This dissertation has been microfilmed exactly as received

KAPFER, Philip Gordon, 1936–
CRITERIA FOR EVALUATING RESEARCH AND THEIR APPLICATION TO SCIENCE EDUCATION.

The Ohio State University, Ph.D., 1964
Education, general

University Microfilms, Inc., Ann Arbor, Michigan
CRITERIA FOR EVALUATING RESEARCH AND THEIR APPLICATION TO SCIENCE EDUCATION

DISSERTATION

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

By

Philip Gordon Kapfer, B.A., M.A.

* * * * * *

The Ohio State University
1964

Approved by

[Signature]
Adviser
Department of Education
ACKNOWLEDGMENTS

I wish to express my gratitude to my adviser, Professor John S. Richardson, for the encouragement and direction he gave in formulating, developing, and carrying out this study.

The expression of my appreciation goes to Professors Earl W. Anderson and Jack G. Calvert for their assistance in planning my graduate program and for serving on the reading committee for the dissertation.

Thanks are extended to those graduate students and science educators who helped in the early development of the research criteria instrument, to the pilot study judges, and to the seventy-two judges whose responses to the lengthy research criteria instrument were included in the study.

Special thanks are given to my wife, Miriam Bierbaum Kapfer, for her encouragement, faith, and assistance throughout my graduate studies, even while meeting requirements for her own doctoral program.
VITA

July 4, 1936  Born - Detroit, Michigan
1958 ....  B.A., State College of Iowa, Cedar Falls, Iowa
1958-1959 . Teaching Assistant, Department of Chemistry, State College of Iowa, Cedar Falls, Iowa
1959 ....  M.A., State College of Iowa, Cedar Falls, Iowa
1959-1961 . Instructor, St. John's College and Academy, Winfield, Kansas
1961-1964 . Teaching Assistant, Department of Chemistry, The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Science Education

Studies in Science Education. Professors John S. Richardson and Fred R. Schlessinger

Major Field: Chemistry

Studies in Physical Chemistry. Professors Jack G. Calvert and William J. Taylor

Studies in Inorganic Chemistry. Professor Sheldon G. Shore

Studies in Organic Chemistry. Professors Harold Shechter and William N. White

Studies in Quantitative Chemistry and in the History of Chemistry. Professor Earle R. Caley

Minor Field: Teacher-Higher Education


Studies in Higher Education. Professor Everett J. Kircher
CONTENTS

ACKNOWLEDGMENTS ............................................. ii
VITA AND FIELDS OF STUDY ........................................... iii
TABLE OF CONTENTS ............................................. iv
LIST OF TABLES .................................................. vii

Chapter

I. FORMULATION AND DEFINITION OF THE PROBLEM .......... 1

Purpose of the Study ........................................... 1
Need for the Study ............................................. 2
Historical Perspective ........................................... 2
Dispersal of Criteria ........................................... 3
Lack of Previous Research ..................................... 5
Process of Research ............................................ 6
Unifying Traits of Research ................................... 7
Creativity in Research ........................................ 11
Presentation of the Problem ................................... 13
Statement of the Problem ..................................... 13
Definition of Terms ............................................ 14
Working Hypotheses .......................................... 15
Organization of the Remainder of the Study ............... 16

II. RESEARCH PROCEDURES ..................................... 17

Development of the Research Criteria ....................... 17
Source of the Research Criteria ............................... 17
Selection of the Research Criteria ........................... 19
Organization of the Research Criteria ....................... 21
Pilot Study ..................................................... 23
Research Criteria Used in the Study ......................... 26
Validation of the Research Criteria ......................... 48
Selection of Expert Judges ................................... 48
Mailing and Follow-up Procedure ............................ 51
Limitations in the Development and Validation of the Research Criteria ........................................... 54
Assumptions in the Development and Validation of the Research Criteria ........................................... 54
## CONTENTS--Continued

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytical Procedures</td>
<td>55</td>
</tr>
<tr>
<td>Ranking of the Criteria</td>
<td>55</td>
</tr>
<tr>
<td>Test for Consistency</td>
<td>56</td>
</tr>
<tr>
<td>Test for Significance of Differences</td>
<td>56</td>
</tr>
<tr>
<td>Summary</td>
<td>57</td>
</tr>
<tr>
<td><strong>III. PRESENTATION AND ANALYSIS OF DATA</strong></td>
<td>59</td>
</tr>
<tr>
<td>Statistical Data</td>
<td>59</td>
</tr>
<tr>
<td>Order of Importance of the Research Criteria</td>
<td>59</td>
</tr>
<tr>
<td>Consistency of Ratings</td>
<td>60</td>
</tr>
<tr>
<td>Significance of Differences</td>
<td>64</td>
</tr>
<tr>
<td>Descriptive Data</td>
<td>68</td>
</tr>
<tr>
<td>Subjective Data</td>
<td>73</td>
</tr>
<tr>
<td>Summary</td>
<td>76</td>
</tr>
<tr>
<td><strong>IV. AN EXAMPLE OF THE APPLICATION OF THE RESEARCH CRITERIA TO A STUDY IN SCIENCE EDUCATION</strong></td>
<td>78</td>
</tr>
<tr>
<td>Introduction</td>
<td>78</td>
</tr>
<tr>
<td>An Overview of the Example Study</td>
<td>80</td>
</tr>
<tr>
<td>Formulation and Definition of the Problem</td>
<td>80</td>
</tr>
<tr>
<td>Procedural Design</td>
<td>81</td>
</tr>
<tr>
<td>Presentation of Evidence</td>
<td>83</td>
</tr>
<tr>
<td>Summary and Conclusions</td>
<td>84</td>
</tr>
<tr>
<td>Application of the Research Criteria to the Example Study</td>
<td>84</td>
</tr>
<tr>
<td>Application of the Criteria to the Formulation and Definition of the Problem</td>
<td>84</td>
</tr>
<tr>
<td>Application of the Criteria to the Procedural Design</td>
<td>90</td>
</tr>
<tr>
<td>Application of the Criteria to the Presentation of Evidence</td>
<td>96</td>
</tr>
<tr>
<td>Application of the Criteria to the Summary and Conclusions</td>
<td>98</td>
</tr>
<tr>
<td>Summary</td>
<td>100</td>
</tr>
<tr>
<td><strong>V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS</strong></td>
<td>102</td>
</tr>
<tr>
<td>Summary</td>
<td>102</td>
</tr>
<tr>
<td>Formulation and Definition of the Problem</td>
<td>102</td>
</tr>
</tbody>
</table>
## CONTENTS--Continued

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Procedures</td>
<td>104</td>
</tr>
<tr>
<td>Presentation and Analysis of Data</td>
<td>107</td>
</tr>
<tr>
<td>An Example of the Application of the</td>
<td></td>
</tr>
<tr>
<td>Research Criteria to a Study in Science Education</td>
<td>108</td>
</tr>
<tr>
<td>Conclusions</td>
<td>108</td>
</tr>
<tr>
<td>Conclusions Related to the Ranking of the Research Criteria</td>
<td>108</td>
</tr>
<tr>
<td>Conclusions Related to the Consistency of Ratings</td>
<td>119</td>
</tr>
<tr>
<td>Conclusions Related to the Significance of Differences</td>
<td>121</td>
</tr>
<tr>
<td>Conclusions Related to the Descriptive Data</td>
<td>129</td>
</tr>
<tr>
<td>Conclusions Related to the Subjective Data</td>
<td>129</td>
</tr>
<tr>
<td>Recommendations</td>
<td>130</td>
</tr>
<tr>
<td>Recommendations for the Use of the Research Criteria</td>
<td>130</td>
</tr>
<tr>
<td>Recommendations for Future Research</td>
<td>131</td>
</tr>
</tbody>
</table>

## APPENDIXES

I. Research Criteria and Related Statements                             | 133  |
II. Correspondence and Materials Related to the Pilot Study             | 165  |
III. Correspondence and Materials Related to the Study                  | 193  |
IV. Data Obtained in the Study                                          | 214  |

BIBLIOGRAPHY                                                            | 220  |
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>Total Group of Selected and Nominated Expert Judges</td>
<td>52</td>
</tr>
<tr>
<td>3.</td>
<td>Returns of the Criteria Instrument According to Groups of Judges</td>
<td>53</td>
</tr>
<tr>
<td>4.</td>
<td>Rank Order of the Research Criteria As Rated by the Five Groups of Judges</td>
<td>61</td>
</tr>
<tr>
<td>5.</td>
<td>Distribution of Standard Deviations from the Combined Mean of All Judges for Each Criterion</td>
<td>64</td>
</tr>
<tr>
<td>6.</td>
<td>F Values for the Five Groups of Judges for Each of the Research Criteria</td>
<td>65</td>
</tr>
<tr>
<td>7.</td>
<td>Analysis of the Significant Differences Among Means of the Five Groups of Judges</td>
<td>67</td>
</tr>
<tr>
<td>8.</td>
<td>Mean Number of Years of Research Experience and Mean Number of Studies Directed and Conducted by the Expert Judges</td>
<td>69</td>
</tr>
<tr>
<td>9.</td>
<td>Academic Qualifications of the Expert Judges</td>
<td>70</td>
</tr>
<tr>
<td>10.</td>
<td>Number of Expert Judges Possessing Experience in Various Types of Research</td>
<td>72</td>
</tr>
<tr>
<td>11.</td>
<td>Numbers of Criteria Ranked Within Specified Mean Score Ranges of Importance by Groups of Judges</td>
<td>118</td>
</tr>
<tr>
<td>12.</td>
<td>Qualifications of Pilot Study Judges</td>
<td>167</td>
</tr>
<tr>
<td>13.</td>
<td>Frequency of Responses and Means for Pilot Study Participants</td>
<td>184</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>14</td>
<td>Means for Each Research Criterion as Rated by the Five Groups of Judges</td>
<td>215</td>
</tr>
<tr>
<td>15</td>
<td>Standard Deviations from the Combined Mean of All Judges for Each Criterion</td>
<td>219</td>
</tr>
</tbody>
</table>
CHAPTER I

FORMULATION AND DEFINITION OF THE PROBLEM

Purpose of the Study

... We simultaneously seek absolutely correct answers and absolutely perfect method; both exist as ideals which we never expect to attain but hope constantly to approach. ... ¹

In a serious effort to improve the quality of research in science education, this dissertation was planned as a means of developing and validating criteria for evaluating research. More specifically, criteria for evaluating research were (1) located and extracted from the literature, (2) classified as to use, (3) validated by a group of expert judges, and (4) demonstrated as to use in evaluating research in science education.

The study has a secondary purpose. The responses of expert judges from three areas have been compared in order to discover differences and similarities. The experts were selected from the physical and biological sciences, education, and the sub-specialty of science education.

Need for the Study

How closely the results of a particular piece of research approximate the truth depends upon the quality of that research. Many factors contribute to the quality of a study. These factors should be considered by the researcher as he conducts the study as well as by the consumer of research as he reads the research report. The goal of the researcher should be the control of his method of inquiry—or the factors that contribute to the quality of his study—at all times. Churchman and Ackoff stated,

Suppose a scientist is forced to take a certain step in his inquiry for which he has no alternative. Then this step "leads him," he does not lead it—or analogously, he does not have it under control. Practically in science, this means that he cannot investigate the advisability of his step. In so far as the scientist can examine the adequacy of his steps and make an efficient selection, then he leads or controls his steps, and they do not lead him.

... A method of inquiry, then, is under complete control when every aspect of the activity is itself subject to inquiry with respect to its adequacy for the problematic purpose.²

The present study was designed to contribute to a greater attainment of the kind of control described above.

Historical perspective

The history of educational research during the last sixty-five years can be divided into three periods. Accord-

²Ibid., pp. 10-11.
ing to Barnes, the first of these periods was from 1900 to 1915. During this early period in educational research the "stress was on the development of measuring instruments and techniques and procedures for collecting and interpreting data." Barnes described the succeeding two periods as follows:

From 1915 to 1935 research studies were conducted in almost every facet of the educational program. Instruments were developed; tools and skills were sharpened.

Since about 1940, educational research has undergone a period of self-criticism. Instead of its attainments one reads about its ailments and shortcomings.

Educational research is still in the period of self-criticism noted by Barnes. Because this self-criticism has been active for nearly twenty-five years, and because a sizeable volume of literature has accumulated, a need exists for synthesizing the criticisms. The development of research criteria represents one manner in which such a synthesis could occur.

Dispersal of criteria

Criteria for judging research in science education are spread throughout the professional literature; isolated

---


4 Ibid.

5 Ibid.
criteria may be gleaned from the more general writing concerning research, while entire lists specifically identified as research criteria also may be found. The following discussion points out the need for a synthesis and evaluation of these research criteria.

An example of an isolated criterion is found in an article by Watson, who stated, "Much of our research is miniscule, if not trivial. Too often it deals with the current state of a dynamic system rather than with the factors influencing changes in the system." In contrast to the preceding example, which actually was neither identified by the author as a research criterion nor stated in the usual form of such criteria, a large grouping of enumerated criteria was given by Dvorak. Dvorak's list of twenty-three criteria was compiled for the stated purpose of aiding the graduate student in evaluating completed research and in planning and executing research problems.

Many of the criteria which were found in the professional literature were fragmentary or exhibited only slight variations in meaning. For example, Crawford suggested that

---


"pure" research should be judged for its contribution to existing knowledge, while Abelson\(^9\) suggested that the problem chosen should have either important practical significance or should serve as a foundation for subsequent important practical research. Relative to the same point, Gaylord and Stunkel\(^10\) wrote that "applied" research should be judged or evaluated with reference to the extent to which it meets some socially determined need. These three criteria have been combined into one criterion in this study: the problem chosen should have either direct significance for people's needs or should serve as a foundation for subsequent research having significance for people's needs. This kind of synthesis is necessary if criteria are to be useful for evaluating research.

Lack of previous research

Although many criteria are available in the literature, none has been found which represent the beliefs of a large group of research experts. In general, research criteria which are found in the literature represent the beliefs of a single person or only a few persons—the author or authors of a particular article or book.


During the 1959 Phi Delta Kappa Symposium on Educational Research, Fattu\textsuperscript{11} suggested that research be done in the area of criteria for distinguishing between good and poor research. He further suggested that each of the research criteria be illustrated by a specific example and submitted first to a jury of peers, and then, if warranted, to further study using a scaling technique. This suggestion, together with the factors already mentioned regarding the history of educational research and the wide dispersal of criteria, point out clearly the need for research of the type illustrated by the present study.

\textbf{Processes of Research}

The research criteria which have been developed and validated in this study frequently are stated in terms which permit their application to research in any field. In some cases, however, the criteria are applicable only to certain kinds of research. These less general criteria indicate that research problems differ in nature and that there are more ways than one of carrying out research. The latter conclusion is further supported by the trend away from the phrase, "the scientific method," in favor of such phrases as "scientific methods" and "processes of research." In a

lecture for non-scientists, Conant stated the following:

> It would be my thesis that those historians of science, and I might add philosophers as well, who emphasize that there is no such thing as the scientific method are doing a public service. To my mind, some of the oversimplified accounts of science and its workings . . . are based on a fallacious reading of the history of physics, chemistry, and biology. . . .12

On the other hand the presence of general criteria indicates that research, regardless of the field of study, has certain unifying traits. In order to understand the significance and potential for use of research criteria, the nature of these unifying traits should be examined further.

**Unifying traits of research**

As pointed out earlier, research in education is a relatively young enterprise. Researchers in education wisely have sought assistance from related research areas and from more firmly established disciplines. The notable success of scientific methods of problem solving in the natural sciences served to stimulate interest in those methods in particular. However, to assume that methods of research in one area are always of use in another is open to question. In order to examine this assumption, several statements on this topic are presented in the following paragraphs.

---

In 1929, Dewey wrote:

... The important thing is to discover those traits in virtue of which various fields are called scientific. When we raise the question in this way, we are led to put emphasis upon methods of dealing with subject-matter rather than to look for uniform objective traits in subject-matter. From this point of view, science signifies, I take it, the existence of systematic methods of inquiry, which, when they are brought to bear on a range of facts, enable us to understand them better and to control them more intelligently, less haphazardly and with less routine.13

The key phrase in this statement, "systematic methods of inquiry," could be considered a unifying trait which is common to research in general.

Two more unifying traits are found in the following statement by Cohen and Nagel:

If we look at all the sciences not only as they differ among each other but also as each changes and grows in the course of time, we find that the constant and universal feature of science is its general method, which consists in the persistent search for truth, constantly asking: Is it so? To what extent is it so? Why is it so?—that is, What general conditions or considerations determine it to be so? . . .

.................................................................

The other methods discussed [the methods of tenacity, authority, and intuition] are all inflexible, that is, none of them can admit that it will lead us into error. Hence none of them can make provision for correcting its own results. What is called scientific method differs radically from these by encouraging and developing the utmost possible doubt, so that what is left after such doubt is always supported by the best available evidence. As new evidence or new doubts arise it is the essence of scientific method to incorporate

them—to make them an integral part of the body of knowledge so far attained. Its method, then, makes science progressive because it is never too certain about its results.14

The phrases in this source which constitute unifying traits in research are (1) "the persistent search for truth," and (2) "it is never too certain about its results."

Conant repeated the previously cited trait concerning the search for truth in research, and also indicated an additional unifying feature—the "conviction of the basic importance of free inquiry." He stated:

... But since I believe that in terms of academic history and present practice, both the driving force and the frame of reference are the same for the scholar in the humanities, the social sciences, and the natural sciences, I hope this audience will receive my words with that belief in mind. The unity of the academic world is the unity of a conviction of the basic importance of free inquiry, of the deep significance of the scholar's pursuit of truth, however elusive this truth may appear to be. It is the quest, not the certainty, which all scholars in their heart of hearts must prize most highly, and the joy of the chase unites them all.15

Through his argument that experimentation is not a necessary element of science, Patterson suggested another unifying trait:

... The criterion of a science is not a particular method, such as experimentation, but that the method can be replicated and that the


method leads to the deriving of inferences of conclusions whose bases are acceptable to other trained investigators. This requires the use of variables which can be measured in some way but which are not necessarily subject to direct manipulation or control.\(^{16}\)

Simply stated, the trait suggested above is that scientific methods of research should be capable of replication by independent qualified investigators.

To summarize, the unifying traits of research which were found in the literature just cited are as follows: (1) systematic methods of inquiry, (2) persistent search for truth, (3) uncertainty about research results, (4) belief in the basic importance of free inquiry, and (5) methods replicable by independent and qualified investigators. This is by no means a complete list, as illustrated by the fact that with proper restatement these five traits would be very similar to some of the criteria which have been developed and validated in this study.

In conclusion, the above five traits are sufficient to demonstrate the validity of seeking criteria for evaluating research in whatever literature they may be found, regardless of the area or discipline represented. In addition, support is given thereby to the selecting of expert researchers for judging the criteria from the biological and physical

Creativity in research

The importance of creativity in research commonly is thought to be considerable. Whether one considers creativity with reference to process or with reference to product, it is the opposite of conformity. Torrance chose to define creativity as, "... the process of sensing problems or gaps in information, forming ideas or hypotheses, testing and modifying these hypotheses, and communicating the results." Torrance stated further that

... creativity has also been defined as a successful step into the unknown, getting away from the main track, breaking out of the mold, being open to experience and permitting one thing to lead to another, recombinining ideas or seeing new relationships among ideas, and so on. ...

The creative process may lead, for example, to verbal or to nonverbal products, or to concrete or abstract products.

A question which is of importance in the present study is the relationship between criteria for evaluating research and creativity. Popper suggested that "there is no such thing as a logical method of having new ideas, or a logical reconstruction of this process." If this is true,


18Ibid.

and it does seem reasonable, then research criteria cannot be expected to contribute in a direct way to creativity. In this regard Brenneman suggested the following:

... When you back off from the traditional problem-solving scientific techniques, ignore the ultra-rational logic of the trained investigator and allow yourself the relaxed atmosphere of a non-scheduled, artful approach you just might increase creativity. ...  

This suggestion for increasing creativity is, in fact, almost opposed to the use of research criteria.

A solution to this apparent conflict between creativity and the use of research criteria can be found in a paragraph written by Dewey. He stated the following:

This digression seems to be justified not merely because those who object to the idea of a science put personality and its unique gifts in opposition to science, but also because those who recommend science urge that uniformity of procedure will be its consequence. So it seems worth while to dwell on the fact that in the subjects best developed from the scientific point of view, the opposite is the case. Command of scientific methods and systematized subject-matter liberates individuals; it enables them to see new problems, devise new procedures, and, in general, makes for diversification rather than for set uniformity. But at the same time these diversifications have a cumulative effect in an advance shared by all workers in the field. 

Although Dewey did not mention the word creativity, it is apparent from the previous definitions that he was referring to this concept.

---


It is the writer's hope that the research criteria in this study will be used to liberate the researcher by providing him with the foresight to avoid unnecessary mistakes. Hopefully, a greater command of scientific methods, which use of the criteria may provide, will enable the researcher to "back off" and speculate more freely as well as more frequently. In this manner, creativity in research can be increased, if only indirectly, through the use of research criteria.

Presentation of the Problem

The purpose of the study and the need for developing and validating research criteria have been explored. The absence of previous studies on research criteria has been pointed out. In addition, a theoretical structure consisting of the process of research has been given, into which the function of research criteria was fitted. Finally, it has been suggested that research criteria can contribute to creativity in research.

Statement of the Problem

Stated in the most general terms, the problem of this study centered around the evaluative procedure of distinguish-

22The word, speculate, is used here to mean "thinking in which the factual basis for the hypotheses propounded is slight. Although often used derogatorily, speculation has an important place in discovery and even in verification." See Horace B. English and Ava Champney English, A Comprehensive Dictionary of Psychological and Psychoanalytical Terms (New York: Longmans, Green and Co., 1958), p. 516.
ing between good and poor research. More specifically, criteria which would serve in evaluating completed research and research in progress were developed and validated.

The validation procedure included two questions:

1. What is the order of importance of each of the research criteria as judged by research experts from the biological and physical sciences, from education, and from the sub-speciality of science education?

2. What is the significance of differences in the judgments regarding the research criteria which appear among the research experts from the biological and physical sciences and the research experts from science education and the other areas of education?

A final aspect of the problem involved an illustration of the use of criteria in evaluating research. It was determined that a research report from the area of science education would be selected for evaluation using the validated research criteria.

Definition of terms

The following three terms required definition at this time. All other terms which seem to need comment will be defined or explained as they are used.

1. Research—"A systematic, detailed, and relatively prolonged attempt to discover or confirm the facts that bear
upon a certain problem or problems and the laws or principles that govern it."\textsuperscript{23}

While accepting the preceding statement as a concise and workable definition of research, it is the writer's conviction that the research criteria themselves provide a more adequate and comprehensive, if less manageable, definition of the term "research."

2. Criterion—"A comparison object, rule, standard, or test for making a judgment, especially a qualitative judgment."\textsuperscript{24}

3. Expert judge—A mature, experienced, and skillful researcher who is capable of deciding on the importance of a given criterion for evaluating research.

Working hypotheses

Broadly stated, this study was based on the hypothesis that research in science education can be improved through the use of research criteria. The reaction of expert judges to the research criteria served, in part, to test this hypothesis. In addition, the demonstration of use of the criteria in evaluating research in science education helped indirectly in this regard.

More specifically, the following null hypothesis was tested statistically using the data gathered in this study:

\textsuperscript{23}Ibid., p. 459.
\textsuperscript{24}Ibid., p. 130.
There is no significant difference among the mean ratings of each criterion for the following groups of judges: (1) science education; (2) education, excluding science education; (3) biological science; (4) chemistry; and (5) physics.

Organization of the Remainder of the Study

The remainder of this study is divided into four chapters. The procedures used in developing and validating the criteria and the analytical procedures for treating the data are considered in Chapter II. The data obtained through the use of the criteria instrument are presented and analyzed in Chapter III. Chapter IV contains an example of application of the criteria to research in science education. The final chapter, Chapter V, is devoted to a summary of the study and to conclusions and recommendations.
CHAPTER II

RESEARCH PROCEDURES

The purpose of the study has been discussed previously. The function of the present chapter is to give a detailed description of the procedures that were followed in order to accomplish those purposes.

Development of the Research Criteria

Source of the research criteria

The periodical literature from 1941 through autumn, 1963, was examined for research criteria. This particular period of time was selected because, as was pointed out earlier, these years encompass the period of self-criticism in educational research. A list of the areas covered in the various indexes will help to indicate the sources of the criteria and the scope of the search. Ten categories in the *Education Index* were searched in order to locate potentially fruitful literature. These ten categories included "action," "biological," "chemical," "higher education," "historical," "physical," "psychological," "scientific," and "social" research as well as all categories under the heading of "research." *The Readers Guide* listed pertinent literature under only two headings, "scientific research" and "research,"
from 1941 to 1960, and under only one heading, "research," since 1960. The union Catalog of The Ohio State University was searched under five headings—"methodology," "methods of research," "research in education," "research procedures in education," and "research in science."

Although no attempt was made to limit the literature selected for study to any one discipline or area, most of the research criteria were found in the literature of education, including science education. In addition, the psychological literature was somewhat useful, as was the field of philosophy and logic. As Masson pointed out, "The strictly technical literature of the various experimental sciences is almost barren of specific information concerning research procedures, except occasionally a vague reference to the scientific method without, however, amplification as to how it is used as part of the research process."\(^1\)

The number of articles devoted to a critical analysis of the methods of research was small in 1941. The rate increased rapidly to the present high level of concern expressed in numerous books and articles. The next problem, after locating the many criteria which were available, involved selecting and functionally organizing them so as to be most useful for evaluating research.

Selection of the research criteria

Each criterion which was extracted from the literature was restated if necessary, placed on a small card, and filed according to a tentative organizational outline. Whenever a given criterion was found to be repeated in a second or succeeding source, this fact was indicated on the card. A simple code system was used for this purpose. When new criteria no longer were being located, the literature search was halted.

Careful examination of all the research criteria showed that many of them could be grouped into sets composed of a single statement followed by one or more examples or closely related statements. To illustrate, the criterion, "The problem, from the beginning, should condition the steps taken to solve it," has four related statements as follows:

1. Research should display a measure of creativity by dividing new techniques and new approaches to problems.

2. "Quantitative," "statistical," and "experimental" should not be set in opposition to "qualitative," "clinical," and "nonexperimental"--a balanced

---


4Fattu, pp. 1-21.
approach in selecting techniques appropriate for solving the problem at hand is needed.\textsuperscript{5}

3. If a unique competency or position possessed by the researcher lends itself to the solution of the problem that competency should be used.\textsuperscript{6}

4. The researcher should recognize the many subtle but fundamental differences in the situations involved before he attempts to apply intact to a new field or problem a design or technique which has been used successfully in or with another.\textsuperscript{7}

A complete list of such related statements, each under the appropriate research criterion, is given in Appendix I.

The process of distilling the research criteria to their important fundamental elements was followed by submitting them to a group of science education graduate students. The resultant comments and responses were used in developing further the research criteria instrument, and in writing a cover letter. In addition, a single page questionnaire was devised to elicit information about the judges and about their reactions to the criteria and the instrument. The revised instrument then was duplicated and used in a pilot study.

\textsuperscript{5}Carter V. Good, "The Methodology of Educational Research and Scholarship," School and Society, 91 (March, 1963), 140-143.


Organization of the research criteria

As stated previously, it was assumed that a functional scheme for organizing the research criteria would be of greatest value to the researcher. A "functional scheme" was defined for the purpose of this study as one which permitted the evaluation of research in progress as well as the evaluation of completed research reports. However, it was pointed out in Chapter I that there is no such thing as the scientific method, but rather scientific methods of research. Thus, no organizational scheme could be expected which would follow the stepwise and, hopefully, creative process which an individual researcher might follow in a particular investigation. Greater success could be expected in an attempt to organize the criteria according to a scheme which permitted the evaluation of research reports. Although research reports differ in their organization, they generally follow, to a greater or lesser extent, the typical "steps" of "the scientific method." The point of view expressed by Brown and Ghiselli in the following paragraph was adopted in deciding upon an organizational scheme:

The scientific method does not consist of a single set of steps followed in some invariable chronological order. To a large extent, the chronology of a scientific experiment is tied to the ups and downs of the scientist's motivations. At some time in the early training of a student of science, however, there is need for a systematic treatment of those steps that are fundamental to any scientific attack upon a problem. In presenting such a
treatment, no claim should be made that the steps are well delineated one from another or that the order of the steps is established regardless of the nature of the scientist's progress. The chronology of the steps given herein is to be interpreted as a serviceable one and as a frequently used one, but not as an invariable one.

With the reservations expressed in this quotation, the "steps" of the scientific method were used as a pattern for evolving the particular elements of the organizational scheme for the research criteria. The final scheme which was used is as follows:

Formulation and Definition of the Problem

Problem Statement
Introduction
Problem Choice
Delimitation and Analysis
Assumptions
Definitions
Related Literature
Hypotheses and Derived Theorems

Procedural Design

Procedure
Design Selection
Design for Replication and Validity
Self-Contained Design
Design for Data Gathering
Special Methods
Instrumental Techniques
Statistical Techniques

Presentation of Evidence

Presentation of Data
Analysis of Data

---

8 Brown and Ghiselli, p. 131.
Summary and Conclusions

Summary
Conclusions
   Formulation and Development
   Limitations
   Applications, Recommendations, and Needed Research

This scheme of organization was devised to enable the beginning researcher to deviate at will from a predetermined research plan. The logical organization of the criteria facilitates referral to them as a study is being conducted. Thus, the researcher can give full expression to his own creative impulses with the knowledge that he can turn to the research criteria for their evaluative assistance when needed. Of more importance, the researcher can use the criteria for evaluating research reports, and perhaps, through this process, increase his research skills.

Pilot study

In order to refine further the research criteria instrument, a pilot study was conducted. A list of pilot study judges together with some of their qualifications may be found in Appendix II. In addition, copies of the two cover letters, pilot study research criteria, judges' information page, letter of thanks, and a summary of the results have been placed in Appendix II.
Five of the nine judges who participated in the pilot study were selected from the science areas, and four from the area of education. Of the five science judges, three were chemists, and two were biologists. Two of the four education judges were science educators. All of the education judges had had experience in experimental classroom research, three had had experience with descriptive studies and one with clinical studies. One of the education judges had had experience in field research in biology. As might be expected, all the science judges were experienced with laboratory research in their fields. One of the science judges also had had experience with classroom and field research, descriptive research, and with the case study approach. Additional information concerning the qualifications of the pilot study judges may be found in Appendix II.

The frequency of response to each level of importance (great importance, high importance, average importance, little importance, and no importance) for each criterion is listed in the summary of the results of the pilot study, found in Appendix II. In addition, the mean response values for the education judges and for the science judges for each criterion are listed. Further analysis of the data was not attempted because of the small number of pilot study respondents. Visual examination of the pilot study data indicated, however, that significant differences between the scale responses of education judges and science judges might
be expected in the main study on a number of the criteria.

In addition, comments made by the pilot study judges indicated that mixed reaction to the usefulness of the criteria might be expected from the expert judges in the study proper. Six of the judges in the pilot study felt that the criteria would be useful for evaluating research. Two of the judges had reservations which might be expressed by the statement of one of them, "In a practical sense, research is evaluated against one's experience subjectively, rather than item by item in criteria." One of the judges did not wish to generalize the research process to the extent called for by the criteria. As expected, several of the judges qualified their responses to particular criteria with reference to the kind of research being evaluated. One of the judges stated that he assumed the researcher would select the criteria which applied to a given research method. The latter assumption was expected and was incorporated into the instructions for the study instrument.

Based on the results of the pilot study, several of the criteria were revised, where clarity appeared to be lacking, and five were omitted. In addition, a revised cover letter and set of instructions were written. It should be pointed out, however, that the form for obtaining responses was unchanged from that of the pilot study. The research criteria themselves will be presented in the next section of this chapter.
Research criteria used in the study

The research criteria which were developed for validation in this study came from many sources. In order to give credit to those sources, the criteria are listed in this section together with proper citations. It should be noted that the criteria listed below may not be stated in a form with which the cited author or authors would agree; in some cases the idea for a criterion, not the criterion itself, was obtained from the cited source. The criteria and their organizational headings are listed in the same order in which they appeared in the study instrument. Only the five forced-response categories of degree of "importance" and the boxes for checking the desired response have been omitted. As stated in the previous section, the form for responding was identical in both the pilot study and the study proper. The criteria are as follows:
FORMULATION AND DEFINITION OF THE PROBLEM

Problem Statement

1. The problem should be stated precisely in clear and understandable language.9,10,11,12

2. The problem statement should include all subordinate problems which must be answered.13,14,15,16

3. The problem statement, including all of its parts, should encompass all elements of the problem.17

4. The location of the problem statement in the research report should be identified so that it can be found and recognized readily.18

---


11Francis D. Curtis, "What Constitutes a Research Investigation in Science Education?" Science Education, 37 (February, 1953), 53-54.


13Abelson, pp. 308-309.


18Ibid.
Introduction

Problem Choice

5. The origin of a research problem should be identified.

6. The body of organized knowledge in an area should be examined for gaps, for discrepancies, and for the directions in which extension is possible. 19, 20

7. A problem should be chosen which avoids unnecessary duplication of research effort. 21

8. A problem should be chosen which can be formulated and defined with sufficient precision to permit a good idea of the goal to be achieved before it is attempted. 22

9. Attention should be given to the theoretical framework of knowledge underlying issues and problems before their resolution or solution is sought directly. 23

10. The problem chosen should have either direct significance for people's needs or should serve as a foundation for subsequent research having significance for people's needs. 24, 25

19 Abelson, pp. 20-21.


22 Crawford, p. 23.


11. Experimentation involving human subjects should be planned and spaced carefully for their protection.  

Delimitation and Analysis

12. The problem should be delimited exactly as to scope, the nature and quantity of data desired, and the methods of procedures to be followed.  

13. In addition to simple delimitation of the problem, the subjects (if any are used) should be described with reference to the variables among them.  

---


27 Abelson, pp. 308-309.


29 Brooks, School and Society 18 (December, 1923), 724-729.


14. The factors which are selected for study should be the decisive ones for solution of the problem.\(^{35}\)

15. Variables which are used in experimental research should differentiate between or among the subjects or objects under study.\(^{36}\)

Assumptions

16. Assumptions upon which the inquiry rests should be identified.\(^{37},^{38},^{39},^{40}\)

17. Assumptions upon which the inquiry rests should be justified.\(^{41},^{42},^{43},^{44}\)

18. The assumptions in a study should be consistent with each other.\(^{45}\)


\(^{36}\)Gaylord and Stunkel, Educational and Psychological Measurement, 14 (Summer, 1954), 294-300.

\(^{37}\)Barnes, p. 87.

\(^{38}\)Crawford, p. 278.


\(^{40}\)Van Dalen, Educational Administration & Supervision, 44 (May, 1958), 174-181.

\(^{41}\)Barnes, p. 87.

\(^{42}\)Crawford, p. 278.


\(^{44}\)Van Dalen, Educational Administration & Supervision, 44 (May, 1958), 174-181.

\(^{45}\)Ibid.
19. It should not be assumed that the non-research literature is an accurate indicator of best or prevalent practice.46

Definitions

20. A definition should be expressed in precise language.47

21. A definition must not be circular; it must not contain, directly or indirectly, the subject to be defined.48

22. A definition should be stated in positive terms, rather than in negative terms, wherever possible.49

23. Important terms, that are employed in an unusual sense, should be defined.50,51

24. Undefined terms should be employed according to the meanings demanded by standard references.52,53

25. Operational definitions of all generalized concepts which will be studied should be specified.54,55


48 Cohen and Nagel, p. 128.

49 Ibid.

50 Abelson, pp. 308-309.


52 Ibid.

53 Abelson, pp. 308-309.


55 Churchman and Ackoff, p. 198.
26. Defined terms and concepts should be used in the entire study according to the given definitions.

Related Literature

27. The related literature should be analyzed for its relevancy to the problem situation.

28. The related literature should be analyzed for any bias which may have significance for the study.

29. The review of related literature should include the previous research which may contribute to the solution of the problem.

30. The problem should be related to an orienting frame of reference or a systematic area of study.


57 Barnes, p. 86.


61 Abelson, pp. 308-309.

62 Brooks, School and Society, 18 (December, 1923), 724-729.


31. Research in a given area should make use of relevant principles and research techniques of related disciplines.67

32. Literature sources should be assigned to definite authors and times,68,69,70,71

Hypotheses and Derived Theorems

33. The researcher should display creativity in his formulation of hypotheses.72,73,74

34. A hypothesis should bring together a given number of observations and interpret the connections between them.75

35. The possible alternative answers which are hypothesized for a given question should be independent of one another (or not overlap).76,77

68Barnes, pp. 110-113.
69Perdew, pp. 82-83.
71Matthew J. Whitehead, "The Place of Subjectivity in Research," The Quarterly Review of Higher Education Among Negroes, 12 (April, 1944), 63-64.
74Pella, Science Education, 45 (December, 1961), 396-399.
76Churchman and Ackoff, p. 226.
77Symonds, Journal of Educational Psychology, 47 (February, 1956), 100-109.
36. Hypotheses should be formulated on the basis of a thorough analysis of the theoretical and factual background of the problem.78

37. Hypotheses should be formulated so that implications can be traced clearly by means of well-established techniques of deduction.79

38. The consequences of alternative hypotheses should be developed for comparison with observable phenomena.80,81,82

39. Hypotheses should be revised on the basis of induction from the data which accumulate as the study progresses.83,84,85

PROCEDURAL DESIGN

Procedure


79Cohen and Nagel, p. 97.

80Ibid., p. 86.


Design Selection

40. The problem, from the beginning, should condition the steps taken to solve it. 86, 87

41. The adequacy of the research methodology chosen for solution of the problem should be explained and defended. 88, 89, 90, 91

42. The research design should provide for collection of the most objective data available, 92, 93, 94

43. The research design should include several planned directions of inquiry in order to increase efficiency and comprehensiveness. 95

86 Brown and Ghiselli, p. 6.
87 Monroe and Engelhart, pp. 444-445.
88 Abelson, pp. 308-309.
89 J. V. Breitwieser, "Essentials of Good Research," Phi Delta Kappan, 24 (December, 1941), 185.
91 Brooks, School and Society, 18 (December, 1923), 724-729.
93 Fox, Phi Delta Kappan, 39 (March, 1958), 284-286.
44. The experimental design should include the replication of one or more variables whose effects have already been assessed as significant in a previous experiment.96,97

Design for Replication and Validity

45. The research design should be described in sufficient detail to permit repetition of the study by independent investigators.98,99

46. The degree of internal and external validity in the research design should be identified and justified.100

47. The data collected should afford some basis for generalization.101,102,103,104

---

96Barnes, p. 191.
102Perdew, pp. 82-83.
103Perdew, Phi Delta Kappan, 32 (December, 1950), 134-136.
Self-Contained Design

48. The self-contained experiment should include a representative sample of the subjects or objects under study.\textsuperscript{105}

49. The self-contained experiment should include a control to exclude, at a known level of probability, a number of alternative interpretations of the experimental results.\textsuperscript{106,107,108}

50. The self-contained experiment should supply independent evidence on the question in dispute.\textsuperscript{109,110}

51. The self-contained experiment should permit an estimate of the experimental errors which affect the comparisons made.\textsuperscript{111,112,113}

Design for Data Gathering


\textsuperscript{106} \textit{Ibid.}

\textsuperscript{107}Symonds, \textit{Journal of Educational Psychology}, 47 (February, 1956), 100-109.


\textsuperscript{110}Anderson, \textit{Science Education}, 37 (February, 1953), 55-61.

\textsuperscript{111} \textit{Ibid.}

\textsuperscript{112}Brown and Ghiselli, p. 88.

52. The design for making observations should establish causal connections between events in which the cause is a sufficient and necessary condition for the effect.\textsuperscript{114}

53. The data of research should approach the reality of the situation being investigated as nearly as possible.\textsuperscript{115}

54. Modern technological instruments should be used to make complete visual and/or oral records of primary data for review by independent investigators.\textsuperscript{116}

55. The data collected should be within the boundaries defined by the hypotheses.\textsuperscript{117}

56. The data included within the defined boundaries of the study should be collected without personal choice.\textsuperscript{118}

57. Data should be gathered in a well organized, highly structured, and yet creative manner.\textsuperscript{119}

58. The collection of data or evidence should include a thorough investigation of primary data.\textsuperscript{120}

Special Methods

Instrumental Techniques

\textsuperscript{114}Churchman and Ackoff, 138-139.

\textsuperscript{115}Cogan, \textit{The Journal of Teacher Education}, 14 (September, 1963), 238-243.

\textsuperscript{116}Ibid.

\textsuperscript{117}Barnes, pp. 117-118.

\textsuperscript{118}Cohen and Nagel, p. 91.

\textsuperscript{119}Barnes, pp. 110-113.

\textsuperscript{120}Ibid.
59. A data-gathering instrument should have validity or truthfulness; it should measure what it purports to measure.\textsuperscript{121,122}

60. A data-gathering instrument should have reliability or consistency; it should agree with itself when used repeatedly with the same individuals, groups, or objects,\textsuperscript{123,124}

61. The degree of reliability and validity which is accepted should be justified with reference to the objectives of the research problem.

62. The most effective instrument for collecting the necessary data should be anticipated in the research design.\textsuperscript{125}

\textbf{Statistical Techniques}

63. The statistical technique should be identified before the data are gathered.\textsuperscript{126,127,128}

\textsuperscript{121}Ibid., pp. 130-131.
\textsuperscript{122}Granville B. Johnson, Jr., \textit{The Journal of Educational Research}, 51 (October, 1957), 149-151.
\textsuperscript{123}Ibid.
\textsuperscript{124}Barnes, pp. 110-113.
\textsuperscript{127}Curtis, \textit{Science Education}, 37 (February, 1953), 53-54.
\textsuperscript{128}Pella, \textit{Science Education}, 45 (December, 1961), 396-399.
64. The statistical hypotheses should be considered at the time the experiment is designed.129,130

65. The statistical design should include as many independent variables in the same design as may seem necessary.131,132,133

66. The statistical design should include as many independent variables in the same design as may seem necessary.134,135,136

67. The basic assumptions underlying statistical methods should be satisfied by the data to which they are applied.137,138

68. The statistical sample should be selected from a carefully defined parent population.139,140

---


133Good, School and Society, 91 (March, 1963), 140-143.

134Ibid.


137Ibid.

138Johnson and Jackson, p. 15.


69. The probability of error estimate should be justified.\textsuperscript{141}

70. The statistical technique should include tests for the reliability and validity of the data.\textsuperscript{142,143,144}

71. The statistical technique should allow for maximum objectivity in evaluation and in estimation.\textsuperscript{145}

72. The statistical technique should provide tests to assure that a sufficient quantity of data is collected.\textsuperscript{146}

73. The statistical technique should allow consideration for tests of significance and problems of estimation.\textsuperscript{147,148}

PRESENTATION OF EVIDENCE

Presentation of Data

74. The data should be logically organized and presented.\textsuperscript{149,150}

\textsuperscript{141}Fox, \textit{Phi Delta Kappan}, 39 (March, 1958), 284-286.
\textsuperscript{142}Curtis, \textit{Science Education}, 37 (February, 1953), 53-54.
\textsuperscript{143}Dvorak, \textit{Peabody Journal of Education}, 34 (November, 1956), 141-144.
\textsuperscript{145}Anderson et al., \textit{Science Education}, 38 (December, 1954), 333-365.
\textsuperscript{146}Dvorak, \textit{Peabody Journal of Education}, 34 (November, 1956), 141-144.
\textsuperscript{147}Anderson, \textit{Science Education}, 38 (December, 1954), 390-397.
\textsuperscript{148}Lindquist, p. 3.
\textsuperscript{149}Perdew, pp. 82-83.
\textsuperscript{150}Perdew, \textit{Phi Delta Kappan}, 32 (December, 1950), 134-136.
75. All data which are collected for the solution of a given problem should be reported.

76. The sources of all the facts which are presented should be identified. 151, 152

77. The proof necessary to validate important alleged facts should be provided. 153

78. The apportionment of space in the research report should be made with reference to the judged relative importance of the topics treated.

79. When graphical methods are used, they should present the data simply and accurately, both in construction and in the impression given to the reader. 154

Analysis of Data

80. The analysis of data should be anticipated and worked through abstractly or in a pilot study before the study itself is begun. 155

81. Preconceived ideas about the results to be expected should not affect the manipulation of the data. 156

82. Accurate statistical or speculative methods of organizing the data should be employed. 157

151 Ibid.
152 Perdew, pp. 82-83.
153 Abelson, pp. 308-309.
154 Johnson and Jackson, p. 58.
157 Abelson, pp. 308-309.
83. Evidence which is collected in a study should be creatively and meaningfully synthesized.\textsuperscript{158}

84. Any accidents of circumstance which might affect the reliability or validity of the data should be reported.\textsuperscript{159}

85. Corrections should be made in the analysis of the data for any distortions which may have occurred in the data gathering process.\textsuperscript{160}

**Interpretation of Data**

**Factors Involved**

86. The interpretation of data should involve intuition and rational subjective evaluation.\textsuperscript{161}

87. The relationship between a hypothesis or a theory and the observational evidence, when both are given, should be judged.\textsuperscript{162}

88. The reliability of a prediction, which is based either on a hypothesis or a theory, should be judged on the basis of the evidence.\textsuperscript{163}

89. Both statistical significance and practical significance should be considered in the interpretation of research results.\textsuperscript{164}


\textsuperscript{159}Barnes, pp. 131-135.


\textsuperscript{161}Churchman and Ackoff, p. 100.

\textsuperscript{162}Carnap, Science, 104 (December, 1946), 520-521.

\textsuperscript{163}Ibid.

90. Observation and theory should serve as checks on one another with neither having priority over the other.165

91. The interpretation of the data should be guided by their source and nature.166

Evaluation of Theories and Hypotheses

92. A theory should encompass a relatively wide range of data.167

93. A theory should provide a basis for successful prediction.168

94. A deductively elaborated hypothesis should be factually substantiated before it is used as an explanatory device.169

95. A necessary and sufficient condition for the rejection of a theory or a hypothesis is the observation of reproducible contradictory evidence.170

96. The assumptions which led to the experimentally testable propositions should be carefully examined before the hypothesis is tentatively accepted or rejected.171

97. A hypothesis should make possible the determination of relationships between the known facts.172

165Churchman and Ackoff, p. 227.
166Johnson and Jackson, p. 15.
168Ibid.
170Popper, p. 86.
171Cohen and Nagel, p. 110.
172Ibid., p. 92.
SUMMARY AND CONCLUSIONS

Summary

98. The summary should be an overview of the entire study.\textsuperscript{173}

99. The summary should be self-contained.\textsuperscript{174}

100. The summary should include only previously reported data.\textsuperscript{175}

101. The summary should avoid unnecessary enumeration of details reported earlier in the manuscript.\textsuperscript{176}

102. The summary should be a final synthesis of the whole study which is greater than the mere process of adding together chapter or section summaries.\textsuperscript{177}

Conclusions

Formulation and Development

103. Conclusions should be drawn with reference to the problematic situations that engendered the inquiry.\textsuperscript{178,179}

104. Conclusions should be tested by examining them for internal consistency,\textsuperscript{180,181}

\textsuperscript{173}Good and Scates, p. 851.
\textsuperscript{174}Ibid.
\textsuperscript{175}Ibid.
\textsuperscript{176}Ibid.
\textsuperscript{177}Ibid.
\textsuperscript{178}Abelson, pp. 108-109.
\textsuperscript{179}Barnes, p. 88.
\textsuperscript{180}Crawford, p. 279.
\textsuperscript{181}Smith and Smith, pp. 107-108.
105. Conclusions should be tested by examining them for external consistency.182,183

106. Conclusions should include inferences relative to all the significant data.184,185

107. When several possible explanations of a phenomenon seem equally correct, the simplest should be favored.186,187,188

108. Skillful grouping and appropriate subordination of conclusions under major headings should be used to avoid encyclopedic enumeration of individual statements.189

109. The findings of research should be stated with simplicity within the context of a total pattern or theory.190

Limitations

110. The limitations of a study should be presented to guide the reader in using the findings, conclusions, and recommendations.191

182Ibid.
183Crawford, p. 279.
184Smith and Smith, pp. 107-108.
185Abelson, pp. 308-309.
186Crawford, p. 282.
187Fletcher Watson, "Research in the Physical Sciences," Phi Delta Kappan, 35 (October, 1953), 4-6.
189Good and Scates, p. 852.
190Bronowski, Scientific American, 199 (September, 1958, 59-65.
111. A study should be thorough enough to arrive at carefully stated results which have a substantial degree of reliability and validity.\(^{192}\)

Applications, Recommendations, and Needed Research

112. Generalizations should be directed to the solution of problems other than those considered in the study.\(^{193}\)

113. Both the data and the conclusions should warrant the recommendations made.\(^{194},^{195},^{196}\)

114. Creative speculation concerning possible applications and recommendations should be made.

115. Applications or recommendations which are based on the creative insight of the investigator should be labeled plainly as such.\(^{197}\)

116. Research should contribute to the identification and to the solution of new problems.\(^{198},^{199},^{200},^{201}\)


\(^{193}\)Brown and Ghiselli, p. 236.

\(^{194}\)Curtis, *Science Education*, 37 (February, 1953), 53-54.


\(^{196}\)Obourn and Boeck, *Science Education*, 44 (December, 1960), 374-399.

\(^{197}\)Good and Scates, p. 852.

\(^{198}\)Crawford, p. 21.

\(^{199}\)Watson, *Phi Delta Kappan*, 35 (October, 1953), 4-6.


\(^{201}\)Whitney, p. 179.
Validation of the Research Criteria

Selection of expert judges

With the assistance of the chairman of the dissertation committee, nineteen judges were selected from the area of science education. With one exception, these nineteen judges were selected from institutions which have formal doctoral programs in each of the areas from which judges were selected, namely, education, science education, biology, chemistry, and physics. The exception, George Peabody College, does not have a graduate program in the physical and biological sciences. However, because of the close relationship between George Peabody College and Vanderbilt University, judges from the science areas were obtained from the latter institution.

In order to obtain nominations of expert judges from the areas of education, biology, chemistry, and physics, letters were sent to the heads of the departments or areas involved. A duplicate letter also was included, which was to be given to the person or persons nominated so that they might be aware of their nomination. Copies of these letters may be found in Appendix III. In order to obtain the same number of judges from the science areas as from the education areas, one nomination each in biology, physics, and chemistry, and two nominations from the education area, excluding science education, were requested. In other words, a total of six judges, three from the sciences and three from all
areas of education, was desired from each of the nineteen selected institutions. The results of this nomination technique appear in Table 1. To summarize briefly, 26/38 or over 68% of the education judges, 10/19 or approximately 53% of the biological science judges, 9/19 or approximately 47% of the chemistry judges, and 6/19 or nearly 32% of the physics judges were nominated by the above process.

In order to complete the nominations in those areas where judges had not been obtained, personal letters were sent to the selected science education judges. A copy of one of these letters, all of which were similar in form, may be found in Appendix III. In addition to being asked to indicate a willingness to serve as a judge, each selected science education person was asked to nominate, from his own institution, expert judges in the areas where nominations had not been received. All nineteen of the selected science education judges responded to this initial letter, with only one indicating unwillingness to participate in the study (due to lack of time). As a result of the nominations provided by the science education judges, the total number of judges nominated in each area was as follows: (1) education, 38/38 or 100%; (2) biology, 18/19 or 95%; (3) chemistry, 18/19 or

202In this case and in future references to the selected institutions, George Peabody College and Vanderbilt University will be referred to as if they were one institution.
TABLE 1

EXPERT JUDGES NOMINATED BY HEADS OF AREAS OF EDUCATION, BIOLOGY, CHEMISTRY, AND PHYSICS

<table>
<thead>
<tr>
<th>Institution</th>
<th>Education</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston University</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Columbia University</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>University of Connecticut</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Cornell University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Florida State University</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Harvard University</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>University of Illinois</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana University</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>State University of Iowa</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Kansas</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Michigan State University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>University of Minnesota</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Washington Square College, N.Y.U.</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Northwestern University</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oregon State University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>George Peabody College and Vanderbilt University</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>University of Pittsburgh</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanford University</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Wisconsin</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

*aEach X represents the nomination of one expert judge.*
95%; and physics, 17/19 or 90% of the number of judges initially requested. In a few cases, more than the requested number of judges were nominated from a given institution. The total selection and nomination results are summarized in Table 2, and a list of the judges may be found in Appendix III.

**Mailing and follow-up procedure**

The cover letter and instrument were duplicated by the offset press method. By using this method of duplication, the 116 criteria, together with the instructions to judges and the judges' information page, were compressed into a brief four-sheet 8 1/2 by 11 inch center-stapled booklet. Each judge received in the first mailing (1) a cover letter, (2) a copy of the research criteria instrument for his own files, and (3) a copy of the research criteria instrument which was stapled into a heavy paper cover complete with return address and postage for mailing. A copy of the cover letter and the research criteria instrument may be found in Appendix III.

The first follow-up letter, a copy of which may be found in Appendix III, was mailed a little over two weeks after mailing the criteria instrument. At the time that the first follow-up letter was mailed, a return of 43.1% had been received. In terms of the various areas involved, this return was analyzed as follows: science education, 66.7%;
TABLE 2
TOTAL GROUP OF SELECTED AND NOMINATED EXPERT JUDGES

<table>
<thead>
<tr>
<th>Institution</th>
<th>Science</th>
<th>Education Education</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston University</td>
<td>X&lt;sup&gt;a&lt;/sup&gt;</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Columbia University</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>University of Connecticut</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Cornell University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Florida State University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Harvard University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Illinois</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Indiana University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>State University of Iowa</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Kansas</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Michigan State University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>University of Minnesota</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Washington Square College, N.Y.U.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Northwestern University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Oregon State University</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>George Peabody College and Vanderbilt University</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>University of Pittsburgh</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Stanford University</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Wisconsin</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

<sup>a</sup>Each X represents the nomination or selection of one expert judge.
education, 52.6%; biology, 33.3%; chemistry, 27.8%; and physics, 23.5%.

A second follow-up letter together with another copy of the research criteria instrument was mailed five weeks after the first mailing. A copy of this follow-up letter also may be found in Appendix III. At the time the second follow-up letter and additional copy of the instrument were mailed, a total return of 73.4% had been received, and was categorized as follows: science education, 77.8%; education, 68.4%; biology, 77.8%; chemistry, 72.2%; and physics, 76.5%.

Approximately two months after the first mailing, analysis of the responses was begun. A summary of the total and usable return at that time, according to groups of judges, is presented in Table 3.

TABLE 3

RETURNS OF THE CRITERIA INSTRUMENT
ACCORDING TO GROUPS OF JUDGES

<table>
<thead>
<tr>
<th>Groups of Judges</th>
<th>Total Returns</th>
<th></th>
<th>Usable Returns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Science Education (18)</td>
<td>18</td>
<td>100</td>
<td>17</td>
<td>94.5</td>
</tr>
<tr>
<td>Education (38)</td>
<td>31</td>
<td>81.6</td>
<td>24</td>
<td>63.2</td>
</tr>
<tr>
<td>Biology (18)</td>
<td>14</td>
<td>77.8</td>
<td>12</td>
<td>66.7</td>
</tr>
<tr>
<td>Chemistry (18)</td>
<td>16</td>
<td>88.9</td>
<td>12</td>
<td>66.7</td>
</tr>
<tr>
<td>Physics (17)</td>
<td>13</td>
<td>76.5</td>
<td>7</td>
<td>41.2</td>
</tr>
<tr>
<td>Total (109)</td>
<td>92</td>
<td>84.4</td>
<td>72</td>
<td>66.1</td>
</tr>
</tbody>
</table>
Limitations in the Development and Validation of the Research Criteria

As the study developed, certain limitations became evident. They were as follows:

1. The process of research and the research report, although related, are not identical. The usefulness of the criteria was limited by the choice which had to be made between process and report as the criteria were written and organized.

2. The validation of the criteria was limited by the difficulty of communication due to discipline-related differences in terminology. Examples of application were not given to the expert judges for each criterion because of the prohibitive amount of time that would have been required to read and evaluate the material.

3. The validation of the criteria was limited by the amount of care which was taken by the expert judges in evaluating each criterion.

Assumptions in the Development and Validation of the Research Criteria

The following assumptions were necessary in the development and validation of the criteria:

1. It was assumed that the criteria should be organized in a pattern which would permit their application first to research reports and only indirectly to the process of conducting research. This assumption was justified previously.
2. It was assumed that the development of the research criteria, as well as their validation, would be more complete and useful if an example of their application to research in science education was given. Because the criteria will be used mainly by inexperienced researchers, the purpose of the example is to clarify the meaning of as many criteria as possible, and thus to promote their application and educational value.

Analytical Procedures

Ranking of the criteria

In order to rank the criteria according to the importance expressed by the rating given by all of the judges and by each group of judges, the arithmetic mean was computed for each criterion. Numerical values of 4 for "great importance," 3 for "high importance," 2 for "average importance," 1 for "little importance," and 0 for "no importance" were used. This arbitrary system of weighting each of the five possible response categories is commonly accepted in preference to more complicated techniques.203,204


Test for consistency

To measure the consistency of the rating by the judges, the standard deviation of the distribution of the ratings from the mean was computed for each criterion as rated by all judges.

Test for significance of differences

The factors in this study indicated the use of an analysis of variance technique. The process of analysis of variance "might better be termed analysis of means for our objective is the comparison of means among groups."205 Analysis of variance, or the F test, is "a method for determining whether the differences (expressed as variance) found in a dependent variable, when it is exposed to the influence of one or more experimental variables, exceed what may be expected by chance."206 In order to discover if the differences between means exceeds what may be expected by chance, it is customary to establish certain levels of significance which must be met before a null hypothesis (that no actual differences between means exist) will be rejected; in the present study these levels were set at 1 percent and 5 percent. For example, if the differences found are at the 5 percent level, then these differences could result by chance alone only five times in one hundred, and consequently the

205Ibid., p. 255.

206English and English, p. 28.
null hypothesis could be rejected at that level. The level of significance is actually "the probability of rejecting a hypothesis when it is, in fact, true." ^207

Summary

The method of development of the research criteria was discussed in detail in this chapter. It was stated that the criteria were extracted from books and from the periodical literature published mainly between 1941 and 1963. The selection, synthesis, and organization of the research criteria into a meaningful system of classification was described. A pilot study was conducted as a final step in the research criteria development process. The resulting 116 research criteria were listed together with proper citations of all sources.

The validation processes included the selection and nomination of 109 expert judges from the various areas of education and from the biological and physical sciences. An attempt was made in this selection process to obtain the services of as many judges from all the areas of professional education combined as from three areas of science combined. Two follow-up letters were used to obtain a greater response. The number of usable returns which were received in time to be included in the study was 72, or 66.1% of the number of

expert judges contacted. This return was analyzed as follows: science education judges, 94.5%; education judges, 63.2%; biology judges, 66.7%; chemistry judges, 66.7%; and physics judges, 41.2%.

Three limitations and two assumptions in the development and validation of the research criteria were listed. Finally, the analytical procedures were described. These procedures included the following: (1) ranking the criteria according to the importance expressed by the mean rating given by all of the judges and by each group of expert judges, (2) a test for consistency of rating using the standard deviation of the distribution of the ratings from the mean, and (3) a test for significance of differences using the analysis of variance technique.
CHAPTER III

PRESENTATION AND ANALYSIS OF DATA

An analysis of the data from the research criteria rating instrument is provided in this chapter. The data were of three types—statistical, descriptive, and subjective. The principal concern in this study was to determine the order of importance of a list of criteria for evaluating research as seen by five groups of judges. In addition, the data were analyzed statistically for the significance of differences in the judgments (about the research criteria) which appeared among the research experts from the biological and physical sciences, from science education, and from the other areas of education. As an aid in understanding and applying the statistical findings, descriptive data about the expert judges, and subjective data consisting of unstructured comments by the judges, were solicited and analyzed.

Statistical Data

Order of importance of the research criteria

In order to rank the criteria according to the importance expressed by the rating given by all of the judges combined and by each of the five groups of judges separately,
the arithmetic means were computed for each criterion. The mean scores for each criterion are listed in Table 14 in Appendix IV. Numerical values of 4 for "great importance," 3 for "high importance," 2 for "average importance," 1 for "little importance," and 0 for "no importance" were used. The mean scores ranged from a high of 4.00 to a low of 0.86.

The rank orders of the research criteria are listed in Table 4. In this table each research criterion is identified by the number assigned to it in the original instrument and in the preceding chapter. To illustrate the use of Table 4, research criterion number 1 is given a rank of 2 by the science education judges, 1 by the education judges, 3 by the biology judges, 8 by both the chemistry and physics judges, and 1 by all five groups combined.

From Table 4 an answer can be found to the question: What is the order of importance of each of the research criteria as judged by research experts from the biological and physical sciences, from education, and from the sub-specialty of science education? The 116 research criteria submitted to the five groups of judges for evaluation have been given a rank order number of importance by each group and by all groups combined.

Consistency of ratings

To obtain a measure of the consistency of the ratings by the judges, the standard deviation of the distribution of
### TABLE 4

**RANK ORDER OF THE RESEARCH CRITERIA AS RATED BY THE FIVE GROUPS OF JUDGES**

<table>
<thead>
<tr>
<th>Criterions</th>
<th>Science Number</th>
<th>Education Education</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88.5</td>
<td>97</td>
<td>114</td>
<td>14</td>
<td>8</td>
<td>105.5</td>
</tr>
<tr>
<td>2</td>
<td>106</td>
<td>107</td>
<td>114</td>
<td>112</td>
<td>112</td>
<td>112</td>
</tr>
<tr>
<td>3</td>
<td>71.5</td>
<td>36</td>
<td>58</td>
<td>69</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>4</td>
<td>112.5</td>
<td>92.5</td>
<td>110</td>
<td>89</td>
<td>101.5</td>
<td>104.5</td>
</tr>
<tr>
<td>5</td>
<td>64</td>
<td>36</td>
<td>37.5</td>
<td>37</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>6</td>
<td>88</td>
<td>80.5</td>
<td>67.5</td>
<td>37</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>7</td>
<td>80</td>
<td>22</td>
<td>84.5</td>
<td>45</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>8</td>
<td>33</td>
<td>58.5</td>
<td>46.5</td>
<td>30</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>115</td>
<td>109.5</td>
<td>116</td>
<td>116</td>
<td>116</td>
<td>116</td>
</tr>
<tr>
<td>10</td>
<td>48.5</td>
<td>36</td>
<td>7.5</td>
<td>8</td>
<td>12.5</td>
<td>17</td>
</tr>
<tr>
<td>11</td>
<td>24</td>
<td>58.5</td>
<td>92</td>
<td>110.5</td>
<td>101.5</td>
<td>86</td>
</tr>
<tr>
<td>12</td>
<td>12.5</td>
<td>72</td>
<td>80</td>
<td>96.5</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>13</td>
<td>56.5</td>
<td>36</td>
<td>76</td>
<td>80</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>14</td>
<td>56.5</td>
<td>36</td>
<td>46.5</td>
<td>89</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>15</td>
<td>9</td>
<td>8</td>
<td>30</td>
<td>8</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>16</td>
<td>84.5</td>
<td>83.5</td>
<td>92</td>
<td>94.5</td>
<td>69.5</td>
<td>90.5</td>
</tr>
<tr>
<td>17</td>
<td>88</td>
<td>52.5</td>
<td>53</td>
<td>61.5</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>18</td>
<td>108.5</td>
<td>97</td>
<td>92</td>
<td>53.5</td>
<td>69.5</td>
<td>97</td>
</tr>
<tr>
<td>19</td>
<td>12.5</td>
<td>12.5</td>
<td>17.5</td>
<td>45</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>38.5</td>
<td>14</td>
<td>53</td>
<td>45</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>21</td>
<td>56.5</td>
<td>99</td>
<td>84.5</td>
<td>102</td>
<td>94.5</td>
<td>95.5</td>
</tr>
<tr>
<td>22</td>
<td>38.5</td>
<td>19.5</td>
<td>7.5</td>
<td>37</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>23</td>
<td>56.5</td>
<td>64.5</td>
<td>17.5</td>
<td>89</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>24</td>
<td>33</td>
<td>58.5</td>
<td>37.5</td>
<td>80</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>4</td>
<td>7.5</td>
<td>37</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>26</td>
<td>24</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>27</td>
<td>71.5</td>
<td>72</td>
<td>46.5</td>
<td>53.5</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>28</td>
<td>24</td>
<td>36</td>
<td>12</td>
<td>4</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>29</td>
<td>101.5</td>
<td>55</td>
<td>103</td>
<td>99.5</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>30</td>
<td>91.5</td>
<td>90.5</td>
<td>37.5</td>
<td>99.5</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>31</td>
<td>80</td>
<td>87</td>
<td>17.5</td>
<td>45</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>32</td>
<td>67.5</td>
<td>111</td>
<td>67.5</td>
<td>45</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>33</td>
<td>84.5</td>
<td>100</td>
<td>46.5</td>
<td>30</td>
<td>69.5</td>
<td>78</td>
</tr>
<tr>
<td>34</td>
<td>108.5</td>
<td>107</td>
<td>112</td>
<td>110.5</td>
<td>109.5</td>
<td>113</td>
</tr>
<tr>
<td>35</td>
<td>67.5</td>
<td>64.5</td>
<td>17.5</td>
<td>12.5</td>
<td>56</td>
<td>56</td>
</tr>
</tbody>
</table>
TABLE 4—Continued

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Science</th>
<th>Number</th>
<th>Education</th>
<th>Education</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>71.5</td>
<td>87</td>
<td>92</td>
<td>80</td>
<td>69.5</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>98.5</td>
<td>94.5</td>
<td>84.5</td>
<td>61.5</td>
<td>86</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>43</td>
<td>113</td>
<td>17.5</td>
<td>8</td>
<td>34</td>
<td>70.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>56.5</td>
<td>72</td>
<td>67.5</td>
<td>99.5</td>
<td>22</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>24</td>
<td>77</td>
<td>76</td>
<td>73.5</td>
<td>69.5</td>
<td>67.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>33</td>
<td>22</td>
<td>12</td>
<td>8</td>
<td>22</td>
<td>12.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>114</td>
<td>115</td>
<td>84.5</td>
<td>89</td>
<td>94.5</td>
<td>108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>112.5</td>
<td>116</td>
<td>84.5</td>
<td>89</td>
<td>94.5</td>
<td>107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>33</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4.5</td>
<td>3.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>24</td>
<td>36</td>
<td>76</td>
<td>21.5</td>
<td>44</td>
<td>33.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>64</td>
<td>12.5</td>
<td>46.5</td>
<td>15</td>
<td>56</td>
<td>23.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>33</td>
<td>25</td>
<td>80</td>
<td>69</td>
<td>94.5</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>56.5</td>
<td>58.5</td>
<td>37.5</td>
<td>45</td>
<td>22</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>43</td>
<td>55</td>
<td>67.5</td>
<td>53.5</td>
<td>69.5</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>24</td>
<td>25</td>
<td>30</td>
<td>21.5</td>
<td>22</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>84.5</td>
<td>102</td>
<td>58</td>
<td>53.5</td>
<td>56</td>
<td>82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>17</td>
<td>55</td>
<td>96</td>
<td>53.5</td>
<td>86</td>
<td>64.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>105</td>
<td>102</td>
<td>103</td>
<td>104</td>
<td>109.5</td>
<td>106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>95</td>
<td>87</td>
<td>114</td>
<td>115</td>
<td>115</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>43</td>
<td>49.5</td>
<td>67.5</td>
<td>109</td>
<td>99</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>95</td>
<td>94.5</td>
<td>67.5</td>
<td>73.5</td>
<td>109.5</td>
<td>95.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>56.5</td>
<td>46</td>
<td>46.5</td>
<td>61.5</td>
<td>105.5</td>
<td>64.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>2</td>
<td>4</td>
<td>7.5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>4.5</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>4.5</td>
<td>10.5</td>
<td>58</td>
<td>15</td>
<td>12.5</td>
<td>10.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>17</td>
<td>27.5</td>
<td>30</td>
<td>80</td>
<td>69.5</td>
<td>35.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>48.5</td>
<td>64.5</td>
<td>106.5</td>
<td>104</td>
<td>94.5</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>24</td>
<td>27.5</td>
<td>92</td>
<td>96.5</td>
<td>69.5</td>
<td>64.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>76</td>
<td>87</td>
<td>98.5</td>
<td>69</td>
<td>56</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>80</td>
<td>92.5</td>
<td>103</td>
<td>80</td>
<td>44</td>
<td>90.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>12.5</td>
<td>46</td>
<td>37.5</td>
<td>53.5</td>
<td>12.5</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>12.5</td>
<td>42.5</td>
<td>67.5</td>
<td>61.5</td>
<td>69.5</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>24</td>
<td>72</td>
<td>84.5</td>
<td>69</td>
<td>22</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>43</td>
<td>49.5</td>
<td>46.5</td>
<td>37</td>
<td>22</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>24</td>
<td>42.5</td>
<td>12</td>
<td>21.5</td>
<td>44</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>71.5</td>
<td>77</td>
<td>30</td>
<td>30</td>
<td>56</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>24</td>
<td>64.5</td>
<td>37.5</td>
<td>45</td>
<td>56</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>24</td>
<td>16.5</td>
<td>7.5</td>
<td>8</td>
<td>34</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>103</td>
<td>114</td>
<td>110</td>
<td>104</td>
<td>101.5</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>38.5</td>
<td>77</td>
<td>37.5</td>
<td>61.5</td>
<td>44</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groups of Judges</td>
<td>All Five Groups Combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>Education</td>
<td>Education</td>
<td>Biology</td>
<td>Chemistry</td>
<td>Physics</td>
<td>Combined</td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>48.5</td>
<td>64.5</td>
<td>24</td>
<td>21.5</td>
<td>22</td>
<td>35.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>108.5</td>
<td>109.5</td>
<td>103</td>
<td>80</td>
<td>79</td>
<td>102.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>33</td>
<td>46</td>
<td>17.5</td>
<td>30</td>
<td>22</td>
<td>23.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>84.5</td>
<td>72</td>
<td>108</td>
<td>107</td>
<td>101.5</td>
<td>99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>56.5</td>
<td>22</td>
<td>7.5</td>
<td>21.5</td>
<td>22</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>48.5</td>
<td>8</td>
<td>58</td>
<td>30</td>
<td>86</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>71.5</td>
<td>64.5</td>
<td>24</td>
<td>21.5</td>
<td>56</td>
<td>47.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>7</td>
<td>4</td>
<td>17.5</td>
<td>12.5</td>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>7</td>
<td>30.5</td>
<td>76</td>
<td>94.5</td>
<td>44</td>
<td>23.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>111</td>
<td>90.5</td>
<td>98.5</td>
<td>61.5</td>
<td>94.5</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>64</td>
<td>16.5</td>
<td>46.5</td>
<td>45</td>
<td>44</td>
<td>30.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>33</td>
<td>10.5</td>
<td>24</td>
<td>30</td>
<td>12.5</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>80</td>
<td>30.5</td>
<td>58</td>
<td>80</td>
<td>105.5</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>108.5</td>
<td>107</td>
<td>106.5</td>
<td>99.5</td>
<td>114</td>
<td>109.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>43</td>
<td>72</td>
<td>30</td>
<td>61.5</td>
<td>44</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>95</td>
<td>105</td>
<td>92</td>
<td>113</td>
<td>56</td>
<td>102.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>48.5</td>
<td>52.5</td>
<td>67.5</td>
<td>53.5</td>
<td>2</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>98.5</td>
<td>64.5</td>
<td>67.5</td>
<td>69</td>
<td>86</td>
<td>79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>71.5</td>
<td>49.5</td>
<td>30</td>
<td>37</td>
<td>44</td>
<td>45.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>64</td>
<td>16.5</td>
<td>24</td>
<td>21.5</td>
<td>69.5</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>88</td>
<td>72</td>
<td>80</td>
<td>69</td>
<td>69.5</td>
<td>77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>91.5</td>
<td>97</td>
<td>76</td>
<td>21.5</td>
<td>94.5</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>31.5</td>
<td>83.5</td>
<td>37.5</td>
<td>89</td>
<td>86</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>101.5</td>
<td>79</td>
<td>110</td>
<td>107</td>
<td>113</td>
<td>104.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>76</td>
<td>36</td>
<td>67.5</td>
<td>80</td>
<td>69.5</td>
<td>69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>38.5</td>
<td>87</td>
<td>67.5</td>
<td>61.5</td>
<td>69.5</td>
<td>70.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>56.5</td>
<td>29</td>
<td>58</td>
<td>80</td>
<td>79</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>64</td>
<td>36</td>
<td>46.5</td>
<td>15</td>
<td>22</td>
<td>30.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>80</td>
<td>49.5</td>
<td>58</td>
<td>30</td>
<td>44</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>91.5</td>
<td>82</td>
<td>98.5</td>
<td>89</td>
<td>94.5</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>98.5</td>
<td>102</td>
<td>98.5</td>
<td>107</td>
<td>34</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>56.5</td>
<td>80.5</td>
<td>53</td>
<td>80</td>
<td>79</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>48.5</td>
<td>42.5</td>
<td>24</td>
<td>45</td>
<td>56</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>12.5</td>
<td>42.5</td>
<td>67.5</td>
<td>37</td>
<td>34</td>
<td>30.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>2</td>
<td>19.5</td>
<td>30</td>
<td>21.5</td>
<td>44</td>
<td>12.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>116</td>
<td>112</td>
<td>103</td>
<td>80</td>
<td>105.5</td>
<td>114.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>12.5</td>
<td>16.5</td>
<td>46.5</td>
<td>21.5</td>
<td>56</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>104</td>
<td>104</td>
<td>92</td>
<td>93</td>
<td>109.5</td>
<td>101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>17</td>
<td>25</td>
<td>88</td>
<td>69</td>
<td>69.5</td>
<td>45.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>76</td>
<td>64.5</td>
<td>17.5</td>
<td>53.5</td>
<td>34</td>
<td>51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the ratings from the combined mean was computed for each criterion. Table 5 shows a frequency distribution of the standard deviations for the five groups of judges combined. A complete list of the standard deviations from the combined mean of all judges for each criterion may be found in Table 15 in Appendix IV. The lowest standard deviation was .43, the highest was 1.52, and the mean was 1.03.

**TABLE 5**

DISTRIBUTION OF STANDARD DEVIATIONS FROM THE COMBINED MEAN OF ALL JUDGES FOR EACH CRITERION

<table>
<thead>
<tr>
<th>Range of Standard Deviations</th>
<th>Number of Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>.43 - .49</td>
<td>1</td>
</tr>
<tr>
<td>.50 - .74</td>
<td>6</td>
</tr>
<tr>
<td>.75 - .99</td>
<td>39</td>
</tr>
<tr>
<td>1.00 - 1.24</td>
<td>55</td>
</tr>
<tr>
<td>1.25 - 1.49</td>
<td>14</td>
</tr>
<tr>
<td>1.50 - 1.52</td>
<td>1</td>
</tr>
</tbody>
</table>

Significance of difference

Analysis of variance, or the F test, was used to test the significance of differences among the ratings of each research criterion given by the five groups of judges. Table 6 contains the F values for each research criterion.

Significant differences at the 1 percent level occurred among the five groups of judges for eleven of the research criteria, and at the 5 percent level for nineteen of the research criteria. As a result, the null hypothesis of no significant differences among means was rejected at the
<table>
<thead>
<tr>
<th>Criterion Number</th>
<th>F</th>
<th>Criterion Number</th>
<th>F</th>
<th>Criterion Number</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.36</td>
<td>2</td>
<td>.57</td>
<td>7</td>
<td>1.39</td>
</tr>
<tr>
<td>2</td>
<td>6.22a</td>
<td>3</td>
<td>2.01b</td>
<td>8</td>
<td>4.09a</td>
</tr>
<tr>
<td>3</td>
<td>2.08</td>
<td>4</td>
<td>2.75b</td>
<td>9</td>
<td>.93</td>
</tr>
<tr>
<td>4</td>
<td>.37</td>
<td>5</td>
<td>2.41</td>
<td>10</td>
<td>1.66</td>
</tr>
<tr>
<td>5</td>
<td>1.45</td>
<td>6</td>
<td>.39</td>
<td>11</td>
<td>.17</td>
</tr>
<tr>
<td>6</td>
<td>2.77</td>
<td>7</td>
<td>8.64</td>
<td>12</td>
<td>1.59</td>
</tr>
<tr>
<td>7</td>
<td>.70</td>
<td>8</td>
<td>1.64</td>
<td>13</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>1.15</td>
<td>9</td>
<td>1.45</td>
<td>14</td>
<td>1.56</td>
</tr>
<tr>
<td>9</td>
<td>2.39</td>
<td>10</td>
<td>2.79b</td>
<td>15</td>
<td>2.86b</td>
</tr>
<tr>
<td>10</td>
<td>1.11</td>
<td>11</td>
<td>5.00a</td>
<td>16</td>
<td>2.15</td>
</tr>
<tr>
<td>11</td>
<td>6.28a</td>
<td>12</td>
<td>1.11</td>
<td>17</td>
<td>.07</td>
</tr>
<tr>
<td>12</td>
<td>4.54a</td>
<td>13</td>
<td>1.56</td>
<td>18</td>
<td>.19</td>
</tr>
<tr>
<td>13</td>
<td>1.92</td>
<td>14</td>
<td>3.26b</td>
<td>19</td>
<td>.27</td>
</tr>
<tr>
<td>14</td>
<td>1.18</td>
<td>15</td>
<td>.62</td>
<td>20</td>
<td>.17</td>
</tr>
<tr>
<td>15</td>
<td>1.94</td>
<td>16</td>
<td>6.70a</td>
<td>21</td>
<td>1.62</td>
</tr>
<tr>
<td>16</td>
<td>1.01</td>
<td>17</td>
<td>4.36a</td>
<td>22</td>
<td>.65</td>
</tr>
<tr>
<td>17</td>
<td>.50</td>
<td>18</td>
<td>.93</td>
<td>23</td>
<td>2.07</td>
</tr>
<tr>
<td>18</td>
<td>.81</td>
<td>19</td>
<td>1.82</td>
<td>24</td>
<td>.60</td>
</tr>
<tr>
<td>19</td>
<td>1.56</td>
<td>20</td>
<td>2.56b</td>
<td>25</td>
<td>1.85</td>
</tr>
<tr>
<td>20</td>
<td>.87</td>
<td>21</td>
<td>2.54b</td>
<td>26</td>
<td>.84</td>
</tr>
<tr>
<td>21</td>
<td>3.35b</td>
<td>22</td>
<td>3.48b</td>
<td>27</td>
<td>.55</td>
</tr>
<tr>
<td>22</td>
<td>.42</td>
<td>23</td>
<td>2.22</td>
<td>28</td>
<td>.81</td>
</tr>
<tr>
<td>23</td>
<td>1.55</td>
<td>24</td>
<td>5.30a</td>
<td>29</td>
<td>2.16</td>
</tr>
<tr>
<td>24</td>
<td>2.13</td>
<td>25</td>
<td>7.48a</td>
<td>30</td>
<td>2.52b</td>
</tr>
<tr>
<td>25</td>
<td>1.37</td>
<td>26</td>
<td>1.98</td>
<td>31</td>
<td>.83</td>
</tr>
<tr>
<td>26</td>
<td>2.10</td>
<td>27</td>
<td>1.86</td>
<td>32</td>
<td>3.51b</td>
</tr>
<tr>
<td>27</td>
<td>.51</td>
<td>28</td>
<td>2.76b</td>
<td>33</td>
<td>3.47b</td>
</tr>
<tr>
<td>28</td>
<td>2.59b</td>
<td>29</td>
<td>2.58b</td>
<td>34</td>
<td>2.24</td>
</tr>
<tr>
<td>29</td>
<td>1.69</td>
<td>30</td>
<td>3.42b</td>
<td>35</td>
<td>1.76</td>
</tr>
<tr>
<td>30</td>
<td>1.92</td>
<td>31</td>
<td>1.26</td>
<td>36</td>
<td>.83</td>
</tr>
<tr>
<td>31</td>
<td>1.26</td>
<td>32</td>
<td>1.52</td>
<td>37</td>
<td>3.55b</td>
</tr>
<tr>
<td>32</td>
<td>4.96a</td>
<td>33</td>
<td>.35</td>
<td>38</td>
<td>.60</td>
</tr>
<tr>
<td>33</td>
<td>2.61b</td>
<td>34</td>
<td>2.32</td>
<td>39</td>
<td>.94</td>
</tr>
<tr>
<td>34</td>
<td>1.53</td>
<td>35</td>
<td>.67</td>
<td>35</td>
<td>.94</td>
</tr>
<tr>
<td>35</td>
<td>1.48</td>
<td>36</td>
<td>1.64</td>
<td>36</td>
<td>.94</td>
</tr>
<tr>
<td>36</td>
<td>1.25</td>
<td>37</td>
<td>1.58</td>
<td>37</td>
<td>.94</td>
</tr>
<tr>
<td>37</td>
<td>.62</td>
<td>38</td>
<td>1.19</td>
<td>38</td>
<td>.94</td>
</tr>
<tr>
<td>38</td>
<td>9.28a</td>
<td>39</td>
<td>.94</td>
<td>39</td>
<td>.94</td>
</tr>
</tbody>
</table>

*a*Significant difference at the 1 percent level.

*b*Significant difference at the 5 percent level.
two designated levels of significance for a total of thirty of the criteria.

The F values indicated only that differences among means existed at a given level of significance. The mean values were then examined in order to ascertain which group or groups of judges differed significantly from the other groups. Table 7 contains the mean values for each of the five groups of judges for those research criteria for which significant differences were found. In addition, the appropriate F values were supplied and combinations of groups of judges who differed significantly from each other were indicated.

The use of Table 7 is indicated in the following illustration. For example, for research criterion number 2, the means of the science education, education, and physics judges were greater, at the 1 percent level of significance, than the means of the biology and chemistry judges. Also evident from Table 7 are the following groupings: (1) the means for the science education judges were significantly greater than the means of the other four groups of judges in eleven of the thirty cases where significant differences existed; (2) the means of the science education and education judges were significantly greater than the means of the biology, chemistry, and physics judges for seven of the research criteria; (3) the means of the science education, biology, chemistry, and physics judges were significantly
### TABLE 7

ANALYSIS OF THE SIGNIFICANT DIFFERENCES AMONG MEANS
OF THE FIVE GROUPS OF JUDGES

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Science</th>
<th>Number</th>
<th>Education</th>
<th>Education</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.76a</td>
<td>2.33a</td>
<td>1.33</td>
<td>1.33</td>
<td>2.00a</td>
<td>2.00</td>
<td>2.00a</td>
<td>6.22b</td>
</tr>
<tr>
<td>10</td>
<td>1.76a</td>
<td>2.04a</td>
<td>1.00</td>
<td>1.00</td>
<td>.86</td>
<td>2.05</td>
<td>2.79c</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3.65a</td>
<td>2.83a</td>
<td>2.33</td>
<td>1.58</td>
<td>2.14</td>
<td>6.28b</td>
<td>6.28b</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3.76a</td>
<td>2.75</td>
<td>2.58</td>
<td>2.33</td>
<td>2.43</td>
<td>4.54b</td>
<td>4.54b</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>3.35a</td>
<td>2.29</td>
<td>2.50</td>
<td>2.08</td>
<td>2.43</td>
<td>3.35c</td>
<td>3.35c</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>3.65a</td>
<td>3.04</td>
<td>3.42a</td>
<td>3.67a</td>
<td>3.43a</td>
<td>2.59c</td>
<td>2.59c</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>3.24a</td>
<td>1.96</td>
<td>2.75a</td>
<td>3.08a</td>
<td>2.71a</td>
<td>4.96b</td>
<td>4.96b</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>3.00a</td>
<td>2.25</td>
<td>3.00a</td>
<td>3.25a</td>
<td>2.86a</td>
<td>2.61c</td>
<td>2.61c</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>3.47a</td>
<td>1.79</td>
<td>3.33a</td>
<td>3.58a</td>
<td>3.29a</td>
<td>9.28b</td>
<td>9.28b</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>3.65a</td>
<td>2.71</td>
<td>2.67</td>
<td>2.75</td>
<td>2.862</td>
<td>2.99c</td>
<td>2.99c</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>2.24a</td>
<td>1.54</td>
<td>2.50a</td>
<td>2.58a</td>
<td>2.43a</td>
<td>2.75c</td>
<td>2.75c</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>3.59a</td>
<td>3.17a</td>
<td>2.58</td>
<td>2.83</td>
<td>2.43</td>
<td>2.60c</td>
<td>2.60c</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>3.71a</td>
<td>2.87</td>
<td>2.25</td>
<td>3.00</td>
<td>2.57</td>
<td>3.26c</td>
<td>3.26c</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>2.82a</td>
<td>2.50a</td>
<td>1.33</td>
<td>1.25</td>
<td>1.00</td>
<td>6.70b</td>
<td>6.70b</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>3.47a</td>
<td>2.92a</td>
<td>2.75a</td>
<td>1.67</td>
<td>2.29</td>
<td>4.36b</td>
<td>4.36b</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>4.00a</td>
<td>3.54</td>
<td>3.50</td>
<td>3.75</td>
<td>4.00a</td>
<td>2.56c</td>
<td>2.56c</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>3.94a</td>
<td>3.42</td>
<td>3.67</td>
<td>3.75</td>
<td>4.00a</td>
<td>2.54c</td>
<td>2.54c</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>3.94a</td>
<td>3.38</td>
<td>2.83</td>
<td>3.42</td>
<td>3.57</td>
<td>3.48c</td>
<td>3.48c</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>3.41a</td>
<td>2.79</td>
<td>2.00</td>
<td>2.00</td>
<td>2.43</td>
<td>5.30b</td>
<td>5.30b</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>3.65a</td>
<td>3.12a</td>
<td>2.33</td>
<td>2.33</td>
<td>2.86</td>
<td>7.48b</td>
<td>7.48b</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>3.76a</td>
<td>2.96</td>
<td>3.08</td>
<td>3.00</td>
<td>3.57a</td>
<td>2.76c</td>
<td>2.76c</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>3.76a</td>
<td>3.00</td>
<td>2.75</td>
<td>2.92</td>
<td>2.86</td>
<td>2.58c</td>
<td>2.58c</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>3.65a</td>
<td>2.75</td>
<td>2.50</td>
<td>2.83</td>
<td>3.43a</td>
<td>3.42c</td>
<td>3.42c</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>3.00a</td>
<td>2.75a</td>
<td>1.83</td>
<td>1.75</td>
<td>2.14</td>
<td>4.09b</td>
<td>4.09b</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>3.88a</td>
<td>3.08</td>
<td>2.67</td>
<td>3.42</td>
<td>3.14</td>
<td>4.00b</td>
<td>4.00b</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>2.82a</td>
<td>2.13a</td>
<td>2.33a</td>
<td>1.42</td>
<td>3.00a</td>
<td>2.86c</td>
<td>2.86c</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>3.35a</td>
<td>2.63</td>
<td>2.92</td>
<td>2.67</td>
<td>2.71</td>
<td>2.52c</td>
<td>2.52c</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>3.76a</td>
<td>3.00</td>
<td>2.75</td>
<td>3.17</td>
<td>3.29</td>
<td>3.51c</td>
<td>3.51c</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>4.00a</td>
<td>3.25</td>
<td>3.17</td>
<td>3.33</td>
<td>3.14</td>
<td>3.47c</td>
<td>3.47c</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>3.71a</td>
<td>3.17a</td>
<td>2.42</td>
<td>2.83</td>
<td>2.86</td>
<td>3.55c</td>
<td>3.55c</td>
<td></td>
</tr>
</tbody>
</table>

aFor a given criterion, means with the superscript "a" are significantly different (greater) from means without a superscript.

bSignificant differences at the 1 percent level.

cSignificant differences at the 5 percent level.
greater than the means of the education judges in five cases; (4) the means of the science education and physics judges were significantly greater than the means of the education, biology, and chemistry judges for four of the research criteria; and (5) in all thirty cases, the means of the science education judges were in the significantly higher group.

From Tables 6 and 7 the answers can be found to the question: What is the significance of differences in the judgments regarding the research criteria which appear among the research experts from the biological and physical sciences and the research experts from science education and the other areas of education? Significant differences in the judgments were found for thirty of the 116 research criteria which were submitted to the expert judges. No significant differences were found among groups of judges, therefore, for eighty-six of the research criteria.

**Descriptive Data**

The descriptive data gathered in this study consisted of information concerning the qualifications of the expert judges. Several aspects of these qualifications were summarized in the three tables which follow. The first of these, Table 8, contains information about the mean amount of research experience of each of the five groups of judges. Included are (1) the mean number of years of research
TABLE 8
MEAN NUMBER OF YEARS OF RESEARCH EXPERIENCE AND
MEAN NUMBER OF STUDIES DIRECTED AND
CONDUCTED BY THE EXPERT JUDGES

<table>
<thead>
<tr>
<th>Research Experience</th>
<th>Science</th>
<th>Education</th>
<th>Education</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Number of Years of Research Experience</td>
<td>20</td>
<td>20</td>
<td>25</td>
<td>25</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Mean Number of Master's Degree Studies Advised</td>
<td>25</td>
<td>35</td>
<td>15</td>
<td>10</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Mean Number of Doctoral Degree Studies Advised</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Mean Number of Research Studies Personally Conducted</td>
<td>10</td>
<td>15</td>
<td>35</td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

experience possessed by each group of expert judges, (2) the mean number of master's degree students for which the judges in each group were major advisers, (3) the mean number of doctoral degree studies for which the judges in each group were major advisers, and (4) the mean number of formal research studies to which the judges in each group made major contributions—excluding research studies included in numbers (2) and (3) above. It will be noted that the mean number of years of research experience for the five groups
of judges ranged from 20 to 25 years. The science education and education judges had somewhat less experience in personally conducted research than the judges from the biological and physical sciences. The reverse was true in research experience related to advising master's and doctoral students in which case the science education and education judges had more experience than the other three groups of judges.

The academic qualifications of the expert judges are listed in Table 9. Many of the judges did not earn master's

| TABLE 9 |
| ACADEMIC QUALIFICATIONS OF THE EXPERT JUDGES |

<table>
<thead>
<tr>
<th>Academic Qualifications</th>
<th>Science</th>
<th>Education</th>
<th>Education</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year Baccalaureate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earned</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1932</td>
<td>1941</td>
<td>1939</td>
<td>1938</td>
<td>1936</td>
<td></td>
</tr>
<tr>
<td>Year Doctorate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earned</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1945</td>
<td>1951</td>
<td>1946</td>
<td>1944</td>
<td>1945</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>1924-56</td>
<td>1928-64</td>
<td>1932-59</td>
<td>1921-58</td>
<td>1931-54</td>
<td></td>
</tr>
<tr>
<td>Number of Judges</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per Academic Rank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecturer</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Instructor</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Asst. Professor</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Assoc. Professor</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Professor</td>
<td>15</td>
<td>15</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>
degrees, and therefore this information was omitted from the table. The mean year and the range of years in which the baccalaureate degrees and the doctoral degrees were received were listed for each group of judges. In addition, the number of judges holding each academic rank was listed. As is evident from the table, the education group had the most recent mean year for receiving both the baccalaureate and the doctorate. The mean year for receipt of the baccalaureate degree by the science education judges was the least recent of the five groups, while the mean year for receipt of the doctorate was the least recent for the chemistry group. The amount of time which elapsed between the mean times at which the baccalaureate degree and the doctorate were received was (1) thirteen years for the science education judges, (2) ten years for the education judges, (3) seven years for the biology judges, (4) six years for the chemistry judges, and (5) nine years for the physics judges. Fifty-three of the seventy-two judges had achieved the rank of full professor.

The third and final table in this section, Table 10, contains a summary of the eight different types of research experience of the expert judges. Totals are given at the right of the table and also at the bottom. The greatest number of the total sample of seventy-two judges were experienced in experimental classroom research (38) and in laboratory research (39). It is interesting that two of the biology judges had experience with experimental classroom
<table>
<thead>
<tr>
<th>Type of Research</th>
<th>Science</th>
<th>Education Education</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
<th>Total Number of Judges Possessing the Various Types of Research Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Classroom</td>
<td>15</td>
<td>21</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>Experimental Field</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Experimental Laboratory</td>
<td>5</td>
<td>6</td>
<td>11</td>
<td>12</td>
<td>5</td>
<td>39</td>
</tr>
<tr>
<td>Case Study</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Clinical</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Descriptive</td>
<td>9</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Historical</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Theoretical (Mathematical and Philosophical)</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Total Number of Research Experiences Per Group of Judges</td>
<td>41</td>
<td>56</td>
<td>24</td>
<td>13</td>
<td>7</td>
<td>141</td>
</tr>
</tbody>
</table>
research, while none of the judges in the physical sciences had such experience. Twenty of the expert judges had experience with experimental field research and twenty-two with descriptive research. Eleven or fewer judges had experience with the other four types of research listed in the table. The seventeen science education judges had the greatest variety of research experience, an average of approximately 2.4 types of experience per judge. The twenty-four education judges had an average of 2.3 types of experience per judge, and the twelve biology judges had an average of two types of experience per judge. Of the five groups of judges, the twelve chemistry and seven physics judges were the most specialized in research experiences, with an average of only 1.1 and 1.0 types of experience per judge respectively. When considering the total sample of seventy-two judges, each judge had an average approximately two types of research experience.

Subjective Data

The expert judges were invited to make any comments they wished concerning the research criteria. Approximately 35 percent of the judges availed themselves of this opportunity, and as a result only a sampling of these comments can be provided here. The comments of the judges generally fell into the following categories: (1) unfavorable comments displaying a lack of appreciation for the use of research
criteria, (2) unfavorable comments based on the difficulty of evaluating the criteria apart from actual examples of research, and (3) favorable comments displaying an understanding of the usefulness of research criteria in evaluating research. The ratio of responses in these three categories was approximately 1:2:2.

Two examples illustrating the first of the categories listed above are as follows:

Education judge:

There are about five rules for writing a report that need to be given to students (with a little elaboration). After that it is a matter of style, and yours is probably as good as mine. You completely omitted one section on the related literature. You assume it is either correct or reasonably so. This is a downright error. Never trust anybody. About 70-90% is in error. If you accept this, what good have all the elaborate rules been anyway?

Chemistry judge:

Perhaps there are some general criteria for effective research evaluation as your long list suggests, but I could not rate their relative importance. Your items all seem somewhat pertinent. However, my rating would be only opinion, not valid data. I suppose that research evaluation often cannot be made at once—it must stand the test of time, and the analysis of one's present and future peers. I have supposed that the art of researching has been known for a century or more—at least, somehow, lots of good research does get done and reported.

The following examples are illustrative of the second category of comments:

Physics judge:

My research in physics has always been directed toward a definite problem whose place in the framework of physics was obvious. We have no difficulty
in the formulation or definition of the problem. I am at a loss to apply this document to my field. It may be useful in the more indefinite scientific fields.

Education judge:

Thank you for sending two copies. I did one at home and one in my office to get an estimate of my reliability—I'd guess it is about .60—not very high. The trouble is your instruction # 4—the assumption of selective application of the research criteria to the appropriate kind of research. Sometimes I think of one kind (type, example, etc.) of research and sometimes another.

The following are examples of favorable comments which were received:

Science education judge:

You have undertaken an important problem area and your work should help to describe good research. It should help to guide graduate students to consider important factors.

Science education judge:

Excellent instrument! My "negative" responses are based on possibilities rather than hopes: i.e., my response is largely normative. I would like to see a section added on non-parametric studies.

Biology judge:

You have prepared a most interesting series of statements. Your attempts to give some sense of perspective through an ordering in terms of significance is clearly desirable.

Although attention to these details is necessary, I am bored and frustrated by papers in which the author laboriously and obviously tries to follow all the rules. A more skilled writer is able to meet the criteria in a more unobtrusive manner.

You may be interested in the enclosed brief summary of some criteria for studies in human genetics.
Education judge:

You have here a very fine list of criteria—stated usually in idealistic form. I am inclined to believe that you should not weaken at any stage and permit use of average or little importance. In brief, I don't think we can emphasize too much the need for tightening up all of our research activities. The final instrument should be idealistic in the sense that we have something to shoot at.

Other examples of comments in each of the three categories illustrated above could of course be presented. This was not done, however, because the general feelings represented in each of the three types of reactions to the criteria were surprisingly similar.

Summary

Three types of data were presented in this chapter—statistical data, descriptive data, and subjective data. The statistical data included the mean ratings of each criterion by each group of judges and by all five groups of judges combined. These means were used to rank the research criteria according to their order of importance and to calculate an F value for each criterion. Judgments concerning thirty of the 116 research criteria were found to have F values such that the judgments were identified as significantly different.

The descriptive data were obtained in order to indicate the qualifications of the expert judges. These data were summarized from the responses which were requested on the last page of the research criteria instrument.
Finally, the expert judges were invited to write their comments concerning the research criteria. Samples of this subjective type of data, both unfavorable and favorable, were presented in the last section of this chapter.
CHAPTER IV

AN EXAMPLE OF THE APPLICATION OF THE RESEARCH CRITERIA TO A STUDY IN SCIENCE EDUCATION

Introduction

It was stated previously that a final aspect of the problem of this study involved illustration of the use of the criteria in the process of evaluating research. It was felt that the research criteria would be most useful to science educators if an example of the application of the criteria was taken from the field of science education.

One study was chosen for evaluation, a Master of Arts thesis written in 1963 at The Ohio State University.¹ Any one of a large number of theses would have been satisfactory for the purpose. Although the present choice was somewhat arbitrary, the selection was made from a list suggested by a member of The Ohio State University faculty who advises graduate students writing master's degree theses in science education. The rather unique topic and the relative simplicity of the procedural and statistical techniques employed in this thesis made it readily understandable by the non-

statistically trained researcher. Thus, it was hoped that an evaluation of this thesis, rather than a more complicated one, would be more useful as a learning device with respect to illustrating the use of the research criteria.

The choice of a Master's level study for examination was conditioned by the desire that students at both the master's and the doctoral levels might make use of the present study in their own evaluations of research. It was felt that the master's level student might derive more from an evaluation which was nearer his more immediate goal, whereas the advanced graduate student should profit from an illustration of the use of the criteria in evaluating either a thesis or a dissertation— the thesis, of course, ideally being identical to the dissertation except in the degree of delimitation of the problem. In addition, it was felt that the evaluation which was presented as an example should be thorough, and that all applicable and/or highly ranked criteria should be utilized in the evaluation example. Due to the lengthy process involved in such a thorough treatment of the material, the number of studies evaluated was limited to one.

The following two sections of this chapter contain an overview of the example study, followed by the application of the criteria to that study. Illustrations of the use of the criteria are indicated by footnotes which refer the reader to the list of criteria given in Chapter II.
An Overview of the Example Study

Formulation and definition of the problem

The need for the study, according to the author, related to (1) the expansion of school enrollments, (2) increased extracurricular duties, (3) new science programs, (4) increased emphasis on individualized instruction, (5) the explosion of scientific knowledge, (6) advance placement courses taught in high schools, and (7) the use of professional time for non-professional duties. These and other similar considerations, it was stated, make it imperative that full use be made of the teaching resources of this country. Thus, according to the author, "in order to more fully utilize these resources it is necessary that careful studies be conducted on the present utilization in each school system to determine what changes, if any, would produce a more efficient utilization of the professional manpower available."2

The purpose of the study was stated as follows:

1. To determine whether or not it is possible to study the work of the secondary science teacher by means of work-sampling methods. Referred to in text as "in school" part of the study.
2. To define the work of the secondary science teacher in terms of elements which will be meaningful to the teacher and descriptive of the teachers work.
3. To ascertain with reasonable accuracy the percentage of time which a group of teachers spend on these elements.

2Ibid., p. 1.
4. To determine the group utilization of non-teaching time.
5. To determine the amount and type of work a teacher does outside of school on school duties. Referred to in text as "out of school" part of study.3

Procedural design

The procedure consisted of two parts, the "in school" portion and the "out of school" portion. The "in school" part of the study involved the method of work sampling to determine the percentage of time spent in each of twenty-one "elements," into which the work of teaching had been divided. The twenty-one elements were as follows: (1) "lecture and/or lecture demonstration," (2) "audio-visual," (3) "class discussion," (4) "testing," (5) "other" classroom duties, (6) "laboratory," (7) "study hall," (8) "hallways," (9) "assemblies," (10) "other" supervisory duties, (11) "lunch and personal activity," (12) "course preparation," (13) "course follow-up" grading duties, (14) "course follow-up" class discussion, (15) "time spent by science teachers on other areas of instruction," (16) "meetings," (17) "informal discussion and helping other teachers," (18) "personal development," (19) "absent," (20) "individual research," and (21) "group research."4 These twenty-one elements were described further in terms of the associated behavior which would be

3Ibid., pp. 3-4.
4Ibid., pp. 6-8.
expected for each element. The work sampling elements then were used as a check-list during class visitation times. It usually was necessary merely to look through the window of the classroom door in order to determine which of the work sampling elements was in progress. Thus, the author stated that the classes were not disturbed by the observer. The observation schedule included morning and afternoon periods each day. The starting times of each of the observation periods, as well as the order of the schools visited, were determined by random selection.

The "out of school" portion of the study was conducted by the telephone survey technique. Each teacher was assigned to a group which was called each evening by telephone for one week. An attempt was made to determine by this means the amount of time spent by each teacher outside of school on school duties.

In order to establish the reliability of the percentage of time values obtained for each "in school" element, the following formula, based on the normal distribution curve, was used:  

$$\sigma_{100p} = \sqrt{\frac{100\bar{p} (100-100\bar{p})}{n}}$$

where $100\bar{p} =$ average percentage

$n =$ sample size

Ibid., p. 12.
This formula was used to calculate the number of observations which were necessary for the size of the sample. The author stated the following concerning the statistical technique:

When the limits of accuracy are predetermined to within two standard errors or at the 5% level of confidence, it can be stated that in ninety-five cases out of one hundred the estimated true percent value of an element would occur within the limits of \( \pm 1.96 \) errors from the observed percentage. The key to the accuracy of work sampling is in the number of observations.\(^6\)

The sample of teachers studied consisted of twenty high school science teachers from schools located in a mid-western metropolitan area. The twenty teachers who participated in the study taught an average of six classes a day with one supervisory assigned period and one free period.\(^7\) Group time utilization rather than individual time utilization was determined.

Presentation of evidence

The data were presented in tables, graphs, and discussion. It is unnecessary to present the data in detail here, but examples and summarizing comments will be given to illustrate both the "in school" and the "out of school" portions of the data.

As an example of the "in school" part of the data, the element, "lecture and/or lecture demonstration," was observed 223 times. This represented 23.69 percent of the total

\(^6\)Ibid., p. 12.

\(^7\)Ibid., p. 4.
number of observations. After submitting this percentage figure to statistical treatment, it was established, at the 5 percent level of confidence, that the true value for this element would lie between 20.93 and 26.45 percent of the total "in school" time.

The "out of school" part of the data consisted of simply an identification of professionally related activities which were engaged in during "out of school" time, and a tabulation of the amounts of time involved in these types of activities. The author found that teachers spent an average of 9.25 hours per week on school work during "out of school" time, most of which was devoted to course preparation and grading of papers.8

Summary and conclusions

Conclusions and recommendations were presented concerning the method of work sampling. A summary of the steps involved in doing work sampling also was provided. Finally, conclusions concerning the study of twenty high school teachers were given.

Application of the Research Criteria to the Example Study

Application of the criteria to the formulation and definition of the problem

Neither the Table of Contents nor the body of the thesis contained a section bearing the heading, "Problem of

8Ibid., p. 45.
A series of statements was found under the heading, "Purpose of the Study," which, although lacking somewhat in grammatical polish, were sufficiently clear to communicate the broad outline of the study. (These statements of purpose, which were five in number, were listed in a previous section of this chapter under "An Overview of the Example Study.") A problem statement, or in this case the statement of purpose, in addition to being clear and understandable, should be expressed precisely. This was not accomplished in the study being examined, as will be shown in the following paragraphs.

The author stated that the purpose of the study was to determine the possibility of the use of work-sampling methods. However, after careful examination of the entire thesis and the related studies which were reported, the purpose appears to have been one of demonstrating the usefulness of the work sampling technique. The lack of precision in stating the purpose of the study gave the remainder of the study a certain lack of clarity and direction. One indication that the author was not attempting to solve a problem was that he did not present any hypotheses. One might assume as a possible hypothesis, based on the first of the five statements of purpose, that "it is possible to study the work of the secondary

9See criterion 4, p. 27.

10See criterion 1, p. 27.
science teacher by means of work sampling methods." As a matter of fact, however, earlier studies in industrial and educational settings already had tested this hypothesis. Although one might argue that the method had not been proved effective for time utilization studies of science teachers, this line of reasoning does not seem very practical; the widely differing tasks of both elementary teachers and university engineering faculty members had been studied successfully by the method. Thus, the "purpose of the study" might have more accurately indicated the resultant purpose, if not the intended purpose, if it had been worded as follows:

The purposes of this study was as follows:
1. To demonstrate the work sampling method of studying the "in school" work of the secondary science teacher.
   a. To define the work of a selected group of secondary science teachers in terms of elements which were meaningful to these teachers and descriptive of their work.
   b. To ascertain with reasonable accuracy the percentage distribution of time which a selected group of science teachers spent on these elements.
2. To determine the group utilization of non-teaching time by a selected group of secondary science teachers in order to demonstrate a method of determining the amount and type of work a teacher does "out of school" on school duties.

In addition to an increase in precision brought about by use of the word, "demonstrate," the precision of the statement of purpose also would be increased by showing that a selected group of science teachers was used. The data which were collected could not be generalized to the total population of secondary science teachers, as the author's statement of
purpose implied. This fact, it should be added, is another indication that the author's purpose was one of demonstration rather than of problem solving.

With regard to the problem choice, the author stated that "a review of the multitude of studies in teacher load suggests that there seems to be no consistent philosophy or criteria by which appraisals of teaching load are made."\(^{11}\) He concluded that extension was possible in the direction of a statistical method of determining teacher load. In these statements, he implied that a thorough study of the literature on teacher load had been made and that this body of organized knowledge had been examined for gaps, for discrepancies, and for the directions in which extension was possible.\(^{12}\) However, the author did not cite any literature which would describe the kind of non-statistical studies of teacher load which had been conducted in the past. Thus, although the foundation of the study was identified, the study was not related to an orienting frame of reference or a systematic area of study except by implication.\(^{13}\) Because of this lack of orientation, it is difficult for the reader to understand the significance of the author's demonstration of a statistical approach to studying teacher load. In addition, by this omission the author did not show that unnecessary duplication

\(^{11}\)Stonecipher, p. 3.

\(^{12}\)See criterion 6, p. 28.

\(^{13}\)See criterion 30, p. 32.
of research effort had been avoided. A theoretical framework of knowledge underlying the problem of examining teacher load also was neglected by omitting a discussion of previous non-statistical studies.

Additional criticism must be directed toward the lack of adequate statements of the assumptions which were basic to the thesis topic. Only the assumption of normality required by the work sampling method was identified. If there were no other assumptions, which is doubtful, then this fact should have been stated.

Definitions generally were adequate in the study. Except for the author's use of the word, "accuracy," which will be discussed later, most of the terms used in the study were employed according to the meanings demanded by accepted usage. Operational definitions of all concepts, or "elements" involved in the teacher's work-load, were given and were used throughout the study as defined.

The discussion of related literature in the thesis generally was inadequate. Two studies were reported by the

---

14See criterion 7, p. 28.
15See criterion 9, p. 28.
16See criteria 16 and 17, p. 30.
17See criterion 24, p. 31.
18Stonecipher, pp. 6-8.
19See criteria 25 and 26, pp. 31 and 32.
author in which work sampling methods were applied to time utilization in the teaching profession.\textsuperscript{20,21} The conclusions of these two studies were listed, but no attempt was made to analyze the studies for their relevancy to the problem situation or for any bias which may have existed.\textsuperscript{22} In addition, there were a few instances in which the author failed to cite his literature source.\textsuperscript{23} One of these was in his discussion of the reliability formula and calculations.\textsuperscript{24} Examination of a literature source by Williams,\textsuperscript{25} which was cited in an earlier section of the thesis, revealed that most of the reliability section of the thesis was based on the Williams source. A second failure to cite possible literature sources was in the author's development of the work sampling elements and the associated behavior which one might expect to observe for each element. The two previous studies cited by the author in which work sampling methods were used undoubtedly were of considerable assistance in developing the list for secondary science teachers, although he did not state that such was the case.

\textsuperscript{20}Stonecipher, pp. 20-25.
\textsuperscript{21}See criterion 29, p. 32.
\textsuperscript{22}See criteria 27 and 28, p. 32.
\textsuperscript{23}See criterion 32, p. 33.
\textsuperscript{24}Stonecipher, pp. 12-15.
\textsuperscript{25}The Williams' article was cited on p. 12 of the Stonecipher thesis.
An important positive quality of the thesis was in the use of relevant literature from other disciplines.\textsuperscript{26,27} The work sampling technique, for example, originally was designed for industrial use in studying the time utilization of manual laborers. Thus, in formulating the problem, the author made useful applications of ideas from a discipline other than education.

**Application of the criteria to the procedural design**

Several positive qualities in the area of procedural design were evident in the thesis. The study was delimited carefully in several of its aspects—scope, the nature and quantity of data desired, and the methods of procedure to be followed.\textsuperscript{28,29} The "in school" and "out of school" teaching-related activities of the teachers were listed carefully. For example, one work sampling element was "class discussion," which was to be observed by the behavior pattern of "students talking (not undisciplined) or students raising hands for permission to talk."\textsuperscript{30} One further example of the careful delimitation was the discussion of statistical procedures; the number of observations which would be necessary for

\begin{itemize}
  \item \textsuperscript{26}Stonecipher, pp. 11-12.
  \item \textsuperscript{27}See criterion 31, p. 33.
  \item \textsuperscript{28}Stonecipher, pp. 4-19.
  \item \textsuperscript{29}See criterion 12, p. 29.
  \item \textsuperscript{30}Stonecipher, p. 6.
\end{itemize}
reliable conclusions based on a sample of twenty teachers was predetermined by the author.\textsuperscript{31,32}

The thesis must be criticized, however, for an inadequate description of the duties of the teachers who participated in the study; teaching schedules, for example, were not provided. Also, although it was stated that seven of the teachers had student teachers under their supervision, the relationship of this fact to the observation schedule was unknown.\textsuperscript{33} In addition, some of the teachers were not full-time science teachers. These are examples of the fact that the teachers were not described adequately with reference to the variables among them.\textsuperscript{34} This fact is pointed out because the study, as a demonstration of the use of statistical sampling of science teacher activities, should have illustrated the kinds of information needed about the sample chosen for study. In addition, the parent population from which the sample was selected should have been defined so that generalizations could have been made with reference to that parent population.\textsuperscript{35} Such an omission in a study based on the demonstration of a method would considerably decrease the number of directions of inquiry possible, as well as the

\textsuperscript{31}Ibid., pp. 15-16.
\textsuperscript{32}See criteria 63, 64, and 80, pp. 39, 40, and 42.
\textsuperscript{33}Stonecipher, p. 5.
\textsuperscript{34}See criterion 13, p. 29.
\textsuperscript{35}See criteria 47, 48, and 68, pp. 36, 37, and 40.
efficiency and comprehensiveness of the study. Thus, it can be concluded that the author did not fully allow the apparent purpose of his study, that of demonstrating the work sampling technique, to condition the procedural design which was followed.\(^{36}\)

Several positive qualities were evident in the use of the work sampling technique. For example, the technique itself was, at least in a replicative and demonstrative sense, an object of study.\(^{37}\) Further, the objectivity of the work sampling method for collecting data was a primary reason for promoting that method over other reportedly less reliable methods.\(^{38}\) These positive characteristics of the study were strengthened by the author's adequately detailed description of the work sampling method itself, thus facilitating use of the technique by other investigators.\(^{39}\)

The internal and external validity of the study was examined as a measure of its adequacy.\(^{40}\) The external validity of a study can be measured by the amount of agreement which exists between facts known by means which are external to the study and facts demonstrated in the study itself. The internal validity is determined by the extent to which the various

\(^{36}\)See criterion 40, p. 35.

\(^{37}\)See criteria 41 and 44, pp. 35 and 36.

\(^{38}\)See criterion 42, p. 35.

\(^{39}\)See criterion 45, p. 36.

\(^{40}\)See criterion 46, p. 36.
aspects of a study agree with each other. An examination of the internal validity of a study cannot be separated from a consideration of the data-gathering instrument. According to one of the research criteria, the data-gathering instrument should have validity or truthfulness; it should measure what it purports to measure.\textsuperscript{41}

An indication of the internal validity of the study being evaluated, as well as the validity of the data-gathering instrument used in that study, was lacking entirely. For example, no attempt was made to show that the "elements" into which the work of the teacher had been divided, and the expected associated observable behavior, actually corresponded to the "true" activities of the teacher. How was the list of elements and observable behavior developed? Did the author find it necessary to revise or add to the list during trial runs using the data-gathering instrument? How does each element relate to the kind of learning which takes place in the classroom? Answers to questions such as these would have helped in establishing the internal validity of the study and the validity of the data gathering instrument.

It was stated earlier that the author's use of the word "accuracy" was inadequately defined. His use of the word "accuracy" was related to the validity of the study, and hence a discussion of the use of the word was postponed until the validity was considered. The author commented that "no study

\textsuperscript{41}\textsuperscript{41}See criterion 59, p. 39.
has been made with the teaching profession to determine the accuracy of the work sampling method.\textsuperscript{42} He probably was not referring to statistical validity in this statement, but rather to the relationship between what is observed and what learning actually is occurring in the classroom.\textsuperscript{43} The lecture-demonstration method can be authoritarian, for example, or it can be used to stimulate a questioning, reasoning attitude on the part of students. If the latter relationship is what the author meant when he referred to the "accuracy" of the work sampling method, then perhaps this relationship would have been more productive as the object of study rather than the demonstration of the work sampling method as a research tool. In any case, the degree of validity which was accepted for the data-gathering instrument should have been justified with reference to the objectives of the study.\textsuperscript{44}

A consideration of external validity was completely lacking in the study being evaluated. For example, the author found, by the work sampling method, that the teachers who participated in his study spent an average of 18.9 hours in class, plus 2.4 hours in the laboratory, for a total of 21.3 hours per week "in front of students."\textsuperscript{45} (The time values for each element were calculated by multiplying the percentage of

\begin{itemize}
  \item \textsuperscript{42} Stonecipher, p. 18.
  \item \textsuperscript{43} See criterion 62, p. 39.
  \item \textsuperscript{44} See criterion 61, p. 39.
  \item \textsuperscript{45} Stonecipher, pp. 30-31.
\end{itemize}
time spent in each element by the required hourly work week of the teacher.) The author did not compare this 21.3 hour value, however, with the teaching schedules of the teachers. In addition, as mentioned previously, the teaching schedules of the participants were not provided; hence, a reader could not make this comparison either. Such a check, of course, would have provided some indication of the external validity of the study. Other checks of this sort could have been made, but were omitted.

With regard to reliability or consistency, a data-gathering instrument, according to one of the research criteria, should agree with itself when used repeatedly with the same individuals, groups, or objects. An indication of the reliability of the data-gathering instrument in the thesis being evaluated may be found through an examination of three graphs. According to the author, these three graphs "illustrate one of the important principles of work sampling, namely, the leveling off of percentages as the total number of observations increases." The graphs were plots of percentage of time spent on a given element or study, such as "lecture and/or lecture demonstration," versus the total number of observations that had been made. The graph of this

46 See criterion 60, p. 39.
47 Stonecipher, pp. 42-44.
48 Ibid., p. 41.
particular element indicated that the percentage of time spent on lecture and/or lecture demonstration varied over a range of approximately 20 percent until six hundred observations had been made. Then the percentage reached a plateau which held fairly constant up to nine hundred observations. These graphs, although somewhat inadequately labeled on the percentage axis, served to indicate the reliability of the data-gathering instrument because of the consistency which was achieved in the percentage of time spend on each element. In addition to the graphs, the reliability of the statistical method of gathering data was given by the standard deviation of each percentage value.\(^{49,50}\)

**Application of the criteria to the presentation of evidence**

Evaluvative comments concerning the manner in which the evidence was presented in the example thesis will be directed toward two aspects of the study, the actual **presentation** of the data and the subsequent **discussion** of these data. Both positive and negative qualities can be found in both aspects of the presentation of the evidence by the author.

The data were logically organized and presented in tables and in graphs.\(^{51}\) In general the graphs were excellent,

\(^{49}\)Ibid., pp. 27-29.

\(^{50}\)See criteria 53 and 73, pp. 38 and 41.

\(^{51}\)See criterion 74, p. 41.
both in construction and in the impression given to the
reader, and several represented a rather creative synthesis
of the data. These included a pie graph of professional
and nonprofessional work, a chart showing the results of
the randomizing of the observation period starting times,
and the graphs, already mentioned, illustrating some of the
element percentages which were found. The main criticism
which might be made of the graphical techniques which were
used is that in some cases information and/or discussion which
would have been helpful for the reading of the graphs and
tables was omitted. A further minor point of criticism
involves a judgment made by the author—he chose to include
the element of "laboratory teaching" with the supervisory
synthesis rather than with the in-class synthesis, but perhaps
reasons can be found for placing this element in either
synthesis category.

In general, the discussion of the data was adequate in
terms of amount, but was lacking in terms of logical organiza-
tion and grammatical style. Also, the author frequently
interpreted the data, while presenting it, in terms of his
own personal experience and biases. This was due in part to

52See criterion 79, p. 42.
53See criterion 83, p. 43.
54Stonecipher, p. 33.
55Ibid., p. 40.
56Ibid., pp. 42-43.
a tendency to oversimplify the cause and effect relationships for the data in the study.\textsuperscript{57} For example, the author interpreted at considerable length the fact that no individual or group research was being carried out by the sample of teachers in the study. A small section of this interpretation was as follows:

... These teachers are not given any opportunity to carry on any respectable program of research. In examination of their work week at school the time demands placed on them by the school system excluded any but very minor and rudimentary experimentation.

\textsuperscript{58}

This type of interpretation of the data concerning the lack of research activities, although probably containing part of the truth, is certainly incomplete. The author assumed that the sample group of teachers possessed the necessary knowledge and/or interest for carrying out formal classroom research—an assumption which possibly is not a valid one, although there is no way of determining its validity from the data presented in the study. In any case, an interpretation of this sort might better be placed elsewhere than in the chapter on the analysis of data.

Application of the criteria to the summary and conclusions

The thesis did not include a summary of the entire study,\textsuperscript{59} a section which might have added somewhat to its

\textsuperscript{57}See criteria 86 and 88, p. 43.

\textsuperscript{58}Stonecipher, pp. 32-34.

\textsuperscript{59}See criteria 98 through 102, p. 45.
clarity. Instead, the final chapter was devoted to (1) conclusions concerning the method of work sampling, (2) a summary of steps involved in work sampling, and (3) conclusions concerning the study.

If one accepts as the primary purpose of this study the demonstration of work sampling as a method of studying the "in school" work of the secondary science teacher (which was identified previously as the author's probable although not clearly stated purpose of the study), then the major conclusions included in the study were drawn properly with reference to the goal which engendered the study. On the more positive side, an examination of the conclusions revealed that they were both internally and externally consistent. Conclusions were reached with reference to all or most of the data, and were grouped under appropriate major headings.

The limitations of the method employed in gathering the data were listed. In all cases the limitations were in logical order and were not unreasonable in limiting the usefulness of the method demonstrated. The author also listed an interesting and insightful group of problems that lend themselves to solution by the work sampling method. His

60 See criterion 103, p. 45.
61 See criteria 104 and 105, pp. 45 and 46.
62 See criteria 106 and 108, p. 46.
63 See criterion 110, p. 46.
recommendations for future research were warranted by both the data and the conclusions of the study.\textsuperscript{64}

\textbf{Summary}

A master's degree thesis was selected in order to illustrate the use of the criteria in evaluating research. An overview of the thesis was presented, followed by the evaluation itself. In the process of applying as many as possible of the research criteria to the study, both positive and negative aspects of the study were pointed out. In any thesis or dissertation, one would expect to find both strengths and weaknesses; these should be considered a result of the learning process in conducting research.

The illustration of the use of the research criteria which was presented in this chapter should help the researcher to use the criteria as a tool in conducting his own research and in evaluating research reports. The use of fifty-three of the criteria was illustrated in the evaluation of the example study. The fifty-three criteria represented all sections of the criteria instrument except two. These two sections included criteria 33 through 39 under "Hypotheses and Derived Theorems," and criteria 92 through 97 under the heading, "Evaluation of Theories and Hypotheses." Sixty-three of the criteria could not be applied to the example study.

\textsuperscript{64}See criterion 116, p. 47.
thesis and, of these, twelve were removed from the research criteria instrument, as will be seen in Chapter V.

It will be recalled from Chapter II that many of the criteria were found to be associated with one or more closely related statements. A complete list of such related statements, each under the appropriate research criterion, was given in Appendix I. This list may be examined for further clarification of the use of the criteria, particularly those which were not applicable to the example study.
Summary

Formulation and definition of the problem

The purpose of the study was identified as a serious effort to improve the quality of research in science education through the development and validation of criteria for evaluating research. It was proposed that such an effort would fulfill at least three needs. First, the use of research criteria for the purpose of evaluating research reports and research design could enable the researcher to gain better control of a method of inquiry. As a result, he could investigate more carefully the advisability of a proposed step in the case of a research design, or the adequacy of a completed step in the case of a research report. Second, the approximately sixty-five year history of educational research had produced a large and widely dispersed volume of self-criticism, especially during the last twenty-five years. It appeared that a synthesis of these evaluative comments as research criteria would be profitable for furthering the quality of research in science education. The third and final need for the study was found in the lack of research of
the kind undertaken in this study. No similar synthesis and
ranking of research criteria by expert judges could be found
in the literature.

The rationale for selecting research criteria for
science education from whatever literature such statements
could be found was presented. The rationale for selecting
expert judges from the physical and biological sciences,
science education, and the broad area of education was con­
sidered. It was concluded that although research in one area
may differ considerably from research in another area, there
are unifying traits which are common to research in all areas.
In addition, these unifying traits were found to be very
similar to the research criteria. Thus, the interdisciplin­
ary approach to the study seemed justified.

The relationship between the use of research criteria
and creativity in research was found to be related to the
first need, which was mentioned above, namely, that of
increasing the researcher's control over a research method.
By so doing, the researcher could find more freedom for
speculative thought. Thus, the research criteria could con­
tribute, although indirectly, to greater creativity in
research.

The problem which was investigated in this study
involved the development and validation of criteria which
would serve in evaluating completed research and research in
progress. The validation procedure included two questions:
(1) What is the order of importance of each of the research criteria as judged by research experts from the biological and physical sciences, from education, and from the subspecialty of science education? (2) What is the significance of differences in the judgments regarding the research criteria which appear among the research experts from the biological and physical sciences and the research experts from science education and the other areas of education? A final aspect of the problem involved an illustration of the use of criteria in evaluating research. It was determined that a research report from the area of science education would be selected for evaluation using the validated research criteria.

Broadly stated, this study was based on the hypothesis that research in science education can be improved through the use of research criteria. More specifically, the following null hypothesis was tested statistically using the data gathered in this study: There is no significant difference among the mean ratings of each criterion for the following groups of judges: (1) science education; (2) education, excluding science education; (3) biology; (4) chemistry; and (5) physics.

**Research procedures**

The research criteria were extracted mainly from the literature which has accumulated since approximately 1940.
Thus, the period of self-criticism in the history of educational research, which was described in Chapter I, was included. The research criteria were selected, organized, and then submitted to a group of graduate students. Then, after revision, the criteria were submitted to a group of college faculty members who served as pilot study judges. Representatives both from the sciences and from education served in the latter capacity. The research criteria which resulted from the process of development briefly described here were listed in Chapter II.

Expert judges were selected personally from the area of science education, and were selected by nominating processes from the areas of education, biology, chemistry, and physics. Each judge received in the first mailing (1) a cover letter, (2) a copy of the research criteria instrument for his own files, and (3) a copy of the research criteria instrument which was stapled into a heavy paper cover complete with return address and postage for mailing. Two follow-up letters were used, the second of which was accompanied by another copy of the research criteria instrument. A total return of 84.4% and a usable return of 66.1% were received.

Three limitations in the development and validation of the research criteria were identified. They were as follows:

1. The process of research and the research report, although related, are not identical. The usefulness of the criteria was limited by the choice which had to be made
between process and report as the criteria were written and organized.

2. The validation of the criteria was limited by the difficulty of communication due to discipline-related differences in terminology. Examples of application were not given to the expert judges for each criterion because of the prohibitive amount of time that would have been required to read and evaluate the material.

3. The validation of the criteria was limited by the amount of care which was taken by the expert judges in evaluating each criterion.

The following three assumptions in the development and validation of the research criteria were identified:

1. It was assumed that the criteria should be organized in a pattern which would permit their application first to research reports and only indirectly to the process of conducting research.

2. It was assumed that the development of the research criteria, as well as their validation, would be more complete and useful if an example of their application to research in science education was given.

In order to rank the criteria according to the importance expressed by the rating given by all of the judges and by each group of judges, the arithmetic mean was computed for each criterion. Numerical values of 4 for "great importance," 3 for "high importance," 2 for "average importance," 1 for
"little importance," and 0 for "no importance" were used. To measure the consistency of the rating by the judges, the standard deviation of the distribution of ratings from the mean was computed for each criterion as rated by all judges. The factors in this study indicated the use of an analysis of variance technique, or the F test, in order to test for significance of differences among the mean ratings of the judges.

**Presentation and analysis of data**

Three types of data were collected in this study: statistical, descriptive, and subjective. The statistical data included the ratings given by the expert judges for each criterion, and were expressed as means. In addition, the standard deviations from the combined mean of all judges for each criterion were given. The means were used to calculate the F value for each criterion. Groups of judges which differed from each other in their mean ratings then could be ascertained at the 1 percent and at the 5 percent level of significance. The standard deviations were used as an indication of the consistency of the ratings by the expert judges.

The descriptive data were presented in three tables, and represented the qualifications of the expert judges. Three classes of descriptive data were presented: (1) the mean number of years of research experience and the mean number of studies directed and conducted by the judges, (2)
the academic qualifications of the judges, and (3) the number of expert judges possessing experience in various types of research.

The third type of data which were collected, the subjective data, included the written comments of the judges. Three categories of comments were identified and examples of each type were presented.

An example of the application of the research criteria to a study in science education

A final aspect of the problem of this study involved illustration of the use of the criteria in the process of evaluating research. A Master of Arts thesis written in 1963 at The Ohio State University was selected for evaluation. An overview of the example study was presented, using the same four major divisions as were used in classifying the research criteria. A similar method of division was used in the evaluation itself. The purpose of illustrating the application of the research criteria was to help the researcher use the criteria as a tool in conducting his own research and in evaluating research reports.

Conclusions

Conclusions related to the ranking of the research criteria

The first problem in the validation of the research criteria was the following: What is the order of importance
of each of the research criteria as judged by research experts from the biological and physical sciences, from education, and from the sub-specialty of science education? The rank order of the research criteria as rated by the five groups of judges was given in Table 4.

One of the aspects of ranking was the anticipated removal of a number of the lowest ranking criteria from the instrument. It will be recalled that the mean scores for the five groups of judges were found to be significantly different for thirty of the one hundred and sixteen criteria. In addition, careful examination of the mean scores (Table 14 in Appendix IV) which were used to find the rank order of importance of the research criteria, revealed that twenty-two of the criteria had received mean ratings of less than "average importance" by one or more of the five groups of judges. Ten of these twenty-two criteria also were in the group of thirty criteria for which significant differences among mean scores were obtained. These two factors, a mean score of less than average importance, and significant differences among mean scores, were used in deciding which of the twenty-two criteria should be removed from the research criteria instrument. Only those criteria were removed for which (1) ratings of less than average importance were given by one or more groups of judges, and (2) no statistically significant differences were found among the five groups of judges in rating the criteria. The cut-off point of "less
than average importance," although somewhat arbitrary, was selected because of the large difference in the number of criteria rated at average or greater importance and the number rated at less than average importance. The second factor, the lack of significant differences among mean scores, was added because it was felt that a low rated criterion should be retained only if one or more of the groups of judges disagreed significantly in rating the criterion.

Five of the twenty-two criteria were considered to be of less than average importance by three or four of the five groups of judges. Three of these five criteria should be removed from the research criteria instrument. They were as follows:

1. Criterion number 3: "The problem statement, including all of its parts, should encompass all elements of the problem." This criterion was considered to be of less than average importance by the biology, chemistry, and physics judges. Although the science education and education judges rated this criterion higher, there was no statistically significant difference in ratings among the five groups of judges.

2. Criterion number 35: "The possible alternative answers which are hypothesized for a given question should be independent of one another (or not overlap)." This criterion was rated at less than average importance by the biology, chemistry, and physics judges. The mean scores of
the science education and education judges, although higher, were not significantly higher.

3. Criterion number 100: "The summary should include only previously reported data." This criterion was ranked at less than average importance by the biology, chemistry, and physics judges. The mean score of the five groups of judges were not significantly different from each other.

Two of the five criteria mentioned above should be retained. They were as follows:

1. Criterion number 10: "The problem chosen should have either direct significance for people's needs or should serve as a foundation for subsequent research having significance for people's needs." This criterion received a rating of less than average importance by the science education, biology, chemistry, and physics judges. The mean scores of the science education and education judges were higher at the 5 percent level of significance than the mean scores of the other three groups of judges.

2. Criterion number 55: "The data collected should be within the boundaries defined by the hypotheses." This criterion was ranked at less than average importance by the biology, chemistry, and physics judges. The mean scores of the science education and education judges were higher at the 1 percent level of significance than the mean scores of the other three groups of judges.
Four of the twenty-two criteria were considered to be of less than average importance by any two of the five groups of judges. Two of these four criteria should be removed from the instrument. They were the following:

1. Criterion number 75: "All data which are collected for the solution of a given problem should be reported." This criterion was rated at less than average importance by the education and biology judges. The mean scores of the five groups of judges were not significantly different from each other.

2. Criterion number 112: "Generalizations should be directed to the solution of problems other than those considered in the study." This criterion was rated at less than average importance by the science education and education judges. The mean scores of the five groups of judges were not significantly different from each other.

The remaining two criteria from the group of four above should be retained. They were the following:

1. Criterion number 2: "The problem statement should include all subordinate problems which must be answered." This criterion was rated at less than average importance by the biology and chemistry judges. The mean scores of the science education, education, and physics judges were higher at the 1 percent level of significance than the mean scores of the other two groups of judges.
2. Criterion number 80: "The analysis of data should be anticipated and worked through abstractly or in a pilot study before the study itself is begun." This criterion was rated at less than average importance by the biology and chemistry judges. The mean scores of the science education and education judges were higher at the 1 per cent level of significance than the mean scores of the other three groups of judges.

Thirteen of the research criteria were considered to be of less than average importance by one of the five groups of judges. Seven of these thirteen criteria should be removed from the research criteria instrument. They were as follows:

1. Criterion number 5: "The origin of a research problem should be identified." This criterion was rated at less than average importance by the biology judges. The mean scores of the five groups of judges were not significantly different from each other.

2. Criterion number 43: "The research design should include several planned directions of inquiry in order to increase efficiency and comprehensiveness." This criterion was rated at less than average importance by the education judges. The mean scores of the five groups of judges were not significantly different from each other.

3. Criterion number 54: "Modern technological instruments should be used to make complete visual and/or oral records of primary data for review by independent
investigators." This criterion was rated at less than average importance by the physics judges. The mean scores of the five groups of judges were not significantly different from each other.

4. Criterion number 57: "Data should be gathered in a well organized, highly structured, and yet creative manner." This criterion was rated at less than average importance by the physics judges. The mean scores of the five groups of judges were not significantly different from each other.

5. Criterion number 90: "Observation and theory should serve as checks on one another with neither having priority over the other." This criterion was rated at less than average importance by the physics judges. The mean scores of the five groups of judges were not significantly different from each other.

6. Criterion number 107: "When several possible explanations of a phenomenon seem equally correct, the simplest should be favored." This criterion was rated at less than average importance by the chemistry judges. The mean scores of the five groups of judges were not significantly different from each other.

7. Criterion number 114: "Creative speculation concerning possible applications and recommendations should be made." This criterion was rated at less than average importance by the physics judges. The mean scores of the five
groups of judges were not significantly different from each other.

The remaining six criteria from the group of thirteen mentioned above should be retained. They were the following:

1. Criterion number 12: "The problem should be delimited exactly as to scope, the nature and quantity of data desired, and the methods or procedure to be followed." This criterion was rated at less than average importance by the chemistry judges. The mean scores of the science education and education judges were higher at the 1 percent level of significance than the mean scores of the other three groups of judges.

2. Criterion number 33: "The researcher should display creativity in his formulation of hypotheses." This criterion was rated at less than average importance by the education judges. The mean scores of the science education, biology, chemistry, and physics judges were higher at the 1 percent level of significance than the mean score of the education judges.

3. Criterion number 39: "Hypotheses should be revised on the basis of induction from the data which accumulates as the study progresses." This criterion was rated at less than average importance by the education judges. The mean scores of the science education, biology, chemistry, and physics judges were higher at the 1 percent level of significance than the mean scores of the education judges.
4. Criterion number 44: "The experimental design should include the replication of one or more variables whose effects have already been assessed as significant in a previous experiment." This criterion was rated at less than average importance by the education judges. The mean scores of the science education, biology, chemistry, and physics judges were higher at the 5 percent level of significance than the mean score of the education judges.

5. Criterion number 56: "The data included within the defined boundaries of the study should be collected without personal choice." This criterion was rated at less than average importance by the chemistry judges. The mean scores of the science education, education, and biology judges were higher at the 1 percent level of significance than the mean scores of the other two groups of judges.

6. Criterion number 92: "A theory should encompass a relatively wide range of data." This criterion was rated at less than average importance by the chemistry judges. The mean scores of the science education, education, biology, and physics judges were higher at the 5 percent level of significance than the mean scores of the education judges.

To summarize, ten of the twenty-two criteria were retained because, although one to four groups of judges rated each of them at less than average importance, the remaining groups of judges rated them higher at the 1 or the 5 percent level of significance. Twelve of the twenty-two criteria
were removed because one to three groups of judges rated them at less than average importance and because no statistically significant differences among the ratings or mean scores of the five groups of judges were found.

The conclusions relating to the ranking of the research criteria have dealt, up to this point, with criteria that received low mean scores. The fact that only twenty-two of the one hundred and sixteen criteria received ratings at less than average importance by one to four groups of judges, and the fact that only twelve of these were finally removed from the research criteria instrument is an indication of the high rating which most of the criteria received. Table 11 has been compiled to facilitate presenting conclusions based on mean scores listed in Table 14, Appendix IV. The following conclusions are evident from Table 11: (1) most of the criteria were ranked at more than average importance for evaluating research, and (2) the science education judges found a greater number of criteria to be of more than high importance in evaluating research than did any of the other groups of judges. Although there were differences among the other four groups of judges in their distribution of ratings, none of these differences was as pronounced as that between the science education judges and the other groups of judges.
TABLE 11
NUMBER OF CRITERIA RANKED WITHIN SPECIFIED MEAN SCORE RANGES OF IMPORTANCE BY GROUPS OF JUDGES

<table>
<thead>
<tr>
<th>Mean Score Ranges of Importance(^a)</th>
<th>Science Education</th>
<th>Education</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
<th>All Five Groups Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00 - 0.99</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.00 - 1.99</td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>2.00 - 2.99</td>
<td>28</td>
<td>66</td>
<td>56</td>
<td>45</td>
<td>46</td>
<td>52</td>
</tr>
<tr>
<td>3.00 - 4.00</td>
<td>86</td>
<td>44</td>
<td>51</td>
<td>60</td>
<td>61</td>
<td>60</td>
</tr>
</tbody>
</table>

\(^a\)Numerical values of 0 for "no importance," 1 for "little importance," 2 for "average importance," 3 for "high importance," and 4 for "great importance" were used.
Conclusions related to the consistency of ratings

To obtain a measure of the consistency of the ratings by the judges, the standard deviation of the distribution of the ratings from the combined mean was computed for each criterion. Table 5 in Chapter III contained a frequency distribution of the standard deviations for the five groups of judges combined. It is difficult, on the basis of Table 5 alone, to draw conclusions relative to the magnitude of consistency in the ratings. If one examines the frequency distribution of ratings for a selected sample of criteria, however, conclusions become more evident. The frequency distributions of ratings for three of the criteria will be examined. Two of these three distributions resulted in the lowest and the highest standard deviations, and the remaining one was near the mean of the standard deviations.

Criterion number 1 received the lowest standard deviation value of 0.43. Fifty-nine of the seventy-two judges rated it at "great importance" (4), twelve judges rated it at "high importance" (3), and one judge rated it at "average importance" (2). No ratings of less than average importance were given. There was a high level of consistency in the rating of this criterion, and this fact was reflected in the low standard deviation of 0.43.

 Criterion number 62 possessed a standard deviation near the mean standard deviation. The mean standard devia-
tion was 1.03, and the standard deviation for criterion number 62 was 1.02. Thirty-four judges rated this criterion at "great importance" (4), twenty-five at "high importance" (3), eight at "average importance" (2), two at "little importance" (1), and three at "no importance" (0). This distribution of ratings, although wider than that of criterion number 1, still indicated a fairly high level of consistency.

Criterion number 100 received the highest standard deviation, 1.52. Twenty-one judges rated it at "great importance" (4), fifteen at "high importance" (3), twelve at "average importance" (2), nine at "little importance" (1), and fifteen at "no importance" (0). Thus, a large number of judges (twenty-one) strongly approved of this criterion while a large number (fifteen) strongly disapproved. Because of this difference of opinion a broad distribution of ratings resulted, with the two extremes weighted rather heavily. Thus, a relatively large standard deviation, and a low level of consistency, was obtained.

On the basis of these three examples, and the data in Table 5, it seems reasonable to conclude that the ratings of well over half of the criteria (standard deviation range of from 0.43 to about 1.24) were consistent enough to have meaning with respect to the ranking of the criteria. As the standard deviations approached 1.52, however, the rank order of the criteria became somewhat meaningless. Because examples of research were not associated with each of the criteria at
the time judgments were made, complete consistency could not be expected.

Conclusions related to the **significance of differences**

Examination of Table 7 in Chapter III revealed several groupings of the data. Each of these groupings will be discussed in the following paragraphs.

The first grouping of data was that in which the means for the science education judges were significantly greater than the means of the other four groups of judges in eleven of the thirty cases where significant differences existed. These eleven criteria were the following:

1. Criterion number 13: "In addition to simple delimitation of the problem, the subjects (if any are used) should be described with reference to the variables among them."

2. Criterion number 22: "A definition should be stated in positive terms, rather than in negative terms, wherever possible."

3. Criterion number 41: "The adequacy of the research methodology chosen for solution of the problem should be explained and defended."

4. Criterion number 53: "The data of research should approach the reality of the situation being investigated as nearly as possible."

5. Criterion number 61: "The degree of reliability and validity which is accepted should be justified with reference to the objectives of the research problem."
6. Criterion number 63: "The statistical technique should be identified before the data are gathered."

7. Criterion number 68: "The statistical sample should be selected from a carefully defined parent population."

8. Criterion number 85: "Corrections should be made in the analysis of the data for any distortions which may have occurred in the data gathering process."

9. Criterion number 108: "Skillful grouping and appropriate subordination of conclusions under major headings should be used to avoid encyclopedic enumeration of individual statements."

10. Criterion number 110: "The limitations of a study should be presented to guide the reader in using the findings, conclusions, and recommendations."

11. Criterion number 111: "A study should be thorough enough to arrive at carefully stated results which have a substantial degree of reliability and validity."

There does not appear to be anything unique about the above list that would have caused the science education judges to rate these eleven criteria higher than did any of the other groups of judges. It will be recalled that the science education judges generally rated the research criteria higher than did the other groups of judges, and that in all thirty cases where significant differences occurred, the means of the science education judges were in the significantly higher group. It also will be recalled that the
science education judges were selected, with the assistance of the chairman of the dissertation committee, rather than being nominated. It is possible that a more sympathetic group of judges was obtained, because of the selection process, from the science education area. It also is possible that the greater variety of research experiences possessed by the science education judges, as reflected in Table 10, Chapter III, predisposed these judges to favor the research criteria more highly than did the other groups of judges. In any case, it seems reasonable to conclude that the significantly higher ratings given to eleven of the criteria by the science education judges reflected their generally more favorable reaction to all of the criteria.

The second grouping of data from Table 7 was that in which the means of the science education and education judges were significantly greater than the means of the biology, chemistry, and physics judges for seven of the research criteria. There does not appear to be anything inherent in two of the seven criteria that would have caused the science education and education judges to rate them significantly higher than did the other three groups of judges. These two criteria were as follows:

1. Criterion number 48: "The self-contained experiment should include a representative sample of the subjects or objects under study."
2. Criterion number 115: "Applications or recommendations which are based on the creative insight of the investigator should be labeled plainly as such."

One of the seven criteria in the second grouping of data from Table 7 was particularly interesting. This criterion, number 10, was as follows: "The problem chosen should have either direct significance for people's needs or should serve as a foundation for subsequent research having significance for people's needs." This criterion was given a low rating by all the judges, including the science education and education judges, even though the latter two groups of judges rated it significantly higher than did the other three groups. With regard to the idea expressed in this criterion, it would seem that all research in any area and, indeed, all human activity eventually should be justified with reference to human needs. Evidently the judges interpreted the words, "people's needs," much more narrowly than these words might have been interpreted. It is true that much worthwhile research is not based on human needs which are readily apparent, but eventually, as experience has shown, human needs frequently are served.

The remaining four of the seven criteria which received significantly higher ratings from the science education and education judges were as follows:

1. Criterion number 12: "The problem should be
 delimited exactly as to scope, the nature and quantity of data desired, and the methods of procedure to be followed."

2. Criterion number 55: "The data collected should be within the boundaries defined by the hypotheses."

3. Criterion number 64: "The statistical hypotheses should be considered at the time the experiment is designed."

4. Criterion number 80: "The analysis of data should be anticipated and worked through abstractly or in a pilot study before the study itself is begun."

The above four criteria all serve to help clarify, limit, define, or specify the area of study. It is possible that the science education and education judges, who are more familiar with the difficulties of conducting research in the less tangible educational research area, for this reason judged the above four criteria to be helpful.

The third grouping of data from Table 7 was that in which the means of the science education, biology, chemistry, and physics judges were significantly greater than the means of the education judges in five cases. These five criteria were as follows:

1. Criterion number 29: "The review of related literature should include the previous research which may contribute to the solution of the problem."

2. Criterion number 33: "The researcher should display creativity in his formulation of hypotheses."
3. Criterion number 34: "A hypothesis should bring together a given number of observations and interpret the connections between them."

4. Criterion number 39: "Hypotheses should be revised on the basis of induction from the data which accumulate as the study progresses."

5. Criterion number 44: "The experimental design should include the replication of one or more variables whose effects have already been assessed as significant in a previous experiment."

The above five criteria all are important in research in the more established fields such as the biological and physical sciences. The science education judges also would be expected to have a greater understanding of these criteria due to their education in science. For example, the review of related literature is more important in fields which have much longer histories of significant and useful research than has the field of education. Also, although creativity in formulating hypotheses, as suggested in criterion number 33, should be important in all areas of research, it is possibly more important in fields where a vast amount of factual information is available for the stimulation of thought.

Similar conclusions, although somewhat speculative in nature, could be made concerning all five of the above criteria.

The fourth and last grouping of data from Table 7 was that in which the means of the science education and physics
judges were significantly greater than the means of the education, biology, and chemistry judges for four of the research criteria. The four criteria were as follows:

1. Criterion number 59: "A data-gathering instrument should have validity or truthfulness; it should measure what it purports to measure."

2. Criterion number 60: "A data-gathering instrument should have reliability or consistency; it should agree with itself when used repeatedly with the same individuals, groups, or objects."

3. Criterion number 67: "The basic assumptions underlying statistical methods should be satisfied by the data to which they are applied."

4. Criterion number 69: "The probability of error estimate should be justified."

There does not appear to be anything unique concerning the above four criteria that would have produced the significant differences which were observed. Thus, no conclusions beyond the one just stated will be made.

Three of the thirty criteria for which significant differences among means were obtained could not be fitted into any of the four groupings of data from Table 7 which were just discussed. These criteria were as follows:

1. Criterion number 2: "The problem statement should include all subordinate problems which must be answered."
2. Criterion number 56: "The data included within the defined boundaries of the study should be collected without personal choice."

3. Criterion number 92: "A theory should encompass a relatively wide range of data."

Criterion number 2 was rated significantly higher by the science education, education, and physics judges. Criterion number 56 was rated significantly higher by the science education, education, and biology judges. Criterion number 92 was rated significantly higher by the science education, education, biology, and physics judges. A possible explanation for the distribution of ratings given above was that these three criteria lacked in clarity. For example, some of the judges were troubled by the word "all" in criterion number 2. In criterion number 56, the apparent exclusion of serendipity, or fortuitous discovery, was a source of disagreement among some of the judges. In criterion number 92, the word "relatively" also was questioned. In each of these cases, a more precise choice of wording might have resulted in a distribution of ratings with more apparent meaning.

In summary, the significant differences which occurred among the mean scores of the five groups of judges who rated the research criteria were discussed. Thirty criteria were found for which significant differences among means existed. Possible explanations for the significant differences were
given for as many of the ratings of the thirty criteria as permitted reasonable speculation.

Conclusions related to the descriptive data

The descriptive data gathered in this study consisted of information concerning the qualifications of the expert judges. The following conclusions are evident from this data: (1) Only minor differences in the quantity of research experience possessed by the expert judges were apparent. (2) The academic qualifications of the expert judges were similar, with the possible exception of the education judges who were, on the average, less well qualified. (3) In general, the judges were limited in the number of types of research experience which they had had (an average of 1.0 to 2.4 types per judge.)

Conclusions related to the subjective data

The subjective data gathered in this study consisted of solicited, unstructure comments written by the judges. A narrow point of view about the processes of research was at times evident in these comments. This may be a reflection of the third conclusion which was listed in the preceding section, that the number of types of research experience possessed by many of the judges was limited. This lack of a variety of kinds of research experience may account for the low ratings of individual criteria, a lack of appreciation of
the study in general, and difficulty in understanding the meaning of certain of the criteria. Most of the judges, however, provided constructive comments, and displayed a genuine interest in the study.

Recommendations

Two classifications of recommendations will be given. The first will relate to possible uses of the research criteria; the second will contain suggestions for future research.

Recommendations for the use of the research criteria

The research criteria may be used by beginning graduate students for assistance in evaluating completed research in professional journals. After the experience of selecting the appropriate criteria for application to short research reports of this sort, longer reports such as theses and dissertations might profitably be evaluated. An entire class of graduate students might benefit from a thorough discussion of a thesis or dissertation which all members of the class had evaluated separately. Experiences of this sort would be of considerable value in developing some of the attitudes and skills which are necessary in research.

A student who had become familiar with the research criteria through experiences of the type described in the preceding paragraph would be in an excellent position to make maximum use of the criteria when conducting his own research.
The criteria would be of considerable assistance as reminders of some of the factors which should be examined and decided upon in the research prospectus. If a thorough consideration of the research design was presented in the research prospectus, then further reference to the criteria probably would be unnecessary until the research report was being written. Then the research criteria again could serve as reminders of the many factors which must be remembered in the final reporting of research. An experienced researcher also might find the criteria useful if employed in a manner similar to that described in this paragraph.

Recommendations for future research

The present study has served to develop and validate a rather lengthy list of research criteria. It was pointed out that examples of application of the research criteria to research reports were not given in the research criteria instrument because of the excessive amount of time that would have been required for judging. It is obvious from the comments of the judges, however, that examples would have increased the validity and reliability of the ratings. Thus, a study of research criteria in which examples of application were given would be of value. It might be possible to divide the criteria into groups and ask a given judge to evaluate only one group of criteria. Such a procedure would have the
distinct advantage of decreasing the time demand placed on each judge, while still allowing the use of examples of application.

Another interesting area of research would involve the reasons why the judges differed significantly in their ratings of some of the criteria. Speculative conclusions were reached concerning some of the thirty criteria, but justifiable conclusions could not be proposed for all of them.

The relationships which existed between the qualifications of the judges and their ranking of the criteria also would be of interest as a subject of study. If such relationships were known, the research criteria instrument could be used as a measure of the level of achievement of researchers at various stages in their development. The ratings given by a random sample of science educators, for example, could be compared to the ratings given by expert judges.

If studies of the type just indicated were conducted, involving various groups and numbers of people, a greater general awareness of the need for increased quality in research would develop more readily. Studies involving research criteria, in addition to their stimulating effect, would have the added advantage of immediately providing a means for improving research, namely, through the criteria themselves.
APPENDIX I

RESEARCH CRITERIA AND RELATED STATEMENTS
FORMULATION AND DEFINITION OF THE PROBLEM

Problem Statement

1. The problem should be stated precisely in clear and understandable language.

2. The problem statement should include all subordinate problems which must be answered.

3. The problem statement, including all of its parts, should encompass all elements of the problem.

4. The location of the problem statement in the research report should be identified so that it can be found and recognized readily.

   The problem statement should appear early in the research report.1

Introduction

Problem Choice

5. The origin of a research problem should be identified.

6. The body of organized knowledge in an area should be examined for gaps, for discrepancies, and for the directions in which extension is possible.

   Previously completed research should be evaluated for adequacy and repeated or extended if necessary.2

   Conclusions of research studies which were conducted at different points in time, location, and social situation should be re-examined.3

7. A problem should be chosen which avoids unnecessary duplication of research effort.

---

1Van Dalen, Educational Administration & Supervision, 44 (May, 1958), 174-181.

2Ibid.

The review of related literature should show that unnecessary duplication of research effort has been avoided.\textsuperscript{4,5}

8. A problem should be chosen which can be formulated and defined with sufficient precision to permit a good idea of the goal to be achieved before it is attempted.

9. Attention should be given to the theoretical framework of knowledge underlying issues and problems before their resolution or solution is sought directly.

Rival methods of achieving a demonstrably desirable aim should be tested only if readiness for achieving the aim normally is displayed by the learner or if the aim is of such importance that it should be externally imposed by society regardless of pupil readiness.\textsuperscript{6}

Such questions as the following should be asked concerning "method A" before the relative effectiveness of "method A" and "method B" is tested: (a) Why do some students progress rapidly using "method A" and others slowly? (b) Why do learning difficulties arise with "method A," what are they, and at what point in the sequence of learning do they arise? (c) Why is "method A" effective or ineffective?

10. The problem chosen should have either direct significance for people's needs or should serve as a foundation for subsequent research having significance for people's needs.

An applied research problem should be judged or evaluated for the extent to which it meets some


\textsuperscript{5}Van Dalen, Educational Administration & Supervision, 44 (May, 1958), 174-181.


\textsuperscript{7}Brownell, California Journal of Educational Research 14 (March, 1963), 51-66.
socially determined need, and, therefore, contributes to societal development.8,9,10

Research in education should be judged for its contribution to existing knowledge rather than for any immediate service which it might render in a particular educational situation.11

The problem, when solved, may help to define, refine, and clarify concepts and objectives which were previously more or less nebulous although generally accepted as fundamental and essential.12

A problem should be chosen which may yield results of significance to as large a number of people, over as large a territory, and for as long a duration of time, as possible.13,14,15,16,17

The problem, when solved, may introduce and evaluate new and highly effective instructional procedures.18

8Gaylord and Stunkel, Educational and Psychological Measurement, 14 (Summer, 1954), 294-300.


10Whitney, p. 311.

11Crawford, p. 20.


13Abelson, pp. 20-21.

14Crawford, p. 21.


16Richey and Richey, Educational Horizons, 41 (Winter, 1962), 63-68.


The problem, when solved, may clarify the nature and relative values of various functional materials in teaching.19

The problem, when solved, may bring about marked changes in instructional materials and/or practices.20,21,22

The problem, when solved, may implement the attainment of accepted objectives by establishing the relative values and practicability of certain types of organization, methods, or procedures.23

The problem, when solved, may implement the attainment of accepted objectives by refining and improving methods and devices of instruction already in common use.24

The problem, when solved, may reveal and explore promising fields of curricular expansion.25

A study or a group of studies may result in the following kind of statement: teachers who are X accomplish learning outcomes Y through procedure A with pupils S.26

---


24Ibid.

25Ibid.

The principal task of educational research is to maximize learning in the schools through the discovery and development of reliable predictive systems which are based upon laws and principles.27

11. Experimentation involving human subjects should be planned and spaced carefully for their protection.

Delimitation and Analysis

12. The problem should be delimited exactly as to scope, the nature and quantity of data desired, and the methods of procedure to be followed.

13. In addition to simple delimitation of the problem, the subjects (if any are used) should be described with reference to the variables among them.

14. The factors which are selected for study should be the decisive ones for solution of the problem.

15. Variables which are used in experimental research should differentiate between or among the subjects or objects under study.

Assumptions

16. Assumptions upon which the inquiry rests should be identified.

The assumptions which underlie initial and subsequent hypotheses should be listed.28

17. Assumptions upon which the inquiry rests should be justified.

18. The assumptions in a study should be consistent with each other.

19. It should not be assumed that the non-research literature is an accurate indicator of best or prevalent practice.


28Pella, Science Education, 45 (December, 1961), 396-399.
An instrument which is designed to survey practices and content should not assume that present practices and content were selected on a sound basis.  

Definitions

20. A definition should be expressed in precise language.

21. A definition must not be circular; it must not contain, directly or indirectly, the subject to be defined.

22. A definition should be stated in positive terms, rather than in negative terms, wherever possible.

23. Important terms, that are employed in an unusual sense, should be defined.

24. Undefined terms should be employed according to the meanings demanded by standard references.

25. Operational definitions of all generalized concepts which will be studied should be specified.

26. Defined terms and concepts should be used in the entire study according to the given definition.

Related Literature

27. The related literature should be analyzed for its relevancy to the problem situation.

The interpretation of related literature should demonstrate the investigator's familiarity with the integrating concepts possessed by the subject matter under investigation.  

28. The related literature should be analyzed for any bias which may have significance for the study.

The author of an historical document should not be judged ignorant of certain events simply because he failed to mention them.

30 Cohen and Nagel, p. 92.
31 Good, p. 138.
Facts which are taken from primary sources should be verified by at least two independent witnesses.  

29. The review of related literature should include the previous research which may contribute to the solution of the problem.

30. The problem should be related to an orienting frame of reference or a systematic area of study.

A conceptual framework should be used to assist in organizing the knowledge which is available into a meaningful set of relations.

The conceptual framework may enable the researcher to get a clear perspective of the variables at work as well as their inter-relations.

The conceptual framework may be used to reveal modes of attack for collecting additional information.

The conceptual framework may assist the researcher in correctly choosing and elaborating upon significant hypothesis.

31. Research in a given area should make use of relevant principles and research techniques of related disciplines.

32. Literature sources should be assigned to definite authors and times.

32 Ibid.

33 Whitney, p. 200.

34 Brown and Ghiselli, p. 135.

35 Churchman and Ackoff, p. 217.

36 Ibid.

37 Gaylord and Stunkel, Educational and Psychological Measurement, 14 (Summer, 1954), 294-300.

38 Brown and Ghiselli, p. 135.

39 Ibid.

40 Churchman and Ackoff, p. 217.

41 Brown and Ghiselli, p. 167.
Hypotheses and Derived Theorems

33. The researcher should display creativity in his formulation of hypotheses.

34. A hypothesis should bring together a given number of observations and interpret the connections between them.

35. The possible alternative answers which are hypothesized for a given question should be independent of one another (or not overlap).

36. Hypotheses should be formulated on the basis of a thorough analysis of the theoretical and factual background of the problem.

Hypotheses should be derived from the problem statement itself, because no situation for which resolution is possible is ever completely indeterminate, confused, obscure, or conflicting.42

37. Hypotheses should be formulated so that implications can be traced clearly by means of well-established techniques of deduction.

38. The consequences of alternative hypotheses should be developed for comparison with observable phenomena.

The development of the hypothesis should serve to clarify the issues at stake and to crystallize the problem for investigation.43

39. Hypotheses should be revised on the basis of induction from the data which accumulate as the study progresses.

PROCEDURAL DESIGN

Procedure

Design Selection

40. The problem, from the beginning, should condition the steps taken to solve it.

42Churchman and Ackoff, p. 198.

43Van Dalen, Educational Administration & Supervision, 42 (December, 1956), 457-460.
Research may display a measure of creativity by devising new techniques and new approaches to problems.  

"Quantitative," "statistical," and "experimental" should not be set in opposition to "qualitative," "clinical," and "nonexperimental," but rather, a balanced approach in selecting techniques appropriate for solving the problem at hand is needed.  

If a unique competency or position possessed by the researcher lends itself to the solution of the problem, that competency or position should be used.  

The researcher should recognize the many subtle but fundamental differences in the situations involved before he attempts to apply intact to a new field or problem a design or technique which has been successfully used in or with another.  

41. The adequacy of the research methodology chosen for solution of the problem should be explained and defended.  

The research design should rely on empirical data more than on assumptions.  

The study, if of the questionnaire type, is justified when precise description is not material to a valid conclusion.  

---

44Fattu, pp. 1-21.  
45Good, School and Society, 91 (March, 1963), 140-143.  
47Lindquist, p. vi.  
The study, if comparative, should have similar or even identical primary objectives for the experimental and the control groups. 50

The speculative method of problem solving should be applied in those areas (questions of ultimate reality, truth, origin, and values) where there is no efficient method of checking conclusions through observation. 51

Measurements, which have as their aim to rank or order certain properties or other objects, should satisfy the formal demands of measurement. 52

Because human behavior and educational objectives are more complex than univariate experimental designs can adequately handle, more sophisticated and fruitful experimental designs should be employed in educational research. 53

The researcher should justify his decision as to whether it is better to control the variables experimentally or to test the contributions of the variables statistically. 54, 55

---

51 Churchman and Ackoff, pp. 118-120.
52 Ibid., p. 224.
54 Symonds, Journal of Educational Psychology, 47 (February, 1956), 100-109.
Evaluation of variants in the educational structure under appraisal must be done either with reference to the effects upon the behavior of children, or with reference to some secondary criteria which must, in turn, be validated against child behavior.56,57

42. The research design should provide for collection of the most objective data available.

Questionnaires should not be used to obtain data that is easily accessible from official records.58

Because learning is likely to be less when the expectations of pupils about the teacher's role differ from that of the teacher himself, there should be congruence between the teacher's idea of his role and the idea that the pupils have of the proper role for the teacher.59

The research design should include the development of a clear plan of attack on the problem, and the methodical following of procedures appropriate to the plan.60

The experimental design which is selected should exhibit a clear idea in advance of the investigation as to what effects would be expected.61

56Hoyt, Phi Delta Kappan, 35 (October, 1953), 59-62.


60Obourn and Boeck, Science Education, 44 (December, 1960), 374-399.

The most promising (i.e., most likely to be relevant) variables should be built into the design so that gains can be assessed separately for each variable. 62

43. The research design should include several planned directions of inquiry in order to increase efficiency and comprehensiveness.

Research should be designed not only to discover new and isolated facts, but also to discover wholes or patterns, and consequences thereof. 63,64

Research should be designed to supply general knowledge of student development and institutional functioning which is useful not only in restricted contexts, but also in providing dynamic descriptions as to how and why the status or achievement or opinions changed. 65,66

Designs of the factorial type in which several variables are manipulated simultaneously and then evaluated independently of each other should replace single variable designs in controlled classroom research. 67

44. The experimental design should include the replication of one or more variables whose effects have already been assessed as significant in a previous experiment.

Design for Replication and Validity

45. The research design should be described in sufficient detail to permit repetition of the study by independent investigators.


64 Monroe and Engelhart, pp. 444-445.


66 Watson and Cooley, pp. 297-312.

46. The degree of internal and external validity in the research design should be identified and justified.

Nearness to the true average experimental effect, regardless of the degree of replicability or generalizability, is the only criterion of an experiment's usefulness. 68

Generalizability or relevance to what actually occurs in the school is the only criterion of an experiment's usefulness. 69

If solution of the problem requires action research in actual school situations, no attempt should be made to isolate out a factor and study it alone, divorced from the environment which gave it meaning. 70

The statistical technique should provide for evaluation of what happened to the individual in the process by which he arrived at the end product. 71

47. The data collected should afford some basis for generalization.

Generalizations should be restricted to those warranted by the data. 72

The experimental design should allow generalizations to be made from the experimental situation to comparable non-experimental situations. 73, 74

69 Ibid.
71 Barnard, pp. 1-29.
74 Good, p. 372.
Generalizations based on descriptive research should be limited completely to the situation under study and to the areas in which substantial information has been obtained and verified.\(^75\)

Self-Contained Design

48. The self-contained experiment should include a representative sample of the subjects or objects under study.

49. The self-contained experiment should include a control to exclude, at a known level of probability, a number of alternative interpretations of the experimental results.

A random sampling technique should be used whenever possible in order to remove the possible bias which may result from any personal selection by the researcher.\(^6,77,78\)

In any research where the object being studied is variable both for the perceiver and for the perceived, random selection of both groups should be practiced if conclusions are to be made for both.\(^79\)

In order for data to be representative, a sufficiently large and unbiased sample must be examined with a known and acceptable degree of precision.\(^80\)

\(^75\)Ray, p. 172.


\(^80\)Breitwieser, Phi Delta Kappan, 24 (December, 1941), 185.
In order to provide a check on possible bias, research which is designed to judge the teaching of academic subjects should have on the evaluating team representatives from the academic field in question.81

In research which is designed to judge teacher competency, provision should be made for removing possible bias introduced by the teacher himself.82

50. The self-contained experiment should supply independent evidence on the question in dispute.

The collection, treatment, and presentation of data should display qualitative care and accuracy and should be reasonably sufficient in quantity for adequate solution of the problem.83,84

A good experimental design will make possible an objective test of a specific hypothesis.85

51. The self-contained experiment should permit an estimate of the experimental errors which affect the comparisons made.

The research design should state clearly the degree of precision desired.86,87

The selection of valid "error" terms and the use of "pooled" error terms which are provided for in the experimental design should be defended.88

81Koerner, pp. 64-68.
82Ibid.
83Abelson, pp. 308-309.
85Lindquist, pp. 6-7.
86Ibid.
87Pella, Science Education, 45 (December, 1961), 396-399.
88Lindquist, p. vii.
A good experimental design will permit a quantitative description of the precision of the observed treatment effects regarded as estimates of the "true" effects.89

Design for Data Gathering

52. The design for making observations should establish causal connections between events in which the cause is a sufficient and necessary condition for the effect.

Historical research in education should provide the proper social, political, economic, and cultural setting for the period covered as well as any other pertinent processes and institutions which might effect school practices.90,91,92

The significance of the experiment and the errors of the observations should be taken into account in determining the number of observations which should be taken.93

53. The data of research should approach the reality of the situation being investigated as nearly as possible.

Factual sources should be selected for their pertinence to the solution of the problem.94,95

Data should be selected on the basis of operational definitions of the generalized concepts which are under study.96

89Ibid., pp. 6-7.


91Perdew, pp. 82-83.

92Perdew, Phi Delta Kappan, 32 (December, 1950), 134-136.

93Churchman and Ackoff, p. 235.

94Perdew, pp. 82-83.

95Perdew, Phi Delta Kappan, 32 (December, 1950), 134-136.

The research design should allow objective observations that are unbiased estimates of true effects. 97, 98

Procedures should be used that will make possible an accurate description of all the variables operating in the situation. 99

Operations should be specifiable by which any competent observer can tell whether or not the proposal has in fact been carried out. 100

Tests should be made to be sure that the phenomena to be observed have not been destroyed, missed, or misinterpreted by the observer. 101

Self-descriptions by extremely ego-involved subjects should be avoided. 102

If student progress is not used as the criterion of teacher competency, then the relationship between student progress and the criteria which are used to judge teacher competency should be demonstrated. 103

If judgments are made in a study, care must be taken that the judges have the necessary intelligence, information, background and other qualifications to permit them to make the judgments. 104

97 Lindquist, p. 3.
98 Pella, Science Education, 45 (December, 1961), 396-399.
99 Brown and Ghiselli, p. 177.
102 Ibid.
103 Koerner, pp. 64-68.
104 Symonds, Journal of Educational Psychology, 47 (February, 1956), 100-109.
If judgments are made in a study, the basis on which the judgments were made should be specified. 105

A subject should not report interactions of which he is a part and on which the culture has placed distinct values. 106

Generally, all the data should be collected at the time the experiment is conducted. 107

Personal experiences which were not subject to the controls under which the research data were gathered should be omitted. 108

The reliability of retrospective data should be tested before they are accepted. 109

Attention should be focused on the interactions among relevant variables as well as on the variables themselves. 110

Any peculiar samplings of persons or materials involved should be accounted for adequately. 111

54. Modern technological instruments should be used to make complete visual and/or oral records of primary data for review by independent investigators.

55. The data collected should be within the boundaries defined by the hypotheses.

105 Ibid.
107 Johnson and Jackson, pp. 335-336.
111 Abelson, pp. 308-309.
The hypotheses should help in determining what facts to locate, and what procedures and methods to employ in conducting the study.¹¹²

56. The data included within the defined boundaries of the study should be collected without personal choice.

Only those "basic statements" should be accepted which are logically connected through their discovery in the course of testing theories or hypotheses; a basic statement asserts that an observable event is occurring in a certain individual region of space and time.¹¹³

57. Data should be gathered in a well organized, highly structured, and yet creative manner.

Care must be exercised so that the data are recorded with accuracy.¹¹⁴

58. The collection of data or evidence should include a thorough investigation of primary data.

Special Methods

Instrumental Techniques

59. A data-gathering instrument should have validity or truthfulness; it should measure what it purports to measure.

In sampling procedures, the assumption that the part (which is measured and determined by the data-gathering instrument) reflects the greater whole must be justified.¹¹⁵,¹¹⁶

¹¹²Van Dalen, Educational Administration & Supervision 42 (December, 1956), 457-460.

¹¹³Popper, p. 106.

¹¹⁴Noble, High School Journal, 12 (October, 1929), 197-206.


¹¹⁶Barnes, pp. 130-131.
The data-gathering instrument must be examined for built-in biases which may affect the data in an unknown amount.\textsuperscript{117,118}

The validity of a data-gathering instrument might be jeopardized by the choice of questions, the order in which questions are asked, the use of emotionally charged words and phrases, and by the choice of sponsorship for the questionnaire.\textsuperscript{119}

60. A data-gathering instrument should have reliability or consistency; it should agree with itself when used repeatedly with the same individuals, groups, or objects.

61. The degree of reliability and validity which is accepted should be justified with reference to the objectives of the research problem.

62. The most effective instrument for collecting the necessary data should be anticipated in the research design.

A questionnaire should be constructed only after a thorough study of the field and related fields.\textsuperscript{120}

The background or nature of the individuals, groups, or objects being questioned or measured must be investigated and the use of the instrument justified.\textsuperscript{121}

The data-gathering apparatus or instrument should enable the researcher to collect data in the form required for checking the implications of the hypothesis under study.\textsuperscript{122}

\textsuperscript{117}Ibid., pp. 131-135.

\textsuperscript{118}Obourn et al., \textit{Science Education}, 41 (December, 1957), 375-411.

\textsuperscript{119}Barnes, pp. 131-135.

\textsuperscript{120}Harry Huffman, "Improving the Questionnaire as a Tool of Research," \textit{The National Business Education Quarterly}, 17 (October, 1948), 15-18 and 55-61.

\textsuperscript{121}Barnes, pp. 131-135.

\textsuperscript{122}Brown and Ghiselli, p. 159.
The nature and limitations of the data-gathering instrument should be known.\textsuperscript{123}

Statistical Techniques

63. The statistical technique should be identified before the data are gathered.

Statistics is a tool of educational research and should be considered useful only as it reveals better practices, procedures, and content.\textsuperscript{124,125}

64. The statistical hypotheses should be considered at the time the experiment is designed.

65. The statistical design should include as many independent variables in the same design as may seem necessary.

66. The statistical design should include as many dependent variables in the same design as may seem necessary.

67. The basic assumptions underlying statistical methods should be satisfied by the data to which they are applied.

The data must always be examined carefully to determine whether or not the underlying variation is, or may be assumed to be, continuous before statistical methods are employed which have been devised and are appropriate, only for data which vary continuously.\textsuperscript{126}

The assumptions underlying each type of transformation and any conclusions derived should be judged valid only to the extent to which these assumptions are satisfied.\textsuperscript{127}

\textsuperscript{123}Cohen and Nagel, p. 107.

\textsuperscript{124}Seymour, The American School Board Journal, 144 (June, 1962), 17-19.

\textsuperscript{125}Walker and Lev, p. 7.

\textsuperscript{126}Johnson and Jackson, p. 19.

\textsuperscript{127}Ibid., pp. 222-223.
68. The statistical sample should be selected from a carefully defined parent population.

The statistical sample must give an accurate cross section of the population under study.\(^{128}\)

A statistical method of measure should have the smallest possible sampling error.\(^{129}\)

69. The probability of error estimate should be justified.

The statistical technique should allow a given source of error to be eliminated both from the experimental results and from the estimates of error, or not at all.\(^{130}\)

70. The statistical technique should include tests for the reliability and validity of the data.

71. The statistical technique should allow for maximum objectivity in evaluation and in estimation.

72. The statistical technique should provide tests to assure that a sufficient quantity of data is collected.

73. The statistical technique should allow consideration for tests of significance and problems of estimation.

PRESENTATION OF EVIDENCE

Presentation of Data

74. The data should be logically organized and presented.

75. All data which are collected for the solution of a given problem should be reported.

Data which are not germane, or which are in excess of that necessary for solution of the problem should not be collected or reported.\(^{131}\)

\(^{128}\) Ibid., p. 323.

\(^{129}\) Ibid., p. 123.

\(^{130}\) Lindquist, p. 3.

\(^{131}\) Barnes, pp. 131-135.
76. The sources of all the facts which are presented should be identified.

77. The proof necessary to validate important alleged facts should be provided.

78. The apportionment of space in the research report should be made with reference to the judged relative importance of the topics treated.

79. When graphical methods are used, they should present the data simply and accurately, both in construction and in the impression given to the reader.

A graph or a diagram should be given a clear and complete title.\(^{132}\)

The axes of a graph should be labeled and the units which are used should be given.\(^{133}\)

The original data which was used in constructing a graph should be given with necessary discussion in the text.\(^{134}\)

The origin of a graph should be taken so that increasing magnitudes read from left to right and from bottom to top.\(^{135}\)

All letters and figures in a graph should be arranged for the convenience of the reader so that turning more than once is unnecessary.\(^{136}\)

Analysis of Data

80. The analysis of data should be anticipated and worked through abstractly or in a pilot study before the study itself is begun.

81. Preconceived ideas about the results to be expected should not affect the manipulation of the data.

\(^{132}\)Johnson and Jackson, p. 59.

\(^{133}\)Ibid.

\(^{134}\)Ibid.

\(^{135}\)Ibid.

\(^{136}\)Ibid.
82. Accurate statistical or speculative methods of organizing the data should be employed.

83. Evidence which is collected in a study should be creatively and meaningfully synthesized.

The data should be classified in ways that assist the researcher to reach pertinent conclusions.\textsuperscript{137,138}

84. Any accidents of circumstance which might affect the reliability or validity of the data should be reported.

85. Corrections should be made in the analysis of the data for any distortions which may have occurred in the data gathering process.

Flaws in the procedural design should be identified and an estimate of their effect upon the findings determined.\textsuperscript{139}

The investigation should take into account the possibility of unconscious signaling or previous practice influencing the results.\textsuperscript{140}

Because the motivation of the subjects may vary while the study is in progress, the effect of motivation on the data should be examined.\textsuperscript{141}

**Interpretation of Data**

**Factors Involved**

86. The interpretation of data should involve intuition and rational subjective evaluation.

87. The relationship between a hypothesis or a theory and the observational evidence, when both are given, should be judged.

\textsuperscript{137}Brown and Ghiselli, p. 206.


\textsuperscript{139}Ibid.


\textsuperscript{141}Symonds, *Journal of Educational Psychology*, 47 (February, 1956), 100-109.
A knowledge of the subject matter should be demonstrated before judgments of the relevance of the hypothesis to the problem can be considered well founded.142

88. The reliability of a prediction, which is based either on a hypothesis or a theory, should be judged on the basis of the evidence.

Educational research should take into consideration contemporary social, political, and educational movements which are related to the problem being studied.143

89. Both statistical significance and practical significance should be considered in the interpretation of research results.

90. Observation and theory should serve as checks on one another with neither having priority over the other.

The readings of data-gathering instruments must be "corrected and interpreted in the light of a comprehensive theoretical system.144

The hypothesis should agree with and explain the evidence, and should agree with satisfactorily proved generalizations.145

91. The interpretation of the data should be guided by their source and nature.

Adequate recognition must be given to any faults in the data and to the limitations of the statistical procedure employed in handling the data.146

Historical documents must be interpreted according to conceptions of the time at which they were written.147

142 Cohen and Nagel, p. 92.
143 Noble, High School Journal, 12 (October, 1929), 197-206.
144 Cohen and Nagel, p. 107.
145 Good, p. 79.
147 Good, p. 138.
Evaluation of Theories and Hypotheses

92. A theory should encompass a relatively wide range of data.

A theory should be free from internal contradiction, or be logically consistent.\textsuperscript{148}

93. A theory should provide a basis for successful prediction.

Research may result in a new tool or a new technique.\textsuperscript{149}

Research may result in a new method of analyzing data that can be used in attacking other problems.\textsuperscript{150}

A theory should lead to deductions which can be credited or discredited by empirical test.\textsuperscript{151}

94. A deductively elaborated hypothesis should be factually substantiated before it is used as an explanatory device.

95. A necessary and sufficient condition for the rejection of a theory or a hypothesis is the observation of reproducible contradictory evidence.

A hypothesis should be rejected or revised if the results of the test do not correspond to the consequences stated in the derived theorem.\textsuperscript{152}

96. The assumptions which led to the experimentally testable proposition should be carefully examined before the hypothesis is tentatively accepted or rejected.

97. A hypothesis should make possible the determination of relationships between the known facts.


\textsuperscript{149}Crawford, p. 21.

\textsuperscript{150}Ibid.

\textsuperscript{151}Willower, The Journal of Educational Research, 56 (December, 1962), 210-213.

\textsuperscript{152}Brown and Ghiselli, pp. 170-171.
SUMMARY AND CONCLUSIONS

Summary

98. The summary should be an overview of the entire study.

99. The summary should be self-contained.

100. The summary should include only previously reported data.

101. The summary should avoid unnecessary enumeration of details reported earlier in the manuscript.

102. The summary should be a final synthesis of the whole study which is greater than the mere process of adding together chapter or section summaries.

Conclusions

Formulation and Development

103. Conclusions should be drawn with reference to the problematic situations that engendered the inquiry.

104. Conclusions should be tested by examining them for internal consistency.

The conclusions should follow logically from the assumptions and postulates of the hypothesis.153,154

105. Conclusions should be tested by examining them for external consistency.

The conclusions, whenever possible, should be formulated in the light of similar findings from earlier studies, so that trends may be discovered and extrapolations made.155

---


The conclusions should be tested by comparing them with objective facts or external realities from the present research and with the body of facts and conclusions arrived at by other investigators in the field.\footnote{156,157}

106. Conclusions should include inferences relative to all the significant data.

Because the obvious conclusions to which the data seem to point are quite frequently erroneous, the researcher must rely heavily upon an intimate knowledge of the situation, past experience, and common sense.\footnote{158}

Correlation should not be mistaken for causation; two social phenomena may show an apparent relationship because they are related to a third.\footnote{159,160,161,162}

Because two things are similar in a few particulars, it should not be assumed that they are similar in all.\footnote{163}

107. When several possible explanations of a phenomenon seem equally correct, the simplest should be favored.

Simple hypotheses or theories should be prized more highly than less simple ones because they tell us more, because their empirical content is greater, and because they are more easily testable.\footnote{164}

\footnote{156}{Crawford, p. 279.}
\footnote{157}{Smith and Smith, pp. 107-108.}
\footnote{158}{Johnson and Jackson, p. 12.}
\footnote{159}{Ibid., pp. 12-13.}
\footnote{160}{Crawford, pp. 280-282.}
\footnote{162}{Smith and Smith, pp. 107-108.}
\footnote{163}{Crawford, pp. 280-282.}
\footnote{164}{Popper, pp. 140-142.
An explanation which explains all the facts with the smallest number of assumptions should be favored.165, 166, 167

108. Skillful grouping and appropriate subordination of conclusions under major headings should be used to avoid encyclopedic enumeration of individual statements.

109. The findings of research should be stated with simplicity within the context of a total pattern or theory.

Limitations

110. The limitations of a study should be presented to guide the reader in using the findings, conclusions, and recommendations.

Qualifying assumptions which are basic to the conclusions should be clearly and concisely expressed.168, 169

The number of returns in a questionnaire type study should permit a high order of confidence in the conclusions derived.170

111. A study should be thorough enough to arrive at carefully stated results which have a substantial degree of reliability and validity.

The conclusions of a study should not be so qualified by the use of inadequately developed techniques that they are of little or no value.171

165Crawford, p. 282.
166Watson, *Phi Delta Kappan*, 35 (October, 1953), 4-6.
171Abelson, pp. 20-21.
The conclusions of a study should not be so qualified by the application of unsubstantiated assumptions that they are of little or no value.\textsuperscript{172}

The conclusions reached in a study should be amply and incontrovertably supported by the data.\textsuperscript{173}

A fundamental limitation of the descriptive method is that the findings indicate norms, not standards, causes, reasons, meanings, or possibilities.\textsuperscript{174}

It should not be assumed that because a study includes the total population (a 100\% sample) the conclusions drawn are more valid than if a sample study had been carried out.\textsuperscript{175}

Applications, Recommendations, and Needed Research

112. Generalizations should be directed to the solution of problems other than those considered in the study.

The conditions under which the experiment was conducted should be taken into account in generalizing conclusions.\textsuperscript{176}

Generalizations should be based upon representative data from an adequate number of cases.\textsuperscript{177,178}

113. Both the data and the conclusions should warrant the recommendations made.

\textsuperscript{172}Ibid.\textsuperscript{173}Breitwieser, \textit{Phi Delta Kappan}, 24 (December, 1941), 185.


\textsuperscript{175}Johnson and Jackson, p. 324.


\textsuperscript{177}Crawford, pp. 280-282.

\textsuperscript{178}Smith and Smith, pp. 107-108.
No attempt should be made to compare one set of data with another which has been gathered in a very different form or whose elements are so different as to have little in common with it.179

114. Creative speculation concerning possible applications and recommendations should be made.

115. Applications or recommendations which are based on the creative insight of the investigator should be labeled plainly as such.

116. Research should contribute to the identification and to the solution of new problems.

A few carefully identified problems for future investigations which are derived directly from the study reported, should be suggested.180,181,182

Research may contribute to the solution of new problems by making possible greater precision than was possible before.183

Theories which are developed as explanations should be fruitful of further experimentation and inquiry.184

179Ibid.


181Good, p. 852.

182Whitney, p. 179.

183Crawford, p. 21.

184Watson, Phi Delta Kappan, 35 (October, 1953), 4-6.
APPENDIX II

CORRESPONDENCE AND MATERIALS RELATED TO THE PILOT STUDY
Pilot Study Judges

Dr. Glenn W. Blaydes
Professor of Botany and Plant Pathology
The Ohio State University

Dr. Jack G. Calvert
Professor of Chemistry
The Ohio State University

Dr. Frederick R. Cyphert
Associate Professor of Education and
Associate Chairman, Education
The Ohio State University

Dr. Jack Frymier
Association Professor of Education
The Ohio State University

Dr. Thomas W. Lippincott
Professor of Chemistry
The Ohio State University

Dr. Devon Meek
Assistant Professor of Chemistry
The Ohio State University

Dr. Earl Jack Montague
Ball State Teachers College
Muncie, Indiana

Dr. Fred R. Schlessinger
Professor of Education
The Ohio State University

Dr. Grant L. Stahly
Professor of Microbiology
The Ohio State University
<table>
<thead>
<tr>
<th>Judge Number</th>
<th>Academic Degrees Received</th>
<th>Academic Year Received</th>
<th>Academic Rank</th>
<th>Years of Research Experience</th>
<th>Number of Research Studies Participated in</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bachelor 1935, Master 1941, Ph.D. 1957</td>
<td>Professor 1963</td>
<td></td>
<td>&lt;10</td>
<td>&gt;30</td>
</tr>
<tr>
<td>2</td>
<td>Bachelor 1951, Master 1957, Ph.D. 1963</td>
<td>Asst. 1959, Prof. 1963</td>
<td></td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td>3</td>
<td>Bachelor 1949, Master 1950, Ed.D. 1957</td>
<td>Assoc. 1962, Prof. 1962</td>
<td></td>
<td>&lt;5</td>
<td>&lt;5</td>
</tr>
<tr>
<td>4</td>
<td>Bachelor ?, Master ?, Ed.D. 1957</td>
<td>Assoc. 1959, Prof. 1960</td>
<td></td>
<td>&lt;10</td>
<td>&lt;20</td>
</tr>
<tr>
<td>5</td>
<td>Bachelor 1944, Ph.D. 1949</td>
<td>Professor 1960</td>
<td></td>
<td>&lt;15</td>
<td>&gt;30</td>
</tr>
<tr>
<td>6</td>
<td>Bachelor 1958, Master 1960, Ph.D. 1961</td>
<td>Asst. 1961, Prof. 1961</td>
<td></td>
<td>5</td>
<td>&lt;10</td>
</tr>
<tr>
<td>7</td>
<td>Bachelor 1948, Ph.D. 1954</td>
<td>Professor 1961</td>
<td></td>
<td>&lt;20</td>
<td>&gt;30</td>
</tr>
<tr>
<td>8</td>
<td>Bachelor 1923, Master 1925, Ph.D. 1935</td>
<td>Professor 1948</td>
<td></td>
<td>&gt;20</td>
<td>&gt;30</td>
</tr>
<tr>
<td>9</td>
<td>Bachelor ?, Master ?, Ph.D. 1931</td>
<td>Professor 1945</td>
<td></td>
<td>&gt;20</td>
<td>&lt;30</td>
</tr>
</tbody>
</table>
Thank you very much for agreeing to help by participating in this pilot study. The covering letter and the instrument which will be used in the study itself are attached.

We hope you will find it possible to complete the instrument and return it by Tuesday, January 14th.

Very sincerely yours,

Philip G. Kapfer
This letter is written to solicit your assistance with a doctoral study in science education. In a serious effort to improve the quality of research in the field, we are proposing to develop and validate criteria which may be used for evaluating research.

As an experienced researcher, you undoubtedly realize the importance of being able to evaluate research studies. In fact, the process of evaluating research probably has become such a part of your thinking that you no longer bother to verbalize the criteria which you apply. The attached instrument has been designed to tap your rich background of knowledge and experience in order to validate a list of logically organized research criteria. The criteria are stated in terms which permit their application to research in most areas, although they are intended primarily for evaluating research in science education. The list of research criteria in the attached instrument represents a substantial condensation of the sometimes conflicting and frequently repetitive literature on this topic which has accumulated during the past twenty to thirty years. However, full usefulness of this instrument by the less experienced researcher cannot be realized without validation by a number of experienced researchers. Judges have been selected from three areas—the physical and biological sciences, education, and the sub-specialty of science education.

Let me emphasize three important points:

1. You need only indicate with a check mark in the appropriate box whether you think a given criterion is of "great importance," "high importance," "average importance," "little importance," or "no importance," for evaluating research.

2. A single page of easily answered questions about you and about your impressions of the research criteria and their organization has been placed at the end of the instrument.

3. You may add criteria, suggest that a particular criterion be placed in a different group, or alter any of the criteria as you see fit.

The answers provided will not be identified with specific institutions or individuals in the report of this study. If a summary of the report is desired, please so
indicate in the space provided at the end of the instrument. A self-addressed stamped envelope is provided for your convenience in returning the completed instrument. Due to a time factor involved, we hope you will find it possible to complete the instrument and return it by Friday, January 31st. Thank you very much.

Very sincerely yours,

Philip G. Kapfer
PILOT STUDY
CRITERIA FOR JUDGING RESEARCH

FORMULATION AND DEFINITION OF THE PROBLEM

Problem Statement

1. The problem should be stated precisely in clear and understandable language.

2. The problem statement should include all subordinate problems which must be answered.

3. The problem statement, including all of its parts, should encompass all elements of the problem.

4. The problem statement should be identified so that it can be located and recognized readily.

Introduction

Problem Choice

5. The origin of a research problem should be identified.

6. The body of organized knowledge in an area should be examined for gaps, for discrepancies, and for the directions in which extension is possible.

7. A problem should be chosen which avoids unnecessary research effort.

8. A problem should be chosen which can be formulated and defined with sufficient precision to permit a good idea of the goal to be achieved before it is attempted.

9. Attention should be given to the theoretical framework of knowledge underlying issues and problems before their resolution or solution is sought directly.

10. The problem chosen should have either direct significance for people's needs or should serve as a foundation for subsequent research having significance for people's needs.

11. Experimentation involving human subjects, especially school children, should be planned and spaced carefully for the protection of the subjects.
Delimitation and Analysis

12. The problem should be delimited exactly as to scope, the nature and quantity of data desired, and the methods or procedure to be followed.

13. In addition to simple delimitation of the problem, the subjects (if any are used) should be described with reference to the variables among them.

14. The factors which are selected for study by the process of delimitation and analysis should be the decisive ones for solution of the problem.

15. When variables are used in experimental research, they should be shown to differentiate between or among the subjects or objects under study.

Assumptions

16. Assumptions upon which the inquiry rests should be identified.

17. Assumptions upon which the inquiry rests should be justified.

18. The assumptions in a study should be consistent with each other.

19. The assumption that the non-research literature is an accurate indicator of best or prevalent practice should be avoided.

Definitions

20. A definition must be expressed in precise language.

21. A definition must not be circular; it must not contain, directly or indirectly, the subject to be defined.

22. A definition should be stated in positive terms, rather than in negative terms, wherever possible.

23. Important terms, that are employed in an unusual sense, should be defined.

24. Undefined terms should be employed according to the meanings demanded by standard references.

25. A distinct symbol should be provided for each of the attributes, phases, qualities, aspects, elements, or the like, that can be found in a given class of events.
26. Operational definitions of all generalized concepts which will be studied should be specified.

27. Defined terms and concepts should be used in the entire study according to the given definitions.

Related Literature

28. The related literature should be analyzed for its relevancy to the problem situation.

29. The related literature should be analyzed for any bias which may have significance for the study.

30. The review of related literature should include the previous research which may contribute to the solution of the problem.

31. The problem should be related to an orienting frame of reference or a systematic area of study.

32. Research in a given area should make use of relevant principles and research techniques of related disciplines.

33. Literature sources should be assigned to definite authors, places, and times.

Hypotheses and Derived Theorems

34. The researcher should display creativity in his formulation of hypotheses.

35. A hypothesis should bring together, in a simple manner, a given number of observations and interpret the connections between them.

36. The possible alternative answers which are hypothesized for a given question should be independent of one another (or not overlap).

37. Hypotheses should be formulated on the basis of a thorough analysis of the theoretical and factual background of the problem.

38. Hypotheses should be formulated so that implications can be traced clearly by means of well-established techniques of deduction.
39. The consequences of alternative hypotheses should be developed for comparison with observable phenomena.

40. Hypotheses should be revised on the basis of induction from the data which accumulates as the study progresses.

PROCEDURAL DESIGN

Procedure

Design Selection

41. The problem, from the beginning, should condition the steps taken to solve it.

42. The adequacy of the research methodology chosen for solution of the problem should be explained and defended.

43. The research design should provide for collection of the most objective data available.

44. The research design should include several directions of inquiry in order to increase efficiency and comprehensiveness.

45. The experimental design should include the replication of one or more variables whose effects have already been assessed as significant in a previous experiment.

Design for Replication and Validity

46. The research design should be described in sufficient detail to permit repetition of the study by independent investigators.

47. The degree of internal and external validity in the research design should be identified and justified.

48. The data collected should afford some basis for generalization.

Self-Contained Design

49. The self-contained experiment should include a representative sample of the subjects or objects under study.
50. The self-contained experiment should include a control to exclude, at a known level of probability, a number of alternative interpretations of the experimental results.

51. The self-contained experiment should supply independent evidence on the question in dispute.

52. The self-contained experiment should contain within itself the possibility of making an estimate of the experimental errors which affect the comparison made.

Design for Data Gathering

53. The design for making observations should establish causal connections between events in which the cause is a sufficient and necessary condition for the effect.

54. The data of research should approach the reality of the situation being investigated as nearly as possible.

55. Modern technological instruments should be used to make complete visual and/or oral records of primary data for review by independent investigators.

56. The data collected should be within the boundaries defined by the hypotheses.

57. The data included within the defined boundaries of the study should be collected without personal choice.

58. Data should be gathered in a well organized, highly structured, and yet creative manner.

59. The collection of data or evidence should include a thorough investigation of primary data.

Special Methods

Instrumental Techniques

60. A data-gathering instrument should have validity or truthfulness; it should measure what it purports to measure.

61. A data-gathering instrument should have reliability or consistency; it should agree with itself when used repeatedly with the same individuals, groups, or objects.
62. The degree of reliability and validity which is accepted should be justified with reference to the objectives of the research problem.

63. The most effective instruments for collecting the necessary data should be anticipated in the research design.

Statistical Techniques

64. The statistical technique should be identified before the data are gathered.

65. The statistical hypotheses should be considered at the time the experiment is designed.

66. The statistical design should include as many independent variables in the same design as may seem necessary.

67. The statistical design should include as many dependent variables in the same design as many seem necessary.

68. The basic assumptions underlying statistical methods should be satisfied by the data to which they are applied.

69. The statistical sample should be taken from an explicitly defined parent population.

70. The probability of error estimate should be justified.

71. If a given source of error is eliminated from the experimental results, it should be eliminated from the estimate of error as well.

72. The statistical technique should include tests for the reliability and validity of the data.

73. The statistical technique should allow for maximum objectivity in evaluation and in estimation.

74. The statistical technique should provide tests to assure that a sufficient quantity of data is collected.

75. The statistical technique should allow consideration for tests of significance and problems of estimation.
PRESENTATION OF EVIDENCE

Presentation of Data

76. The data should be logically organized and presented.

77. All data which are collected for the solution of a given problem should be reported.

78. The sources of all the facts which are presented should be identified.

79. The proof necessary to validate important alleged facts should be provided.

80. The apportionment of space in the research report should be made with reference to the judged relative importance of the topics treated.

81. When graphical methods are used, they should present the data simply and accurately, both in construction and in the impression given to the reader.

Analysis of Data

82. The analysis of data should be anticipated and worked through abstractly or in a pilot study before the study itself is begun.

83. Preconceived ideas about the results to be expected should not affect the manipulation of the data.

84. Accurate statistical or speculative methods of organizing the data should be employed.

85. Any accidents of circumstance which might affect the reliability or validity of the data should be reported.

86. Measures should be taken in the analysis of the data to correct for any distortions which may have occurred.

87. Scattered, obscure, or concealed bits of information which are collected in a study should be creatively and meaningfully synthesized.

Interpretation of Data

Factors Involved

88. The interpretation of data should involve intuition and rational subjective evaluation.
89. The relationship between a hypothesis or a theory and the observational evidence, when both are given, should be judged.

90. The reliability of a prediction, which is based either on a hypothesis or a theory, must be judged on the basis of the evidence.

91. Both statistical significance and practical significance should be considered in the interpretation of research results.

92. Observation and theory should serve as checks on one another with neither having priority over the other.

93. The interpretation of the data should be guided by their source and nature.

Evaluation of Theories and Hypotheses

94. A theory should encompass a relatively wide range of data.

95. A theory should provide a basis for successful prediction.

96. A deductively elaborated hypothesis should be factually substantiated before it is used as an explanatory device.

97. A hypothesis should be rejected if the results of the test do not correspond to the consequences stated in the derived theorem.

98. A necessary and sufficient condition for the rejection of a theory or a hypothesis is the observation of reproducible contradictory events.

99. The assumptions which led to the experimentally testable propositions should be carefully examined before the hypothesis is tentatively accepted or rejected.

100. A hypothesis should make possible the determination of relationships between the known facts.

SUMMARY AND CONCLUSIONS

Summary

101. The summary should be an overview of the entire study.

102. The summary should be self-contained.
103. The summary should include only previously reported data.

104. The summary should avoid unnecessary enumeration of details reported earlier in the manuscript.

105. The summary should be a final synthesis of the whole study which is greater than the mere process of adding together chapter or section summaries.

Conclusions

Formulation and Development

106. Conclusions should be drawn with reference to the problematic situations that engendered the inquiry.

107. Conclusions should be tested by examining them for internal consistency.

108. Conclusions should be tested by examining them for external consistency.

109. Conclusions should follow logically from the assumptions and postulates of the hypotheses.

110. Conclusions should include inferences relative to all the significant data.

111. When several possible explanations of a phenomenon seem equally correct, the simplest should be tentatively chosen.

112. Skillful grouping and appropriate subordination of conclusions under major headings should be used to avoid encyclopedic enumeration of individual statements.

113. The findings of research should be stated with simplicity within the context of a total pattern or theory.

Limitations

114. The limitations of a study should be presented to guide the reader in using the findings, conclusions, and recommendations.

115. A study should be thorough enough to arrive at carefully stated results which have a substantial degree of reliability and validity.
Applications, Recommendations, and Needed Research

116. Generalizations should be directed to the solution of problems other than those considered in the study.

117. Both the data and the conclusions should warrant the recommendations made.

118. Creative speculation concerning possible applications and recommendations should be made.

119. Applications or recommendations which are based on the creative insight of the investigator should be labeled plainly as such.

120. Sets of data with very different histories should not be compared.

121. Research should contribute to the identification and to the solution of new problems.
1. Name (please print): ___________________________________  
   last first middle or maiden  
(Your name and institution will not be associated with the information requested.)

2. Academic Degrees Year Received:  
   Bachelor __ ___________
   Master __ ___________
   Ph.D. __ ___________
   Ed.D. __ ___________

3. Present Academic Year Received:  
   Instructor __ ___________
   Asst. Prof. __ ___________
   Assoc. Prof. __ ___________
   Full Prof. __ ___________

4. Number of Years of Research Experience:
   Less than 5 __
   Less than 10 __
   Less than 15 __
   Less than 20 __
   More than 20 __

5. Number of Formal Research Studies in Which You Were a Major Contributor:
   Less than 5 __
   Less than 10 __
   Less than 20 __
   Less than 30 __
   More than 30 __

6. Area of Research Competence:
   Biology __
   Chemistry __
   Physics __
   Education __
   Other __________

7. Type(s) of Research Methods Commonly Employed:
   Experimental
   Classroom __
   Field __
   Laboratory __
   Other __________
   Descriptive __
   Case Study __
   Clinical __
   Other __________

8. Do you wish to receive a summary of the results of this study?  
   yes __
   no __

9. Is the organization scheme for classifying the criteria functional for evaluating research in progress?  
   yes __
   no __
10. Is the organizational scheme for classifying the criteria functional for evaluating completed research? yes __  no __

11. Do the research criteria encourage creativity in research? yes __  no __

12. Are the research criteria useful for evaluating research in progress? yes __  no __

13. Are the research criteria useful for evaluating completed research? yes __  no __

Note: Any comments which you may wish to make may be made on this or on the reverse side. Thank you very much for your generous assistance.
January 30, 1964
115 McPherson Chem. Lab.
The Ohio State University
Columbus, Ohio

Dear ________:

Your help in completing and returning promptly the criterion instrument for my pilot study of criteria for evaluating research is appreciated very much. You also expressed an interest in the results of the study. I will be very happy to prepare a summary of the findings and send it to you after the work is completed. I will include also the instrument which is used in the study. If all goes according to schedule, I hope to complete the project in June.

Very sincerely yours,

Philip G. Kapfer
<table>
<thead>
<tr>
<th>Criterion Number</th>
<th>Judge</th>
<th>Frequency of Response</th>
<th>Mean&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>great importance</td>
<td>high importance</td>
</tr>
<tr>
<td>1</td>
<td>Education</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Education</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Education</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Education</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Education</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Education</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>Education</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>Education</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>Education</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>Education</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>Education</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<sup>a</sup>The mean values were determined by letting great importance = 4, high importance = 3, average importance = 2, little importance = 1, and no importance = 0.
TABLE 13—Continued

<table>
<thead>
<tr>
<th>Criterion Number</th>
<th>Judge</th>
<th>Great Imp.</th>
<th>High Imp.</th>
<th>Average Imp.</th>
<th>Little Imp.</th>
<th>No Imp.</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Education</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Education</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>3.2</td>
</tr>
<tr>
<td>14</td>
<td>Education</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Education</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td>16</td>
<td>Education</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.0</td>
</tr>
<tr>
<td>17</td>
<td>Education</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td>18</td>
<td>Education</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.0</td>
</tr>
<tr>
<td>19</td>
<td>Education</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Education</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>3.4</td>
</tr>
<tr>
<td>21</td>
<td>Education</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td>22</td>
<td>Education</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Education</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.0</td>
</tr>
<tr>
<td>24</td>
<td>Education</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td>25</td>
<td>Education</td>
<td>3</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion Number</td>
<td>Judge</td>
<td>Frequency of Response</td>
<td>Mean¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>--------</td>
<td>------------------------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Education</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Education</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
<td>1</td>
<td>4.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Education</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
<td>1</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Education</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
<td>1</td>
<td>4.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Education</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>1</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Education</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>1</td>
<td>2.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Education</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
<td>1</td>
<td>3.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Education</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>2</td>
<td>3.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Education</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>1</td>
<td>4.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Education</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>2.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td>1</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Education</td>
<td>3</td>
<td>2</td>
<td>10</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td>2</td>
<td>3.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Education</td>
<td>3</td>
<td>2</td>
<td>10</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>3</td>
<td>3.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Education</td>
<td>3</td>
<td>2</td>
<td>10</td>
<td>2.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>1</td>
<td>3.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Education</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
<td>1</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion Number</td>
<td>Judge</td>
<td>Frequency of Response</td>
<td>Mean¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----------------</td>
<td>-----------------------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>great importance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Education</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td>4.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Education</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>Education</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
<td>3.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Education</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>Education</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
<td>3.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Education</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
<td>3.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>Education</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>Education</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>Education</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>Education</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>Education</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>Education</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>Education</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
<td>3.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>Education</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>Education</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ Mean refers to the average frequency of response across all judges.
TABLE 13—Continued

<table>
<thead>
<tr>
<th>Criterion Number</th>
<th>Judge</th>
<th>Frequency of Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>great importance</td>
</tr>
<tr>
<td>56</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
</tr>
<tr>
<td>57</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
</tr>
<tr>
<td>58</td>
<td>Education</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
</tr>
<tr>
<td>59</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
</tr>
<tr>
<td>60</td>
<td>Education</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
</tr>
<tr>
<td>61</td>
<td>Education</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
</tr>
<tr>
<td>62</td>
<td>Education</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
</tr>
<tr>
<td>63</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
</tr>
<tr>
<td>64</td>
<td>Education</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
</tr>
<tr>
<td>65</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
</tr>
<tr>
<td>66</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
</tr>
<tr>
<td>67</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
</tr>
<tr>
<td>68</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
</tr>
<tr>
<td>69</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
</tr>
<tr>
<td>70</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
</tr>
<tr>
<td>Criterion Number</td>
<td>Judge</td>
<td>great importance</td>
</tr>
<tr>
<td>------------------</td>
<td>------------</td>
<td>------------------</td>
</tr>
<tr>
<td>71</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
</tr>
<tr>
<td>72</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
</tr>
<tr>
<td>73</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
</tr>
<tr>
<td>74</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
</tr>
<tr>
<td>75</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
</tr>
<tr>
<td>76</td>
<td>Education</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
</tr>
<tr>
<td>77</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
</tr>
<tr>
<td>78</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
</tr>
<tr>
<td>79</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
</tr>
<tr>
<td>80</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
</tr>
<tr>
<td>81</td>
<td>Education</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
</tr>
<tr>
<td>82</td>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
</tr>
<tr>
<td>83</td>
<td>Education</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>5</td>
</tr>
<tr>
<td>84</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
</tr>
<tr>
<td>85</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
</tr>
</tbody>
</table>
### TABLE 13--Continued

<table>
<thead>
<tr>
<th>Criterion Number</th>
<th>Judge</th>
<th>great importance</th>
<th>high importance</th>
<th>average importance</th>
<th>little no importance</th>
<th>Mean^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>86</td>
<td>Education</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>4.0</td>
</tr>
<tr>
<td>87</td>
<td>Education</td>
<td>1</td>
<td>2</td>
<td></td>
<td>1</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td>88</td>
<td>Education</td>
<td>2</td>
<td>2</td>
<td></td>
<td>1</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>2.3</td>
</tr>
<tr>
<td>89</td>
<td>Education</td>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td>3.6</td>
</tr>
<tr>
<td>90</td>
<td>Education</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td>91</td>
<td>Education</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>92</td>
<td>Education</td>
<td>1</td>
<td>3</td>
<td></td>
<td>1</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1.8</td>
</tr>
<tr>
<td>93</td>
<td>Education</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td></td>
<td>3</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>94</td>
<td>Education</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>3.3</td>
</tr>
<tr>
<td>95</td>
<td>Education</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>2.2</td>
</tr>
<tr>
<td>96</td>
<td>Education</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>2.8</td>
</tr>
<tr>
<td>97</td>
<td>Education</td>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>98</td>
<td>Education</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td>99</td>
<td>Education</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td>100</td>
<td>Education</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>Criterion Number</td>
<td>Judge</td>
<td>Frequency of Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>-----------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>great importance</td>
<td>high importance</td>
<td>average importance</td>
<td>little importance</td>
<td>no importance</td>
</tr>
<tr>
<td>101</td>
<td>Education Science</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>Education Science</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>Education Science</td>
<td>1</td>
<td>2</td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>104</td>
<td>Education Science</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>Education Science</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>Education Science</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>Education Science</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>Education Science</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>Education Science</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>Education Science</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>Education Science</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>Education Science</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>Education Science</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>Education Science</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>Education Science</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion Number</td>
<td>Judge</td>
<td>mean</td>
<td>great importance</td>
<td>high importance</td>
<td>average importance</td>
<td>little importance</td>
</tr>
<tr>
<td>------------------</td>
<td>--------</td>
<td>------</td>
<td>------------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>116</td>
<td>Education</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>1.8</td>
</tr>
<tr>
<td>117</td>
<td>Education</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td>118</td>
<td>Education</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>3.4</td>
</tr>
<tr>
<td>119</td>
<td>Education</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>3.4</td>
</tr>
<tr>
<td>120</td>
<td>Education</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>2.6</td>
</tr>
<tr>
<td>121</td>
<td>Education</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>3.6</td>
</tr>
</tbody>
</table>
APPENDIX III

CORRESPONDENCE AND MATERIALS RELATED TO THE STUDY
January 13, 1964
Division of General Chemistry
The Ohio State University
Columbus, Ohio 43210

This letter is written to solicit your assistance with a doctoral study in science education. In a serious effort to improve the quality of research in the field, we are proposing to develop and validate criteria which may be used for evaluating research. Our request of you is quite brief—would you or a person of your choice serve as a judge? In your opinion, the person whom you nominate should be (1) a truly expert researcher who would be qualified to evaluate research criteria, and (2) he should be willing to spend an hour of his time later this month in concentrated consideration of the criteria.

The research criteria are stated in terms which permit their application to research in most areas, although they are intended primarily for evaluating research in science education. The validation process requires the judgment of mature researchers from three areas—the physical and biological sciences, education, and the sub-specialty of science education.

Please write the name and address of the nominee from your department in the space provided below. If you wish, you may give the enclosed copy of this letter to your nominee so that he will know of his selection. A self-addressed stamped envelope is enclosed for your convenience in returning the form provided below. Due to a time factor involved, we hope you will find it possible to reply within a day or two. Thank you very much for your assistance.

Very sincerely yours,

Philip G. Kapfer

My nominee to serve as a judge is: (Please print or type)

Name ________________________________________________
first middle last
Title __________________________________________________
Department ____________________________________________
Institution ____________________________________________
City _____________ Zone_______ State_______
Signed _______________________________
This letter is written to solicit your assistance with a doctoral study in science education. In a serious effort to improve the quality of research in the field, we are proposing to develop and validate criteria which may be used for evaluating research. Our request of you is quite brief—would you nominate two people from your department, excluding science education faculty members, to serve as judges? You may, if you wish, include yourself as one of the two judges. In your opinion, the persons whom you nominate should be (1) truly expert researchers who would be qualified to evaluate research criteria, and (2) they should be willing to spend an hour of time later this month in concentrated consideration of the criteria.

The research criteria are stated in terms which permit their application to research in most areas, although they are intended primarily for evaluating research in science education. The validation process requires the judgment of mature researchers from three areas—the physical and biological sciences, education, and the sub-specialty of science education.

Please write the name and addresses of the nominees from your department in the space provided below. If you wish, you may give the enclosed copies of this letter to your nominees so that they will know of their selection. A self-addressed stamped envelope is enclosed for your convenience in returning the form provided below. Due to a time factor involved, we hope you will find it possible to
reply within a day or two. Thank you very much for your assistance.

Very sincerely yours,

Philip G. Kapfer

My nominees to serve as judges are: (Please print or type)

1. Name ________________________________
   first          middle          last
   Title ________________________________
   Department __________________________
   Institution __________________________
   City ________ Zone _______ State ______

2. Name ________________________________
   first          middle          last
   Title ________________________________
   Department __________________________
   Institution __________________________
   City ________ Zone _______ State ______
This letter is written to solicit your assistance with a doctoral study in science education. In a serious effort to improve the quality of research in the field, we are proposing to develop and validate criteria which may be used for evaluating research. The research criteria are stated in terms which permit their application to research in most areas, although they are intended primarily for evaluating research in science education. The validation process requires the judgment of mature researchers from three areas—the physical and biological sciences, education, and the sub-specialty of science education.

We have personally nominated nineteen judges, of which you are one, from the area of science education. In addition, letters have been sent to deans or chairmen of schools or departments of education, biology, chemistry, and physics in order to obtain the names of thirty-eight judges from the area of education excluding science education, and nineteen judges from each of the three science areas. Although the response from most institutions has been gratifying, some of the requests for nominees have not been answered and a few have been refused. We have indicated below the nomination(s) which have been received from (NAME OF UNIVERSITY) as well as the area(s) from which (a) nomination(s) have not been received.

<table>
<thead>
<tr>
<th>NAME</th>
<th>AREA</th>
</tr>
</thead>
</table>

Would you supply, on the attached sheet, the name(s) of (a) judge(s) from those areas where nominations have not been made? In your opinion, the persons whom you nominate should be (1) (a) mature researcher(s) who would be qualified to evaluate research criteria, and (2) they (he) should be willing to spend an hour of time in concentrated consideration of the criteria. Finally, would you indicate your willingness to serve as a judge by checking the appropriate space provided on the attached sheet?

A self-addressed stamped envelope is enclosed for your convenience in returning the attached sheet. Due to a time factor involved, we hope you will find it possible to reply within a day or two. Thank you very much for your assistance.

Very sincerely yours,

Philip G. Kapfer
My nominee to serve as a judge is: (Please print or type)

OR

My nominees to serve as judges are: (please print or type)

Education 1. Name
First  Middle  Last
Title

Education 2. Name
First  Middle  Last
Title

Biology 3. Name
First  Middle  Last
Title

Chemistry 4. Name
First  Middle  Last
Title

Physics 5. Name
First  Middle  Last
Title

Are you willing to serve as a judge? yes ___ no ___

Signed
Dr. xxxxx  x. xxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxx x xxxxxxxx
xxxxxxxxxxx, xxxxxxxxxxxxxx
NOMINATED JUDGES

Boston University
Dr. John G. Read
Science Education
Dr. Harlan Philippi
Education
Dr. Robert F. Slechtta
Biology
Dr. Armand Siegel
Physics

Columbia University
Dr. Frederick L. Fitzpatrick
Science Education
Dr. Philip H. Phenix
Education
Dr. Cheves Walling
Chemistry
Dr. Samuel Devons
Physics

University of Connecticut
Dr. David J. Blick
Science Education
Dr. Earl H. Newcomer
Biology
Dr. William L. Masterton
Chemistry
Dr. Charles A. Reynolds
Physics

Cornell University (continued)
Dr. Jason Millman
Education
Dr. D. Bob Gowin
Education
Dr. Lowell Uhler
Biology
Dr. Richard P. Korf
Biology
Dr. L. Todd Reynolds
Chemistry
Dr. Robert A. Plane
Chemistry
Dr. Kenneth Greisen
Physics

Florida State University
Dr. J. Stanley Marshall
Science Education
Dr. Eugene M. Boyce
Education
Dr. Garth K. Blake
Education
Dr. Harry J. Lipner
Biology
Dr. James V. Quagliano
Chemistry
Dr. Bruno Linder
Chemistry
Dr. Robert Andrew Kromhout
Physics

Cornell University
Dr. Philip G. Johnson
Science Education
Dr. Verne N. Rockcastle
Education
<table>
<thead>
<tr>
<th>NOMINATED JUDGES—Continued</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Harvard University</strong></td>
</tr>
<tr>
<td>Dr. Fletcher Watson</td>
</tr>
<tr>
<td>Science Education</td>
</tr>
<tr>
<td>Dr. William W. Cooley</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Dr. John B. Carroll</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Dr. Charles Walcott</td>
</tr>
<tr>
<td>Biology</td>
</tr>
<tr>
<td>Dr. Leonard K. Nash</td>
</tr>
<tr>
<td>Chemistry</td>
</tr>
<tr>
<td>Dr. Victor Guillemin</td>
</tr>
<tr>
<td>Physics</td>
</tr>
<tr>
<td>University of Illinois</td>
</tr>
<tr>
<td>Dr. R. Will Burnett</td>
</tr>
<tr>
<td>Science Education</td>
</tr>
<tr>
<td>Dr. R. Stewart Jones</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Dr. Rupert N. Evans</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Dr. Lester Ingle</td>
</tr>
<tr>
<td>Biology</td>
</tr>
<tr>
<td>Dr. Dean F. Martin</td>
</tr>
<tr>
<td>Chemistry</td>
</tr>
<tr>
<td>Dr. Peter Axel</td>
</tr>
<tr>
<td>Physics</td>
</tr>
<tr>
<td>Indiana University</td>
</tr>
<tr>
<td>Dr. Stanley Brown</td>
</tr>
<tr>
<td>Science Education</td>
</tr>
<tr>
<td>Dr. James Weigand</td>
</tr>
<tr>
<td>Education</td>
</tr>
</tbody>
</table>
### NOMINATED JUDGES--Continued

#### University of Kansas

<table>
<thead>
<tr>
<th>Name</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Kenneth E. Anderson</td>
<td>Science Education</td>
</tr>
<tr>
<td>Dr. Dale P. Scannell</td>
<td>Education</td>
</tr>
<tr>
<td>Dr. Ross O. Armstrong</td>
<td>Education</td>
</tr>
<tr>
<td>Dr. A. Byron Leonard</td>
<td>Biology</td>
</tr>
<tr>
<td>Dr. Arthur W. Davidson</td>
<td>Chemistry</td>
</tr>
<tr>
<td>Dr. A. A. Strassenburg</td>
<td>Physics</td>
</tr>
</tbody>
</table>

#### Michigan State University

<table>
<thead>
<tr>
<th>Name</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. John M. Mason</td>
<td>Science Education</td>
</tr>
<tr>
<td>Dr. Julian R. Brandou</td>
<td>Education</td>
</tr>
<tr>
<td>Dr. Walter R. Stellwagen</td>
<td>Education</td>
</tr>
<tr>
<td>Dr. David Krathwohl</td>
<td>Education</td>
</tr>
</tbody>
</table>

#### University of Minnesota

<table>
<thead>
<tr>
<th>Name</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Clarence H. Boeck</td>
<td>Science Education</td>
</tr>
<tr>
<td>Dr. Donovan A. Johnson</td>
<td>Education</td>
</tr>
<tr>
<td>Dr. Cyril J. Hoyt</td>
<td>Education</td>
</tr>
<tr>
<td>Dr. V. Elving Anderson</td>
<td>Biology</td>
</tr>
</tbody>
</table>

#### University of Minnesota (contd.)

<table>
<thead>
<tr>
<th>Name</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Otto H. Johnson</td>
<td>Chemistry</td>
</tr>
<tr>
<td>Dr. James H. Werntz, Jr.</td>
<td>Physics</td>
</tr>
</tbody>
</table>

#### Washington Square College, N.Y.U.

<table>
<thead>
<tr>
<th>Name</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Darrell Barnard</td>
<td>Science Education</td>
</tr>
<tr>
<td>Dr. Roscoe C. Brown, Jr.</td>
<td>Education</td>
</tr>
<tr>
<td>Dr. James H. Hanscom</td>
<td>Education</td>
</tr>
<tr>
<td>Dr. William J. Crotty</td>
<td>Biology</td>
</tr>
<tr>
<td>Dr. Alvin I. Kosak</td>
<td>Chemistry</td>
</tr>
<tr>
<td>Dr. Morris H. Shamos</td>
<td>Physics</td>
</tr>
</tbody>
</table>

#### Northwestern University

<table>
<thead>
<tr>
<th>Name</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Daniel Tanner</td>
<td>Education</td>
</tr>
<tr>
<td>Dr. Norman Bowers</td>
<td>Education</td>
</tr>
<tr>
<td>Dr. Albert Wolfson</td>
<td>Biology</td>
</tr>
<tr>
<td>Dr. Richard C. Bowers</td>
<td>Chemistry</td>
</tr>
<tr>
<td>Oregon State University</td>
<td>University of Pittsburgh</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------------</td>
</tr>
</tbody>
</table>
| Dr. Stanley E. Williamson  
Science Education         | Dr. Vernon C. Lingren  
Science Education         |
| Dr. T. Antoinette Ryan  
Education                 | Dr. J. William Asher  
Education                |
| Dr. Jack V. Hall  
Education                | Dr. C. Mauritz Lindvall  
Education               |
| Dr. Harold J. Evans  
Biology               | Dr. Ralph Buchsbaum  
Biology                   |
| Dr. Max B. Williams  
Chemistry              | Dr. Jerome L. Rosenberg  
Chemistry                  |
| Dr. James J. Brady  
Physics                 |                          |
|                          |                          |
| George Peabody College | Stanford University     |
| Dr. H. Craig Sipe  
Science Education          | Dr. Paul DeH. Hurd  
Science Education          |
| Dr. Sam P. Wiggins  
Education                | Dr. Nathaniel L. Gage  
Education                 |
| Dr. Robert A. Davis  
Education                | Dr. Arthur P. Coladarci  
Education                |
|                          |                          |
| Vanderbilt University   | University of Wisconsin |
| Dr. Frederick T. Wolf  
Biology                   | Dr. Milton O. Pella  
Science Education         |
| Dr. Arthur William Ingersoll  
Chemistry               | Dr. Chester Harris  
Education                |
| Dr. R. T. Lagemann  
Physics                 | Dr. Julian Stanley  
Education                |
| Dr. Wendell G. Holladay  
Physics              | Dr. Donald H. Bucklin  
Biology                   |
|                          | Dr. John W. Thomson  
Biology                   |
|                          | Dr. John D. Ferry  
Chemistry                |
|                          | Dr. Julian E. Mack  
Physics                  |
This letter is written to solicit your assistance with a doctoral study in science education. In a serious effort to improve the quality of research in the field, we have developed and now are ready to validate criteria which may be used for evaluating research. Judges have been selected from three areas—the physical and biological sciences, education, and the sub-specialty of science education.

As an experienced researcher, the process of evaluating research probably has become such a part of your thinking that you no longer bother to verbalize the criteria which you apply. The accompanying instrument has been designed to tap your rich background of knowledge and experience concerning research. The research criteria which we wish you to evaluate are stated in terms which permit their selective application to research in most areas, although they are intended primarily for evaluating research in science education. The criteria represent a substantial condensation of the sometimes conflicting and frequently repetitive literature on this topic which has accumulated during the past twenty to thirty years.

A brief list of directions is given at the beginning of the instrument. The responses provided will not be identified with specific institutions or individuals in the report of this study. If a summary of the report is desired, please so indicate in the space provided at the end of the instrument. A second copy of the instrument has been enclosed for your personal use. If you have any questions concerning the study, please do not hesitate to write to me. We hope you will find it possible to complete the instrument and return it within the next two weeks. Thank you very much for your assistance.

Very sincerely yours,

Philip G. Kapfer
INSTRUCTIONS

1. YOU NEED ONLY INDICATE WITH A CHECK MARK IN THE APPROPRIATE BOX whether you think a given criterion is of "great importance," "high importance," "average importance," "little importance," or "no importance," for evaluating research.

2. RESPOND ACCORDING TO YOUR UNDERSTANDING OF THE IMPORTANCE OF EACH CRITERION, even if you must qualify your response with a brief comment. Please do not omit a criterion by not responding to it.

3. JUDGE EACH CRITERION SEPARATELY FOR ITS IMPORTANCE IN EVALUATING COMPLETED RESEARCH. If a criterion is not applicable to any type of research in your field, judge it according to your understanding of the criterion's importance for evaluating research in other fields. In those cases where a criterion generally does not apply to the content of a research report, judge the criterion for its importance in helping the researcher as he conducts research.

4. All the criteria are not applicable to every kind of research study--ASSUME THAT THE PERSON USING THE CRITERIA WILL APPLY THEM SELECTIVELY TO THE APPROPRIATE KIND OF STUDY.

5. A single page of easily answered questions has been placed at the end of the instrument. These questions have been kept to a bare minimum, and your additional investment of time will be most helpful.

CRITERIA FOR EVALUATING RESEARCH

FORMULATION AND DEFINITION OF THE PROBLEM

<table>
<thead>
<tr>
<th>Problem Statement</th>
<th>great</th>
<th>high</th>
<th>avg</th>
<th>low</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The problem should be stated precisely in clear and understandable language.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The problem statement should include all subordinate problems which must be answered.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. The problem statement, including all of its parts, should encompass all elements of the problem.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The location of the problem statement in the research report should be identified so that it can be found and recognized readily.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Introduction

<table>
<thead>
<tr>
<th>Problem Choice</th>
<th>great</th>
<th>high</th>
<th>avg</th>
<th>low</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. The origin of a research problem should be identified.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. The body of organized knowledge in an area should be examined for gaps, for discrepancies, and for the directions in which extension is possible.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. A problem should be chosen which avoids unnecessary duplication of research effort.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. A problem should be chosen which can be formulated and defined with sufficient precision to permit a good idea of the goal to be achieved before it is attempted.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Attention should be given to the theoretical framework of knowledge underlying issues and problems before their resolution or solution is sought directly.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prepared by Philip G. Kapfer
Center for Science & Mathematics Education
The Ohio State University, Columbus, Ohio
February, 1964
10. The problem chosen should have either direct significance for people's needs or should serve as a foundation for subsequent research having significance for people's needs.

11. Experimentation involving human subjects should be planned and spaced carefully for their protection.

Delimitation and Analysis

12. The problem should be delimited exactly as to scope, the nature and quantity of data desired, and the methods of procedure to be followed.

13. In addition to simple delimitation of the problem, the subjects (if any are used) should be described with reference to the variables among them.

14. The factors which are selected for study should be the decisive ones for solution of the problem.

15. Variables which are used in experimental research should differentiate between or among the subjects or objects under study.

Assumptions

16. Assumptions upon which the inquiry rests should be identified.

17. Assumptions upon which the inquiry rests should be justified.

18. The assumptions in a study should be consistent with each other.

19. It should not be assumed that the non-research literature is an accurate indicator of best or prevalent practice.

Definitions

20. A definition should be expressed in precise language.

21. A definition must not be circular; it must not contain, directly or indirectly, the subject to be defined.

22. A definition should be stated in positive terms, rather than in negative terms, wherever possible.

23. Important terms, that are employed in an unusual sense, should be defined.

24. Undefined terms should be employed according to the meanings demanded by standard references.

25. Operational definitions of all generalized concepts which will be studied should be specified.

26. Defined terms and concepts should be used in the entire study according to the given definitions.

Related Literature

27. The related literature should be analyzed for its relevancy to the problem situation.
28. The related literature should be analyzed for any bias which may have significance for the study.

29. The review of related literature should include the previous research which may contribute to the solution of the problem.

30. The problem should be related to an orienting frame of reference or a systematic area of study.

31. Research in a given area should make use of relevant principles and research techniques of related disciplines.

32. Literature sources should be assigned to definite authors and times.

Hypotheses and Derived Theorems

33. The researcher should display creativity in his formulation of hypotheses.

34. A hypothesis should bring together a given number of observations and interpret the connections between them.

35. The possible alternative answers which are hypothesized for a given question should be independent of one another (or not overlap).

36. Hypotheses should be formulated on the basis of a thorough analysis of the theoretical and factual background of the problem.

37. Hypotheses should be formulated so that implications can be traced clearly by means of well-established techniques of deduction.

38. The consequences of alternative hypotheses should be developed for comparison with observable phenomena.

39. Hypotheses should be revised on the basis of induction from the data which accumulate as the study progresses.

PROCEDURAL DESIGN

Procedure

Design Selection

40. The problem, from the beginning, should condition the steps taken to solve it.

41. The adequacy of the research methodology chosen for solution of the problem should be explained and defended.

42. The research design should provide for collection of the most objective data available.

43. The research design should include several planned directions of inquiry in order to increase efficiency and comprehensiveness.

44. The experimental design should include the replication of one or more variables whose effects have already been assessed as significant in a previous experiment.
Design for Replication and Validity

45. The research design should be described in sufficient detail to permit repetition of the study by independent investigators.

46. The degree of internal and external validity in the research design should be identified and justified.

47. The data collected should afford some basis for generalization.

Self-Contained Design

48. The self-contained experiment should include a representative sample of the subjects or objects under study.

49. The self-contained experiment should include a control to exclude, at a known level of probability, a number of alternative interpretations of the experimental results.

50. The self-contained experiment should supply independent evidence on the question in dispute.

51. The self-contained experiment should permit an estimate of the experimental errors which affect the comparisons made.

Design for Data Gathering

52. The design for making observations should establish causal connections between events in which the cause is a sufficient and necessary condition for the effect.

53. The data of research should approach the reality of the situation being investigated as nearly as possible.

54. Modern technological instruments should be used to make complete visual and/or oral records of primary data for review by independent investigators.

55. The data collected should be within the boundaries defined by the hypotheses.

56. The data included within the defined boundaries of the study should be collected without personal choice.

57. Data should be gathered in a well organized, highly structured, and yet creative manner.

58. The collection of data or evidence should include a thorough investigation of primary data.

Special Methods

Instrumental Techniques

59. A data-gathering instrument should have validity or truthfulness; it should measure what it purports to measure.

60. A data-gathering instrument should have reliability or consistency; it should agree with itself when used repeatedly with the same individuals, groups, or objects.

61. The degree of reliability and validity which is accepted should be justified with reference to the objectives of the research problem.
62. The most effective instrument for collecting the necessary data should be anticipated in the research design. | great | high | aver- | lit- | tle | no |
|------|------|------|------|------|----|

**Statistical Techniques**

63. The statistical technique should be identified before the data are gathered. | □ | □ | □ | □ | □ |

64. The statistical hypotheses should be considered at the time the experiment is designed. | □ | □ | □ | □ | □ |

65. The statistical design should include as many independent variables in the same design as may seem necessary. | □ | □ | □ | □ | □ |

66. The statistical design should include as many dependent variables in the same design as may seem necessary. | □ | □ | □ | □ | □ |

67. The basic assumptions underlying statistical methods should be satisfied by the data to which they are applied. | □ | □ | □ | □ | □ |

68. The statistical sample should be selected from a carefully defined parent population. | □ | □ | □ | □ | □ |

69. The probability of error estimate should be justified. | □ | □ | □ | □ | □ |

70. The statistical technique should include tests for the reliability and validity of the data. | □ | □ | □ | □ | □ |

71. The statistical technique should allow for maximum objectivity in evaluation and in estimation. | □ | □ | □ | □ | □ |

72. The statistical technique should provide tests to assure that a sufficient quantity of data is collected. | □ | □ | □ | □ | □ |

73. The statistical technique should allow consideration for tests of significance and problems of estimation. | □ | □ | □ | □ | □ |

**PRESENTATION OF EVIDENCE**

**Presentation of Data**

74. The data should be logically organised and presented. | □ | □ | □ | □ | □ |

75. All data which are collected for the solution of a given problem should be reported. | □ | □ | □ | □ | □ |

76. The sources of all the facts which are presented should be identified. | □ | □ | □ | □ | □ |

77. The proof necessary to validate important alleged facts should be provided. | □ | □ | □ | □ | □ |

78. The apportionment of space in the research report should be made with reference to the judged relative importance of the topics treated. | □ | □ | □ | □ | □ |

79. When graphical methods are used, they should present the data simply and accurately, both in construction and in the impression given to the reader. | □ | □ | □ | □ | □ |

**Analysis of Data**

80. The analysis of data should be anticipated and worked through abstractly or in a pilot study before the study itself is begun. | □ | □ | □ | □ | □ |
81. Preconceived ideas about the results to be expected should not affect the manipulation of the data.

82. Accurate statistical or speculative methods of organizing the data should be employed.

83. Evidence which is collected in a study should be creatively and meaningfully synthesized.

84. Any accidents of circumstance which might affect the reliability or validity of the data should be reported.

85. Corrections should be made in the analysis of the data for any distortions which may have occurred in the data gathering process.

**Interpretation of Data**

Factors Involved

86. The interpretation of data should involve intuition and rational subjective evaluation.

87. The relationship between a hypothesis or a theory and the observational evidence, when both are given, should be judged.

88. The reliability of a prediction, which is based either on a hypothesis or a theory, should be judged on the basis of the evidence.

89. Both statistical significance and practical significance should be considered in the interpretation of research results.

90. Observation and theory should serve as checks on one another with neither having priority over the other.

91. The interpretation of the data should be guided by their source and nature.

**Evaluation of Theories and Hypotheses**

92. A theory should encompass a relatively wide range of data.

93. A theory should provide a basis for successful prediction.

94. A deductively elaborated hypothesis should be factually substantiated before it is used as an explanatory device.

95. A necessary and sufficient condition for the rejection of a theory or a hypothesis is the observation of reproducible contradictory evidence.

96. The assumptions which led to the experimentally testable propositions should be carefully examined before the hypothesis is tentatively accepted or rejected.

97. A hypothesis should make possible the determination of relationships between the known facts.

**SUMMARY AND CONCLUSIONS**

**Summary**

98. The summary should be an overview of the entire study.
99. The summary should be self-contained.

100. The summary should include only previously reported data.

101. The summary should avoid unnecessary enumeration of
details reported earlier in the manuscript.

102. The summary should be a final synthesis of the whole study
which is greater than the mere process of adding together
chapter or section summaries.

Conclusions

Formulation and Development

103. Conclusions should be drawn with reference to the
problematic situations that engendered the inquiry.

104. Conclusions should be tested by examining them for
internal consistency.

105. Conclusions should be tested by examining them for
external consistency.

106. Conclusions should include inferences relative to all
the significant data.

107. When several possible explanations of a phenomenon seem
equally correct, the simplest should be favored.

108. Skillful grouping and appropriate subordination of
conclusions under major headings should be used to avoid
encyclopedic enumeration of individual statements.

109. The findings of research should be stated with simplicity
within the context of a total pattern or theory.

Limitations

110. The limitations of a study should be presented to guide
the reader in using the findings, conclusions, and
recommendations.

111. A study should be thorough enough to arrive at carefully
stated results which have a substantial degree of
reliability and validity.

Applications, Recommendations, and Needed Research

112. Generalisations should be directed to the solution of
problems other than those considered in the study.

113. Both the data and the conclusions should warrant the
recommendations made.

114. Creative speculation concerning possible applications
and recommendations should be made.

115. Applications or recommendations which are based on the
creative insight of the investigator should be labeled
plainly as such.

116. Research should contribute to the identification and to
the solution of new problems.
1. Name (please print): ___________________________ first       middle       last

   (Your name and institution
   will not be associated with
   the information requested.)

   less less less less less
   than than than than than
   more
   than than than than than
   30--
   please
   specify

   □ □ □ □ □       □ □ □ □ □       □ □ □ □ □       □ □ □ □ □       □ □ □ □ □

2. Number of years of research experience:

3. Number of Master's Degree studies for which you were
   the major adviser:

4. Number of Doctoral Degree studies for which you were
   the major adviser:

5. Number of formal research studies in which you were a
   major contributor (exclude research studies included
   in numbers 3 and 4 above):

6. Academic Degrees: Year Received:

   Bachelor
   Master
   Ph.D.
   Ed.D.
   Other

   □ □ □ □ □

7. Present Academic Rank: Year Received:

   Instructor
   Asst. Prof.
   Assoc. Prof.
   Full Prof.
   Other

8. Area of Research Competence:

   Education
   Biological Sciences
   Chemistry
   Physics
   Other

   □ □ □ □ □

9. Type(s) of Research Methods Commonly
   Employed:

   Experimental
   Classroom
   Field
   Laboratory
   Other

   □ □ □ □ □

10. In your rating of the criteria, which
    responses indicate that you think a
criterion should not be included in
    the final instrument?

    Great Importance
    High Importance
    Average Importance
    Little Importance
    No Importance

    □ □ □ □ □

11. Do you wish to receive a summary of
    the results of this study?

    Yes □
    No □

Please feel free to make any comments on the opposite page.
Thank you very much for your generous assistance.
Because we have not heard from you yet, we are taking this opportunity to remind you of our previous request. It is now over two weeks since we mailed you the instrument containing criteria which may be used for evaluating research. Although we have a large number of returns, we are making every effort to obtain a reply from each researcher.

As we indicated in the cover letter which accompanied the instrument, you were selected to participate in this study together with other researchers in the educational and scientific fields. In order to obtain the services of the best qualified judges, a careful selection procedure was designed and followed, and a minimum number of experienced researchers were nominated. For this reason, your response is urgently needed. I am sure that you appreciate the importance of each reply for a valid study.

If you have lost or misplaced the criterion instrument, please let us know and we will send you another. If you have recently mailed back the instrument containing your responses, please disregard this reminder and accept our thanks for your cooperation.

Very sincerely yours,

Philip G. Kapfer

Dissertation Committee:
Dr. John S. Richardson, Professor of Education (Science Education)
Dr. Earl W. Anderson, Professor of Education (Teacher-Higher Education)
Dr. Jack G. Calvert, Professor of Chemistry (Physical Chemistry)
THE OHIO STATE UNIVERSITY  
Center for  
Science & Mathematics Education  
Columbus, Ohio 43210  
April 4, 1964

It is now over six weeks since we asked your participation in our study of criteria which may be used for evaluating research. Over 70% of the judges have responded, and the analysis of data has started. However, we have not yet received your return.

We have enclosed another copy of the instrument for your convenience. A brief list of directions is given at the beginning of the instrument, and postage has been provided for return mailing. Won't you please take the necessary time to complete it and mail it to us?

With your cooperation we look forward to completing the study. If you have recently mailed back the instrument containing your responses, please disregard this reminder and accept our thanks.

Very sincerely yours,

Philip G. Kapfer

Dissertation Committee:

Dr. John S. Richardson, Professor of Education (Science Education)
Dr. Earl W. Anderson, Professor of Education (Teacher-Higher Education)
Dr. Jack G. Calvert, Professor of Chemistry (Physical Chemistry)
APPENDIX IV

DATA OBTAINED IN THE STUDY
### TABLE 14

MEANS FOR EACH RESEARCH CRITERION AS RATED BY THE FIVE GROUPS OF JUDGES

<table>
<thead>
<tr>
<th>Groups of Judges</th>
<th>All Five Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Criterion Number</strong></td>
<td><strong>Science Education</strong></td>
</tr>
<tr>
<td>1</td>
<td>4.00</td>
</tr>
<tr>
<td>2</td>
<td>2.76</td>
</tr>
<tr>
<td>3</td>
<td>2.47</td>
</tr>
<tr>
<td>4</td>
<td>3.18</td>
</tr>
<tr>
<td>5</td>
<td>2.24</td>
</tr>
<tr>
<td>6</td>
<td>3.29</td>
</tr>
<tr>
<td>7</td>
<td>2.94</td>
</tr>
<tr>
<td>8</td>
<td>3.06</td>
</tr>
<tr>
<td>9</td>
<td>3.59</td>
</tr>
<tr>
<td>10</td>
<td>1.76</td>
</tr>
<tr>
<td>11</td>
<td>3.41</td>
</tr>
<tr>
<td>12</td>
<td>3.65</td>
</tr>
<tr>
<td>13</td>
<td>3.76</td>
</tr>
<tr>
<td>14</td>
<td>3.35</td>
</tr>
<tr>
<td>15</td>
<td>3.35</td>
</tr>
<tr>
<td>16</td>
<td>3.82</td>
</tr>
<tr>
<td>17</td>
<td>3.00</td>
</tr>
<tr>
<td>18</td>
<td>2.94</td>
</tr>
<tr>
<td>19</td>
<td>2.35</td>
</tr>
<tr>
<td>20</td>
<td>3.76</td>
</tr>
<tr>
<td>21</td>
<td>3.53</td>
</tr>
<tr>
<td>22</td>
<td>3.35</td>
</tr>
<tr>
<td>23</td>
<td>3.53</td>
</tr>
<tr>
<td>24</td>
<td>3.35</td>
</tr>
<tr>
<td>25</td>
<td>3.59</td>
</tr>
<tr>
<td>26</td>
<td>3.88</td>
</tr>
<tr>
<td>27</td>
<td>3.65</td>
</tr>
<tr>
<td>28</td>
<td>3.18</td>
</tr>
<tr>
<td>29</td>
<td>3.65</td>
</tr>
<tr>
<td>30</td>
<td>2.71</td>
</tr>
<tr>
<td>31</td>
<td>2.88</td>
</tr>
<tr>
<td>32</td>
<td>3.06</td>
</tr>
<tr>
<td>33</td>
<td>3.24</td>
</tr>
<tr>
<td>34</td>
<td>3.00</td>
</tr>
<tr>
<td>35</td>
<td>2.35</td>
</tr>
<tr>
<td>Number</td>
<td>Education</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
</tr>
<tr>
<td>36</td>
<td>3.24</td>
</tr>
<tr>
<td>37</td>
<td>3.18</td>
</tr>
<tr>
<td>38</td>
<td>2.76</td>
</tr>
<tr>
<td>39</td>
<td>3.47</td>
</tr>
<tr>
<td>40</td>
<td>3.35</td>
</tr>
<tr>
<td>41</td>
<td>3.65</td>
</tr>
<tr>
<td>42</td>
<td>3.59</td>
</tr>
<tr>
<td>43</td>
<td>2.12</td>
</tr>
<tr>
<td>44</td>
<td>2.24</td>
</tr>
<tr>
<td>45</td>
<td>3.59</td>
</tr>
<tr>
<td>46</td>
<td>3.65</td>
</tr>
<tr>
<td>47</td>
<td>3.29</td>
</tr>
<tr>
<td>48</td>
<td>3.59</td>
</tr>
<tr>
<td>49</td>
<td>3.35</td>
</tr>
<tr>
<td>50</td>
<td>3.47</td>
</tr>
<tr>
<td>51</td>
<td>3.65</td>
</tr>
<tr>
<td>52</td>
<td>3.00</td>
</tr>
<tr>
<td>53</td>
<td>3.71</td>
</tr>
<tr>
<td>54</td>
<td>2.53</td>
</tr>
<tr>
<td>55</td>
<td>2.82</td>
</tr>
<tr>
<td>56</td>
<td>3.47</td>
</tr>
<tr>
<td>57</td>
<td>2.82</td>
</tr>
<tr>
<td>58</td>
<td>3.35</td>
</tr>
<tr>
<td>59</td>
<td>4.00</td>
</tr>
<tr>
<td>60</td>
<td>3.94</td>
</tr>
<tr>
<td>61</td>
<td>3.94</td>
</tr>
<tr>
<td>62</td>
<td>3.71</td>
</tr>
<tr>
<td>63</td>
<td>3.41</td>
</tr>
<tr>
<td>64</td>
<td>3.65</td>
</tr>
<tr>
<td>65</td>
<td>3.12</td>
</tr>
<tr>
<td>66</td>
<td>3.06</td>
</tr>
<tr>
<td>67</td>
<td>3.76</td>
</tr>
<tr>
<td>68</td>
<td>3.76</td>
</tr>
<tr>
<td>69</td>
<td>3.65</td>
</tr>
<tr>
<td>70</td>
<td>3.47</td>
</tr>
<tr>
<td>Criterion Science Number</td>
<td>Education</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>3.65</td>
</tr>
<tr>
<td>72</td>
<td>3.18</td>
</tr>
<tr>
<td>73</td>
<td>3.65</td>
</tr>
<tr>
<td>74</td>
<td>3.65</td>
</tr>
<tr>
<td>75</td>
<td>2.65</td>
</tr>
<tr>
<td>76</td>
<td>3.53</td>
</tr>
<tr>
<td>77</td>
<td>3.41</td>
</tr>
<tr>
<td>78</td>
<td>2.35</td>
</tr>
<tr>
<td>79</td>
<td>3.59</td>
</tr>
<tr>
<td>80</td>
<td>3.00</td>
</tr>
<tr>
<td>81</td>
<td>3.35</td>
</tr>
<tr>
<td>82</td>
<td>3.41</td>
</tr>
<tr>
<td>83</td>
<td>3.18</td>
</tr>
<tr>
<td>84</td>
<td>3.88</td>
</tr>
<tr>
<td>85</td>
<td>3.88</td>
</tr>
<tr>
<td>86</td>
<td>2.29</td>
</tr>
<tr>
<td>87</td>
<td>3.29</td>
</tr>
<tr>
<td>88</td>
<td>3.59</td>
</tr>
<tr>
<td>89</td>
<td>3.06</td>
</tr>
<tr>
<td>90</td>
<td>2.35</td>
</tr>
<tr>
<td>91</td>
<td>3.47</td>
</tr>
<tr>
<td>92</td>
<td>2.82</td>
</tr>
<tr>
<td>93</td>
<td>3.41</td>
</tr>
<tr>
<td>94</td>
<td>2.76</td>
</tr>
<tr>
<td>95</td>
<td>3.18</td>
</tr>
<tr>
<td>96</td>
<td>3.29</td>
</tr>
<tr>
<td>97</td>
<td>2.94</td>
</tr>
<tr>
<td>98</td>
<td>2.83</td>
</tr>
<tr>
<td>99</td>
<td>2.88</td>
</tr>
<tr>
<td>100</td>
<td>2.71</td>
</tr>
<tr>
<td>101</td>
<td>3.12</td>
</tr>
<tr>
<td>102</td>
<td>3.53</td>
</tr>
<tr>
<td>103</td>
<td>3.35</td>
</tr>
<tr>
<td>104</td>
<td>3.29</td>
</tr>
<tr>
<td>105</td>
<td>3.06</td>
</tr>
</tbody>
</table>
TABLE 14—Continued

<table>
<thead>
<tr>
<th>Number</th>
<th>Cri-terion Science</th>
<th>Education</th>
<th>Education</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Physics</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>106</td>
<td>2.88</td>
<td>2.58</td>
<td>2.17</td>
<td>2.58</td>
<td>2.43</td>
<td>2.57</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>2.76</td>
<td>2.21</td>
<td>2.17</td>
<td>1.75</td>
<td>3.29</td>
<td>2.36</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>3.35</td>
<td>2.63</td>
<td>2.92</td>
<td>2.36</td>
<td>2.71</td>
<td>2.86</td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>3.41</td>
<td>3.00</td>
<td>3.25</td>
<td>3.08</td>
<td>3.00</td>
<td>3.15</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>3.76</td>
<td>3.00</td>
<td>2.75</td>
<td>3.17</td>
<td>3.29</td>
<td>3.19</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>4.00</td>
<td>3.25</td>
<td>3.17</td>
<td>3.33</td>
<td>3.14</td>
<td>3.42</td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>1.35</td>
<td>1.83</td>
<td>2.08</td>
<td>2.67</td>
<td>2.00</td>
<td>1.92</td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>3.76</td>
<td>3.29</td>
<td>3.00</td>
<td>3.33</td>
<td>3.00</td>
<td>3.33</td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>2.59</td>
<td>2.17</td>
<td>2.33</td>
<td>2.50</td>
<td>1.86</td>
<td>2.32</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>3.71</td>
<td>3.17</td>
<td>2.42</td>
<td>2.83</td>
<td>2.86</td>
<td>3.08</td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>3.12</td>
<td>2.79</td>
<td>3.33</td>
<td>3.00</td>
<td>3.29</td>
<td>3.04</td>
<td></td>
</tr>
<tr>
<td>Criterion Number</td>
<td>Standard Deviation</td>
<td>Criterion Number</td>
<td>Standard Deviation</td>
<td>Criterion Number</td>
<td>Standard Deviation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.432</td>
<td>40</td>
<td>1.210</td>
<td>79</td>
<td>0.888</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.092</td>
<td>41</td>
<td>1.019</td>
<td>80</td>
<td>1.149</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.253</td>
<td>42</td>
<td>0.900</td>
<td>81</td>
<td>1.048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.904</td>
<td>43</td>
<td>1.251</td>
<td>82</td>
<td>0.992</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.045</td>
<td>44</td>
<td>1.150</td>
<td>83</td>
<td>0.939</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.850</td>
<td>45</td>
<td>0.685</td>
<td>84</td>
<td>0.597</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.043</td>
<td>46</td>
<td>0.924</td>
<td>85</td>
<td>0.934</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.087</td>
<td>47</td>
<td>0.949</td>
<td>86</td>
<td>1.232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.959</td>
<td>48</td>
<td>1.080</td>
<td>87</td>
<td>0.762</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.300</td>
<td>49</td>
<td>1.050</td>
<td>88</td>
<td>0.705</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.906</td>
<td>50</td>
<td>1.094</td>
<td>89</td>
<td>1.104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1.332</td>
<td>51</td>
<td>0.787</td>
<td>90</td>
<td>1.397</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1.142</td>
<td>52</td>
<td>1.238</td>
<td>91</td>
<td>0.963</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.983</td>
<td>53</td>
<td>1.204</td>
<td>92</td>
<td>1.326</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1.053</td>
<td>54</td>
<td>1.146</td>
<td>93</td>
<td>1.120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.749</td>
<td>55</td>
<td>1.342</td>
<td>94</td>
<td>1.068</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>1.054</td>
<td>56</td>
<td>1.308</td>
<td>95</td>
<td>1.110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>1.175</td>
<td>57</td>
<td>1.299</td>
<td>96</td>
<td>1.045</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>1.267</td>
<td>58</td>
<td>1.144</td>
<td>97</td>
<td>1.186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.820</td>
<td>59</td>
<td>0.587</td>
<td>98</td>
<td>1.206</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.949</td>
<td>60</td>
<td>0.620</td>
<td>99</td>
<td>1.091</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1.137</td>
<td>61</td>
<td>0.854</td>
<td>100</td>
<td>1.518</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.924</td>
<td>62</td>
<td>1.021</td>
<td>101</td>
<td>0.939</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1.055</td>
<td>63</td>
<td>1.104</td>
<td>102</td>
<td>1.240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.075</td>
<td>64</td>
<td>0.911</td>
<td>103</td>
<td>1.028</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>0.838</td>
<td>65</td>
<td>1.043</td>
<td>104</td>
<td>0.792</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>0.805</td>
<td>66</td>
<td>1.168</td>
<td>105</td>
<td>0.796</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>0.956</td>
<td>67</td>
<td>0.911</td>
<td>106</td>
<td>1.072</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>0.723</td>
<td>68</td>
<td>1.029</td>
<td>107</td>
<td>1.325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.095</td>
<td>69</td>
<td>1.048</td>
<td>108</td>
<td>0.810</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>0.906</td>
<td>70</td>
<td>0.883</td>
<td>109</td>
<td>0.781</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1.255</td>
<td>71</td>
<td>0.879</td>
<td>110</td>
<td>0.850</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>1.103</td>
<td>72</td>
<td>1.028</td>
<td>111</td>
<td>0.801</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>1.064</td>
<td>73</td>
<td>0.928</td>
<td>112</td>
<td>1.230</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>1.179</td>
<td>74</td>
<td>0.786</td>
<td>113</td>
<td>0.888</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>0.948</td>
<td>75</td>
<td>1.332</td>
<td>114</td>
<td>1.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>1.158</td>
<td>76</td>
<td>1.067</td>
<td>115</td>
<td>1.031</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>1.083</td>
<td>77</td>
<td>1.061</td>
<td>116</td>
<td>1.119</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>1.317</td>
<td>78</td>
<td>1.156</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Breitwieser, J. V. "Essentials of Good Research," Phi Delta Kappan, 24 (December, 1941), 185.


Crawford, Claude C. The Technique of Research in Education. Los Angeles, California: The University of Southern California, 1928.


Watson, Fletcher G. "Research in the Physical Sciences," Phi Delta Kappan, 35 (October, 1953), 4-6.


Whitehead, Matthew J. "The Place of Subjectivity in Research," The Quarterly Review of Higher Education Among Negroes, 12 (April, 1944), 63-64.

