr.           | 4.7434         | 2   | 2.3717      | 0.1143|         |
| B, Course           | 171.1184       | 1   | 171.1184    | 8.2445| .01     |
| AB                  | 11.5891        | 2   | 5.7846      | 0.2787|         |
| Error, S_y          | 7762.5000      | 374 | 20.7553     | ...   |         |

2. For the experimental group.

To be significant at the .01 level, the \( F \)-value for 1 and 374 degrees of freedom must be 6.70 or greater. A value of 8.2445, Table 26, was computed for this part resulting in the null hypothesis being rejected. A value of 8.2445 for \( F \) indicated that the experimental course did make a significant contribution to improving the
critical thinking skills of the students.

F. There is no significant gain pre- to posttest in knowledge of the processes of science as measured by the WISP.

Table 27 contains the statistical results of the analysis of variance for WISP.

TABLE 27
ANALYSIS OF VARIANCE FOR WISP
EXPERIMENTAL GROUP

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Level of Signif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, Instr.</td>
<td>558.0015</td>
<td>2</td>
<td>279.0007</td>
<td>6.6101</td>
<td>1</td>
</tr>
<tr>
<td>B, Course</td>
<td>2574.0015</td>
<td>1</td>
<td>2574.0015</td>
<td>60.9829</td>
<td>.001</td>
</tr>
<tr>
<td>AB</td>
<td>220.9985</td>
<td>2</td>
<td>110.4993</td>
<td>2.6179</td>
<td></td>
</tr>
<tr>
<td>Error, S_y</td>
<td>15786.0000</td>
<td>374</td>
<td>42.2085</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. For the control group.

WISP was not given as a pretest so no calculations were made.

2. For the experimental group.

To be significant at the .001 level, the F-value for 1 and an infinite number of degrees of freedom must equal or exceed 10.83. A value of 60.9829, Table 27, was computed for this part resulting in the null hypothesis being rejected. This very large F-value of 60.9829 clearly in-
dicated that the experimental course made an extremely significant contribution to the students' understanding of the processes of science as measured by the WISP. This value of $F$ was the largest recorded on any of the criterion instruments used.

**Summary for Sub-Problem I.** The values obtained by the analysis of variance, using pretest and posttest scores on the criterion instruments for the control group, resulted in the null hypotheses being accepted for the **TOUS-Parts II and III, TOUS-Total, and the CCTT.** The $F$-values were not significant at the .10 level of significance for each one. The null hypothesis was rejected for the control group for **TOUS-Part I** which had an $F$-value significant at the .05 level. No calculations were possible for **WISP** because no pretest was given.

Results of the analysis of variance for the experimental group gave $F$-values that were significant at the .001 level of significance for the **TOUS-Parts I and III, TOUS-Total, and WISP.** The **CCTT** had an $F$-value significant at the .01 level. The null hypotheses were rejected for all of these tests. The values of $F$ indicated the experimental course made a very important contribution to the students' gains on these tests. The only null hypothesis
accepted for the experimental course was for TOUS-Part II where the F-value was not significant at the .10 level.

Sub-Problem II

Were there significant differences in learning outcomes as measured by the criterion instruments between the control group and the experimental group?

The following null hypotheses were tested.

A. There is no significant difference in learning outcomes in understanding the scientific enterprise as measured by TOUS-Part I between the experimental group and the control group.

The results of the analysis of covariance comparing the experimental group with the control group on TOUS-Part I and using pretest scores on TOUS-Part I, SAT-Verbal scores, and English Achievement scores as covariates are found in Table 28, page 135.

These covariates were selected to compensate for differences between the two groups in: (1) scholastic aptitude as measured by SAT-Verbal, (2) English Achievement, and (3) understanding of the scientific enterprise brought to the course as measured by the pretest TOUS-Part I.

These factors were also selected on the basis of their relatively high correlation coefficients.
## TABLE 28

**ANALYSIS OF COVARIANCE FOR TOUS-PART I**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>YY</th>
<th>Sum-Squares (Due)</th>
<th>Sum-Squares (About)</th>
<th>df</th>
<th>Mean-Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (Between)</td>
<td>1</td>
<td>1.2227</td>
<td>. .</td>
<td>. .</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Error (Within)</td>
<td>332</td>
<td>1723.9922</td>
<td>501.7852</td>
<td>1222.2070</td>
<td>329</td>
<td>3.7149</td>
<td></td>
</tr>
<tr>
<td>Treatment + Error</td>
<td>333</td>
<td>1725.2148</td>
<td>486.6995</td>
<td>1238.5154</td>
<td>330</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16.3083</td>
</tr>
</tbody>
</table>

**For testing adjusted treatment means**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum-Squares</th>
<th>df</th>
<th>Mean-Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference</td>
<td></td>
<td></td>
<td>1</td>
<td>16.3083</td>
<td>4.390</td>
</tr>
</tbody>
</table>

## ADJUSTED MEANS

<table>
<thead>
<tr>
<th>Treatment No.</th>
<th>Treatment Mean</th>
<th>Adjusted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Experimental)</td>
<td>10.9000</td>
<td>11.0414</td>
</tr>
<tr>
<td>2 (Control)</td>
<td>10.7778</td>
<td>10.5912</td>
</tr>
</tbody>
</table>
To be significant at the .05 level, the F-value, with 1 and 329 degrees of freedom, must be 3.87 or greater. Examination of Table 28, page 135, shows an F-value of 4.390; therefore, the null hypothesis was rejected because there was a significant difference in learning outcomes in favor of the experimental group.

B. There is no significant difference in outcomes of understanding of scientists as measured by TOUS-Part II between the experimental group and the control group.

The null hypothesis was accepted because under Sub-Problem I, null hypothesis B, page 125, regarding no significant pre- to posttest gains was accepted for both the control and experimental groups. The analysis of variance for each group gave F-values that were not significant even at the .10 level of significance; hence, covariance analysis was not used for TOUS-Part II.

C. There is no significant difference in outcomes in understanding the methods and aims of science as measured by TOUS-Part III between the experimental group and the control group.

The results of the analysis of covariance comparing the experimental group with the control group
on TOUS-Part III and using pretest scores on TOUS-Part III, SAT-Verbal, and English-Achievement scores as covariates are found in Table 29, page 138.

These covariates were selected to compensate for differences between the two groups in: (1) scholastic aptitude as measured by SAT-Verbal, (2) English-Achievement, and (3) understanding of the methods and aims of science brought to the course as measured by the pretest TOUS-Part III.

These factors were also selected on the basis of their relatively high correlation coefficients. To be significant at the .001 level of significance, the F-value, with 1 and an infinite number of degrees of freedom, must equal or exceed 10.83. No values were given for 1 and 330 degrees of freedom; values (96, p. 915) went from 120 to infinity directly. Examination of Table 29, page 138, shows an F-value of 11.512, so the null hypothesis was rejected. There was a significant difference in outcomes by the two groups, with the experimental group being definitely favored.

D. There is no significant difference in outcomes in understanding the nature of science as measured by TOUS-Total between the experimental group and the con-
## TABLE 29
### ANALYSIS OF COVARIANCE FOR TOUS-PART III

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>YY</th>
<th>Sum-Squares (Due)</th>
<th>Sum-Squares (About)</th>
<th>df</th>
<th>Mean-Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (Between)</td>
<td>1</td>
<td>28.1367</td>
<td>. .</td>
<td>. .</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (Within)</td>
<td>332</td>
<td>2599.2969</td>
<td>612.7407</td>
<td>1986.5562</td>
<td>329</td>
<td>6.0382</td>
<td></td>
</tr>
<tr>
<td>Treatment + Error (Total)</td>
<td>333</td>
<td>2627.4336</td>
<td>571.3640</td>
<td>2056.0696</td>
<td>330</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>. For testing adjusted treatment means</td>
<td>69.5134</td>
<td>1</td>
<td>69.5134</td>
<td>11.512</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### ADJUSTED MEANS

<table>
<thead>
<tr>
<th>Treatment No.</th>
<th>Treatment Mean</th>
<th>Adjusted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Experimental)</td>
<td>12.9263</td>
<td>13.0739</td>
</tr>
<tr>
<td>2 (Control)</td>
<td>12.3403</td>
<td>12.1455</td>
</tr>
</tbody>
</table>
The results of an analysis of covariance comparing the experimental group with the control group on TOUS-Total and using pretest scores on TOUS-Total, SAT-Verbal, and SAT-Math as covariates are found in Table 30, page 140.

These covariates were selected to compensate for differences between the two groups in: (1) scholastic aptitude as measured by SAT-Verbal, (2) SAT-Math, and (3) understanding the nature of science brought to the course as measured by the pretest TOUS-Total.

These factors were also selected on the basis of their relatively high correlation coefficients. To be significant at the .001 level of significance, the F-value, with 1 and an infinite number of degrees of freedom, must be 10.83 or greater. No values were given for 1 and 330 degrees of freedom; values (96, p. 915) went from 120 to infinity directly. Examination of Table 30, page 140, shows an F-value of 12.366, so the null hypothesis was rejected. There was a significant difference in outcomes by the two groups with the experimental group being favored.

E. There is no significant difference in outcomes in critical thinking skills as measured by the CCTT between the experimental and the control groups.
### TABLE 30

**ANALYSIS OF COVARIANCE FOR TOUS-TOTAL**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>YY</th>
<th>Sum-Squares (Due)</th>
<th>Sum-Squares (About)</th>
<th>df</th>
<th>Mean-Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (Between)</td>
<td>1</td>
<td>36.7500</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
</tr>
<tr>
<td>Error (Within)</td>
<td>332</td>
<td>8917.5000</td>
<td>3384.3398</td>
<td>5533.1602</td>
<td>329</td>
<td>16.8181</td>
<td></td>
</tr>
<tr>
<td>Treatment + Error (Total)</td>
<td>333</td>
<td>8954.2500</td>
<td>3213.1096</td>
<td>5741.1367</td>
<td>330</td>
<td>. .</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>331</td>
<td>1207.9766</td>
<td>12.366</td>
</tr>
</tbody>
</table>

**ADJUSTED MEANS**

<table>
<thead>
<tr>
<th>Treatment No.</th>
<th>Treatment Mean</th>
<th>Adjusted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Experimental)</td>
<td>34.6631</td>
<td>35.0692</td>
</tr>
<tr>
<td>2 (Control)</td>
<td>33.9930</td>
<td>33.4572</td>
</tr>
</tbody>
</table>
The results of an analysis of covariance comparing the experimental group with the control group on the CCTT and using pretest scores on CCTT, SAT-Verbal, and SAT-Math as covariates are found in Table 31, page 142.

These covariates were selected to compensate for differences between the two groups in: (1) scholastic aptitude as measured by SAT-Verbal, (2) SAT-Math, and (3) critical thinking skills brought to the course as measured by the pretest CCTT. These factors were also selected on the basis of their relatively high correlation coefficients.

To be significant at the .10 level of significance, the F-value, with 1 and an infinite number of degrees of freedom, must be 2.71 or greater. No values were given for 1 and 330 degrees of freedom; values (96, p. 910) went from 120 to infinity directly. Examination of Table 31, page 142, shows an F-value of 0.407; therefore, the null hypothesis was accepted. The results of the covariance analysis differ drastically from the results obtained in the analysis of variance (pp. 130-131). In the latter case, for the experimental group, an F-value of 8.2445 was obtained which was significant at the .01 level. Using the analysis of covariance, when adjustments are made for
### TABLE 31

**ANALYSIS OF COVARIANCE FOR CCTT**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS (Due)</th>
<th>SS (About)</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (Between)</td>
<td>1</td>
<td>99.8125</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (Within)</td>
<td>332</td>
<td>7099.1875</td>
<td>2169.0603</td>
<td>329</td>
<td>4930.1250</td>
<td>14.9852</td>
</tr>
<tr>
<td>Treatment + Error (Total)</td>
<td>333</td>
<td>7199.0000</td>
<td>2262.7715</td>
<td>330</td>
<td>4936.2266</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>6.1016</td>
<td></td>
<td>1</td>
<td>6.1016</td>
<td>0.407</td>
</tr>
</tbody>
</table>

**ADJUSTED MEANS**

<table>
<thead>
<tr>
<th>Treatment No.</th>
<th>Treatment Mean</th>
<th>Adjusted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Experimental)</td>
<td>29.3263</td>
<td>29.6819</td>
</tr>
<tr>
<td>2 (Control)</td>
<td>30.4305</td>
<td>29.9613</td>
</tr>
</tbody>
</table>
the covariates used, the F-value is not significant at the .10 level, showing that the groups did not differ significantly in outcomes on CCTT.

F. There is no significant difference in outcomes in knowledge of the processes of science as measured by the WISP between the experimental group and the control group.

This hypothesis could not be tested by use of the analysis of covariance in which one of the covariates was the pretest scores because the WISP was not used as a pretest with the control group.

Summary for Sub-Problem II. Analysis of covariance was used to determine if significant differences in learning outcomes existed between the control and experimental groups when adjustments were made for differences in scholastic ability, achievement, and knowledge or skill brought to the course at the beginning of the year. In the covariance analyses, the three covariates used were: (1) the SAT-Verbal means, (2) the pretest means for the specific instrument or part undergoing analysis, and (3) either the English-Achievement means (TOUS-Parts I and III) or the SAT-Math (TOUS-Total and CCTT).

Results of the covariance analyses provided the basis for rejecting the null hypotheses for TOUS-Parts I,
III, and Total. In these three instances, there was a significant difference in learning outcomes between the control and the experimental groups with the latter being favored in each case. The results of the covariance analysis for CCTT necessitated the null hypothesis being accepted. The low F-value of 0.407 indicated no significant difference in learning outcomes between the two groups.

Analysis of covariance was not used with TOUS-Part II because the analysis of variance gave F-values that were not significant even at the .10 level. Analysis of covariance was not used with the WISP because no pretest was given to the control group.

Sub-Problem III

Do the science instructors influence significantly the changes in learning outcomes measured by the criterion instruments?

The following null hypotheses were tested for the control group and for the experimental group. The six null hypotheses are stated and the results are summarized in Table 32, page 146.

A. There is no significant difference among instructors in producing changes in understanding the scientific enterprise as measured by TOUS-Part I.
B. There is no significant difference among instructors in producing changes in understanding of scientists as measured by TOUS-Part II.

C. There is no significant difference among instructors in producing changes in understanding the methods and aims of science as measured by TOUS-Part III.

D. There is no significant difference among instructors in producing changes in understanding the nature of science as measured by TOUS-Total.

E. There is no significant difference among instructors in producing changes in improving the critical thinking skills of students as measured by the CCTT.

F. There is no significant difference among instructors in producing changes in understanding of the processes of science as measured by the WISP.

These six null hypotheses were tested by the Analysis of Variance for Factorial Design with Unequal Cell Frequency, Least Square Solution. In Tables 22 to 27 inclusive, under the Source of Variation, the variation due to the different instructors is the A factor in each analysis. The results are summarized in Table 32, page 146.

For the control group, the first five null hypotheses were accepted. The sixth could not be tested because the WISP pretest was not given.
For the experimental group, null hypotheses 1, 2, 4, and 5 were accepted. Null hypotheses 3 and 6 were rejected.

TABLE 32

ANALYSIS OF VARIANCE FOR DIFFERENCES AMONG INSTRUCTORS IN PRODUCING CHANGES

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Experimental Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of instructors</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>1 and 284</td>
<td>2 and 374</td>
</tr>
<tr>
<td>Minimum F-value, .10 level</td>
<td>2.71</td>
<td>2.30</td>
</tr>
<tr>
<td>Minimum F-value, .05 level</td>
<td>3.88</td>
<td>3.02</td>
</tr>
<tr>
<td>Minimum F-value, .01 level</td>
<td>6.73</td>
<td>4.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TOUS-I</td>
<td>22</td>
<td>124</td>
<td>0.2627</td>
<td>1.0362</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>TOUS-II</td>
<td>23</td>
<td>126</td>
<td>0.3260</td>
<td>1.5642</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>TOUS-III</td>
<td>24</td>
<td>127</td>
<td>0.0023</td>
<td>3.1637</td>
<td>.05</td>
</tr>
<tr>
<td>4</td>
<td>TOUS-Total</td>
<td>25</td>
<td>129</td>
<td>0.1794</td>
<td>1.6617</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CCTT</td>
<td>26</td>
<td>131</td>
<td>0.7688</td>
<td>0.1143</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>WISP</td>
<td>27</td>
<td>132</td>
<td>No pretest</td>
<td>6.6101</td>
<td>.01</td>
</tr>
</tbody>
</table>

Summary for Sub-Problem III. For the control group, the smallest difference attributable to differences in teachers occurred with the TOUS-Part III, methods and aims of science. The greatest difference occurred with the CCTT.

For the experimental group, the smallest difference
attributable to differences in teachers occurred with the CCTT, while the greatest difference was with the WISP which was significant at the .01 level of significance. The F-values for the experimental group were greater on all parts of the TOUS test and on TOUS-Total than for the control group.

The two instructors for the control group were very much alike in their approach to physical science and in their methodology of teaching. The additional instructor for the experimental group differed from the other two in philosophy and methodology.

Analysis of Gains and Losses Made by the Experimental Group

Analysis of the TOUS

Chi square analysis of pretest and posttest scores for each question on the TOUS was used to select the 15 questions, 25% of 60 questions on the test, which showed the greatest gains made by the experimental group. The level of significance on these questions ranged from .001 to .20. Table 33, page 148, shows the results of the analysis using only those questions which were significant at the .10 level or greater. Three questions significant at the .20 level were not included in the detailed analysis. The three questions showing the greatest gains were
### TABLE 33

**ANALYSIS OF TOUS QUESTIONS REGISTERING GREATEST GAINS BY THE EXPERIMENTAL GROUP**

<table>
<thead>
<tr>
<th>Quest. No.</th>
<th>Chi Square</th>
<th>Level of Signif.</th>
<th>Section of TOUS</th>
<th>Freq.</th>
<th>Spacing</th>
<th>Unit or Chapter</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>14.225</td>
<td>.001</td>
<td>III</td>
<td>7</td>
<td>1</td>
<td>1,3(2),4,5, B, 8</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>8.816</td>
<td>.01</td>
<td>III</td>
<td>7</td>
<td>1</td>
<td>6,12(2),14, 15, 16, 19</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>7.211</td>
<td>.01</td>
<td>III</td>
<td>9</td>
<td>1</td>
<td>1(3),4,B, 8,9,15,17</td>
<td>6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>7.031</td>
<td>.01</td>
<td>I</td>
<td>7</td>
<td>2</td>
<td>2,B,6,8,12, 14,19</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>56</td>
<td>7.013</td>
<td>.01</td>
<td>III</td>
<td>3</td>
<td>1</td>
<td>6,24,25</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>5.487</td>
<td>.05</td>
<td>I</td>
<td>3</td>
<td>1</td>
<td>7,15,16</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>49</td>
<td>5.168</td>
<td>.05</td>
<td>I</td>
<td>3</td>
<td>1</td>
<td>7,6,25</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>4.892</td>
<td>.05</td>
<td>III</td>
<td>4</td>
<td>1</td>
<td>B,8,13,14</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>4.231</td>
<td>.05</td>
<td>II</td>
<td>3</td>
<td>3</td>
<td>15,19,23</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>3.881</td>
<td>.05</td>
<td>I</td>
<td>3</td>
<td>2</td>
<td>13,17,19</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>3.056</td>
<td>.10</td>
<td>III</td>
<td>3</td>
<td>4</td>
<td>1,9,20</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>2.851</td>
<td>.10</td>
<td>I</td>
<td>3</td>
<td>1</td>
<td>12,16,17</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
all from Part III—The Methods and Aims of Science. Furthermore, six of the twelve questions were from Part III. Five of the twelve questions were from Part I—The Scientific Enterprise. Only one question was from Part II—The Scientists.

The four questions which showed the greatest gains were referred or alluded to with greater frequency, seven or more times, than the remaining questions (for the majority, three times).

Spacing was categorized as follows: category 1—the item was referred or alluded to two or more times in the same unit or in consecutive units; category 2—reference to the item was separated by one or two units; category 3—reference to the item was separated by three, four, or five units; and category 4—reference to the item was separated by six or more units. The most common spacing for the questions showing the greatest gains was category 1 with eight questions. Two questions were in category 2, one in category 3, and one in category 4.

Examination of the questions showing greatest gains, as related to different units of instruction for the course, revealed that Unit 1—The Solar System, had the greatest frequency, five times. This was followed by the unit on Boyle's Experiments (Harvard Case History) which, with units 6, 8, 12, 15, and 19 (titles are in Appendix A), had
a frequency of four each. A frequency of three was recorded for units 14, 16, and 17. Lesser frequencies were recorded for other units as shown in Table 33, page 148.

The mode of instruction was divided into three categories: 1--lecture and discussion, 2--laboratory exercises and experiments, and 3--assigned questions and problems requiring written answers. The first method of instruction was recorded with the greatest frequency. The laboratory exercises and experiments were next in importance, with their greatest frequency associated with the top four questions showing greatest gains. An over-all frequency of nine was recorded for category 2 and eight for category 3.

Chi square analysis was used to select the 15 questions (25% of the total number of questions) which showed the greatest loss (13 questions) or smallest gain (2 questions) on the TOUS. Levels of significance ranged from .10 to .98. The detailed analysis for the only questions (2) significant at the .10 level is given in Table 34, page 151. After these two questions the level of significance decreased drastically, indicating that the losses could be due to chance factors not related to the course.

The question registering the greatest loss was from Part II--The Scientists. The other question was from
<table>
<thead>
<tr>
<th>Quest. No.</th>
<th>Chi Square</th>
<th>Level of Signif.</th>
<th>Section of TOUS</th>
<th>Freq.</th>
<th>Spacing</th>
<th>Unit or Chapter</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>3.491</td>
<td>.10</td>
<td>II</td>
<td>3</td>
<td>2</td>
<td>16,19,25</td>
<td>1</td>
</tr>
<tr>
<td>59</td>
<td>3.238</td>
<td>.10</td>
<td>I</td>
<td>3</td>
<td>2</td>
<td>8,16,19</td>
<td>3</td>
</tr>
</tbody>
</table>
Part I—The Scientific Enterprise.

Both of these questions were referred or alluded to three times in the course.

Using the spacing system described on page 149, the two questions showing the greatest loss were in category 2.

Examination of the two questions showing significant (.10 level) losses as related to different units of instruction for the course showed Unit 16—Extranuclear Structure and Unit 19—Organic Chemistry present for each question.

The mode of instruction noted for these two questions was primarily category 3—assigned questions and problems.

Analysis of the WISP

Chi square analysis of pretest and posttest scores for each question on the WISP was used to select the 24 questions, 25% of 93 questions on the test, which showed the greatest gains made by the experimental group. The level of significance on these questions ranged from .001 to .25. Table 35, page 153, contains the results of the analysis using only those questions which were significant at the .10 level or greater. Ten questions, significant at the .20 to .25 level inclusive, were not included in the detailed analysis.

The question showing the greatest gain was from sec-
<table>
<thead>
<tr>
<th>Quest. No.</th>
<th>Chi Square</th>
<th>Level of Signif.</th>
<th>Section of WISP</th>
<th>Freq.</th>
<th>Spacing</th>
<th>Unit or Chapter</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>20.033</td>
<td>.001</td>
<td>I-B-3</td>
<td>3</td>
<td>1</td>
<td>13,14,24</td>
<td>3</td>
</tr>
<tr>
<td>69</td>
<td>15.000</td>
<td>.001</td>
<td>II-F-2</td>
<td>4</td>
<td>1</td>
<td>B,7(2),14</td>
<td>3</td>
</tr>
<tr>
<td>40</td>
<td>9.287</td>
<td>.01</td>
<td>II-D-2</td>
<td>3</td>
<td>2</td>
<td>3,15,18</td>
<td>3</td>
</tr>
<tr>
<td>22</td>
<td>8.052</td>
<td>.01</td>
<td>II-E-2</td>
<td>4</td>
<td>1</td>
<td>B,12,13,14</td>
<td>3</td>
</tr>
<tr>
<td>61</td>
<td>6.876</td>
<td>.01</td>
<td>II-A-1</td>
<td>3</td>
<td>2</td>
<td>4,8,12</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>5.875</td>
<td>.05</td>
<td>II-G-4</td>
<td>3</td>
<td>1</td>
<td>6,15,16</td>
<td>3</td>
</tr>
<tr>
<td>47</td>
<td>4.454</td>
<td>.05</td>
<td>I-A-2</td>
<td>3</td>
<td>2</td>
<td>B,7,13</td>
<td>3</td>
</tr>
<tr>
<td>88</td>
<td>4.345</td>
<td>.05</td>
<td>I-A-3</td>
<td>4</td>
<td>1</td>
<td>4,B,6,13</td>
<td>3</td>
</tr>
<tr>
<td>24</td>
<td>3.983</td>
<td>.05</td>
<td>I-A-4</td>
<td>3</td>
<td>2</td>
<td>13,16,22</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3.737</td>
<td>.10</td>
<td>II-B-1</td>
<td>3</td>
<td>3</td>
<td>2,7,14</td>
<td>3</td>
</tr>
<tr>
<td>89</td>
<td>3.563</td>
<td>.10</td>
<td>II-F-2</td>
<td>3</td>
<td>1</td>
<td>7,14(2)</td>
<td>3</td>
</tr>
<tr>
<td>41</td>
<td>3.419</td>
<td>.10</td>
<td>II-D-1</td>
<td>3</td>
<td>1</td>
<td>3,B(2)</td>
<td>3</td>
</tr>
<tr>
<td>72</td>
<td>3.282</td>
<td>.10</td>
<td>II-G-1</td>
<td>4</td>
<td>2</td>
<td>B,7,9,14</td>
<td>3</td>
</tr>
<tr>
<td>75</td>
<td>3.197</td>
<td>.10</td>
<td>II-E-2</td>
<td>4</td>
<td>1</td>
<td>B,17,18,19</td>
<td>4</td>
</tr>
</tbody>
</table>
tion I-B-3, dealing with the scientist assumes the products of science are parsimonious. The following sections were represented by two questions each: II-E-2—the actions of the scientist include communication which is an obligation that makes scientific information available for verification; II-F-2—prediction which is achieved by using deductive logic. The remaining questions each came from a different section of the WISP.

The questions which showed the greatest gains were referred or alluded to either three or four times as shown in Table 35.

The same spacing system used for the TOUS and described on page 149 was used with the WISP. Eight questions were in category 1 (the closest), five in category 2, and one in category 3.

Examination of the questions showing greatest gains, as related to the different units of instruction for the course, revealed the unit on Boyle's Experiments had the greatest frequency, eight times. This was followed by Unit 14—The Periodic Table, Atomic and Molecular Weights, seven times. Unit 7—Procedures of Science, and Unit 13—Basic Chemistry were next with frequencies of six and five respectively. Lesser frequencies were recorded for other units as shown in Table 35, page 153.

The mode of instruction used with the WISP in pro-
ducing the greatest gains was definitely category 1--lecture and discussion. Category 2--laboratory exercises, and category 3--assigned questions and problems were used twice each.

Chi square analysis was used to select the 24 questions, 25% of the total number of questions, which showed the greatest loss (20 questions) or smallest gain (4 questions) on the WISP. Levels of significance ranged from .20 to 1.00. No table of analysis of the WISP losses is included because of the very low levels of significance of the questions showing a loss--over 90% of the twenty-four questions were at the .30 level to 1.00 level. Chance factors not related to the course could be responsible for these slight shifts in pretest to posttest scores. The highly significant gains made by the experimental group on the WISP are not detracted from by losses.

Analysis of the CCTT

Chi square analysis of pretest and posttest scores for each question on the CCTT was used to select the 13 questions, 25% of 52 questions on the test, which showed the greatest gains made by the experimental group. The level of significance on these questions ranged from .10 to .40. These questions were of a lower significance range than the gain questions on either the TOUS or the
Only two of the questions were significant at the .10 level.

Chi square analysis was used to select the 13 questions, 25% of the total number of questions, which showed the greatest losses. Levels of significance ranged from .30 to .90. The very low levels of significance for these questions did not warrant a table of analysis.

There was no attempt made in the course to teach material related to any specific question. The effort was directed to teaching material related to the seven facets of critical thinking listed in the different sections of the CCTT.

The fact that only two questions on the CCTT registered a gain significant at the .10 level, together with the fact that losses were at a very low (.30 to .90) level of significance, indicated that the experimental course made only a slight contribution to the improvement of critical thinking skills.

Summary. Chi square analysis of pretest and posttest scores for each question on the criterion instruments was used to select the 25% of the questions on each test which showed the greatest gains and the 25% which showed the greatest losses or smallest gains. The level of significance for the gain questions ranged from .001 to about .25
on both the TOUS and the WISP. The level of significance for the gain questions on the CCTT ranged from .10 to .40.

The greatest gains on the TOUS were made with Part III—The Methods and Aims of Science, followed by Part I—The Scientific Enterprise. The gain questions on the WISP came from a variety of sections as shown in Table 35, page 153.

Most of the questions on the TOUS and the WISP were referred or alluded to three or four times in the course. The four questions with the greatest gains on the TOUS had a frequency of seven or more.

Most of the high gain questions on the TOUS and the WISP were in category 1 for spacing—at least two references in the same unit or consecutive units.

Although there were some differences between the TOUS and the WISP on the units of instruction that were most productive of high gains, the following units lead on a composite basis: B—Boyle's Experiments, 14—Periodic Table, Atomic and Molecular Weights, 7—Procedures of Science, and 13—Basic Chemistry.

In general, the predominant mode of instruction on all instruments was category 1—lecture and discussion. However, the laboratory exercises and experiments were next in importance with the TOUS gains.

The questions showing losses were at a much lower
level of significance than the gain questions. The level of significance ranged from .10 to 1.00 when considering all three instruments. The very small differences between pretest and posttest scores could be due to chance factors rather than instruction in the experimental course.

Summary for Chapter Four

The control group scored higher on both parts of the Scholastic Aptitude Test and the English-Achievement Test than the experimental group. The control group had higher mean scores on the entire TOUS test and the CCTT. No comparison was possible with the WISP.

The posttest TOUS-Total score showed correlations significant at the .01 level with scores on the other tests except the Math-Achievement. The posttest CCTT score showed correlations significant at the .01 level with the scores on the other instruments except TOUS-Part II which was significant at the .05 level. The posttest WISP score showed correlations significant at the .01 level with scores on the other tests except for SAT-Math and Math-Achievement.

Analysis of variance was used to test the null hypotheses dealing with pre- to posttest gains after a year's study of physical science. Results gave F-values significant at the .001 level for the TOUS-Parts I and III.
TOUS-Total, and WISP for the experimental group. The CCTT had an F-value significant at the .01 level. In general, evidence indicated the experimental course made an important contribution to the students' achievement as measured by the instruments.

Analysis of covariance was used to determine if significant differences in learning outcomes existed between the control and experimental groups when adjustments were made for differences in scholastic ability, achievement, and background knowledge or skill. Results of the analyses indicated a significant difference in learning outcomes between the two groups in favor of the experimental group for TOUS-Parts I, III, and Total. Results were not significant for CCTT.

Analysis of variance was used to test the null hypotheses dealing with differences attributable to different instructors. For the control group, there were no significant differences between the instructors in producing changes on any of the tests. For the experimental group, there was a significant difference attributable to instructor differences on the TOUS-Part III and the WISP.

Chi square analysis of pretest and posttest scores for each question on the criterion instruments was used to select the 25% of the questions on each test which showed the greatest gains and the 25% which showed the
greatest losses or smallest gains. The level of significance was much higher on the TOUS and the WISP than on the CCTT when confined to the gain questions. The questions showing losses were at a much lower level of significance than the gain questions.

The greatest gains on the TOUS were made on Part III--The Methods and Aims of Science, followed by Part I--The Scientific Enterprise.

There was some evidence that greater gains were made on questions on the TOUS and the WISP when material had been referred or alluded to three or more times. In general, questions which showed greater gains had a closer spacing of referred material. Lecture and discussion was used more than the laboratory exercises or assigned questions and problems as a mode of instruction. The laboratory exercises appeared to make their greatest contribution to the gain questions on the TOUS. Although many units of instruction were associated with the greatest gain questions on the TOUS and the WISP, four units were more productive than the others. These were Boyle's Experiments (Harvard Case History); the Periodic Table, Atomic and Molecular Weights; Procedures of Science; and Basic Chemistry.
CONCLUSIONS

Various conclusions, based on this study and described in Chapter IV, will be related to results found in some of the research studies cited in Chapter II.

Critical Thinking

The experimental group, after a year's study of physical science, made a gain significant at the .01 level as determined by an analysis of variance. The control group, after a year's study of physical science, did not make a gain significant at the .10 level as determined by an analysis of variance.

A comparison of the two groups showed that the control group was superior in scholastic aptitude to the experimental group as determined by the SAT-Verbal and the SAT-Math. This is summarized in Tables 36 and 37 on page 162.

Examination of the correlation coefficients of the posttest CCTT scores with SAT scores (p. 118) revealed that the experimental group had the highest correlation
with SAT-Verbal scores, the value of \( r \) being 0.3689. For SAT-Math, \( r \) was 0.3072. The control group had the highest correlation between their CCTT scores and their SAT-Math scores, the value of \( r \) being 0.4241. For SAT-Verbal, \( r \) was 0.3008.

**TABLE 36**

**SUMMARY OF SCORES ON SCHOLASTIC APTITUDE TESTS**

<table>
<thead>
<tr>
<th>Tests</th>
<th>Control Group</th>
<th>Experimental Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*N H L Mean</td>
<td>N H L Mean</td>
</tr>
<tr>
<td>SAT-Verbal</td>
<td>142 669 309 487.514</td>
<td>186 650 267 473.667</td>
</tr>
<tr>
<td>SAT-Math</td>
<td>142 752 328 496.092</td>
<td>186 692 271 490.145</td>
</tr>
</tbody>
</table>

*\( \text{N} \text{=} \text{number, H=} \text{highest, L=} \text{lowest} \)

**TABLE 37**

**PERCENTAGE DISTRIBUTION OF HIGH AND LOW SCORES**

<table>
<thead>
<tr>
<th>Tests</th>
<th>Control Group</th>
<th>Experimental Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>550 or Above</td>
<td>Below 400</td>
</tr>
<tr>
<td></td>
<td>550 or Above</td>
<td>Below 400</td>
</tr>
<tr>
<td>SAT-Verbal</td>
<td>20.42% 9.86%</td>
<td>10.75% 16.14%</td>
</tr>
<tr>
<td>SAT-Math</td>
<td>26.75% 11.27%</td>
<td>19.36% 12.37%</td>
</tr>
</tbody>
</table>
Examination of the studies cited in Chapter II revealed that the experimental groups of Rickert (92) in physical science and Montague (75) in chemistry made significant gains over the control groups in critical thinking skills. Rickert found a coefficient of correlation between the results on the academic ability test (School and College Ability Test) and the critical thinking test (American Council on Education Test of Critical Thinking, Form G) of 0.57, which a t-test showed was significant at the .05 level of confidence. The mean score on the critical thinking pretest was 31.82. The mean gains at the end of one semester was 6.36 for the experimental group and 2.00 for the control physical science survey group; the difference between means being significant at the .05 level. Thus, Rickert, with his instruments, found a greater positive correlation between critical thinking and academic ability than the present investigator did. The present investigator recorded a more significant gain (.01 level) based on analysis of variance for his experimental course than Rickert (.05 level).

Montague found that his experimental group in general chemistry laboratory did significantly better than his control group on the Watson-Glaser Critical Thinking Appraisal. There was no significant difference between his
two groups on Burmester's *A Test of Aspects of Scientific Thinking*; both groups made significant gains. Apparently, the two instruments did not measure the same characteristics or qualities.

The present investigator's experimental group did make a gain which was significant at the .01 level using analysis of variance. Covariance analysis which used scores on the SAT-Verbal, the SAT-Math, and the pretest CCTT as covariates indicated that the experimental group did not make a gain significant at the .10 level of significance over the control group. This result caused one to ask the question, "What caused the difference between the results on the analysis of variance and the covariance analysis?"

First, chi square was computed to determine the significance of the difference in means between the control group and the experimental group for each covariate on the CCTT. The values of chi square ranged from 0.197, (significant at the .70 level of significance) for the SAT-Verbal, to 0.033 (significant at the .90 level) for SAT-Math. These low levels of significance indicated no very significant differences existed between the two groups on the basis of the covariates used.

Second, it should be emphasized that the experimental
group made a significant gain (.01) using analysis of variance.

Third, it should be noted that the experimental group did not make a gain significant at the .10 level over the control group based on covariance analysis. Possibly, there are factors involved which are more important than the covariates used. Teichman reported that students with high mental ability are more likely to demonstrate initial skill in the ability to make conclusions than are less able students. He also concluded that mental ability and reading ability are very poor indications of the ability to improve one's skill in making conclusions. Burton, Kimball, and Wing concluded that correlations reported from different studies indicate that less than half the variability in problem-solving skills may be accounted for by scores on intelligence tests.

Burton, Kimball, and Wing reported that research has indicated that personal-emotional factors play an important role in critical thinking. Emotional stress generally inhibits critical thinking. It should be noted that the experimental group took their posttests at a time of great emotional stress because of campus unrest. The posttest period started May 11, 1970 following the May 8th Day of Mourning for the Kent State University incident.
This might have resulted in a poorer performance by some students, thus lowering class performance.

Other studies which reported that their experimental groups made significant gains over their control groups include Curtis (28), Teichman (100), Kastrinos (59), and Howe (53). The latter two involved biology.

Several other investigations at the college level reported no significant differences between their experimental and control groups. Studies cited included Yudin (120), Novak (81) with general botany, Olson (82) with biological science, and Lyle (71) with general psychology.

Understanding of the Nature of Science

In this section, studies using the TOUS will be considered first, and those using either the Wisconsin Inventory of Science Processes or the Welch Science Process Inventory next.

The control group in this study did not make a gain significant at the .10 level of significance on the TOUS-Parts II and III and TOUS-Total. The control group did make a gain significant at the .05 level on the TOUS-Part I. The experimental group in the present study made a gain significant at the .001 level of significance on the TOUS-Parts I and III and TOUS-Total. The experimental group did not make a gain significant at the .10 level on
TOUS-Part II but it did have greater understanding of scientists than the control group as shown by the shift in F-values from 0.3103 (control) to 1.1903 (experimental). All of these results are based on the analyses of variance. Covariance analysis was used with all parts of the TOUS except TOUS-Part II. The TOUS-Part I showed a gain of the experimental group over the control group significant at the .05 level. The TOUS-Part III and TOUS-Total both showed gains significant at the .001 level for the experimental group over the control group.

Chi square was computed to determine the significance of the difference in means between the control group and the experimental group for each covariate in TOUS-Part I, III, and Total. The values of chi square ranged from 0.197, significant at the .70 level of significance, for the SAT-Verbal, to 0.013, significant at the .95 level, for the TOUS-Part III pretest. These low levels of significance indicated no very significant differences existed between the two groups on the basis of the covariates involved.

Crawford and Backhus (26) found that their experimental group in college physical science, for the second experimental period, made gains significant at the .01 level on the TOUS-Part II and TOUS-Total. The gains on TOUS-
Part III were significant at the .05 level while on TOUS-
Part-I they were classed as not significant. These were
based on an analysis of covariance.

Durst (34) used analysis of variance with data col-
lected with his experimental and control groups in college
general biology. His experimental groups, in general, at-
tained higher scores on the criterion tests (the TOUS was
one) than did students in traditional classes. None of
the differences was significant.

Crumb's (27) investigation of PSSC vs traditional
high school physics groups, using the TOUS, led him to
conclude that the PSSC group did significantly better (.01
level) at the end of one semester and two semesters than
the traditional group did. Covariance analysis, with the
student's mental age, chronological age, and prior science
background as dependent variables, was used.

Trent's (102) study also was of PSSC vs traditional
high school physics. He found, utilizing an analysis of
variance, that the experimental group achieved significant-
ly (.01 level) higher scores on the TOUS than the control
group. However, when an analysis of covariance was used,
with scholastic aptitude and prior science achievement as
covariates, it was found there was no significant differ-
ence (0.20 level) in performance on the TOUS.

Other high school studies reporting significant gains
for students in experimental or special courses included those of Klopfer and Cooley (64), Kleppinger (63), and Cossman (24).

To summarize regarding the TOUS, careful attention should be given to the study of Crawford and Backhus and that of the present investigator. Both involved physical science classes in colleges whose primary purpose is teacher preparation; both were contemporaneous and utilized covariance analysis. The greatest contrast was in the scope of the purpose; the Kansas State effort was confined to the laboratory part of the course, while the Gorham State effort involved the entire course but did not have as diversified a laboratory approach. One additional difference should be noted— the Kansas State course was for one semester but the Gorham State course was two semesters in length. Both studies reported significant (various levels) gains on the TOUS-Total and TOUS-Part III. In contrast, Crawford and Backhus reported a highly significant (.01 level) gain on TOUS-Part II and no significant gain on TOUS-Part I, while the present investigator reported no gain at the .10 level on TOUS-Part II and a highly significant (.001 level) gain on TOUS-Part I.

The experimental group in the present study made a highly significant (.001 level) gain in understanding the processes of science as measured by the WISP and utilizing
analysis of variance. The control group did not take the WISP as a pretest so no comparisons of performance of the two groups are possible.

Larson found, by use of the Wilcoxon test, that a science methods course was effective in increasing knowledge of science processes, as measured by the WISP, when used with prospective elementary teachers. He found the highest correlation coefficient among his different variables was +0.36 between a knowledge of science processes and overall grade point.

Wittwer found that a group of high school science teachers who participated in a research science program scored "significantly higher" on a special Science Process Inventory than teachers without such experience.

Miller, using the WSPI, found significant (.05 level) differences between prospective elementary teachers and prospective high school science teachers in their understanding of science processes. The latter group scored higher.

Niman, with a one-semester college physical science course, found no difference in the means of his experimental and control groups.

At the high school level, Walberg found that physics classes characterized by high IQ's, high grades, and non-authoritarian students gained the most on both the WSPI
RECOMMENDATIONS

1. The objectives and general plan of the experimental course should be continued while the staff tries to develop better approaches and techniques.

2. There should be additional studies made to determine the major factors that are involved in improving critical thinking skills. The results of such studies should be used to improve instruction.

3. There should be a follow-up study of the experimental group in the spring of 1971 to determine the permanency of the learning outcomes in critical thinking and in understanding of the nature of science and its processes. Consideration should be given to investigating factors that affect permanency.

4. An effort should be made to utilize the laboratory more effectively in providing opportunities to promote the objectives of the experimental course, especially skills in critical thinking.

5. The science department should actively seek the cooperation of other departments in planning jointly to improve critical thinking skills.

6. Science departments should have an annual review of objectives for each course. Strategy should be planned
for meeting the objectives. End-of-course evaluations should be made to determine how effectively the objectives have been met. The evaluations should be utilized in reexamining the objectives and teaching procedures.

7. Analysis should be made of the teaching techniques of different physical science instructors who have objectives similar to those of the experimental course. The most effective procedures should be disseminated in an effort to improve instruction.

8. The TOUS and the WISP should be administered as a pretest in the elementary science methods and the secondary science methods classes to provide a basis for correcting misunderstandings about the nature of science and its processes. Posttesting could be used to measure gains and also used for the further guidance of the prospective teachers.
This appendix contains a syllabus for each semester for the control group (1968-69) and for the experimental group (1969-70). The list of laboratory exercises for each semester is considered a part of the syllabus for that semester. Each student received a copy of the syllabus at the first class meeting of a semester. It should be noted that for the experimental group, Unit 12, Electricity and Magnetism, is listed as the last unit for the first semester and the first unit for the second semester. The unit was taught in the second semester only; the instructors were not able to include it in the first semester's schedule.

The same syllabuses and instructional methods, materials, and criterion instruments were used with the physical science classes in the evening division. The results in the evening classes were not used because of: (1) a lack of SAT scores for most of the students, (2) a greater range in ages of the students, and (3) a greater diversity of educational backgrounds and goals.

Specimen pages from the teachers' outlines for the
course are included for illustrative purposes. The coding in the left margin indicates where the instructional material under consideration is related to questions and/or sections of the criterion instruments. For example:

*T-II-42 refers to the TOUS-Part II--The Scientists, and question number 42 on the test.

W-I-C-3-68 refers to the WISP-Part I--The Scientist Assumes, sub-part C--He and His Fellow Scientists Are: Section 3--motivated by a desire to understand the physical universe. There were three questions on the WISP dealing with this item; one was number 68.

C-5 refers to the CCTT-Section 5--Selecting the best prediction.

The specimen pages also include a number in parenthesis immediately below the coding number described above. The number in parenthesis indicates the mode of instruction.

(1)--lecture and discussion
(2)--laboratory exercises and experiments
(3)--assigned problems and questions.

*More detailed information may be obtained by contacting the investigator.

Maurice M. Whitten, Physical Science Department,
University of Maine, Portland-Gorham,
Gorham, Maine 04038
PHYSICAL SCIENCE, SCIENCE 100 and 100e

First Semester, 1968-69

INSTRUCTORS: Mr. James W. Pendleton  
Mr. Maurice M. Whitten

by Krauskopf and Beiser, 1966.

LABORATORY MANUAL: Laboratory Exercises in Physical  
Science by Lange, 1962.  (Some outlines are dittoed.)

EXAMINATIONS: One hour as scheduled. Quiz every two  
weeks. Two-hour final exam.

METHODS: Lectures, discussions, problems, films, dem­
onestrations, individual laboratory work.  
Laboratory work is discussed at the end of  
the laboratory period or at the following  
lecture session.

TENTATIVE SCHEDULE

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Chap.</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sep. 9</td>
<td>1</td>
<td>The Physical Universe</td>
</tr>
<tr>
<td>2</td>
<td>Sep. 16</td>
<td>2</td>
<td>Motion</td>
</tr>
<tr>
<td>3</td>
<td>Sep. 23</td>
<td>3</td>
<td>The Laws of Motion</td>
</tr>
<tr>
<td>4</td>
<td>Sep. 30</td>
<td>4</td>
<td>Gravitation</td>
</tr>
<tr>
<td>5</td>
<td>Oct. 7</td>
<td>5</td>
<td>Energy</td>
</tr>
<tr>
<td>6</td>
<td>Oct. 14</td>
<td>8</td>
<td>Atoms and Molecules</td>
</tr>
<tr>
<td>7</td>
<td>Oct. 21</td>
<td>9</td>
<td>Kinetic Molecular Theory of Matter</td>
</tr>
<tr>
<td>8</td>
<td>Oct. 28</td>
<td>10</td>
<td>Wave Motion and Sound</td>
</tr>
<tr>
<td>9</td>
<td>Nov. 4</td>
<td>11</td>
<td>Electricity</td>
</tr>
<tr>
<td>10</td>
<td>Nov. 12</td>
<td>12</td>
<td>Electricity and Matter</td>
</tr>
<tr>
<td>11</td>
<td>Nov. 18</td>
<td>13</td>
<td>Magnetism</td>
</tr>
<tr>
<td>12</td>
<td>Nov. 25</td>
<td>14</td>
<td>Properties of Light</td>
</tr>
<tr>
<td>13</td>
<td>Dec. 2</td>
<td>17</td>
<td>Atomic Structure</td>
</tr>
<tr>
<td>14</td>
<td>Dec. 9</td>
<td>18</td>
<td>The Nucleus</td>
</tr>
<tr>
<td>15</td>
<td>Dec. 16</td>
<td>19</td>
<td>The Theory of the Atom</td>
</tr>
<tr>
<td></td>
<td>Jan. 6</td>
<td></td>
<td>Reading Day</td>
</tr>
<tr>
<td></td>
<td>Jan. 13 - 17</td>
<td></td>
<td>Examination Week</td>
</tr>
</tbody>
</table>

EXAM.

EXAM.
### PHYSICAL SCIENCE, SCIENCE 100 and 100e

#### Laboratory Exercises, First Semester, 1968-69

<table>
<thead>
<tr>
<th>Week</th>
<th>Starting Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sep. 9</td>
<td>Math Review</td>
</tr>
<tr>
<td>2</td>
<td>Sep. 16</td>
<td>Metric Measurement</td>
</tr>
<tr>
<td>3</td>
<td>Sep. 23</td>
<td>Determination of Density</td>
</tr>
<tr>
<td>4</td>
<td>Sep. 30</td>
<td>The Pendulum</td>
</tr>
<tr>
<td>5</td>
<td>Oct. 7</td>
<td>Acceleration</td>
</tr>
<tr>
<td>6</td>
<td>Oct. 14</td>
<td>Diffusion</td>
</tr>
<tr>
<td>7</td>
<td>Oct. 21</td>
<td>Boyle's Law</td>
</tr>
<tr>
<td>8</td>
<td>Oct. 28</td>
<td>Charles' Law</td>
</tr>
<tr>
<td>9</td>
<td>Nov. 4</td>
<td>Wavelength and Velocity of Sound in Air</td>
</tr>
<tr>
<td>10</td>
<td>Nov. 12</td>
<td>Electrostatics and Magnetism</td>
</tr>
<tr>
<td>11</td>
<td>Nov. 18</td>
<td>Electrical Resistance</td>
</tr>
<tr>
<td>12</td>
<td>Nov. 25</td>
<td>Refraction and Reflection</td>
</tr>
<tr>
<td>13</td>
<td>Dec. 6</td>
<td>The Convex Lens</td>
</tr>
<tr>
<td>14</td>
<td>Dec. 9</td>
<td>Specific Heat of Metals</td>
</tr>
<tr>
<td>15</td>
<td>Dec. 16</td>
<td>Spectroscope</td>
</tr>
<tr>
<td></td>
<td>Jan. 6</td>
<td>Reading Day</td>
</tr>
<tr>
<td></td>
<td>Jan. 13-17</td>
<td>Examination Week</td>
</tr>
</tbody>
</table>
PHYSICAL SCIENCE, SCIENCE 100 and 100e

Second Semester, 1968-69

INSTRUCTORS: Mr. James W. Pendleton
Mr. Maurice M. Whitten

TEXTBOOK: Fundamentals of Physical Science,

LABORATORY MANUAL: Laboratory Exercises in Physical
Science by Lange, 1962. (Some
outlines are dittoed.)

EXAMINATIONS: One hour as scheduled. Quiz every two
weeks. Two-hour final exam.

METHODS: Lectures, discussions, problems, films,
demonstrations, individual laboratory work.
Laboratory work is discussed at the end of
the laboratory period or at the following
lecture session.

TENTATIVE SCHEDULE

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Chap.</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan. 28</td>
<td>20</td>
<td>Periodic Table</td>
</tr>
<tr>
<td>2</td>
<td>Feb. 3</td>
<td>21</td>
<td>The Chemical Bond</td>
</tr>
<tr>
<td>3</td>
<td>Feb. 10</td>
<td>22</td>
<td>The Solid State</td>
</tr>
<tr>
<td>4</td>
<td>Feb. 17</td>
<td>23</td>
<td>Basic Chemistry</td>
</tr>
<tr>
<td>5</td>
<td>Feb. 24</td>
<td>24</td>
<td>Ions and Solutions</td>
</tr>
<tr>
<td>6</td>
<td>Mar. 3</td>
<td>25</td>
<td>Acids, Bases, and Salts</td>
</tr>
<tr>
<td>7</td>
<td>Mar. 10</td>
<td>26</td>
<td>Chemical Reactions</td>
</tr>
<tr>
<td>8</td>
<td>Mar. 17</td>
<td>28</td>
<td>Organic Chemistry</td>
</tr>
<tr>
<td>9</td>
<td>Mar. 24</td>
<td>29</td>
<td>Earth Materials</td>
</tr>
<tr>
<td>10</td>
<td>Apr. 7</td>
<td>30</td>
<td>Rocks</td>
</tr>
<tr>
<td>11</td>
<td>Apr. 14</td>
<td>31</td>
<td>The Atmosphere</td>
</tr>
<tr>
<td>12</td>
<td>Apr. 21</td>
<td>32</td>
<td>Erosion and Sedimentation</td>
</tr>
<tr>
<td>13</td>
<td>Apr. 28</td>
<td>33</td>
<td>Vulcanism and Diastrophism</td>
</tr>
<tr>
<td>14</td>
<td>May 5</td>
<td>35</td>
<td>Interpreting the Rock Record</td>
</tr>
<tr>
<td>15</td>
<td>May 12</td>
<td>37</td>
<td>The Solar System</td>
</tr>
<tr>
<td>16</td>
<td>May 19</td>
<td>39</td>
<td>The Stars</td>
</tr>
<tr>
<td></td>
<td>May 22-29</td>
<td></td>
<td>Examination Week</td>
</tr>
</tbody>
</table>

EXAM.
EXAM.
EXAM.
<table>
<thead>
<tr>
<th>Week</th>
<th>Starting Date</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan. 28</td>
<td>Writing Formulas</td>
</tr>
<tr>
<td>2</td>
<td>Feb. 3</td>
<td>Hydrates and Water of Hydration</td>
</tr>
<tr>
<td>3</td>
<td>Feb. 10</td>
<td>Chemical Reactions</td>
</tr>
<tr>
<td>4</td>
<td>Feb. 17</td>
<td>Writing Equations</td>
</tr>
<tr>
<td>5</td>
<td>Feb. 24</td>
<td>pH</td>
</tr>
<tr>
<td>6</td>
<td>Mar. 3</td>
<td>Vitamin C Analysis</td>
</tr>
<tr>
<td>7</td>
<td>Mar. 10</td>
<td>Analysis of Baking Powder</td>
</tr>
<tr>
<td>8</td>
<td>Mar. 17</td>
<td>Preparation of Organic Compounds</td>
</tr>
<tr>
<td>9</td>
<td>Mar. 24</td>
<td>Identification of Common Minerals</td>
</tr>
<tr>
<td>10</td>
<td>Apr. 7</td>
<td>Identification of Common Rocks</td>
</tr>
<tr>
<td>11</td>
<td>Apr. 14</td>
<td>Introduction to Weather Maps</td>
</tr>
<tr>
<td>12</td>
<td>Apr. 21</td>
<td>Moisture in the Atmosphere</td>
</tr>
<tr>
<td>13</td>
<td>Apr. 28</td>
<td>Topographic Maps</td>
</tr>
<tr>
<td>14</td>
<td>May 5</td>
<td>Earth-Sun Relationships</td>
</tr>
<tr>
<td>15</td>
<td>May 12</td>
<td>Longitude and Time</td>
</tr>
<tr>
<td></td>
<td>May 22-29</td>
<td>Examination Week</td>
</tr>
</tbody>
</table>
PHYSICAL SCIENCE, SCIENCE 100 and 100e

First Semester, 1969-70

INSTRUCTORS: Mr. George H. Ayers
Mr. Parnell S. Hare
Mr. James W. Pendleton
Mr. Maurice M. Whitten

TEXTBOOK: College Physical Science, 2nd edition,
by Miles, Sherwood, and Parsons, 1969.

LABORATORY: The staff will distribute mimeographed
instructions for the laboratory exercises.

EXAMINATIONS: Three one-hour exams and a two-hour final
are scheduled. Every exam will contain
questions dealing with the laboratory exercises.

TENTATIVE SCHEDULE

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Chap.</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sep. 17</td>
<td>1</td>
<td>Course Objectives; The Solar Sys.</td>
</tr>
<tr>
<td>2</td>
<td>Sep. 22</td>
<td>1 cont.</td>
<td>The Solar System</td>
</tr>
<tr>
<td>3</td>
<td>Sep. 29</td>
<td>1 cont.</td>
<td>The Solar System</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Our Planet, the Earth</td>
</tr>
<tr>
<td>4</td>
<td>Oct. 6</td>
<td>3</td>
<td>Motion and Force</td>
</tr>
<tr>
<td>5</td>
<td>Oct. 13</td>
<td>3 cont.</td>
<td>Motion and Force</td>
</tr>
<tr>
<td>6</td>
<td>Oct. 20</td>
<td>4</td>
<td>Law of Universal Gravitation EXAM.</td>
</tr>
<tr>
<td>7</td>
<td>Oct. 27</td>
<td>5</td>
<td>Work, Energy, Power</td>
</tr>
<tr>
<td>8</td>
<td>Nov. 3</td>
<td></td>
<td>Boyle's Experiments in Pneumatics</td>
</tr>
<tr>
<td></td>
<td>Harvard</td>
<td></td>
<td>Case History</td>
</tr>
<tr>
<td>9</td>
<td>Nov. 10</td>
<td>6</td>
<td>Mechanics of Fluids</td>
</tr>
<tr>
<td>10</td>
<td>Nov. 17</td>
<td>7</td>
<td>Procedures in Science EXAM.</td>
</tr>
<tr>
<td>11</td>
<td>Nov. 24</td>
<td>8</td>
<td>Heat and Meteorology</td>
</tr>
<tr>
<td>12</td>
<td>Dec. 1</td>
<td>8 cont.</td>
<td>Heat and Meteorology</td>
</tr>
<tr>
<td>13</td>
<td>Dec. 8</td>
<td>9</td>
<td>The Sun, the Moon, and the Planets</td>
</tr>
<tr>
<td>14</td>
<td>Dec. 15</td>
<td>10</td>
<td>Wave Motion, Sound and Light EXAM.</td>
</tr>
<tr>
<td>15</td>
<td>Jan. 5</td>
<td>10 cont.</td>
<td>Wave Motion, Sound and Light</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>Electricity and Magnetism</td>
</tr>
<tr>
<td>16</td>
<td>Jan. 12</td>
<td>12 cont.</td>
<td>Electricity and Magnetism</td>
</tr>
<tr>
<td></td>
<td>Jan. 15</td>
<td></td>
<td>Reading Day</td>
</tr>
<tr>
<td></td>
<td>Jan. 16 - 24</td>
<td></td>
<td>Examination Week</td>
</tr>
<tr>
<td>Week</td>
<td>Starting Date</td>
<td>Topic</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Sep. 17</td>
<td>TOUS and CCTT, Pretest</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Sep. 24</td>
<td>Earth-Sun Relationships</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Oct. 1</td>
<td>Longitude and Time</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Oct. 8</td>
<td>Math and Metric System</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Oct. 15</td>
<td>Composition and Resolution of Forces</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Oct. 22</td>
<td>Film-Gravitational Constant (g); Water Drop Experiment to Determine &quot;g&quot;.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Oct. 29</td>
<td>Potential Energy (Gravitational and Configurational) of Spring. Determination of each Individual's Horsepower.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Nov. 5</td>
<td>Determination of Water Vapor Pressure</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Nov. 12</td>
<td>Density</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Nov. 19</td>
<td>Determination of Mass</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Dec. 1</td>
<td>Relative Humidity; Weather Instruments</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Dec. 8</td>
<td>Films - Wave Motion</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Dec. 15</td>
<td>Resistance of Conductors</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Jan. 5</td>
<td>Electrical Energy Transformations</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Jan. 12</td>
<td>No Laboratory Exercise (3 days only)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jan. 15</td>
<td>Reading Day</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jan. 16-24</td>
<td>Examination Week</td>
<td></td>
</tr>
</tbody>
</table>
PHYSICAL SCIENCE, SCIENCE 100 and 100e

Second Semester, 1969-70

INSTRUCTORS: Mr. George H. Ayers
               Mr. Parnell S. Hare
               Mr. James W. Pendleton
               Mr. Maurice M. Whitten

TEXTBOOK: College Physical Science, 2nd edition,
           by Miles, Sherwood, and Parsons, 1969.

LABORATORY: The staff will distribute mimeographed
             instructions for the laboratory exercises.

EXAMINATIONS: Three one-hour exams and a two-hour final
               are scheduled. Every exam will contain
               questions dealing with the laboratory ex­
               ercises.

---

TENTATIVE SCHEDULE

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Chap.</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>Electricity and Magnetism</td>
</tr>
<tr>
<td>1</td>
<td>Feb. 3</td>
<td>12</td>
<td>Electricity and Magnetism</td>
</tr>
<tr>
<td>2</td>
<td>Feb. 9</td>
<td>12 cont.</td>
<td>Electricity and Magnetism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>Basic Chemistry</td>
</tr>
<tr>
<td>3</td>
<td>Feb. 16</td>
<td>13 cont.</td>
<td>Basic Chemistry</td>
</tr>
<tr>
<td>4</td>
<td>Feb. 23</td>
<td>14</td>
<td>Atomic and Molecular Weights</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The Periodic Table</td>
</tr>
<tr>
<td>5</td>
<td>Mar. 2</td>
<td>15</td>
<td>Nuclear Atom, Mass Energy</td>
</tr>
<tr>
<td>6</td>
<td>Mar. 9</td>
<td>15 cont.</td>
<td>Nuclear Atom, Mass Energy</td>
</tr>
<tr>
<td>7</td>
<td>Mar. 16</td>
<td>16</td>
<td>Extranuclear Structures and Chemical Bonding</td>
</tr>
<tr>
<td>8</td>
<td>Mar. 23</td>
<td>17</td>
<td>Chemical Order in Acids, Bases, Salts</td>
</tr>
<tr>
<td>9</td>
<td>Apr. 6</td>
<td>18</td>
<td>Reaction Rate and Equilibrium</td>
</tr>
<tr>
<td>10</td>
<td>Apr. 13</td>
<td>19</td>
<td>Organic Chemistry</td>
</tr>
<tr>
<td>11</td>
<td>Apr. 20</td>
<td>20</td>
<td>The Earth and Its Material</td>
</tr>
<tr>
<td>12</td>
<td>Apr. 27</td>
<td>21</td>
<td>Igneous Activity and Metamorphism</td>
</tr>
<tr>
<td>13</td>
<td>May 4</td>
<td>22</td>
<td>Weathering, Sedimentary Rocks, and Geologic Time</td>
</tr>
<tr>
<td>14</td>
<td>May 11</td>
<td>23</td>
<td>Landscape Processes and Forms</td>
</tr>
<tr>
<td>15</td>
<td>May 18</td>
<td>24</td>
<td>Earthquakes and Earth's Interior</td>
</tr>
<tr>
<td>16</td>
<td>May 25</td>
<td>25 cont.</td>
<td>Mountains and Mountain Building</td>
</tr>
<tr>
<td></td>
<td>May 27 - June 4</td>
<td>25 cont.</td>
<td>Mountains and Mountain Building</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Examination Week</td>
</tr>
<tr>
<td>Week</td>
<td>Starting Date</td>
<td>Topic</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Feb. 3</td>
<td>Electrical Resistance</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Feb. 9</td>
<td>Electricity and Magnetism. 3 parts-Electrical Fields, Magnetic Fields and Strength, Oersted Effects</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Feb. 16</td>
<td>Chemical and Physical Changes</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Feb. 25</td>
<td>Hydrates and Water of Hydration</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Mar. 2</td>
<td>Demonstrations with Geiger Counter</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Mar. 9</td>
<td>Atoms, Molecules, Ions</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Mar. 16</td>
<td>Determination of Percentage of Acetic Acid in Vinegar</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Mar. 23</td>
<td>Vitamin C Analysis</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Apr. 6</td>
<td>Reaction Rates</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Apr. 13</td>
<td>Organic Chemistry</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Apr. 20</td>
<td>Minerals</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Apr. 27</td>
<td>Rocks</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>May 4</td>
<td>Topographic Maps</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>May 11</td>
<td>TOUS and CCTT Posttests</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>May 18</td>
<td>Topographic Maps</td>
<td></td>
</tr>
<tr>
<td></td>
<td>May 26</td>
<td>Reading Day</td>
<td></td>
</tr>
<tr>
<td></td>
<td>May 27-June 4</td>
<td>Examination Week</td>
<td></td>
</tr>
</tbody>
</table>
4. Law of Universal Gravitation

Sir Isaac Newton: Three laws of motion, law of universal gravitation, reflecting telescope, study of light (spectrum), developed the calculus, master of the mint. Went home to Woolsthorpe and stayed for 18 months during the Plague of 1665-1666.

Most scientists have more than average intelligence and ability. Newton's was exceptional; his I.Q. was 190 as estimated by Dr. Cox.

Newton's motivation evidently was curiosity, the desire to know. He did not have any technological application in mind. Most scientists are not motivated by practical applications. This is one of the differences between scientists and engineers or technicians.

Newton assumed and believed, as do modern scientists, that the physical universe is consistent, so the law is called the Law of Universal Gravitation. He also assumed the universe was real. What knowledge was prerequisite for Newton before he could formulate the Law of Universal Gravitation?


This law is regarded as one of the greatest inductive steps in all of science.

Is the development of theories and laws primarily an inductive or deductive process?

Is the use of scientific theories an inductive or deductive process?

In Newton's Law, objects are treated as if their entire mass is concentrated at their geometrical centers. This is true of homogeneous spheres. We assume so for most other objects.

4. Law of Universal Gravitation (Page 2)

Difference between a proportion and an equation.

Important problem unsolved by Newton: What was the numerical value for G? Why didn't he determine its value?

Determination of the value of G.
1. The Cavendish Experiment. Describe briefly.

2. Modified von Jolly experiment used at M.I.T. Improved instruments are necessary as science advances.


\[ G = 6.67 \times 10^{-8} \text{ cm}^3/\text{g}.\text{-sec}^2 \]

Constant effort is made to gain accuracy and decrease error; this gives rise to the frequently used expression that science is a self-correcting enterprise.

In conducting research in physics, the investigator uses any method he can think of that will help solve his problem. This calls for originality and ingenuity.

Do scientists select their observations or do they observe indiscriminately? Why?

Does a scientist limit the number of variables that he observes at any given time? Why?

One of the purposes of setting up a laboratory experiment is to provide conditions that can be controlled to a greater degree than under "natural conditions."

Once the value for G had been determined, it was possible to "weigh the earth." See problem 5.

Relation between gravitational and absolute units of force. Difference between weight and mass.
7. Procedures in Science; Special Readings

Bridgman: . . . "there is no scientific method as such . . ." Cite examples from the work of scientists studied thus far to support Bridgman's idea. If there is no set scientific method, what then is common to the endeavors of scientists, if anything?

Induction and Deduction in Science.

Induction: a generalization from large number of observations, facts, etc.
Kepler: 3 Laws of Planetary Motion from Brahe's data.
Newton: Law of Universal Gravitation from work of others and his own.

Deduction: a specific application of a general principle.
Deduce the distance of Pluto from the Sun by Kepler's 3rd Law.
Deduce an object must weigh more at the North Pole.
Bernoulli deduced from his kinetic model the relationship found experimentally by Boyle (his Law).

Which type of reasoning--deductive or inductive--is more creative, imaginative, synthesizing? Which yields more valid conclusions?

General Procedures. (Not steps).
Start with a problem, a "snarl."

Formulate a hypothesis. Conant speaks of "limited hypothesis" and "broad working hypothesis."
What is the difference?

Collect information, observational, experimental.

How does the data agree or gibe with the hypothesis?
7. Procedures in Science; Special Readings (Page 2)

T-III-45
(3)
If . . . then reasoning. More experimentation. Modification of hypothesis, confirmation, or discard.

Conant uses the term "conceptual scheme." What does he mean?

Roger Bacon: "All sciences are connected, they lend each other aid as parts of a whole . . ."

T-III-17
(3)
This statement was made about 700 years ago.
How specialized was science then?
Does this statement hold today with our high degree of specialization?
Cite pros and cons. (Research teams).

C. Darwin: "Have to give up a hypothesis as soon as facts are shown opposed to it."

Is this true of all scientists?
Do they cling to their "favorite notions"? Cite examples.
Is a hypothesis more readily adopted if it does not run counter to an existing hypothesis?
(Priestley was a phlogistonist until he died, 1804)
(Ptolemaic Hypothesis taught at Harvard until c. 1800)

W-II-G-1-36
(1)
What is the origin of hypotheses?
Frequently "inspired or educated guess," or intuitive hunch."

W-II-G-1-72
(1)
How does a hypothesis differ from a theory or a law?

E. Mach: "The factor which most promotes scientific thought is the gradual widening of the field of experience." Explain.
Purpose:

To make observations, formulate hypotheses, investigate the hypotheses, and draw conclusions.

Discussion:

In dictionary definitions and in common usage, deduction is the process of reasoning, from general principles to specific instances; in contrast, induction is reasoning from a large number of particular cases to a generalization. It is difficult to separate the various aspects of the scientists' activities, and one could raise the question here as to when, if ever, a conclusion is completely proven by either process alone. Nevertheless, one can recognize either induction or deduction predominating at various times.

In deduction, one starts from a general principle which may be either already proven, or an assumed principle. Hypotheses are tentatively assumed general principles. The scientist then designs an experiment which he believes will test the assumed general principle or hypothesis on the basis of relevant data. This part of his work is fraught with many possible difficulties and hazards. Can one, for example, be sure that the data are relevant and that they can be used to answer the question being investigated? Can one be sure that one is investigating a single cause-effect variation; in other words, that the results obtained are due to the cause being observed, or whether the results might be produced by a secondary or related cause which varied at the same time? Can one be sure that he has collected enough data to cover all the possibilities of this single cause-effect relationship? Many erroneous theories and conclusions have been drawn in the past because of errors in experimentation or in interpretation of data. These difficulties must be overcome if conclusions are to be satisfactorily drawn, for the essence of the scientific meth-
7. Procedures in Science; Laboratory Exercise (Page 4)

Procedure:

Each laboratory section will be divided into four groups of six students each.

Each group will be given a numbered object.

Each student in a group will weigh the same object individually, using a different balance, and record his results.

Balances will include: triple-beam single pan (sensitivity 0.1 gram), triple-beam single pan (sensitivity 0.05 gram), single-beam trip two pan with loose weights (sensitivity 0.1 gram), double-beam trip two pan (sensitivity 0.1 gram) spring type (5 gram divisions), and "Cent-O-Gram" (sensitivity 0.01 gram).

When all weighings are completed, the results obtained by each student are posted on the blackboard. There are variations.

How do you (students) account for the variations?

Groups develop hypotheses to account for the variations.

Different students select one or more hypotheses to investigate, as time permits.

Describe procedures used to investigate hypothesis. Record all data.

C-1, C-4 (2)

Draw conclusions based on your investigation.
13. Basic Chemistry

Alchemy preceded "modern" chemistry which started about the time of the American Revolution. Alchemists made many important discoveries: HCl, HNO₃, H₂SO₄, NH₄Cl, "sulfuric ether", improved balances, retorts, distilling apparatus.

The chemists, from the latter part of the 18th century on, differed quite drastically from the alchemists. How did they differ?

1. Alchemists - secretive, didn't publish or communicate.
   Chemists - published the results of their experiments so that others might evaluate their work.

   Today the American Chemical Society (ACS) with over 110,000 members is one of the largest scientific societies in the world. It carries on a great variety of activities including sponsoring seminars and meetings for the exchange of ideas, establishing standards of purity, publishing chemical journals and books. Journals are a forum for discussion of new theories, provide information on research in progress, and print papers given at scientific meetings. Reports are usually accurate and honest. The Society promotes chemical research.

2. Alchemists - mystical, preferred supernatural to logical explanation.
   Chemists - believe that natural phenomena can be understood by man. They assume natural phenomena have natural causes. They prefer simple explanations of phenomena.

   W-I-A-4-24 (l)
   W-I-B-3-8 (l)

3. Alchemists - had limited goals which were quite specific, such as preparing a universal solvent, transmuting base metals to gold, etc.
14. Atomic and Molecular Weights; The Periodic Table

Quotation: "...we are perhaps not far removed from the time when we shall be able to submit the bulk of chemical phenomena to calculation."
Gay-Lussac 1809

Law of Definite Composition
Joseph Proust, 1799, formulated the law after extensive experimentation. Claude Berthollet, reached the opposite conclusion at about the same time; namely, elements combine in weight ratios which are variable within limits. Long controversy. How were opposite conclusions reached?

Berthollet observed that copper and tin when heated in air both gave the appearance of forming a "continuous series of compounds" of varying color and composition. He maintained that solutions, alloys, and glasses were compounds of varying compositions.

Proust concluded from his experiments that copper and tin each form two different compounds with oxygen. This is an example of two different scientists examining the same phenomena, but making different observations and reaching different conclusions. How account for this happening? Many apparently objective observations are interpretations of what is seen.

How does a controversy serve a useful purpose? It frequently stimulates a large amount of careful experimentation to check predictions made from observations. This serves to detect errors and is one reason science is referred to as a "self-correcting enterprise."


This law had important implications: the definite ratios between volumes implied that a definite number of particles of a gas were involved.

PREDICTED AND OBSERVED PROPERTIES OF GERMANIUM

<table>
<thead>
<tr>
<th>Mendeleyev's prediction (1871) for the undiscovered element he called eka-silicon (Es)</th>
<th>Observed properties of germanium, discovered by Winkler in 1885</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Atomic weight = 72.</td>
<td>1. Atomic weight = 72.60.</td>
</tr>
<tr>
<td>2. Es a dark gray metal, with high melting point and density = 5.5.</td>
<td>2. Ge is dark gray; melting point = 958°C, density = 5.36.</td>
</tr>
<tr>
<td>3. Es only slightly attacked by acids, resistant to alkalies such as NaOH.</td>
<td>3. Ge not attacked by HCl, but dissolved by concentrated HNO₃; not attacked by NaOH.</td>
</tr>
<tr>
<td>4. Es will form oxide EsO₂ on heating; EsO₂ will have high melting point and density = 4.7.</td>
<td>4. Ge forms oxide GeO₂, with melting point 1100°C, density = 4.70.</td>
</tr>
<tr>
<td>5. Es will form a sulfide EsS₂ which is insoluble in water but soluble in ammonium sulfide.</td>
<td>5. Ge forms sulfide GeS₂, which is insoluble in water but soluble in ammonium sulfide.</td>
</tr>
<tr>
<td>6. Es will form a chloride EsCl₄, with boiling point a little less than 100°C and density = 1.9.</td>
<td>6. Ge forms chloride GeCl₄ with boiling point 83°C and density = 1.88.</td>
</tr>
<tr>
<td>7. Es will be formed upon reaction of EsO₂ or K₂EsF₆ with sodium metal.</td>
<td>7. Ge is formed by reaction of K₂GeF₆ with sodium.</td>
</tr>
</tbody>
</table>
Questions

C-l 1. Briefly describe any similarities between Mendeleyev's prediction of unknown elements and the predictions of Leverrier and Adams (independently) of an unknown planet, Neptune?

W-II-F-2-69 2. Did Mendeleyev make his predictions primarily by use of inductive or deductive reasoning? Justify your answer.

W-II-F-2-89 3. Which process is more likely to yield valid conclusions? Why?

W-II-F-3-9 4. Can scientists, by following the "scientific method" step by step, answer any questions concerning natural phenomena? Cite evidence in support of your answer.

5. Examine carefully the sheet which lists the "Predicted and Observed Properties of Germanium."

C-l What procedures did Mendeleyev use to predict the properties of undiscovered elements?

Itemize each individual characteristic of germanium and its compounds for which he made a prediction, i.e. analyze the 7 categories shown and list each characteristic.

6. Astatine, At, which is at the bottom of the halogen column, has been prepared artificially in minute amounts. Predict the properties of the element with regard to:

(a) State at room temp.: solid, liquid, gas? Why?

(b) How many atoms in a molecule of its vapor? Why?

(c) Slightly, moderately, or extremely soluble in H₂O? Why?

(d) Formula for its compound with H?
The Harvard Case History, *Boyle's Experiments in Pneumatics*, was used as an instructional unit in the course. Paperback copies were sold by the college bookstore so that each student could have a copy. The following study guide was prepared and distributed to the students one week in advance of starting the discussions in class.

**STUDY GUIDE**

1. What radical change took place between the first quarter and the last quarter of the 17th century concerning science? How did this change come about?

2. Why does mercury stay up in the inverted glass tube?

3. Why use mercury instead of water in the tube?

4. What was Torricelli's "new concept"? Why was this new concept important?

5. Is objective observation less important in science today than in the past when instruments were less sophisticated? Cite reasons for your answer.

6. What was Pascal's hypothesis? Describe the experiment of September, 1648, to test the hypothesis.

7. Boyle wanted to test a deduction from Torricelli's concept, "Would mercury fall in a tube if air is removed above the mercury reservoir?" How was this deduction tested?

8. What is the difference between a hypothesis and a theory?
9. Boyle kept accurate records of his experimental work. List at least three reasons why this is so important, even today.

10. Who owns scientific knowledge?

11. Often scientists are confronted with problems which require new or better instruments or equipment than is currently available.
a. Cite an illustration of the above statement from Boyle's investigations.
b. Is this, the problem, science or technology?

12. What is deduction?

13. Mercury was found to stand at 29.5 inches in England but 26 or 27 inches in France or Italy. What possible explanations (hypotheses) were offered?

14. a. What were the results of Boyle's experiments on air as a medium for transmitting sound?
b. What conditions did he attempt to control?

15. What is the advisibility of running a single trial of an experiment to test a hypothesis? Explain.

16. What is the significance of getting negative results in experiments?

17. Does a scientist have some idea of the kinds of observations he expects to make during an experiment? Explain.

18. What is the role of controversy in experimental science?

19. What is a scientific law?

20. What are the variables and constants in Boyle's Law?

21. What are the limitations of Boyle's Law?

22. What does Conant mean by a "conceptual scheme"?

23. How do scientific societies and scientific journals affect the advance of science?
BIBLIOGRAPHY


83. Pella, Milton O. Professor of Science Education, The University of Wisconsin, Madison, Wisconsin, September 10, 1968. Personal communication to investigator.


