

This dissertation has been 62-3987
microfilmed exactly as received

HASTINGS, Glen Richard, 1922-
AN ANALYSIS OF TEACHER-MADE OBJECTIVE
TEST ITEMS DEVELOPED BY A SELECTED
SAMPLE OF IOWA PUBLIC HIGH SCHOOL
TEACHERS.

The University of Oklahoma, Ed.D., 1962
Education, psychology

University Microfilms, Inc., Ann Arbor, Michigan

Copyright by
Glen Richard Hastings

1962

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN ANALYSIS OF TEACHER-MADE OBJECTIVE TEST
ITEMS DEVELOPED BY A SELECTED SAMPLE OF
IOWA PUBLIC HIGH SCHOOL TEACHERS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF EDUCATION

BY

GLEN R. HASTINGS

Norman, Oklahoma

1962

AN ANALYSIS OF TEACHER-MADE OBJECTIVE TEST
ITEMS DEVELOPED BY A SELECTED SAMPLE OF
IOWA PUBLIC HIGH SCHOOL TEACHERS

APPROVED BY

O. J. Ruppel
Claude Kelley
Charles E. Hickey
W. M. J. J.

DISSERTATION COMMITTEE

ACKNOWLEDGMENT

Appreciation is extended to all who have contributed to this study. Sincere appreciation is expressed to Dr. Omer J. Rupiper, chairman of the dissertation committee for his helpful suggestions and encouragement. Also, to members of the committee: Dr. Arthur W. Heilman, Dr. Claude Kelley, and Dr. Charlyce King for their beneficial suggestions.

The assistance and cooperation of Dr. Gordon Rhum, for his helpful suggestions, and Dr. Merrit Melberg, for the use of his test and his Teacher and the Child television class is acknowledged.

Individual indebtedness to the many teachers who participated in this study is recognized. Only through their combined efforts and cooperation was this study possible.

Special appreciation is extended to Winifred for her patience and encouragement, and to Richard, Ronald, and Glenneva for their willingness to sacrifice their Father's time during the period of time required for this study.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENT.....	iii
LIST OF TABLES.....	v
Chapter	
I. THE PROBLEM.....	1
Introduction	
Statement of the Problem	
Hypotheses	
Assumptions	
Limitations	
Importance of the Problem	
Definition of Terms	
Review of Related Literature	
Organization of the Study	
II. EVALUATION OF OBJECTIVE EXAMINATION ITEMS...	14
Rules for Writing Objective Examination Items	
Examples of Errors in Item Writing	
Effect of Errors on Examinee Performance	
Consistency Among Evaluators	
Consistency Between Evaluators	
III. SAMPLING TECHNIQUES AND PROCEDURES.....	37
Sample	
Questionnaire	
Response to Requests	
Selection of Items for Evaluation	
IV. ANALYSIS OF DATA AND CONCLUSIONS.....	57
Tests of Hypotheses	
Findings	
Conclusions	
V. SUMMARY AND RECOMMENDATIONS.....	69
Summary	
Recommendations	
APPENDICES.....	74
BIBLIOGRAPHY.....	99

LIST OF TABLES

Table	Page
1. Comparison of Student Performance on Two Forms of a Test Over a Unit in an Introductory Education Course.....	29
2. Violations of Objective Examination Item Writing Rules Observed in Fifty Teacher-Made Test Items by Four Independent Evaluators.....	33
3. Number of Errors of Objective Item Writing Rules Reported for the Same Items on Two Different Occasions.....	35
4. Distribution of Teachers in Iowa Public High Schools, 1960-1961.....	41
5. Distribution of Teachers in the Sample According to School District Classification.....	42
6. A Comparison of the Geographic Representation of Returns with the Original Sample Population Distribution.....	47
7. Responses by School District Classification and Subject Matter Area.....	48
8. Comparison of the Geographic Representation of the Second Mailing Returns with the Original Sample Population Distribution.....	49
9. Second Mailing Responses by School District Classifications and Subject Matter Area.....	49
10. Comparison of the Number of First Mailing Returns with the Number of Second Mailing Returns on the Basis of Course Training, Subject Matter Teaching Area, and Teaching Experience.....	51
11. Comparison of the Number of Errors Found in Tests Submitted by Teachers Responding to the First Mailing Request with the Number of Errors Found in Tests Submitted by Teachers Responding to the Second Mailing Request.....	54
12. Total Responses by School District and Subject Matter Classification.....	54

Table	Page
13. Total Responses by Geographic Distribution...	55
14. Comparison of the Number of Errors of Item-Writing Rules Found in Tests Submitted by Members of the Sample.....	59
15. Comparison of the Proportion of Errors of Item Writing Rules in Tests Submitted by Members of the Sample.....	60
16. Comparison of the Number of Errors of Item Writing Rules with Form and Style Rule Violations Omitted.....	61
17. Comparison of Proportion of Errors of Item Writing Rules with Form and Style Rule Violations Omitted.....	62
18. Comparison of the Number of Errors Committed by Members of the Sample on the Basis of Course Training and Sex, Time Elapsed, College Preparation and College Attended.....	63
19. Comparison of the Proportion of Errors Committed by Members of the Sample on the Basis of Course Training and College Attended	64

AN ANALYSIS OF TEACHER-MADE OBJECTIVE TEST
ITEMS DEVELOPED BY A SELECTED SAMPLE OF
IOWA PUBLIC HIGH SCHOOL TEACHERS

CHAPTER I

THE PROBLEM

Introduction

During the past two decades increased interest in the evaluation of pupil performance has been exhibited by classroom teachers and by members of various educational associations. As a result of this interest, numerous objective methods have been utilized in an effort to make the evaluation of pupil performance more exact. Comprehensive testing programs, including standardized tests of intelligence, aptitude, achievement, attitude, personality, and interests, have become an integral part of many schools' evaluation and guidance programs.

Lindquist states that "there is a real need in instruction for examinations based on the immediate objectives and upon the peculiar organization and content

of specific courses of instruction."¹ Since there is considerable variation from school to school in course content, it is evident that good locally constructed tests are highly desirable for the measurement of these objectives.

The learning outcomes expected of pupils are usually defined by the classroom teacher. Therefore, the classroom teacher should assess the extent to which each pupil understands and can apply the facts, concepts, and skills required for mastery of the unit of instruction. This requires the development of instruments for measuring pupil achievement. The objective examination, consisting of various item types, is one of these devices.

Teacher education institutions generally provide courses in which the theory and principles of objective teacher-made test construction are considered. However, in many instances, such a course is not required for graduation. Also, many such courses do not provide for the development of skill in teacher-made test construction. If the teacher is to be proficient in constructing classroom tests, the necessary skill must be developed outside the formal classroom situation.

¹E. F. Lindquist, (ed.), Educational Measurement, (Washington, D. C.: American Council on Education, 1951), p. 138.

Statement of the Problem

The purpose of this study was to determine if teachers who had completed a course or a unit in teacher-made test construction committed fewer errors in writing test items than did teachers who were not course or unit trained. A further purpose was to determine if the number of errors was affected by teaching experience and if teachers of one subject-matter field committed more errors in writing teacher-made test items than did teachers of another subject-matter teaching field.

Hypotheses

The following null hypotheses were formulated and tested:

1. There is no significant difference between the number of violations of objective examination item-writing rules made by course trained teachers and teachers who were not course trained.
2. There is no significant difference between the number of violations of objective examination item writing rules made by experienced teachers and limited experience teachers.
3. There is no significant difference between the number of violations of objective examination item writing rules by social science teachers and natural science teachers.

Assumptions

The basic assumption necessary for the pursuance of this study was that item writing rules listed in Tests and Measurements textbooks and related publications reflect a summarization of the experience of those who have been successful in test construction. These are the best available rules for effective item writing. The application of these rules in writing objective examination items is essential to the development of good teacher-made tests. When a test has been developed that is essentially free from violations of these rules and is based on the goals of instruction, subsequent statistical analyses will provide useful evidence regarding the difficulty and discriminating level of items in the test.

Limitations

This study was limited to the analysis of objective teacher-made test items obtained from a selected sample of Iowa public high school teachers. Items for the study were obtained from teachers of the natural and social sciences. Therefore, only those two subject matter areas were involved in the study.

Importance of the Study

Evaluation of pupil achievement has frequently been stressed as one of the most important steps in the

teaching process. In spite of the rather general recognition by classroom teachers of the skills and abilities needed for assessing pupil growth, emphasis has not been given to the development of these skills and abilities in the teacher preparation program. Emphasis given to the development of skills requisite for the construction of teacher-made tests has been limited. When teacher education programs have included courses in which the development of teacher-made test construction skills were emphasized, it is likely that courses have been handled as if there were little or no difference in the skills necessary for the construction of teacher-made tests in different subject-matter fields. Accompanying the problems of teacher preparation, teacher certification agencies have been faced with the question of requiring preparation in teacher-made test construction for certification. The experimental evidence provided by this study can be used as a guide in planning the evaluation phase of the teacher preparation and certification requirements.

Definition of Terms

The following definitions are applicable to the terms used in this study:

Course trained - teachers who have successfully completed a course or unit in another course in which the theory and

application of objective teacher-made test construction is explored and developed,

Examination - objective measurement devices developed by classroom teachers for measuring student achievement in terms of the goals of instruction,

Experienced teacher - a public high school teacher who has completed more than three years of public high school teaching,

Limited experience teacher - a public high school teacher who has completed three years or less of public high school teaching,

Item evaluation - is the evaluation of a single test item through a comparison with predetermined objective item-writing rules,

Item - a single question or exercise in a test; the smallest unit of test construction,

Item form - any of several arrangements in which an item can be presented in an examination,

Tests and Measurements - any of several courses which refer to a collegiate course of study containing units which present the background of evaluation, construction of the teacher-made test and its use, standardized tests and methods of test score interpretation.

Review of Related Literature

Rules or suggestions for writing objective

examination items have appeared in the literature for more than three decades. Greene and Jorgensen, in 1929, stated that "One of the most frequently discussed modern educational topics is the so called 'new examination.'"¹ They listed only four suggestions for writing the objective test item. However, during the last three decades, authors of Tests and Measurements textbooks have written about objective test development and have offered rules for their construction. Although much has been written on the subject and numerous rules offered, there does not appear to be any research in which the application of these rules by the classroom teacher has been experimentally tested.

A search of the literature reveals that many studies of objective type examinations have been reported during the last decade. However, emphasis has been given to the use of item statistics and other statistical techniques in most of these studies. This use of item statistics should be limited to the refinement of items, rather than a means of determining if an item is well written. Ebel takes the position that:

--extensive use of statistical methods for the analysis of responses to test items has seemed to imply that test production can be made a statistical science in which the skill of the

¹Harry A. Greene and Albert N. Jorgensen, The Use and Interpretation of Educational Tests, (New York: Longmans, Green and Co., 1929), p. 90.

item writer is of secondary importance. This notion is based on a misconception of the role of item analysis.¹

The position held by Ebel is supported by Dressel since item statistics are based upon actual groups of students. He contends that "item statistics do not tell us much about how good an item really is."² Studies dealing with statistical analysis, consequently should follow research delving into objective examination item writing.

The published research did not reveal studies that investigated the application of objective examination item writing skills of teachers. Several studies based on the measurement of pupil achievement were reported. Two classes of studies were represented in this category: High school achievement as a basis for college admissions and the evaluation of teaching effectiveness. In the first category, Anderson reported several studies which indicated that some fixed average of secondary school marks or rank in class was one of the criteria used as a basis for admission by a majority of collegiate institutions.³ In

¹Robert L. Ebel, "Writing the Test Item," Chapter VI in Educational Measurement, ed. E. F. Lindquist, (Washington, D. C.: American Council on Education, 1951), p. 186.

²Paul L. Dressel, "The Way of Judgment," paper in the 15th. Yearbook of the National Council on Measurements Used in Education, ed. Edith M. Hudleston, (New York: National Council on Measurements Used in Education, 1958), p. 4.

³G. Lester Anderson, "Colleges and Universities--Organization and Administration," article in The Encyclopedia of Educational Research, Third edition, ed. Chester W. Harris, (New York: The Macmillan Company, 1960), p. 257.

the second category, Orleans advocated measuring student growth as the most effective approach to evaluating teaching effectiveness.¹ McCall attempted to establish a merit index through the measurement of student growth.² Bendig reported that the rating of teachers bears a significant relationship to the student's achievement.³ Travers and Gronlund reported the results of a survey which revealed that, on the average, fifty per cent of a student's marks are determined through examinations.⁴ Consequently, the value of research based on student's marks is questionable.

When students' marks are used as the basis for college admission and the evaluation of teaching effectiveness, skill in objective examination item writing becomes a grave concern for the student as well as for the teacher. Noll stressed the importance of competence in measurement and evaluation when he stated that:

¹Jacob S. Orleans, et. al. "Some Preliminary Thoughts on the Criteria of Teacher Effectiveness," Journal of Educational Research, 45:648, December 1952.

²William A. McCall, "Measurement of Teacher Merit," Publication number 294, State Superintendent of Public Instruction, (Raleigh: State of North Carolina, 1952), p. 34.

³Albert W. Bendig, "The Relation of Level of Course Achievement to Student's Instructor and Course Ratings in Introductory Psychology," Educational and Psychological Measurement, 13:438, 1953.

⁴Robert M. W. Travers and Norman E. Gronlund, "The Meaning of Marks," Journal of Higher Education, 21:373, October 1950.

---measurement and evaluation are essentials of teaching. The teacher must make judgments, measure, appraise and report. To do this he must use tests, examinations, inventories, rating scales and other types of measuring instruments and devices. He must know how to construct such instruments for his own purposes---¹

The importance of the preparation stressed by Noll was substantiated by Cook when teachers polled in his study listed the knowledge of constructing objective tests first in a series of twenty-two teacher's needs.² This clearly demonstrated that teachers have recognized the need for training in the construction of classroom or teacher-made tests.

There are two ways of obtaining this training: preparation in the teacher education program and through in-service training programs. The teacher preparation and certification program offers some hope for the development of these skills. However, Noll reported in 1958 that "only sixty-six of eighty-eight institutions surveyed offered an introductory course in Educational Measurements, but that only seven required prospective teachers to take it."³

¹Victor H. Noll, "Problems in the Pre-service Preparation of Teachers in Measurement," paper read before the annual meeting of the National Council on Measurements in Education, Chicago, Illinois, February 22, 1961.

²Walter W. Cook, "What the Teacher Should Know About Measurements," paper in the 15th. Yearbook of the National Council on Measurements Used in Education, ed. Edith M. Huddleston, (New York: National Council, 1958), p. 17.

³Victor H. Noll, "Should a Course in Measurement be Required for Teacher Certification?" paper in the 15th. Yearbook of the National Council on Measurements Used in Education, ed. Edith M. Huddleston, (New York: National Council, 1958), p. 37.

Woellner and Wood reported the certification requirements of states in 1953 and in 1960.¹ Their survey disclosed that there has been little change in the number of states requiring specific preparation for secondary certification. The overall number was three in 1953 and four in 1960. Investigation into the item writing skills required of the classroom teacher seemed necessary for progress in formal teacher education programs.

The second alternative by which the teacher may develop skill in writing objective examinations is in-service training programs. Although a search of the literature did not reveal any studies in this particular area, the possibility of such research was under study by a committee of the National Council on Measurements in Education. Hastings, a former member of that committee, stated that "most of the appropriate training of teachers in the measurement area will have to be done on an in-service basis."² The absence of experimental evidence on which to base in-service measurement training programs was apparent.

¹R. A. Woellner and A. M. Wood, Requirements for Certification of Teachers, Counselors, Librarians and Administrators. (Chicago: University of Chicago Press, 1960), p. 19.

²J. Thomas Hastings, "How Can In-Service Training in Measurement Techniques be Provided Most Effectively?", paper in the 15th. Yearbook of the National Council of Measurements Used in Education, ed., Edith M. Huddleston, (New York: National Council, 1958), p. 28.

Some preliminary investigations are needed to provide the framework for future research. The need for a study of this kind was implied, though not explicitly stated by Ebel and Damrin and by Hagen and Thorndike.¹ Ebel states that "The amount of reflection, discussion, and publication directed to the problem of writing objective test items has fallen far short of the requirements of the task."² Investigation of the classroom teachers' application of item-writing rules would provide a point of departure for research designed to fill the gap pointed out by Ebel.

Method of Procedure

The collection of rules for writing objective examination items was the first step necessary for the furtherance of this study. The suggested rules for writing objective examination items were compiled through content analysis of all available literature. The relevance of these rules was not substantiated by experimental evidence, however, Ebel states that the suggested rules represent a distillation of the years of experience of those who have

¹Robert L. Ebel and Dora E. Damrin, "Tests and Examinations," article in The Encyclopedia of Educational Research, Third Edition, ed. Chester W. Harris, (New York: The Macmillan Company, 1960), p. 1513; and Elizabeth P. Hagen and R. L. Thorndike, "Evaluation," article in The Encyclopedia of Educational Research, Third Edition, ed. Chester W. Harris, (New York: The Macmillan Company, 1960), p. 484.

²Op. cit., Ebel, "Writing the Test Item," p. 190.

been successful in examination construction.¹ Therefore, the compiled rules were established as the criteria against which examination items used in this study were evaluated.

Since the principal investigator evaluated the examination items used in this study, it was necessary to determine the consistency in evaluating objective examination items between the investigator and different evaluators. Also it was necessary to demonstrate the effect of errors of the item writing rules on examinee performance since experimental evidence was not available.

The selection of the sample was made from Iowa public school teachers. Objective examination items developed by each member of the sample were then requested for the purposes of this study. Items submitted by members of the sample were then evaluated against the criteria of the study. Every error of the criteria was recorded and summed for each member of the sample. The total number of errors of the criteria became the data used in testing the stated hypotheses. The findings were reported and conclusions were reached as a result of the analysis of the data. The study was then summarized and recommendations for further research were made.

¹Ibid.

CHAPTER II

EVALUATION OF OBJECTIVE EXAMINATION ITEMS

Rules for Writing Objective Examination Items

In general, Tests and Measurements textbooks contained listings of rules or suggestions for writing objective examination items. These were presented in order to fulfill a particular author's organization. Consequently, the majority of these lists were incomplete. This made it necessary to collect and organize a complete list for the purposes of this study.

The objective techniques with which this study was concerned, were those in which the procedure for marking and scoring can be specified in exact terms so that different scorers would assign the same or essentially the same score to a given performance.¹ The item forms that were commonly classified as objective were variations of the multiple-choice, alternate response, matching, completion and short-answer forms. Although completion and short-answer forms were not as objective as the other

¹Henry D. Rinsland, Constructing Tests and Grading, (New York: Prentice-Hall, Inc., 1937), p. 15.

forms, they were more objective than ordinary essay forms. For this reason, most authorities in the measurement field included completion and short-answer items as objective item forms.¹ The collection and organization of rules for writing objective examination items was limited to these item forms.

The first step in the compilation of rules for writing objective examination items was the development of a complete reference list of textbooks and materials containing suggestions for writing objective examination items. This reference list is included as Appendix I. These references were thoroughly searched and each suggestion for writing the objective examination item was recorded on a card. When all references had been thoroughly searched and all suggestions recorded, the cards were sorted into categories for the different item types and for related rules. These were then arranged in the order presented below:

I. General rules applicable to all item forms:

A. Involved sentence structure with many dependent clauses, many qualifications and complex ideas should be avoided.

B. Items should be clear and definite, free from ambiguities.

¹Dorothy Adkins Wood, Test Construction, (Columbus, Ohio: Charles E. Merrill Books, Inc., 1960), p. 29.

C. The item should be stated in language which is adapted to the educational level of students for which the item is intended.

D. The use of textbook or stereotyped language in phrasing items should be minimized as much as possible or avoided.

E. An item should be realistic and practical; that is, it should call for knowledge that the pupil must use.

F. Authority must be given for statements that might be considered controversial.

G. One item should not be based upon the response to another statement, nor should one item provide the response clue to another item.

H. Irrelevant clues should direct examinees away from correct responses.

I. Directions and instructions should be clear and complete.

J. The simplest possible method of indicating a response should be provided.

K. Single space within an item, double space between items.

II. Rules applicable to the writing of multiple-choice type items:

A. General:

1. The length of the item stem and choices should be determined by the purpose of the item.

2. A collection of alternate response items should not be presented as a multiple-choice item.

3. Understanding of definitions is better tested by furnishing the word and requiring a choice between alternative definitions.

B. Item Stem:

1. The problem should be stated briefly and accurately but completely.

2. The item stem must present a single central problem.

3. A direct question or an incomplete statement may be used as the item stem.

4. The item stem should be stated positively except where the student is expected to select the alternative which is not (least) like the remaining alternatives.

5. In general, include in the stem any words that must otherwise be repeated in each response.

C. Alternatives (choices):

1. Alternative choices should represent errors commonly made by students.

2. Alternatives should deal with similar ideas or data which are expressed in parallel form.

3. One choice should be clearly best, but the others should appear plausible to the uninformed.

4. There should be only one correct response and it should be unquestionably correct.

5. Each choice should be grammatically correct and consistent with the item stem.

6. The correct choice should not contain the same wording as the stem.

7. Use "none of these" as a response only in items to which an absolutely correct answer can be given. It should be used as an obvious answer several times early in the test, sparingly thereafter.

8. There should be at least four choices. The generally preferred number is five.

9. Alternatives should be listed one under the other, not in successive choices on the same line.

10. The correct punctuation marks should be used at the end of each of the item choices.

11. Alternatives should be about the same in length.

12. The correct response should occur in random order.

III. Rules applicable to the writing of alternate-response type items:

A. Statements should appear practical and relevant to the student.

B. Items should be restricted to one idea.

C. Statements should be true or false without qualification.

D. Statements should not be broad generalizations.

E. Trivial and meaningless statements should be avoided.

F. Opinionated and trick items should be avoided.

G. The meaning of a statement should be the only clue to the correct response. Avoid specific determiners.

1. Strongly worded statements containing such words as all, always, never, no, totally, entirely, completely, solely, fully, absolutely, exclusively, cause, reason, only, every, none or nothing are more often false than true.

2. Statements containing such words as sometimes, maybe, as a rule, should, may, most, some, often, and generally are more often true than false.

3. Avoid a consistent pattern of response.

4. Words or phrases that are repeated often become clues to the correct answer.

5. Generally, true statements are consistently longer than false statements.

6. Double negatives should be avoided.

7. Avoid negative statements whenever possible.

H. False statements should represent an incorrect fundamental idea or principle, not a trivial detail.

I. There should be approximately equal numbers of true and false (yes-no, right-wrong, plus-minus) statements.

IV. Rules applicable to the writing of matching type items:

A. Directions should clearly explain the intended basis for matching.

B. There should be a high degree of homogeneity in every set of matching items.

C. The problem statements and responses should be arranged for maximum clarity and convenience to the examinees.

D. A group of problem statements and their matching responses should all be on one page.

E. Matching exercises should contain at least five and not more than fifteen items.

F. The number of items in the response column should always exceed the number of items in the opposite primary column.

G. Items in the response column should be arranged in some systematic manner: chronologically or alphabetically.

H. Only one item in the response column should be the correct response-for each problem statement in the primary column.

I. The response column should contain shorter statements than the problem column.

J. Do not capitalize words in the response column unless they should be capitalized normally.

K. Avoid clues such as singular and plural items with response agreement in number.

L. Avoid making the matching statements too easy.

M. Check both the primary column and the response column for unwarranted clues.

V. Rules applicable to the writing of completion and short-answer type items:

A. Rules common to both item types.

1. The item should be stated in such a way that it can be answered with a word or short phrase.

2. The item should be phrased so that one or a very small number of answers can be correct.

3. The item should be phrased so that it will indicate the terms in which the response should be given.

4. The following clues that lead to the correct responses should be avoided.

a. Information that points too directly to the correct answer.

b. One item that contains the answer to another item.

c. Grammatical expressions which indicate the correct answer.

5. Blanks should be arranged so that answers will be in a column, preferably at the right of the item. Sufficient space should be allowed for examinees to write their response.

B. Short answer:

1. Use a direct question.
2. Items measuring knowledge of definitions should give the term and require the definition.

C. Completion:

1. Omitted words or phrases should be at or near the end of the statement.
2. A statement should not be mutilated until the meaning is almost lost. Only significant words should be omitted.
3. Enough information should be supplied to make the meaning clear.

Examples of Errors in Item Writing

The illustration and discussion of typical violations of the suggested item writing rules will demonstrate the application of the rules to a few actual items. The method used in this study for determining frequency of violations of item writing rules was used in the analysis of the following illustrative items.

The first item, submitted by a member of the sample cooperating in the study, was selected from a tenth

grade biology test labeled "unit test over plants."

"The Fungi are a group of Thallophytes lacking chlorophyll that count among their number the lichens. They are a prime example of symbiosis and are made up of?

1. Mushrooms and Molds.
2. Mildews and Rusts.
3. Smuts and Blights.
4. Yeasts and Bacteria.
5. Algae and Fungus."

Twelve violations of the suggested rules are illustrated in this one item. Beginning with the general rules: rule I-A is violated by the complex sentence structure; rule I-B is violated because the antecedent of "they" is indefinite. Review of the textbook used by this teacher for the course revealed that this statement was copied directly from the textbook, therefore, rule I-D is violated. In view of the ambiguity of the item stem, it is difficult to determine what knowledge is called for, but if the problem is merely to select an example of symbiosis, it is knowledge which appears to have little apparent practicality for the student, therefore, rule I-E is violated. There is one definite clue to the correct response; students would be quick to recognize the singular of fungi found in the fifth alternative. This clue to the correct response violates rule I-H. The rules violated that are applicable to the multiple-choice item type specifically are rules II-A-1, II-A-3, II-B-1, II-B-2, II-B-3, and II-C-5. The length of the

item stem has little or no relation to the purpose of the item. If a definition of lichens is the purpose of the item, it would be preferred to state alternative definitions and have the respondent select the correct definition. The wording of the problem is not clear, nor does it present a single central idea. The stem is stated positively and then is placed in the interrogative form through the use of the question mark. Finally, the use of the word Fungi in the stem and its singular in the correct alternative repeats the violation of rule I-H. Such a violation would be counted only once in the frequency of errors in the item.

Alternative response items are perhaps the easiest items in which to commit violations of commonly accepted rules for objective item writing. The following items were selected from examination items developed to accompany a college level educational psychology textbook.

In the item quoted below, the respondents must make a decision as to the applicability of the accuracy of the range of report as well as the applicability of the accuracy of report. The inclusion of two ideas is a violation of rule III-B.

"Both range and accuracy of report are greater in the narrative form of report than in the interrogatory form."

The following item is an example of a violation of rule III-C. The qualification necessary to make the item

understandable produces some cause for confusion on the part of the respondent.

"Geometry courses designed primarily to develop everyday reasoning attain as good results on the subject matter of geometry as traditional courses."

The student who is truly "test-wise" would look forward to responding to a complete test composed of items with specific determiners that clue the correct response as does the following item. This item also violates rule III-D, since it is a broad generalization.

"Some learning takes place without activity."

From the three items quoted, it is apparent that training in objective test item writing is necessary before the item writer can develop skill which is demanded by the task. Although the items illustrated were developed as a supplement to a textbook, it is obvious that the items were not carefully developed by a skilled item writer. The book is widely used in college level courses. However, if the items are accepted for use without correction, resulting scores will misrepresent actual student achievement.

The following items were selected from a sixth-grade science test. Thorough review of the test shows that the items selected as an illustration are a good sample of the entire test.

"Matching:

- | | |
|-----------------------------------|--|
| 12. _____ algae | a. Plants that lack chlorophyll |
| 13. _____ invertibrates [sic] | b. animal with backbone |
| 14. _____ protozoa | c. Rellular [sic] specialization is found in this |
| 15. _____ microscope | d. Coral is a member of what group of animals? |
| 16. _____ coelenterate | e. Phyla refers to |
| 17. _____ grouping or classifying | f. simple plant containing chlorophyll |
| 19. _____ sponges | g. substance man has extracted from various kinds of molds |
| 20. _____ vertibrate [sic] | h. animal without backbone |
| 21. _____ antibiotics | i. Name means "first animals" |
| | j. Anthony van Leeuwenhock [sic] |

The items were examined for violations of the rules for writing matching items. The directions do not specify the intended basis for matching. Secondly; the response column violated rule IV-B, since the responses are heterogeneous rather than homogeneous. Thirdly; placing of problem statements on the left and the response column on the right side of the page would have enabled the respondents to consider each problem before considering the alternatives. Rule IV-E states that the number of items in the response column should always exceed the number of items in the primary column. This rule is violated since there is equal matching in the item. Items in the response column do not appear to have any system in their arrangement which violates rule IV-G. Problem "b" refers to an animal with backbone which can only be matched with vertebrate.

This becomes an unwarranted clue for other items in this exercise. The intended response to problem "h" is invertebrates; however, the wording of the stem in the singular and the response as plural is misleading. Also, there are two correct responses to the problem: invertebrates and coelenterates. There are many unwarranted clues in the exercise; for example, problem "g" deals with a substance and the only substance in the response column is antibiotics. Grades were assigned on the basis of student response to the items in this test, but the accuracy of the assigned grades is questionable.

Short-answer items have appeared in measurement literature under varying titles. Some authors refer to the short answer type as simple recall items. This is not an appropriate title under some circumstances. For example; a short answer item may be worded so that the respondent is required to establish relationships involving mental skills of a level higher than simple recall. Other authorities frequently employ the term "sentence completion" which is not always appropriate. When a statement has omitted words such terminology would be appropriate, however, some short answer items are questions that, do in fact, require as the response a phrase or a few words.

The violation of rules applicable to the completion type is illustrated by the classic example

below. This statement has been mutilated to the extent that the respondent would be unable to locate the general problem of the item, therefore, response would be difficult, if at all possible.

"A hay _____ affords another _____ of the
 _____ existing between _____ and
 _____." ¹

Effect of Item-Writing Errors on Examinee Performance

Experimental evidence reflecting differences in examinee performance resulting from violations of suggested objective item-writing rules was lacking. Therefore, it seemed appropriate to obtain data reflecting the efficacy of the application of the suggested rules. In order to obtain experimental evidence, two forms of a test covering materials presented as one unit of an introductory course to education, were developed. One form was developed in which care was exercised in order to avoid violation of the rules for objective examination item writing. This form of the test, included as Appendix II, was called form "A." A second form of the test was developed in which numerous errors of item writing rules were committed. This form, included as Appendix III, was labelled form "B."

¹Robert L. Ebel, "Writing the Test Item," Chapter VI in Educational Measurement, E. F. Lindquist, ed., (Washington, D. C.: American Council on Education, 1951), p. 228.

Two hundred seventy-four students enrolled in a closed circuit television course at the state college of Iowa were selected for the purpose of comparing the two forms of the test. The course, The Teacher and the Child, is an introduction to education, open to students who are interested in following the teacher education program. All students who desired to take The Teacher and the Child course during the semester were enrolled. No other section of the course was available. Instruction was received via closed circuit television in five separate classrooms. Students were assigned to a particular classroom section at the beginning of the semester alphabetically. That is, beginning with students whose surname began with A, assignments were made to the same classroom until the capacity of that room was reached. Beginning with the first name on the roll following the last student assigned to the first classroom, the next and succeeding rooms were filled until all students were assigned.

Tests were randomly distributed in each classroom. Alternate students received form A while the remainder received form B. Those receiving form A were designated as group II, while those receiving form B were designated as group I. The mean and standard deviation of the performance on the test for each group is reported in Table 1.

TABLE 1. -- COMPARISON OF STUDENT PERFORMANCE
ON TWO FORMS OF A TEST OVER A UNIT IN AN
INTRODUCTORY EDUCATION COURSE

	No.	Mean	Standard deviation
Group I	137	71.3	5.64
Group II	137	65.8	5.51

A hypothesis of no significant difference between the means was tested by the "t" test for independent groups. The formula is:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{\sum x_1^2 - \sum x_2^2}{N_1 + N_2 - 2} \right] \left[\frac{1}{N_1} + \frac{1}{N_2} \right]}}$$

where \bar{X}_1 = mean of group I scores.

\bar{X}_2 = mean of group II scores

$\sum x_1^2$ = variance of group I scores

$\sum x_2^2$ = variance of group II scores

N = number of cases in each group

The obtained value of t was 8.209 (P 0.01) which exceeds the interpolated table value of t for 272 degrees of freedom (t = 2.594). This suggests that there was a real difference in the performance of the two groups of students on the two forms of the test. The mean performance on form B (\bar{X} = 71.3), containing numerous errors of objective examination item writing rules, was greater than the mean

performance on form A ($X = 65.8$). The conclusion reached from this evidence was that violation of the rules in writing objective examination items caused an overestimate of examinee performance.

Consistency Among Evaluators

Subjective judgments, on the part of the individual evaluating an item enters into the comparison of an item with item writing rules. Since this judgment may vary from one evaluator to another, it was necessary to determine the degree of consistency between the investigator and other evaluators.

In order to compare the frequency of violation of item writing rules counted by more than one evaluator, the same items must be analyzed. Items submitted by teachers cooperating in the study were selected and duplicated for this purpose. Five tests were randomly selected from among all tests submitted for the study. Ten items were then selected from each test by using a table of random numbers. Thus, fifty items were duplicated in a form identical to the original submitted items. A copy of the items is included as Appendix IV.

Three authorities, trained and experienced in the field of examination construction and evaluation, were requested to evaluate the fifty items for violations

of the suggested objective examination item writing rules.¹ They were asked to record each violation of the item writing rules found in each of the fifty items. When their summary sheet was returned, the total frequency of item writing errors for each of the five tests was tabulated. Table 2 shows the number of errors reported by each evaluator for each test. The number of violations of the item writing rules counted by the investigator was entered in Table 2 for comparison.

Chi square was used to test the statistically significant differences of the categorical data. The formula employed was

$$\chi^2 = \sum \left[\frac{(o-e)^2}{e} \right]$$

where o = the observed frequency of errors
and e = the expected frequency of errors.

Marginal totals were used for determining expected cell frequencies for the total number of errors counted and a chi square value of 3.301 (P = .99) was obtained. The number of errors of the general item writing

¹The three members of the State College of Iowa staff who provided the data for the inter-evaluator study were Dr. E. C. Denny, Dr. Paul Kelso, and Dr. Herbert Silvey. Dr. Denny, Professor of Education, Emeritus, is co-author of the Nelson-Denny Reading Test and the Nelson-Denny-Brown revision of the Nelson-Denny Reading Test. Dr. Kelso, Coordinator of Student Counseling, is co-author of the Kelso-Lamke revision of the Henmon-Nelson Tests of Mental Ability. Dr. Herbert M. Silvey, Director of Examinations and Research, is the author of Masters Theses in Education.

TABLE 2.--VIOLATIONS OF OBJECTIVE EXAMINATION ITEM WRITING RULES OBSERVED IN FIFTY TEACHER-MADE TEST ITEMS BY FOUR INDEPENDENT EVALUATORS

	Types of Errors										Sub-total		Total
Test Number	1		2		3		4		5				
Evaluator	General	Specific	General	Specific	General	Specific	General	Specific	General	Specific	General	Specific	244 264 237 250
A	13	31	28	16	17	29	30	17	30	28	118	131	
B	20	26	17	24	17	31	32	24	34	39	120	144	
C	18	30	23	19	16	24	28	16	26	37	111	126	
D (Investigator)	18	30	15	24	14	26	25	30	30	38	102	148	
Sub-total	69	117	83	83	64	110	115	87	120	152	451	549	
Total	186		166		174		202		272		1000		1000

rules reported by each evaluator was similarly tested. The obtained value of chi square was 7.208 ($P = .81$). Testing for statistically significant differences between evaluators based on the number of specific errors of the item writing rules resulted in a chi square value of 9.971 ($P = .68$). These results indicate that when different evaluators analyze the same items, the variation in number of item writing errors counted is too small to reduce the accuracy of the item evaluation.

The observed discrepancies of reported errors were due largely to the counting of a general rule error twice when a similar specific item type rule was violated. For example, a general rule requires that an item writer should make directions and instructions clear and complete, while a rule for constructing matching items states that directions should clearly explain the intended basis for matching. Thus, when such an error occurred in the same item, the same error could be counted twice by an evaluator. In counting the errors for this study, however, such errors were counted only once.

Consistency Between Evaluations

The data for this study were derived by determining the number of item writing rule violations found in tests submitted by members of the sample. Since this process involves subjective judgment, some evidence of evaluator

consistency was required.

This evidence was provided by evaluating one hundred items on one occasion and recording the error frequency for each item by the investigator. Three months later, the same items were evaluated and the error frequency was recorded. Table 3 shows the number of errors counted on both occasions.

TABLE 3.--NUMBER OF ERRORS OF OBJECTIVE ITEM WRITING RULES REPORTED FOR THE SAME ITEMS ON TWO DIFFERENT OCCASIONS

No. Items	First Analysis	Second Analysis	Total
100	258	290	548
Value of chi square = 1.868 (P = .13)			

The statistical significance of the difference was computed by use of the chi square test for correlated data. The obtained value of chi square was 1.868 (P = .13). Since the observed differences were not statistically significant, it seemed reasonable to conclude that the consistency of the evaluator was adequate for the purposes of this study.

The observed discrepancies were individually compared. The primary differences were errors in punctuation and spacing that had been overlooked in the first evaluation.

These errors were randomly distributed throughout the one hundred items which indicated that such errors would have a tendency to cancel observed differences when different classes of data were compared.

CHAPTER III

SAMPLING TECHNIQUES AND PROCEDURES

Sample

The data for this study were derived from teacher-made test items submitted by a selected sample of natural and social science teachers of Iowa public four-year high schools. The selection of natural and social science teachers as the sample for the comparison of two subject matter fields was based on information contained in the Iowa Educational Directory.¹ An examination of the Directory showed that all public four-year high schools in Iowa offer subjects in both natural and social science. A preliminary survey indicated that teachers of these two subject matter areas generally employed objective type teacher-made tests for student evaluation.

Some major obstacles were encountered in obtaining tests from teachers sampled: first, teachers generally regarded tests as intimate possessions which should be

¹State Department of Public Instruction, Iowa Educational Directory. (Des Moines, Iowa: State of Iowa, 1960), p. 267.

handled discretely; secondly, teachers traditionally used tests during their busiest work periods--near the close of a pupil-progress reporting period. Therefore, a request for copies of their teacher-made tests must necessarily be made at, or near, a time when teachers were most sedulously engaged; consequently, a reduced return would be expected. Finally, since tests were used for purposes of student evaluation, teachers were cautious about who was given access to their examinations.

In an attempt to reduce these problems, a letter of request was developed in which the purpose of the study was stated. The letter pointed out the precautions that would be followed in order to insure that the tests would not be available to high school students. Plans for keeping the teachers personal information and test items separate, but identifiable, were stated. The letter is included as Appendix V.

Size of Sample

The number of teachers included in the sample was dependent upon several factors. Among these, was the purpose of the study, type of study, and the method of sample selection.

In order to satisfy the requirements imposed by the purpose of the study, the sample must be representative of the total parent population since the study is a descriptive survey. The stratified random sampling

technique was selected in order to provide representativeness and to insure randomness. Stratification was based on high school enrollment and subject matter teaching fields. This insured proportionate numbers of natural and social science teachers from each school enrollment classification in the sample.

The total 1959-60 population of public four-year high school teachers in Iowa was 8,318. Of these, 1,500 were listed in the Iowa Educational Directory as teachers of natural and social science.¹ The degree of accuracy desired in the sample was arbitrarily set at the 95% confidence interval. The sample size for a reliability of $\pm 5\%$ was then determined, using a Table of Sample Size Required for Finite Population, for Confidence Limits and Specified Reliability Limits in Sampling Attributes in Per Cent.² Since this table was based on a proportion of 50% it was necessary to adjust the sample size for proportion of 0.91. This proportion was determined through a preliminary survey of teacher-made tests. Using the formula

$$4 [p(1-p)]$$

where p is the proportion of natural and social science teachers in the population who committed violations of the

¹Ibid.

²Herbert Arkin and Raymond R. Colton, Tables for Statisticians, (New York: Barnes and Noble, Inc., 1950), p. 136.

suggested rules for writing objective test items. This technique indicated that 120 teachers were required for the sample.

Sampling Procedure

The teachers who were requested to furnish items from which the data were obtained for this study, were selected in proportion to the total distribution of high school teachers in the public four-year high schools of Iowa. The classification of high schools and the number of teachers as reported in the November 1960 Educational Bulletin was used to determine the percentage of teachers from each school classified according to high school enrollment.¹

Table 4 shows the classification of high schools according to pupil enrollment, total number of districts in each class, total number of high school teachers in these districts and the percentages of the total number of high school teachers represented in each school district classification. Since it was assumed that all schools offered courses in both subject matter fields and that the percentages of natural and social science teachers were distributed in approximately the same proportion, the sample distribution was determined by using the percentages

¹State Department of Public Instruction, Educational Bulletin. (Des Moines, Iowa: State of Iowa, 1960), November 1960.

in Table 4.

TABLE 4. -- DISTRIBUTION OF TEACHERS IN IOWA
PUBLIC HIGH SCHOOLS, 1960-1961

High School Enrollment	Number of Districts	Number of Teachers	Percentage* of Teachers
25 - 49	32	156	1.9
50 - 74	72	435	5.2
75 - 99	73	550	6.6
100 -149	137	1280	15.3
150 -199	72	845	10.0
200 -299	81	1196	14.5
300 -399	31	614	7.4
400 -499	18	423	5.0
500 -599	18	479	6.0
600 -above	28	2340	28.0
TOTALS	562	8318	100.0

*Percentages rounded to nearest one-tenth

It was necessary to adjust the number of teachers included in the sample in some of the strata because of the fractional proportions resulting from uneven percentages. Table 5 shows the number and percentage of teachers in the sample from each school district enrollment classification.

The names of all high school teachers whose teaching assignment for the 1960-61 school year was listed as social studies or a combination including social studies, and those whose assignment was listed as natural sciences or a combination including natural sciences, were obtained from the Educational Directory with complete school address and enrollment classification

of the school.¹ These were typed on cards. The cards were then grouped according to the enrollment classification of high schools. All cards representing teachers of social sciences for a particular school district classification were placed in a box and thoroughly mixed. Name cards were then drawn at random until the appropriate number for that school district classification was obtained. This procedure was repeated for each of the different school district classifications and for each different classification of natural science teachers.

TABLE 5. -- DISTRIBUTION OF TEACHERS IN THE SAMPLE
ACCORDING TO SCHOOL DISTRICT CLASSIFICATION

High School Enrollment	No. of Natural Science	No. of Social Science	Percentage of Sample*
25 - 49	1	1	1.7
50 - 74	5	5	8.3
75 - 99	6	6	10.0
100 - 149	8	8	13.3
150 - 199	6	6	10.0
200 - 299	10	10	16.7
300 - 399	4	4	6.7
400 - 499	3	3	5.0
500 - 599	3	3	5.0
600 - above	14	14	23.3
TOTALS	60	60	100.0

*Percentages rounded to nearest one-tenth.

¹Loc. cit., Iowa Educational Directory.

The school in which each member of the selected sample was teaching was identified on a map of the State. An examination of the map provided a quick visual check on the distribution of members in the selected sample. Requests with the accompanying personal information questionnaire and a stamped return envelope were mailed to each member of the sample.

Questionnaire

The central purpose of the letter of request to members in the sample was to secure objective test items for analysis. In order to appropriately classify each response into the major categories basic to the study, it was necessary to include a short questionnaire. Basic information needed for classification of the responses into the following categories was: (1) teachers of natural or social science, (2) those having had a course or unit in objective test item construction and those without a course or unit, and (3) number of years of teaching experience. In addition to these categories, additional factors were considered. Those factors that might influence the objective test item writing skill of members of the sample were incorporated into the questionnaire. These were: (1) college subject-matter major, (2) college subject-matter minor, (3) name of the college that offered the course or unit in which objective item writing was presented,

(4) year when the course was completed, (5) sex, and (6) were the items that were submitted developed by the respondent? A preliminary form of the questionnaire including these items was developed.

Tryout

Before the final form was prepared for use in the study, the preliminary form was given to fifty graduate students. They were asked to respond to the questions, however, opportunity was given for them to ask questions in order to clarify items in the questionnaire that were not clear to them. The data thus obtained were classified according to categories basic to the study as well as additional categories which might influence item-writing skill. The questionnaire was revised, using the information obtained in the initial tryout to aid in its improvement. The questionnaire was then duplicated on edge-punched cards. The final form of the questionnaire is included in Appendix VI.

Validity

The validity of the questionnaire was judged on the following types of evidence:

1. Is the question relevant?
2. Is the question perfectly clear?
3. Is the response consistent with what is known?
4. Is the item sufficiently inclusive?

5. Can an external criterion be found for
evaluation of the questionnaire?¹

A careful analysis of the tryout responses had provided information which enabled an affirmative response to questions one, two, and four. It was necessary to obtain information from other sources in order to validate items three and five. A sample of responses was randomly selected. The administrative official of the school in which each of the members of the sample was teaching was personally contacted. After an explanation of the purpose of the study was given and with permission of the teacher, each respondent's personal file was examined and compared with responses on the questionnaire. The cooperation and attitudes of teachers sampled and their administrative officers were excellent. This procedure provided evidence for an affirmative reply to question three. The response to question five was answered affirmatively since responses to the questionnaire could be compared with manuals, study guides, course syllabi, stated reference materials, and college statements of course content. Since each of the types of evidence for judging the validity of the questionnaire had been provided, the questionnaire was judged valid for the intended purpose.

¹Douglas E. Scates and Alice V. Yeomans, The Effect of Questionnaire Form on Course Requests of Employed Adults, quoted in Carter V. Good, Introduction to Educational Research (New York: Appleton-Century-Crofts, Inc., 1959), p. 201.

Responses to Requests

Fifty-seven responses, representing 47 per cent of the total number of teachers contacted in the original mailing were received. Three questions should be considered in evaluation of the return. First, are the returns representative of Iowa social and natural science teachers? Secondly, are the school classifications represented in the returns? Finally, are the returns proportionately distributed from the various areas of the state?

One hundred and two schools were represented by the 120 teachers constituting the original sample. Forty-nine schools were represented in the responses. Social science teachers submitted 27 of the 57 responses, which constituted 47 per cent of the total original responses. Table 6 compares the sample distribution and the distribution of responses. Geographic areas of the state, as they appear in Table 6, were determined by using the two major highway systems to divide the state into four quadrants. The close relationship existing between the percentage of teachers in the total sample and the percentage of teachers responding, shows that the returns are representative of the distribution of teachers in the original sample population. A visual comparison of the percentage of schools represented in the original sample with those represented in the returns showed a high degree of agreement. Therefore, it was

concluded that the original returns were representative of the geographical distribution of teachers in the sample.

TABLE 6. -- A COMPARISON OF THE GEOGRAPHIC REPRESENTATION OF RETURNS WITH THE ORIGINAL SAMPLE POPULATION DISTRIBUTION*

Area	TEACHERS			SCHOOLS		
	No. in Sample	No. of Returns	Per Cent Returns	No. in Sample	No. of Returns	Per Cent Returns
NE	37	16	27	29	12	25
SE	35	17	30	31	19	38
SW	26	15	27	20	10	20
NW	22	9	16	22	8	17
TOTAL	120	57	100	102	49	100

*Percentages have been rounded to equal 100%.

Table 7 shows the number of returns from each school district classification. The only classification that was not represented in the returns was in the area of social sciences from schools having an enrollment between 25 and 49. All other classifications were represented by at least one return in each subject-matter area. Therefore, it was concluded that the original returns were representative of each school district enrollment classification.

Although the original returns were representative of the geographical distribution of teachers and schools as well as the pupil-enrollment classification of school districts, the number of responses appeared insufficient for the purposes of the study. A control that would reduce

the possibility of bias caused by the high non-response rate was necessary.

TABLE 7. -- RESPONSES BY SCHOOL DISTRICT
CLASSIFICATION AND SUBJECT-MATTER AREA

Teachers of	Social Science		Natural Science	
Class of District	No. in Sample	No. of Returns	No. in Sample	No. of Returns
25 - 49	1	0	1	1
50 - 74	5	3	5	2
75 - 99	6	3	6	2
100 - 149	8	2	8	2
150 - 199	6	2	6	4
200 - 299	10	6	10	4
300 - 399	4	2	4	2
400 - 499	3	1	3	2
500 - 599	3	1	3	2
600 - above	14	7	14	9
TOTAL	60	27	60	30

The deficiencies of the high non-response rate were overcome by sending a second request to all non-respondents in the original sample in order to obtain additional returns. The responses on these returns could then be compared with the original responses and if no differences existed, the evidence would support the contention that no bias was introduced by the large non-response rate.¹

¹Research Design and Analysis, Second Annual Phi Delta Kappa Symposium on Educational Research, (Bloomington, Indiana: 1961), p. 91.

The second mailing consisted of 63 requests. Thirty-one responses were received. Tables 8 and 9 show the distribution of returns from the second mailing.

TABLE 8. --COMPARISON OF THE GEOGRAPHIC REPRESENTATION OF THE SECOND MAILING RETURNS WITH THE ORIGINAL SAMPLE POPULATION DISTRIBUTION*

Area	TEACHERS			SCHOOLS		
	No. in Sample	No. of Returns	Per Cent Returns	No. in Sample	No. of Returns	Per cent Returns
NE	37	13	42	29	12	42
SE	35	9	29	31	5	17
SW	26	3	10	20	5	17
NW	22	6	19	22	7	24
Total	120	31	100	102	29	100

*Percentages have been rounded to total 100 per cent.

TABLE 9. --SECOND MAILING RESPONSES BY SCHOOL DISTRICT CLASSIFICATION AND SUBJECT MATTER AREA

Class of District	SOCIAL SCIENCE		NATURAL SCIENCE	
	No. in Sample	No. of Returns	No. in Sample	No. of Returns
25 - 49	1	1	0	0
50 - 74	2	1	3	1
75 - 99	3	1	4	2
100 - 149	6	2	6	4
150 - 199	4	3	2	0
200 - 299	4	2	6	4
300 - 399	2	1	2	1
400 - 499	2	1	1	0
500 - 599	2	1	1	0
600 - above	7	4	5	2
Total	33	17	30	14

The returns from the second mailing were then compared on each of the variables with the original sample returns. The "z" test for independent sample proportions was used. The formula is

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2}}}$$

where: p_1 = proportion of the attribute in the original sample returns,

p_2 = proportion of the attribute in the returns from the second mailing,

q_1 = $1 - p_1$,

q_2 = $1 - p_2$,

N_1 = total number of returns in the responses to the original mailing,

N_2 = total number of returns in the responses to the second mailing.

A comparison of the number of respondents in the original returns with the number of respondents in the returns of the second mailing on the basis of course training was made. The computed value of "z" was 0.197 which was not significant at the 0.01 level. The hypothesis of no significant difference was accepted. Comparing the number in each return on the basis of subject matter field yielded a "z" value of 0.258 which

is not significant at the 0.01 level of significance. Therefore, the hypothesis of no significant difference was accepted. When the number in the second mailing returns was compared with the number in the first mailing returns on the basis of years of teaching experience, the resulting value of "t" was 0.367. The hypothesis of no significant difference was accepted at the 0.01 level. Table 10 shows the summary of these comparisons.

TABLE 10. --COMPARISON OF THE NUMBER OF FIRST MAILING RETURNS WITH THE NUMBER OF SECOND MAILING RETURNS ON THE BASIS OF COURSE TRAINING, SUBJECT MATTER TEACHING AREA, AND TEACHING EXPERIENCE

Variables	First Mailing Returns		Second Mailing Returns		Value of t
	No.	%	No.	%	
Course Trained	33	(63.46)	19	(61.29)	0.197 ^a
Not Course Trained	19	(36.54)	12	(38.71)	
Social Science	27	(51.92)	17	(54.84)	0.258 ^a
Natural Science	25	(48.08)	14	(45.16)	
Experienced	29	(55.77)	16	(51.63)	0.367 ^a
Limited experience	23	(44.23)	15	(48.37)	

^aNot significant.

Before the returns could be pooled, it was necessary to determine if there was a difference in the data obtained

from the two sets of returns. Therefore, the number of errors committed by the teachers in the two sets of returns were counted and then compared by means of the chi square test. The null hypothesis for this purpose was a proportionate division of errors. The proportion of the total number of respondents in each set of returns was applied to the total number of errors to determine the expected number of errors for each category tested.

The number of errors committed by members of the original response group was compared with the number of errors committed by members of the second-mailing response group on the basis of course training. The obtained value of chi square was 2.296 ($P = 0.14$). The hypothesis of no statistically significant difference between the number of errors committed by members of the two sets of returns when compared on the basis of course training was accepted.

The number of item writing rule violations committed by teachers in the first- and second-mailing returns were compared on the basis of the subject-matter teaching field. The computed value of chi square was 0.307 ($P = 0.60$), which does not reach the 0.05 level of significance. Therefore, the hypothesis of no statistically significant difference between the two return groups on the basis of the number of errors committed by teachers of

different subject matter teaching fields was accepted.

The number of errors found in examinations submitted by the first- and second-mailing respondents was compared on the basis of teaching experience. The value of chi square resulting from this comparison was 2.793 ($P = 0.10$). This result indicated that there was no statistically significant difference in the number of violations of item writing rules committed by members of the first- and second-mailing respondents when compared on the basis of teaching experience. A summary of these comparisons is presented in Table 11.

Since none of the comparisons made between the two response groups were significantly different, it was concluded that the two groups were not significantly different on the basis of the number of violations of the rules for writing objective examination items.

Since none of the comparisons of the number of responses to the two mailings resulted in a significant "t", and none of the comparisons between the number of errors committed by respondents to the two mailings resulted in a significant chi square, it was concluded that the two samples were not significantly different. This evidence supports the contention that the high non-response rate did not introduce a bias into the returns and the data from the two sets of returns were pooled. The results of the distribution of returns from the State appear as shown in Tables 12 and 13.

TABLE 11. --COMPARISON OF THE NUMBER OF ERRORS FOUND IN TESTS SUBMITTED BY TEACHERS RESPONDING TO THE FIRST MAILING REQUEST WITH THE NUMBER OF ERRORS FOUND IN TESTS SUBMITTED BY TEACHERS RESPONDING TO THE SECOND MAILING REQUEST

Basis for Comparison	Errors in Original Returns	Errors in Second Returns	Value of Chi Square	Proba- bility
Course trained	945	499		
Not course trained	377	231	2.296	0.14
Social Science	679	384		
Natural Science	643	346	0.307	0.60
Experienced	755	389		
Limited experience	567	341	2.793	0.10

TABLE 12. --TOTAL RESPONSES BY SCHOOL DISTRICT AND SUBJECT MATTER CLASSIFICATION

Class of District	SOCIAL SCIENCE		NATURAL SCIENCE	
	No. in Sample	No. in Returns	No. in Sample	No. in Returns
25 - 49	1	1	1	1
50 - 74	5	4	5	3
75 - 99	6	4	6	4
100 - 149	8	4	8	6
150 - 199	6	5	6	4
200 - 299	10	8	10	8
300 - 399	4	3	4	3
400 - 499	3	2	3	2
500 - 599	3	2	3	2
600 - above	14	11	14	11
Totals	60	44	60	44

TABLE 13. --TOTAL RESPONSES BY GEOGRAPHIC DISTRIBUTION *

Area	TEACHERS			SCHOOLS		
	No. in Sample	No. of Returns	Per cent Returns	No. in Sample	No. of Returns	Per cent Returns
NE	37	29	33	29	24	31
SE	35	26	30	31	24	31
SW	26	18	20	20	15	19
NW	22	15	17	22	15	19
Totals	120	88	100	102	78	100

*Percentages have been rounded to equal 100 per cent.

The pooled returns were 73 per cent of the original sample. Inspection of Tables 12 and 13 reveals that the requirements imposed on the sample for representativeness of returns was satisfied. The pooled sample consisted of 88 returns which were equally divided between the two subject matter teaching areas selected for this study. Seventy-eight different school systems were represented in the pooled returns.

Selection of Items for Evaluation

Many of the teachers cooperating in the study submitted a complete test or, in many instances, several tests, and five members of the sample did not submit any items for the study. The problem of insuring equal representation for each cooperating teacher was met by evaluation of only 10 items from the total number submitted by each teacher. The selection of the particular items

for evaluation was accomplished through the use of a table of random numbers.¹ Where a teacher had submitted more than one test, each test was assigned a two-digit numeral and the test from which the items were selected for evaluation was determined through the use of a table of random numbers. In some cases the test submitted consisted of several parts, each beginning with the same item number. The selection of the part of the test from which to select the item was also determined through the use of a table of random numbers.

The items were then evaluated against the rules for item writing and the number of each type of error was recorded. The total number of errors for each item writing rule violated was recorded on the reverse side of the Personal Information Questionnaire. The recording of errors on the reverse side of the questionnaire provided for quick assimilation of the total number of errors for any particular category. Data were then tabulated and subjected to analysis.

¹E. F. Lindquist, Design and Analysis of Experiments in Education and Psychology, (Boston: Houghton-Mifflin Company, 1956), pp. 385-387.

CHAPTER IV

ANALYSIS OF DATA AND CONCLUSIONS

An alphabetical listing of all members of the sample was compiled and as responses were received, the Personal Information Questionnaire and the objective test items were assigned a code number. The prefix "S" with a numeral was the code used for identifying responses of social science teachers. A similar identification system, using the prefix "N" was used to identify responses of natural science teachers.

The Personal Information Questionnaire was edge-punched for each classification of information. The identifying code for all classes of information is included as Appendix VII. When the number of item writing errors for each test had been determined, this information was recorded on the reverse side of the questionnaire. Recording the errors on the reverse side of the questionnaire made it possible to rapidly compile the number of errors found in each classification of data. The collection of all questionnaires in any particular classification of data was accomplished by the key-sort technique.

Data were then subjected to tests of significance. The scale underlying the data was categorical and discontinuous. The chi square test was selected to test the hypotheses since it is appropriate for categorical and discontinuous data. The data were arranged in two by two contingency tables and marginal totals were used to determine expected cell frequencies and to obtain the relevant statistical tests of significance. These data are presented in Appendix VIII.

The number of errors was tested for statistically significant difference between subject matter teaching fields and course training. The computed value of chi square was 6.824 ($P = 0.006$) which is greater than the 0.01 level of significance. To determine which of the classifications of errors caused the observed difference, the number of errors was further tested for significant differences between subject matter teaching fields and teaching experience. The computed value of chi square was 0.388 ($P = 0.60$). Since this was not significant, the significance of difference between teaching experience and course training was tested. The value of chi square was 3.790 ($P = 0.06$) which was not significant. The results of these comparisons are summarized in Table 14.

Since the classifications of data, when tested in a contingency table did not indicate the cause for the observed discrepancy, it was necessary to test each of the

two variables individually. To do this, the critical ratio¹ for independent proportions was computed.

TABLE 14. --COMPARISON OF THE NUMBER OF ERRORS OF ITEM-WRITING RULES FOUND IN TESTS SUBMITTED BY MEMBERS OF THE SAMPLE

Basis of Comparisons	Value of Chi Square	Level of Probability
Subject Matter Teaching Fields vs Course Training	6.824	0.006
Subject Matter Teaching Fields vs Teaching Experience	0.388	0.60
Course Training vs Teaching Experience	3.790	0.06

When the proportion of errors committed by social science teachers was compared with the proportion of errors committed by natural science teachers, the computed value of the critical ratio was 1.644 which was not significant. However, when the proportion of objective examination item writing errors committed by course trained teachers was tested against the proportion of errors committed by teachers who were not course trained, the obtained value of the critical ratio was 18.603. This critical ratio was

¹Quinn McNemar, Psychological Statistics. (New York: John Wiley and Sons, Inc., 1955), pp. 60-61.

significant beyond the 0.001 level of significance. These comparisons are summarized in Table 15.

TABLE 15. --COMPARISON OF THE PROPORTION OF ERRORS OF ITEM-WRITING RULES FOUND IN TESTS SUBMITTED BY MEMBERS OF THE SAMPLE

Basis of Comparisons	Value of Critical Ratio	Level of Probability
Social Science Teachers vs Natural Science Teachers	1.644	Not Significant
Course Trained Teachers vs Teachers Not Course Trained	18.603	0.001

During the course of the investigation it was observed that a large number of violations of rules treating the form and style of examination construction occurred. It seems logical that form and style are critical in examination construction, however, if these types of violations caused the observed discrepancy, a comparison with item writing rules in which form and style rules were omitted would be misleading. To avoid this possibility, a comparison of the number of violations with the form and style rule violations omitted from the total number of violations seemed warranted.

When each of the comparisons on the total number of errors was made where the number of errors with violations of form and style were omitted, the significance of the

results remained relatively unchanged. Table 16 shows these comparisons. The only noticable difference was observed when the proportion of errors committed by social science teachers was compared with the proportion of errors committed by natural science teachers. The value of the computed critical ratio was 2.845 which was significant at the 0.01 level of significance. The significance found between course trained teachers and teachers not course trained was unchanged since the level of significance extended beyond the 0.001 level. These data are summarized in Table 17.

TABLE 16. --COMPARISON OF THE NUMBER OF ERRORS OF ITEM-WRITING RULES WITH FORM AND STYLE RULE VIOLATIONS OMITTED

Basis of Comparisons	Value of Chi Square	Level of Probability
Subject Matter Teaching Fields vs Course Training	11.811	0.001
Subject Matter Teaching Fields vs Teaching Experience	1.159	0.20
Course Training vs Teaching Experience	2.494	0.10

TABLE 17. --COMPARISON OF PROPORTION OF ERRORS OF ITEM-WRITING RULES WITH FORM AND STYLE RULE VIOLATIONS OMITTED

Basis of Comparisons	Value of Critical Ratio	Level of Probability
Social Science Teachers vs Natural Science Teachers	2.845	0.01
Course Trained Teachers vs Teachers Not Course Trained	10.869	0.001

Additional Variables

Evidence that factors other than course training might cause the significant differences was present. Testing the most important of these factors appeared advisable. The factors included sex, number of years since a course in Tests and Measurements was completed, the teacher teaching in the subject matter area of major or minor preparation while in college, and the college attended. The results of these tests are summarized in Table 18.

A statistical test between the number of errors committed by the members of each sex and course training resulted in a value of chi square which was significant beyond the 0.01 level. A critical ratio computed from the proportion of errors committed by each of the sexes was 32.875 ($P = 0.001$). The critical ratio for course trained

TABLE 18. --COMPARISON OF THE NUMBER OF ERRORS COMMITTED BY MEMBERS OF THE SAMPLE ON THE BASIS OF COURSE TRAINING AND SEX, TIME ELAPSED, COLLEGE PREPARATION AND COLLEGE ATTENDED

Course Training vs	Value of Chi Square	Level of Probability
Sex.	19.510	0.001
Time Elapsed Since Course. . .	1.607	0.21
Field of College Preparation .	6.533	0.02
College Attended	19.357	0.001

and teachers not course trained was presented in Table 15. Testing against the number of errors committed by teachers on the basis of the years that have elapsed since a course or unit in Tests and Measurements was completed resulted in a value of chi square that was not significant. The comparison between course training and the number of errors committed by teachers on the basis of area of subject matter preparation when in college was significant. A critical ratio determined by the proportion of errors committed by teachers teaching in the area of their major subject-matter preparation while in college and the proportion of errors committed by teachers teaching in their minor area of subject matter preparation while in college was significant beyond the 0.001 level. The number of errors committed by teachers trained at various colleges and the number of errors found in tests submitted by course

trained teachers and teachers not course trained was tested. A significant difference was observed. The chi square value was 19.357 ($P = 0.001$). Critical ratios were computed for each college and university. These data are summarized in Table 19. All of the critical ratios were significant beyond the 0.05 level with the exception of University B. The critical ratio for University B was 3.465 ($P = 0.07$).

TABLE 19. --COMPARISON OF THE PROPORTION OF ERRORS COMMITTED BY MEMBERS OF THE SAMPLE ON THE BASIS OF COURSE TRAINING AND COLLEGE ATTENDED

Course Training vs	Value of Critical Ratio	Level of Probability
University A	7.960	0.01
University B	3.465	0.07
University C	6.195	0.014
College A	6.439	0.012
Other Iowa Colleges	12.320	0.001
Out-Of-State Colleges	7.210	0.01

Findings

The first hypothesis stated that there is no statistically significant difference between the number of objective examination item writing rules violated by course trained teachers and teachers not course trained. The obtained value of chi square was beyond the 0.001 level of

significance, therefore the hypothesis was rejected. A critical ratio computed on the basis of course trained teachers and teachers not course trained was also significant beyond the 0.001 level. The proportion of errors was greater for teachers who had been course trained. This indicated that teachers who have had course training committed more errors of item writing rules than do teachers who have not had course training. Further analysis of the data with errors or rules pertaining to form and style of examination item writing omitted and analysis of other variables substantiated the findings reported above.

The second hypothesis stated that there is no statistically significant difference between the number of objective examination item writing rules violated by experienced and limited experienced teachers. The results of statistical analysis indicated that there was no significant difference, therefore the hypothesis was accepted.

The third hypothesis stated that there is no statistically significant difference between the number of objective examination item writing rules violated by social science and natural science teachers. Comparisons made during the course of the analysis of the data did not result in any significant differences. Further analysis, with violations of rules pertaining to form and style of

objective examination item writing, omitted did not result in any statistically significant differences. Since there were no statistically significant differences between subject matter teaching fields, the hypothesis was accepted.

Conclusions

The major purpose of this study was to determine if teachers who had completed a course or unit in teacher-made test construction committed fewer errors in writing teacher-made test items than did teachers who were not course or unit trained. The results of statistical analysis of the data clearly indicate that a significant difference does exist between teachers who have been course or unit trained and teachers who have not been course or unit trained. Further analysis indicated that teachers who were course trained committed more errors than those who were not so trained. Two conclusions are possible from the results of this study. These are: (1) teachers have not been aware of the commonly accepted rules for writing objective examination items while in college courses and have not developed skill in their application, and (2) teachers were made aware of the commonly accepted rules, but did not develop skill in their application or if the skill was developed, it was not applied in the development of the items that were the source of data for this study.

Since teachers have full teaching schedules, it

is reasonable to conclude that even if teachers possessed adequate competency in writing objective examination items, that the pressures of time and a heavy teaching schedule could exert a negative influence on the application of the skill with which examination items are developed. A generalization that might be drawn from the results of this study is that teachers who have been course trained in writing teacher-made objective examination items are careless in developing their classroom examination items while those teachers who have not been trained in the area of objective examination construction are aware of their deficiencies and exercise caution in construction of their classroom examinations.

A further purpose of this investigation was to determine if teachers of one subject-matter teaching field committed more errors than did teachers teaching in another subject-matter field when writing objective examination items. Statistical analysis of the data did not produce any evidence to indicate that teachers of one subject-matter teaching field commit more errors in writing objective examinations than do teachers of another subject-matter field.

Finally, this study sought to determine if there was a difference in the number of violations of commonly accepted objective examination item writing rules committed by teachers who had three years or more of teaching

experience and those who had three years or less of teaching experience. The evidence developed during the course of this investigation did not indicate that experienced teachers committed more errors than did teachers who were of limited experience.

In terms of the results of the statistical analysis of the data it appears that the following theoretical considerations should be made: (1) since some of the criteria were concerned with a complete test; perhaps evaluation of complete examinations instead of randomly selected items would have provided more accurate and applicable data for the study, (2) examinations from other subject-matter areas would have provided more representative data which would have added needed credence to the findings of the study, and (3) handling the suggested rules as if they were of equal importance could have introduced a bias in the data which consequently could have influenced the conclusions reached in the study. Recommendations for future research in the following chapter are based principally on these theoretical considerations.

CHAPTER V

SUMMARY AND RECOMMENDATIONS

Summary

The purpose of this study was to determine if teachers who had completed a course or unit in teacher-made test construction committed fewer errors in writing teacher-made test items than did teachers who were not course or unit trained. A further purpose was to determine if the number of errors was affected by teaching experience and if teachers of one subject matter field committed more errors in writing teacher-made test items than did teachers of another subject matter teaching field.

Commonly accepted rules for writing objective test items were collected from available literature in the field of Tests and Measurements. These rules were then organized for the purposes of this study. Since experimental evidence that indicated the effect of item-writing rule violations in a test on student performance was not available, two forms of a test were developed and administered to two equated groups of students. The results indicated

that errors in writing objective examination items produced a statistically significant difference between the performance of the two groups.

The sample used in this study was composed of a selected sample of Iowa public high school teachers of natural and social sciences. The data used in the analysis were errors of the objective examination item writing rules found in tests submitted by each member of the sample. Ten items were randomly selected from among the items submitted by each member of the sample. These items were compared with the item writing rules for violations. When an item violated one of the rules, the violation was recorded. Since different judges might find differing numbers of errors, data which demonstrated the agreement among evaluators and between evaluations were presented.

The number of errors committed by each member of the sample was organized into the necessary groups for testing the stated hypotheses. The chi square test of independence was used to test the categories of data for significant differences.

On the basis of statistical tests of the data the following conclusions were reached:

1. There was a significant difference in the number of item writing rule violations committed by course trained teachers and teachers not course trained.

2. Teachers of social science and teachers of natural science showed no statistically significant difference between the number of objective examination item writing violations of the rules.

3. There was no statistically significant difference between the number of errors committed by teachers who were experienced and teachers whose experience was limited.

Recommendations

The following recommendations for future research became evident as a result of the findings presented:

1. Studies of this type should be extended to populations in other geographic areas. Such studies would add data that could be combined to determine answers to the stated hypotheses when applied to a general, wide-scale teaching population of the United States. Also, similar studies might compare errors committed by teachers of subject matter areas other than the areas investigated in this study.

2. The research reported in the literature on item writing rules does not provide experimental evidence for their use. Or, in many instances, the evidence is too out-dated to be applicable in modern classroom testing practice. The rules collected and organized for

the purposes of this study should be individually subjected to experimental investigation in order to develop evidence of their applicability under present-day teaching and testing conditions. Such a study might incorporate violations of one general area of the rules in a test and develop another form of the same test in which the errors have been eliminated. These two tests could be administered to different sample populations and item statistics used to determine experimentally the value of the rules as a contributor to good item statistics.

3. There seems to be a general need in the field of Tests and Measurements to determine the nature and content of various courses in which the development of skills in writing the teacher-made test should logically be learned. Content research in Tests and Measurements is lacking or is obsolescent. Research in the content area would enable teachers of Tests and Measurements courses to strengthen areas of weaknesses as disclosed by experimental evidence.

4. Research has been quoted in which it was disclosed that a great portion of a pupil's grade is determined through teacher-made tests. Therefore, teachers should develop skill in construction of the teacher-made test. Results of this study clearly demonstrated that the classroom teachers in the sample violated the commonly accepted rules for writing the objective test item. The

conclusion reached from this evidence is that a definite need exists for research in the general area of assessment of pupil growth by the classroom teacher.

APPENDIX I

REFERENCES

- Anastasi, Anne, Psychological Testing (New York: The Macmillan Company, 1954) xiii & 682 pp.
- Adkins, Dorothy C., Construction and Analysis of Achievement Tests (Washington 25, D.C.: U.S. Government Printing Office, 1947) xvii & 292 pp.
- Ahmann, J. Stanley and Marvin D. Glock, Evaluating Pupil Growth (Boston: Allyn and Bacon, 1959) xiii & 605 pp.
- Ahmann, J. Stanley, Marvin D. Glock and Helen L. Wardeberg, Evaluating Elementary School Pupils (Boston: Allyn and Bacon, Inc., 1960) xviii & 435 pp.
- Baron, Dennis and Harold W. Bernard, Evaluation Techniques for Classroom Teachers (New York: McGraw-Hill Book Company, Inc., 1958) xv & 297 pp.
- Bean, Kenneth L., Construction of Educational and Personnel Tests (New York: McGraw-Hill Book Company, Inc., 1953) viii & 231 pp.
- Bradfield, James M. and H. Stewart Moredoc, Measurement and Evaluation in Education (New York: The Macmillan Company, 1957) xiv & 509 pp.
- Cronbach, Lee J., Essentials of Psychological Testing (New York: Harper and Brothers, 1960) xxi & 650 pp.
- Downie, N. M., Fundamentals of Measurement; Techniques and Practices (New York: Oxford University Press, 1958) xi & 413 pp.
- Ebel, Robert L., "Writing the Test Item," chapter VII in Educational Measurement. E. F. Lindquist, ed., (Washington, D.C., American Council on Education, 1951) xix and 819 pp.

- Garrett, Henry E., Testing for Teachers (New York: American Book Company, 1959) x & 262 pp.
- Gerberich, J. R., Specimen Objective Test Items (New York: Longmans, Green and Company, 1956) ix & 436 pp.
- Goodenough, Florence L., Mental Testing (New York: Rinehart and Company, Inc., 1949) xix & 609 pp.
- Greene, Edward B., Measurement of Human Behavior (New York: The Odyssey Press, 1952) xxiv & 790 pp.
- Greene, Harry A., A. N. Jorgensen, The Use and Interpretation of Educational Tests (New York: Longmans, Green and Company, 1929) xxii & 383 pp.
- Jordan, A. M., Measurement in Education (New York: McGraw-Hill Book Company, Inc., 1953) xi & 533 pp.
- Micheels, William J. and M. Ray Karnes, Measuring Educational Achievement (New York: McGraw-Hill Book Company, 1950) vii & 496 pp.
- Remmers, H. H. and N. L. Gage, Educational Measurement and Evaluation (New York: Harper and Brothers, 1943) ix & 580 pp.
- Remmers, H. H. and N. L. Gage and J. Francis Rummel, A Practical Introduction to Measurement and Evaluation (New York: Harper and Brothers, 1960) xiii & 370 pp.
- Rinsland, Henry D., Constructing Tests and Grading (New York: Prentice-Hall, Inc., 1937) xvi & 323 pp.
- Ross, C. C. and Julian C. Stanley, Measurement in Today's Schools (3rd Edition) (New York: Prentice-Hall, Inc., 1954) xv & 485 pp.
- Thorndike, Robert L. and Elizabeth Hagen, Measurement and Evaluation in Psychology and Education. (New York: John Wiley and Sons, Inc., 1955) viii & 575 pp.
- Tiegs, Ernest W., Tests and Measurements in the Improvement of Learning (New York: Houghton Mifflin Co., 1939) xxi and 490 pp.
- Torgerson, Theordore L., Georgia Sachs Adams, and Albert J. Harris, Measurement and Evaluation (New York: The Dryden Press, 1954) xiii & 489 pp.

Travers, Robert M. W., Educational Measurement (New York: The Macmillan Company, 1955) xix & 420 pp.

Travers, Robert M. W., How to Make Achievement Tests, (New York: The Odyssey Press, 1950) ix & 180 pp.

Wandt, Edwin and Gerald W. Brown, Essentials of Educational Evaluation (New York: Henry Holt and Company, 1957) vii and 117 pp.

Weitzman, Ellis and Walter J. McNamera, Constructing Classroom Examinations (Chicago: Science Research Associates, 1949) xvi & 153 pp.

Wrightstone, J. Wayne, Joseph Justman, and Irving Robbins, Evaluation in Modern Education (New York: American Book Company, 1956) xi & 481 pp.

APPENDIX II

UNIT TEST Form A

Alternate-Response Items:

Directions: Below are a number of statements; some are true and some are false. If a statement is true, with the special pencil provided, darken the first space on the answer sheet which corresponds to the item number. If the item is false, with the special pencil darken the second space on the answer sheet which corresponds to the item number. Each item will count one point.

1. A teacher cannot be paid below the floor established by the state minimum-salary schedule.
2. Minimum-salary is guaranteed for all teachers teaching in a state that follows a minimum-salary law.
3. Minimum-salary schedules provide salary increases for preparation beyond the bachelors degree.
4. The same salaries for men and women doing the same work are provided by a single-salary schedule.
5. The trend in regard to teachers' salaries is away from the state minimum-salary schedule.
6. The single-salary schedule provides the same salary for elementary and secondary teachers in the same school system.
7. Teacher tenure is consistent with more recent theories of social justice.
8. Since, in the beginning, each school system developed its own system, there are many differences in tenure provisions.
9. Provisions for teacher tenure in a school system saves time for school employment officials.

10. Teacher tenure laws promote child welfare indirectly.
11. A survey of teacher tenure laws indicates that all states do not have provisions that are in harmony from state to state with regard to tenure.
12. Tenure laws that place a teacher under contract from year to year without provisions for annual release have proven unsatisfactory.
13. The State Board of Education may revoke the certificate of a teacher who quits during a school term without notice to the employing school district.
14. The continuing-contract type of tenure is the most common type of tenure law.
15. College teachers need some form of tenure protection other than continuing-contract.
16. Refusal to obey an extra-curricula requirement established by the principal is a reason for dismissal under continuing-contract tenure.
17. States are compelled by retirement legislation to contribute to the welfare of the teacher.
18. Retirement systems are more important to the older teacher than to the teacher who is just beginning his teacher career.
19. Teacher security is provided through teacher tenure laws and teacher contracts.
20. Sick leave may be granted to an individual teacher for sickness in the teacher's immediate family.
21. Under the "justifiable" clause in continuing contracts, teachers cannot be dismissed during the school term except those teachers who are dismissed through a hearing before the board of education.
22. Sabbatical leave may be granted an individual teacher who desires to teach in another school system for one term.
23. Sabbatical leave may be granted to a teacher for death in his immediate family.
24. Leave granted for professional study is sometimes termed sabbatical leave.

25. Teacher credit unions are organized for the benefit of teachers and may make loans to teachers.
26. A significant rise in birth rate increases the demand for potential teachers.
27. The development of credit reports of members is a major duty of an organized teacher credit union.
28. Credit unions extend loans to teacher members at interest rates below the normal interest rate of other financial institutions.
29. More women than men teach in public elementary schools.
30. More women than men teach in universities and colleges.
31. The most likely area of subject-matter preparation to have an oversupply of teachers is Natural Science.
32. The college teacher-placement officer is responsible for developing the teaching combinations for elementary teachers.
33. Subject-matter teaching combinations are designed by authorities in education who are members of Colleges of education at state universities.
34. A definite pattern of subject-matter teaching combinations exist in public schools.
35. Reorganization of school districts result in better teaching positions.
36. New teaching positions are created when school districts reorganize and consolidate.
37. Teachers who are teaching in a school district that is consolidated with another district are not guaranteed a teaching position in the newly consolidated district.
38. Conditions that decrease demand for teachers encourage an oversupply of teachers.
39. Conditions that increase the demand for teachers, encourages an increase in the number of potential teachers entering teacher training institutions.
40. Employment officials are concerned with information on a prospective teachers information folder that indicates the teachers extent of travel.

41. The consideration given to a persons academic record by a school district employment official is negligible.
42. The primary purpose of teachers' organizations is to provide professional leadership and service.
43. A sincere interest in people will facilitate good adjustment.
44. The American Federation of Teachers is a legally organized association of teachers.
45. The National Education Association is the largest professional organization of teachers.
46. Sponsoring national conventions is evidence that teachers organizations have reached the professional level.
47. There is evidence to indicate that discrimination against married female teachers is practiced in public education.
48. Minimum salary laws are in effect in the school districts of Iowa.
49. Large cities have higher property evaluations and therefore are capable of paying larger salaries than rural schools.
50. The revenue of school districts is obtained through property taxes.

MULTIPLE-CHOICE-ITEMS

The following forty items are multiple choice with only one correct response. Each item is followed by four choices, indicate your selection of the one that completes the item and makes it true by darkening the corresponding numbered space on the answer sheet with the special pencil provided. Each item has a value of two points each.

51. Programs of in-service training are designed to
 - 1) keep teachers intellectually alive and professionally alert.
 - 2) force teachers to get master's degrees.
 - 3) enhance the prestige of the local school system.
 - 4) increase the salaries of teachers.

52. The number engaged in the teaching profession in continental United States is approximately
- | | | | |
|----|---------|----|-----------|
| 1) | 100,000 | 3) | 1,000,000 |
| 2) | 500,000 | 4) | 2,000,000 |
53. Which statement does not exemplify characteristics exhibited by the well-adjusted person?
- 1) He is poised and secure.
 - 2) He uses objective self-criticism.
 - 3) He accepts his weaknesses and keeps a rein on his emotions.
 - 4) He has never experienced failure.
54. One indication of student growth is
- 1) rigid acceptance of textbook material.
 - 2) complete reliance upon lecture notes.
 - 3) the assumption of some responsibility for self-guidance.
 - 4) eagerness to recite in the classroom.
55. Growing up means, among other things,
- 1) steadfast pursuit of goals set up in childhood.
 - 2) accepting personal responsibilities for activity in the adult world.
 - 3) willingness to compromise character for the sake of achievement.
 - 4) continuous defense of one's rights.
56. Students who expect to be teachers need to include in their preparation definite training in mental-health principles because such training will
- 1) help them understand the school community.
 - 2) be one of the subjects which they will teach.
 - 3) help them understand their own problems.
 - 4) become one of the requirements for certification.
57. The study of personal growth and development is most concerned with
- 1) group activities in which the individual finds himself.
 - 2) individual adjustment within the group.
 - 3) the individual apart from the group.
 - 4) individual maladjustment.
58. Good adjustment is best facilitated by
- 1) happy physical surroundings.
 - 2) a wise choice of motives and activities.
 - 3) keen understanding of others.
 - 4) sincere interest in other people.

59. The development of desirable personal characteristics is largely
- 1) something about which one can do nothing.
 - 2) a matter of chance.
 - 3) a matter of wanting to improve.
 - 4) a matter of habit forming.
60. Teachers from southern Indiana would find some problem in teaching in Boston because of
- 1) more advanced children.
 - 2) puritanical regulations in the schools.
 - 3) concentration of population.
 - 4) deviation in speech.
61. Which of the following attributes is most vital to teaching success?
- 1) A pleasant voice and correct speech.
 - 2) The proper social background.
 - 3) A private income to supplement one's salary.
 - 4) Straight "A" grades throughout college.
62. Building personal resources involves
- 1) the suppression of emotional expression.
 - 2) having a good hobby at which to work.
 - 3) solving the personal problems of others.
 - 4) participating in community fund drives.
63. The happiest teachers are usually those who
- 1) devote their waking hours to their profession.
 - 2) do their job with a minimum of effort.
 - 3) like to teach, but have other interests and hobbies.
 - 4) fight continuously to improve working conditions.
64. Which of these qualities does NOT characterize a good group member?
- 1) He treats all ideas put forth by the group as worthy of attention.
 - 2) He pleads earnestly and consistently to win support for his own plan of action.
 - 3) He does not monopolize the conversation.
 - 4) He tries to make his special experience contribute to the thinking of the group.
65. In earlier days the common method of ascertaining an individual's fitness for teaching was
- 1) by his setting up educational requirements.
 - 2) successful completion of an education.
 - 3) certification based on grade school completion.
 - 4) recommendation by the county superintendent.

66. The authority to coordinate and supervise certification procedures belong to
- 1) the local school district.
 - 2) a division of the state department of instruction.
 - 3) the placement division of the teacher-education colleges.
 - 4) the National Education association.
67. Teacher certification requirements
- 1) are widely different from state to state.
 - 2) are uniform in all states.
 - 3) are rather similar from state to state.
 - 4) do not exist in some states.
68. In the United States, the amount of academic preparation for elementary and secondary school teachers is
- 1) generally the same.
 - 2) usually higher for elementary teachers than for secondary teachers.
 - 3) usually lower for elementary teachers than for secondary teachers.
 - 4) decreasing the number of prospective teachers who are available.
69. In general, teachers' salaries are higher in
- 1) areas where standards for certification are low.
 - 2) areas that accept "modern" education.
 - 3) states where standards for certification are high.
 - 4) states having small school enrollments.
70. As certification requirements increase
- 1) the number of prospective teachers tends to decrease.
 - 2) many experienced teachers are forced out of the profession.
 - 3) the teaching profession becomes less attractive to young people.
 - 4) the number entering the profession tends to increase.
71. Local school systems may
- 1) require more training than state laws prescribe.
 - 2) hire teachers who do not meet state requirements.
 - 3) hire teachers who hold licenses from other states.
 - 4) release a teacher from his contract without previous notice.

72. The chief difference between old and new plans for certification of secondary school teachers is that the new plans
- 1) are based on teaching areas rather than on teaching subjects.
 - 2) are easier to complete.
 - 3) provide for no elective subjects.
 - 4) demand greater subject-matter specialization.
73. An elementary teacher needs
- 1) a broader educational background than does the secondary teacher.
 - 2) a greater degree of specialization than the secondary teacher.
 - 3) advanced training in art and music.
 - 4) about the same kind of preparation as secondary teachers.
74. A supervisor in today's schools is regarded as one who
- 1) provides friendly, constructive, and cooperative help.
 - 2) inspects the classroom to rate and supervise the teacher.
 - 3) gives direction and expects the teachers to meet identical standards.
 - 4) assumes the administrative responsibilities of the school.
75. In the nation as a whole, the percentage of teachers who are college graduates approximates
- | | |
|----------------|-----------------|
| 1) 25 percent. | 3) 75 percent. |
| 2) 45 percent. | 4) 100 percent. |
76. The reason which is most often given by men and women as an explanation for selecting teaching career is
- 1) economic security.
 - 2) desire to serve society.
 - 3) the prestige that is associated with teaching.
 - 4) the influence of friends or parents.
77. As a group, teachers tend to be relatively
- | | |
|-----------------|------------------|
| 1) pessimistic. | 3) liberal. |
| 2) radical. | 4) conservative. |
78. Studies of the family background of teachers indicate that they come from families
- 1) with incomes in the upper 10 percent.
 - 2) of average income.
 - 3) with resources in the lower 10 percent.
 - 4) proportionately representative of all income brackets.

79. A survey of high school pupils indicated that generally teachers should be allowed to
- 1) drink and smoke in public.
 - 2) gamble.
 - 3) lead normal and balanced lives.
 - 4) lead normal but conservative lives.
80. Politics and political responsibilities make demands upon many people. A teacher should
- 1) take a firm stand and support his side with utmost vigor.
 - 2) refuse to take a stand publicly.
 - 3) assume full political responsibility as a citizen.
 - 4) never mention politics in the classroom.
81. Education in a democracy requires that every teacher handle controversial issues by
- 1) remaining completely neutral.
 - 2) refusing to consider them in class.
 - 3) analyzing the fact in terms of our moral values.
 - 4) encouraging careful study of both sides of issues.
82. With respect to political behavior, teachers are
- 1) almost uniformly restricted.
 - 2) completely unpredictable.
 - 3) probably more conscientious than most occupational groups.
 - 4) probably less conscientious than most occupational groups.
83. A profession is best distinguished from a vocation by which of the following statements?
- 1) It is a group of individuals who do the same kind of work and organize for the protection of its members.
 - 2) It enforces certain standards for the protection of its members and the public.
 - 3) It prohibits certain kinds of personal conduct.
 - 4) It renders a specialized type of social service and follows a code of ethics.
84. One of the most important factors in making teaching a true profession is
- 1) improved teaching conditions.
 - 2) a higher standard of training.
 - 3) an improved system of teacher evaluation.
 - 4) a better spirit among classroom teachers.

85. The largest and most important professional organization for teachers is the
- 1) National Association of Teachers.
 - 2) National Education Federation.
 - 3) National Education Association.
 - 4) American Federation of Education.
86. The American Federation of Teachers
- 1) is affiliated with the NEA.
 - 2) follows the labor policies of the AFL-CIO.
 - 3) has a policy of striking for its just demands.
 - 4) is a legal entity in its own right.
87. Among the following, the factor which makes it especially difficult to evaluate the salaries of teachers in relation to salaries of other occupations is
- 1) amount of preparation.
 - 2) length of service.
 - 3) variation in age.
 - 4) location of work.
88. The state which pays teachers the highest average salary is
- 1) Mississippi
 - 2) Michigan
 - 3) New York
 - 4) Indiana
89. Of the gainfully employed people in the United States, teachers constitute about 1 out of every
- 1) 50 people
 - 2) 150 people
 - 3) 200 people
 - 4) 250 people
90. About what percentage of all public school teaching positions are vacated each year?
- 1) 2 percent
 - 2) 5 percent
 - 3) 10 percent
 - 4) 25 percent

MATCHING EXERCISES

Three groups of matching exercises follow. Each of the problems in the left hand column are related to the ideas in the right hand column in some way. When you have determined the item in the right hand column that matches the item in the left hand column, darken the corresponding number on the answer sheet opposite the item number.

91. The right a teacher has under contract to continue teaching in the same school system year after year is called
92. A teacher may discontinue to teach and still draw benefits from the former employing school district is
93. A means of determining the salaries of the teachers in a school system is a

94. The chief supervisor of the functions of a school district is the
95. The supervisor of an elementary school and the teachers of that building is called a
96. The person on the school faculty and staff that is responsible for the coordination of teaching is often called the
97. The supervisor of a group of teachers who are teaching related subjects is often called a
- 1) retirement.
2) salary base.
3) salary schedule.
4) dismissal.
5) tenure.
- 1) Curriculum Coordinator.
2) Principal.
3) Superintendent.
4) Attendance officer.
5) Department head.

The following subjects are to be presented during the following semester. You are the principal and wish to determine which of your present staff is most likely qualified to teach each of the subjects. Match the subject with the teacher's major field of college preparation that would most likely qualify them to teach the subjects.

<u>Subjects</u>	<u>College Major</u>
98. Ninth grade science	1) Vocational agriculture
99. Commercial Geography	2) English
100. Forensics	3) Government and Politics
	4) French and Music
	5) Industrial Arts

APPENDIX III

UNIT TEST Form B

1. State minimum-salary schedules often provide a floor below which a teacher cannot be paid.
2. Salary increases for additional preparation beyond the bachelors degree are provided by all state minimum-salary schedules.
3. The same salary for all teachers is guaranteed by state minimum salary schedules.
4. The trend in regard to teachers' salaries is toward a single salary schedule but away from a state minimum-salary program.
5. The same salaries for men and women doing the same work are provided by a single salary schedule.
6. The single-salary schedule always provides the same salary for elementary and secondary teachers.
7. Teacher tenure is consistent with more recent theories of social justice.
8. Teacher tenure saves time for school employment officials.
9. The promotion of child welfare is not an outcome of teacher tenure laws.
10. Differences in tenure laws exist because in the beginning each school system developed its own policies.
11. An analysis of teacher tenure laws indicate that all states have tenure laws although many differences in the provisions have been made.
12. Absolutely all tenure laws have proven worthless.
13. When a teacher is under contract to a school district, there is nothing to prevent him from quitting during the term without notice except the State Board of Education may revoke his certificate.
14. Permanent tenure is the most common type of tenure law.
15. College teachers need more tenure protection than teachers at any other level.
16. Teachers who have established tenure in a school district under tenure laws cannot be dismissed without "justifiable" cause. Immorality, insubordination and/or inefficiency are examples of such "justifiable" causes.

17. Retirement legislation encourages young people to enter the profession.
18. Retirement legislation is valuable because it compels the state to contribute to teacher welfare.
19. Security of teachers is provided through tenure laws and retirement systems.
20. Sick leave, when available, cannot be granted for illness in the immediate family.
21. No schools with continuing contracts cannot dismiss teachers under tenure except for "justifiable" cause.
22. Sabbatical leave may be granted for professional improvement.
23. Death in a teacher's immediate family is sufficient reason for granting sabbatical leave.
24. All colleges grant sabbatical leave to faculty members for professional study.
25. Teacher credit unions make loans to participating teachers.
26. Usually a significant rise in birth rate tends to increase the demand for teachers.
27. Credit unions are generally a benefit to teachers.
28. All credit reports for teachers are always prepared by teacher credit unions.
29. Since increased birth rates are usually accompanied by increased demand for trained teachers, the need for teachers can be approximated in advance.
30. Generally, more women than men teach in elementary schools.
31. There is always an oversupply of elementary school teachers.
32. Subject matter teaching combinations tend to follow definite patterns.
33. Subject matter teaching combinations are not made by the college teacher-placement office.
34. Colleges of Education of State Universities design all subject matter teaching combinations.
35. Consolidation of schools are apt to provide better teaching positions.
36. Consolidation of schools should create new teaching positions.
37. Consolidation of schools always cause many good teachers to lose their positions.
38. Generally, conditions that decrease demand for teachers encourage an oversupply of teachers.
39. Usually, conditions that decrease demand for teachers, encourages an increase in the number of potential teachers.
40. Employment officials are never concerned about a persons success in student teaching.
41. A school district's personnel officers is concerned with a teachers' academic achievement.

42. Teacher organizations provide leadership for members benefit.
43. Good adjustment is not facilitated by sincere interest in people.
44. The American Federation of Teachers is a legal organization.
45. The largest professional organization for teachers is the National Education Association.
46. Sponsoring conventions and establishing codes of ethics are evidences of professional statue for teachers organizations.
47. Discrimination against women teachers is not as great as it may seem.
48. Minimum salary laws will probably be in effect in all states before many years.
49. Every city has higher property evaluations than rural areas and therefore can pay higher teachers salaries.
50. A school districts revenue is obtained from different sources; that is, not all revenue for a school district is obtained through real property taxes.

Multiple choice:

51. Programs of in-service training are designed as: (1) keep teacher intellectually alive and alert so that their instruction of boys and girls will be continually improved; (2) force teachers to get master's degrees; (3) enhance prestige of the local school system; (4) increase the salaries of teachers so that they may maintain an adequate standard of living.
52. In the continental United States the Membership of the teaching profession is approximately: (1) 25,000; (2) 50,000 (3) 1,000,000 (4) 64,000,000.
53. What statement exemplifies characteristics exhibited by the well-adjusted person? (1) shy and insecure; (2) resents criticism; (3) uses objective self-criticism; (4) has never experienced failure.
54. One evidence of student growth is: (1) rigid acceptance of textbook material; (2) complete reliance upon lecture, notes; (3) the growth of responsibility for self-guidance; (4) has never experienced failure.
55. Accepting personal responsibility for activities in the adult world means: (1) voting; (2) growing up; (3) becoming a parent; (4) qualifying for teachers certificate.
56. Students who expect to be teachers need to include in their preparation definite training in mental-health principles, philosophy, history, sociology, anthropology and techniques of evaluation, because such training will: (1) help them understand the school community; (2) will be one of the subjects which they will teach; (3) help them understand their problems; (4) become one of the requirements for certification.

57. The study of personal growth and development is concerned with: (1) group activities in which the individual finds himself; (2) individual adjustment within the group; (3) the individual apart from the group; (4) the individual maladjustment.

58. Good adjustment is best facilitated by: (1) happy physical surroundings; (2) a wise choice of motives and activities; (3) keen understanding of others; (4) sincere interest in people.

59. The development of desirable personal characteristics is largely: (1) something about which one can do nothing; (2) a matter of chance; (3) motivation; (4) a matter of habit forming.

60. Teachers from southern Indiana would find some problem in teaching in Boston because of: (1) something about the climate is uncomfortable; (2) puritanical regulations; (3) concentration of population; (4) deviation in speech.

61. Which of the following attributes is most vital to teaching success and the assumption of satisfying community responsibilities: (1) a pleasant voice and correct speech; (2) the proper social background; (3) a private income to supplement one's salary; (4) a straight A grade point average through college.

62. Building personal resources involves: (1) owning your own home; (2) having a good hobby at which you work; (3) comfortable savings account; (4) a supplementary job with above-average income.

63. The happiest teachers are usually those who: (1) devote their waking hours to their profession; (2) do their job with a minimum of effort; (3) like to teach, but have other interests and hobbies; (4) fight continuously to improve their working conditions.

64. A teacher who tries to make his special experience contribute to the thinking of the group is a (1) Egotist; (2) team leader; (3) group member; (4) dulcimer.

65. Examination was the common method of ascertaining an individual's fitness for teaching and setting up educational requirements: (1) in early America; (2) in secondary schools; (3) in elementary schools.

66. The authority to coordinate and supervise certification procedures belongs to: (1) the local school district; (2) the state department of instruction; (3) the parent-teacher association; (4) the National Federation of Teachers.

67. Teacher certification requirements: (1) include a course in American Government; (2) include a physical examination (3) are rather similar from state to state; (4) includes one year of teaching experience.

68. In the United States, the requirements for elementary and high school teachers are: (1) difficult to meet; (2) easy to meet; (3) usually lower for elementary teachers than for secondary teachers; (4) logical.

69. Of the gainfully employed people in the United States, teachers constitute about 1 out of every (1) 50 people; (2) 1000 people; (3) 1500 people; (4) 2500 people.

70. As certification requirements increase, instructional quality increases, salaries increase, therefore: (1) the number of prospective teachers tends to decrease; (2) many experienced teachers, especially on the elementary level, are forced out of the profession; (3) the teaching profession becomes less attractive to young people; (4) the number entering the profession becomes increased.

71. The local school systems always: (1) require more training than state laws prescribe; (2) hire teachers, if they wish, who do not meet state requirements; (3) hire teachers who hold licenses from other states without reference to their particular state requirements; (4) release a teacher from his contract without previous notice.

72. The older plan for certification of secondary teachers was based on training in a specific subject matter field or a combination of one or more subject matter fields. The current plan is: (1) based on training in teaching areas rather than on teaching subjects; (2) easier to complete; (3) to provide for no elective courses; (4) to demand greater subject-matter specialization.

73. An elementary teacher needs:; (1) adequate supplementary textbooks; (2) ample art materials; (3) training in science; (4) about the same kind of preparation as secondary teachers.

74. About what percentage of all public school teaching positions are vacated each year? (1) 10 percent; (2) 5 percent; (3) 75 percent; (4) 15 percent.

75. In the nation as a whole, the percentage of teachers who are college graduates approximates: (1) 25 percent; (2) 60 percent; (3) 75 percent; (4) 15 percent.

76. Reasons which are most often given by men and women as an explanation for selecting teaching as a career are: (1) an economic security; (2) expressed desires to serve society; (3) the prestige that is associated with teaching; (4) the influence of friends or parents.

77. As a group, teachers tend to be relatively; (1) ambitious; (2) low intellectually; (3) emotionally stable; (4) conservative.

78. Studies of the family backgrounds of teachers indicate that they come from families (1) with incomes in the upper 50 percent; (2) of average income; (3) with resources in the lower 50 percent; (4) proportionately representative of all income brackets.

79. A profession is best distinguished from a vocation by which of the following statements; (1) teachers organizations enforce certain standards for the protection of its members and the public; (2) Teachers are a group of individuals who do the same kind of work and organize for the protection of members; (3) code of ethics prohibit certain kinds of personal conduct; (4) A profession renders a high and specialized type of social service and follows a professional code of ethics.

80. A higher standard of teaching is a characteristic which: (1) indicates teaching is a profession; (2) is resented by most teachers; (3) not encouraged by local boards of education; (4) conducive to better teacher interpersonal relations.
81. The largest and most important professional organization for teachers is the (1) National Society for the Study of Education; (2) National Congress of Parents and Teachers; (3) National Education Association; (4) National Association of School Administrators.
82. Among the following, the factor which makes it especially difficult to evaluate the salaries of teachers in relation to salaries of other occupations is: (1) amount of preparation; (2) services rendered to the community; (3) ratings by supervisors; (4) place of work.
83. The state which pays teachers the highest average salary are: (1) Mississippi; (2) Michigan; (3) New York; (4) Indiana.
84. The purchasing power of teachers' salaries out to: (1) become considerably lower for a number of years; (2) continue the present downward trend for a number of years; (3) continue the present upward trend for a number of years.
85. Which factor tends to cause larger cities to pay better salaries to teachers? (1) Teachers do not wish to teach in the larger cities; (2) the cost of living is high in cities; (3) Cities have higher property valuations from which to secure taxes to support schools.
86. According to the Research Division of the NEA, the gap between the salaries paid classroom teachers and those paid school administrators; (1) is steadily widening; (2) appears to be closing.
87. Discrimination against women teachers is: (1) not so great as it may seem; (2) increasing rather than decreasing.
88. State minimum-salary schedules characteristically: (1) Provide a floor below which a teacher **cannot** be paid; (2) guarantees the same salary for all teachers.
89. The most common type of tenure law is: (1) permanent tenure; (2) annual-contract tenure; (3) three-year contract.
90. One may find more men than women teaching in: (1) elementary schools; (2) Junior high schools; (3) Universities and colleges.

MATCH

- | | |
|---------------------|-------------------------|
| 91. Tenure | 1. continuing contract |
| 92. Security | 2. retirement |
| 93. Employment | 3. interview |
| 94. Professionalism | 4. code of ethics |
| 95. Salary schedule | 5. rate of compensation |

MATCH

- 96. Administration
- 97. Discipline
- 98. Curriculum
- 99. School Law
- 100. Visual Aids

- 1. charts, graphs and
maps
- 2. compulsory attendance
- 3. course of study
- 4. washing blackboards
- 5. Superintendent and
Principals.

APPENDIX IV

Physics Test Mechanics

Name _____

6. If two identical automobiles are moving around the same curve, one going faster than the other, which one will have the greatest tendency to leave the road?
a) The one going the fastest. b) The one going the slowest. c) both will have equal tendency because they are identical. d) neither one. _____
8. Work is calculated by a) dividing force by distance
b) dividing distance by force c) multiplying rate by time d) multiplying force by distance. _____
9. Two bodies of different mass falling from the same height in a vacuum, a) the heaviest body will accelerate faster b) the lightest body will accelerate faster e) the larger mass will reach the ground first d) they both will reach the ground together. _____
11. Work is done on a cement block when you a) raise it to your shoulder b) support it on your shoulder c) carry it along a level walk d) place it on the ground. _____
12. Kinetic energy is a) directly proportional to velocity b) directly proportional to mass c) inversly proportional to velocity d) inversly proportional to mass. _____
13. When the total force acting upon an object is zero, the object a) will not accelerate b) will accelerate at a constant rate c) will not move d) will move at a constant rate. _____
14. In going from one place to another in space, for example, from the earth to the moon, a) your mass would become smaller on the moon b) the pull of gravity would be greatest on the moon d) it would have no effect upon your mass. _____

16. What law of physics does the operation and flight of a rocket depend upon essentially? a) Bernoulli's principle. b) The second law of motion. c) The law of action and reaction d) the law of gravity. _____
17. Horse power is a) a unit of electrical power b) a unit of chemical power c) the MKS unit of power d) the FPS unit of power. _____
18. Kinetic energy is possessed by a) a wound watch spring b) a bullet leaving the muzzle of a gun c) a rock resting on the edge of a cliff d) a pendulum at the point of maximum displacement. _____

MULTIPLE CHOICE QUESTIONS

2. Election of a President by the direct, popular vote of the people is considered unwise because- (a) the people are not capable of making such an important decision, b) it would be much more expensive than the present electoral college method, (c) states with unrestricted suffrage would have an unfair advantage over States with educational qualifications for suffrage.
5. A major function of the Department of Agriculture is to - (a) supervise the government of the territories of Hawaii and Alaska, (b) establish agricultural schools throughout the states, (c) supervise payments to farmers for cooperating with the government in its soil conservation and acreage allotment program.
7. The power to enact bankruptcy laws resides in (a) Congress alone, (b) the state legislatures alone, (c) both Congress and the state legislatures.
8. Congress has the power to abolish - (a) the United States Supreme Court, (b) the state District Courts, (c) the Federal District Courts.
9. Under its power to regulate interstate commerce, Congress may-(a) forbid child labor in all factories, (b) control radio and TV broadcasting, (c) issue franchises to public utilities.

TRUE-FALSE QUESTIONS

1. Individuals may subscribe to a political party's campaign fund, but corporations may not.

2. Today, most contributions for campaigns come from independent committees.
3. A chief justice and nine associate justices make up the United States Supreme Court.
4. Today, some states give suffrage under 21 years of age.
5. The Secretary of State is responsible for conducting the Armed Forces.

Eighth Grade Social Studies

Geography

Fill in the ~~xxx~~ blank or underline the correct answer

7. If you wished to visit France's greatest iron ore region, you would go to (Rhonê Valley, Alsace-Lorraine, Saar Basin).
9. The capital of France is _____.
10. The important river has its outlet in the Mediterranean Sea. _____
13. These two peninsulas in northwestern France were major beachheads during W.W.II _____ and _____
18. What is included ~~in~~ in the U.K.?
20. The "Silver Streak" refers to the _____.
22. The British Isles are fortunate to have great deposits of _____.
23. Newcastle is at the mouth of the _____ River.
29. The capital of Ireland is _____.
33. Name one important scientific inventor that was an Italian. _____

WORLD HISTORY TEST

Semester 2

2. A strong feeling of patriotism or loyalty toward ones country is called _____.

6. A form of government whereby the government owns and controls the major industry is known as _____.
9. We studied three revolutions which were said to have put man on the road to his struggle for democracy. The three are _____ and _____.
10. The French Revolution was said to be one having the greatest influence on mankind and also the bloodiest. A period of the French Revolution in which there was much bloodshed and complete Chaos were known as the _____.
11. The English Revolution was more or less a peaceful one and because of this it is given the name the _____.
14. A French phrase meaning "stroke of state" is _____.
16. A set of laws established by Napoleon I is called _____.
17. One of the slogans taken up by the Congress of Vienna was one which declared that no single nation would ever again have the power to completely dominate Europe again. This theory is known as _____.
20. England seized American sailors during the war of 1812 and forced them to serve in their own navy. This forced seizure is known as _____.
21. An area set aside for the Pope's residence in Rome is known as _____.

7th SCIENCE 2nd Semester

Part III. Multiple choice: Choose the correct letter and place on your answer sheet.

26. What is the shorthand for elements called? A. solid B. symbols C. elements D. physical
27. Which one of the three states of matter has a specific shape? A. gas B. chemicle C. solid D. liquid
28. What kind of change do we call it when matter changes state? A. Physical B. Chemical C. Solid D. gas
33. What type of energy is moving energy? A. Potential B. Kinetic C. Force D. vibration.
35. How many inches in a meter stick? A. 24 B. 10 C. 33 d. 39.37

Part III. Matching: Place correct letter on the answer sheet.

36. What method of heating does the simple hot air furnace depend upon?
37. What method of travel does steam heat use in heating a room?
38. What method of heating the earth does the sun use?
39. What method of heating does electricity use in heating your home?
40. What do you call the part on a furnace that lets air in for burning?
41. What is the place in a furnace called where fuel is burned?
42. What is the reason we hear rumbling in the steam pipes?
43. How many pipes are needed to be used when heating with steam?
44. What do you call the point when a fuel will burst into flame?
45. What is the best color to wear in the hot sun?

- | | |
|---------------|-----------------------|
| A. radiant | H. white |
| B. convection | I. combustion |
| C. Conduction | J. 1 |
| D. Kindling | K. 2 |
| E. damper | L. 3 |
| F. valve | M. rapid condensation |
| G. black | N. slow condensation |
| | P. constoration |

*Evaluate only those items that are boxed

These items have been copied directly from teacher made tests. I have tried to maintain errors in spelling, spacing, strike-overs, et cetera. Please evaluate the items as though they were a part of a total test. Directions have been copied where presented.

Not all sections of this test are alike, because ten items have been selected from five different tests submitted by five different teachers. (These items were selected through the use of a table of random digits.)

APPENDIX V

Cedar Falls, Iowa
April 1, 1961

Dear Teacher,

I am requesting your help in a research study that involves the analysis of objective test items used by classroom teachers. The results of this study may be used as a basis for improving teacher education preparation in measurement and evaluation.

Would you participate by sending twenty or twenty-five objective type (multiple choice, true-false, matching or completion) test items which you have developed for use in your classes? (Ten items are acceptable, however, twenty or more items are preferred.) Please send items which you developed personally, since the study is concerned with your own teacher made test items.

You may rest assured on two counts. First, your items will be safeguarded so that they will not be available to students. Secondly, your name and the items you submit will NOT be disclosed to anyone under any circumstances.

You are requested to complete the enclosed personal information card and return it with your items. When these are received, the test items will be separated from the information card. Thus, even I will be unable to readily identify the source of the items. I will be able to identify you by means of the information card.

Let me urge you to complete the personal information card now. When you have it completed, enclose it with twenty or more of your test items in the stamped, addressed envelope and mail it at your earliest convenience.

When this study is completed you will receive an abstract of the findings.

Thank you for your kind cooperation.

Very truly yours,

Glen R. Hastings, Assistant Professor
Department of Education and Psychology
State College of Iowa

APPENDIX VI

PERSONAL INFORMATION
QUESTIONNAIRE

INFORMATION CARD

- | | | | |
|----|----|--|----|
| 32 | | | 1 |
| 33 | 1. | Your name _____ | 2 |
| 34 | | | 3 |
| 35 | 2. | School address _____ | 4 |
| 36 | | | 5 |
| 37 | 3. | Subject(s) _____ | 6 |
| 38 | | presently teaching? _____ | 7 |
| 39 | 4. | College Major _____ Minor _____ | 8 |
| 40 | 5. | College(s) _____ | 9 |
| 41 | | attended _____ | 10 |
| 42 | 6. | Have you ever had a course in Tests and | 11 |
| 43 | | Measurements? (Circle) YES NO | 12 |
| 44 | | a. What college offered | 13 |
| 45 | | the course? _____ | 14 |
| 46 | | b. When was the course | 15 |
| 47 | | completed? _____ | 16 |
| 48 | | c. Was a unit of the course devoted | 17 |
| 49 | | to subjective test development? | 18 |
| 50 | | YES NO | 19 |
| 51 | | d. Have you ever had any other course | 20 |
| 52 | | in which a unit was devoted to | 21 |
| 53 | | objective test development? | 22 |
| 54 | | YES NO | 23 |
| 55 | | e. If YES, what college offered the | 24 |
| 56 | | course? _____ | 25 |
| 57 | | f. What was the title of the course? | 26 |
| 58 | | _____ | 27 |
| 59 | | g. How many credit hours were granted | 28 |
| 60 | | for successful course completion? | 29 |
| 61 | 7. | How many years of teaching experience have | 30 |
| 62 | | you completed including this (1960-61) | 31 |
| | | school year? _____ | |
| | 8. | Did you develop the items you are submit- | |
| | | ting? (Circle) YES NO | |
| | 9. | If NO, what is the source of the items? | |
| | | _____ | |

APPENDIX VII

PERSONAL INFORMATION QUESTIONNAIRE CODING CHART

Hole #	Response	Hole #	Response
1	Soc. St. teacher	32	Reserve
2	Male sex	33	Math minor
3	Math major	34	Science minor
4	Science major	35	English minor
5	English major	36	Social Studies minor
6	Social studies major	37	Foreign lang. major
7	Foreign languages	38	Industrial Arts minor
8	Industrial Arts	39	Voc. H.E. or Agri.
9	Voc. H.E. or Agriculture	40	Other minors
10	Other majors	41	Colleges attended
11	3 years or less teaching experience (including current yr)	42	S.C.I.
12	over 3 - less than 5	43	I.S.U.
13	over 5 - less than 10	44	S.U.I.
14	over 10 years experience	45	Drake
15	Credit for course in T&M	46	Cornell
16	Unit in Obj. Exam Const.	47	Coe
17	Unit in Obj. Exam Const. in another course	48	Grinnell
18	Age of course, 1 yr	49	Luther
19	2 years	50	Dubuque
20	3 years	51	Upper Iowa
21	4 years	52	Central
22	5 years	53	Wartburg
23	over 5 years	54	Loras
24	over 10 years	55	Other Iowa Senior Colleges
25		56	Iowa Junior Colleges
--		57	Out-of-State Colleges
--	Reserve	58	
--		59	
30		--	Reserve
31	Items submitted are not original	--	
		62	

*When the response was as indicated, the edge of the card was punched, making it possible to select only those cards without the given response by inserting a needle and lifting upward.

APPENDIX VIII

CONTINGENCY TABLES SHOWING THE NUMBER OF ERRORS OF ITEM WRITING
RULES FOUND IN TESTS SUBMITTED BY MEMBERS OF THE SAMPLE

Teachers	Course Trained	Not Course Trained	Total
Social Science	775	288	1063
Natural Science	669	320	989
Total	1444	608	2052

Teachers	Course Trained	Not Course Trained	Total
Experienced	785	359	1144
Limited Experience	659	249	908
Total	1444	608	2052

Teachers	Social Science	Natural Science	Total
Experienced	600	463	1063
Limited Experience	544	445	989
Total	1144	908	2052

Teachers	Course Trained	Not Course Trained	Total
<hr/>			
Sex			
Male	1205	553	1758
Female	239	55	294
Total	1444	608	2052
<hr/>			
Experience			
5 years or less	758	312	1070
More than 5 years	686	296	982
Total	1444	608	2052
<hr/>			
Teaching in			
Major	826	385	1211
Minor	618	223	841
Total	1444	608	2052
<hr/>			
Graduated from			
University A	239	94	333
University B	83	44	127
University C	116	39	155
College A	325	181	506
Other Iowa	441	144	585
Out-Of-State	240	106	346
Total	1444	608	2052
<hr/>			

CONTINGENCY TABLES OF NUMBER OF ERRORS WITH FORM AND STYLE
 ERRORS FOUND IN TESTS SUBMITTED BY MEMBERS OF THE SAMPLE
 REMOVED.

Teachers	Course Trained	Not Course Trained	Total
Social Science	366	131	497
Natural Science	251	161	412
Total	617	292	909

Teachers	Experienced	Limited Experience	Total
Course Trained	363	254	617
Not Course Trained	156	136	292
Total	519	390	909
Social Science	276	221	497
Natural Science	243	169	412
Total	519	390	909

BIBLIOGRAPHY

- Adkins, Dorothy C. Construction and Analysis of Achievement Tests. (Washington, D.C.: U.S. Government Printing Office, 1947) pp. 55-64.
- Anderson, G. Lester "Colleges and Universities--Organization and Administration," article in The Encyclopedia of Educational Research, Third Edition. ed. Chester W. Harris, (New York: The Macmillan Company, 1960) pp. 252-265.
- Arkin, Herbert and Raymond R. Colton Tables for Statisticians. (New York: Barnes and Noble, Inc., 1950).
- Best, John W. Research in Education. (Englewood Cliffs, N.J.: Prentice Hall, Inc., 1959) 320 pp.
- Cook, Walter W. "What the Teacher Should Know About Measurements," paper in the 15th Yearbook of the National Council on Measurement Used in Education, ed. Edith M. Huddleston, (New York: National Council on Measurements used in Education, 1958) pp. 16-18.
- Dressel, Paul L. "The Way of Judgment," paper in the 15th Yearbook of the National Council on Measurement Used in Education. ed. Edith M. Huddleston, (New York: National Council on Measurements Used in Education, 1958) p. 4.
- Ebel, Robert L. "Writing the Test Item" Chapter VII in Educational Measurement. ed. E. F. Lindquist (Washington, D.C.: American Council on Education, 1951) pp. 185-249.
- _____, and Dora E. Damrin "Tests and Examinations," article in The Encyclopedia of Educational Research, Third Edition. ed. Chester W. Harris (New York: The Macmillan Company, 1960) pp. 1503-1513.
- Greene, Harry A. and Albert N. Jorgensen The Use and Interpretation of Educational Tests. (New York: Longmans, Green and Co., 1929).

- Good, Carter V. Dictionary of Education. (New York: McGraw-Hill Book Company, Inc., 1945) 495 pp.
- _____. Introduction to Educational Research. (New York: Appleton-Century-Crofts, Inc., 1950) 424 pp.
- _____. Methods of Research. (New York: Appleton-Century-Crofts, Inc., 1959) 920 pp.
- Hagen, Elizabeth P. and Robert L. Thorndike, "Evaluation" article in The Encyclopedia of Educational Research. ed. Chester W. Harris, (New York: The Macmillan Company, 1960) pp. 482-85.
- Hastings, J. Thomas "How Can In-Service Training In Measurement Techniques be Provided Most Effectively" paper in the 15th Yearbook of the National Council on Measurements Used in Education. ed. Edith M. Huddleston, (New York: National Council on Measurements Used in Education, 1958) p. 28.
- Lindquist, E. F. "Preliminary Considerations in Objective Test Construction," Chapter V in Educational Measurement. ed. E. F. Lindquist, (Washington, D.C.: American Council on Education, 1951) pp. 119-158.
- Lindquist, E. F. Design and Analysis of Experiments in Educational and Psychology. (Boston: Houghton-Mifflin Company, 1956).
- Noll, Victor H. "Should a Course in Measurement be Required for Teacher Certification?" paper in the 15th Yearbook of the National Council on Measurement Used in Education. ed. Edith M. Huddleston, (New York: National Council on Measurements Used in Education, 1958) pp. 33-38.
- Rinsland, Henry D. Constructing Tests and Grading. (New York: Prentice-Hall, Inc., 1937).
- Rummel, J. Francis An Introduction to Research Procedures in Education. (New York: Harper and Brothers, 1958) 413 pp.
- Travers, Robert M. W. An Introduction to Educational Research. (New York: The Macmillan Company, 1958) 466 pp.
- Woellner, R. A. and A. M. Wood Requirements for Certification of Teachers, Counselors, Librarians and Administrators. (Chicago: University of Chicago Press, 1960).

Wood, Dorothy Adkins Test Construction. (Columbus, Ohio: Charles E. Merrill Books, Inc., 1960).

Periodicals

Bondig, Albert W. "The Relation of Level of Course Achievement to Students; Instructor and Course Ratings in Introductory Psychology," Educational and Psychological Measurements, 13: 437-38, 1953.

Davis, Frederick B. "Item Analysis in Relation to Educational and Psychological Testing," Psychological Bulletin, 49: 97-121, 1952.

Ebel, Robert L. "Procedures for the Analysis of Classroom Tests," Educational and Psychological Measurements, 14: 352-64, 1954.

Findley, Warren G. "Rationale for the Evaluation of Item Discrimination Statistics," Educational and Psychological Measurements, 16: 175-80, 1958.

Johnson, A. Pemberton "Notes on a Suggested Index of Item Validity: The V. L. Index," Journal of Educational Psychology, 62: 499-504, 1951.

Lorge, Irving and Lorraine K. Diamond, "Prediction of Absolute Item Difficulty by Ranking and Estimating Techniques," Educational and Psychological Measurements, 14: 365-72, 1954.

Mosier, Charles I. et. al. "Suggestions for the Construction of Multiple Choice Test Items," Educational and Psychological Measurements, 5: 261-70, 1945.

Orleans, Jacob S. et. al. "Some Preliminary Thoughts on the Criteria of Teacher Effectiveness," Journal of Educational Research, 45: 641-48, 1952.

Torbet, David P. "The Attitude of a Select Group of Colorado Secondary School Teachers Toward Informal Teacher Made Tests as Measured by a Projective Interview," Journal of Educational Research, 50: 691-700, May 1957.

Travers, Robert M. W. and Norman E. Gronlund, "The Meaning of Marks," Journal of Higher Education, 21: 369-74, October, 1950.

Pamphlets

Davis, Frederick B. Item Analysis Data. Harvard Education Papers, Number 2, Harvard School of Education, (Cambridge: Harvard University, 1946) 42 pp.

Davis, Frederick B. "Criteria for the Evaluation of Achievement Tests," Proceedings of the Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1951. pp. 73-81.

Lennon, Roger T. et. al. "A Glossary of 100 Measurement Terms," Test Service Notebook Number 13, (Tarrytown-on-Hudson, N.Y.: World Book Company)

McCall, William A. "Measurement of Teacher Merit," Publication Number 294, State Superintendent of Public Instruction, (Raleigh: State of North Carolina, 1952).

Noll, Victor H. "Problems in the Pre-Service Preparation of Teachers in Measurement" paper read before the meeting of the National Council of Measurements in Education, Chicago, Illinois, February 22, 1961.

Phi Delta Kappa, "Research Design and Analysis" Second Annual Phi Delta Kappa Symposium on Educational Research (Bloomington, Indiana: Phi Delta Incorporated, 1961).

Educational Bulletin of the Department of Public Instruction, Des Moines, Iowa, 32: November 1960.

Educational Directory of the Department of Public Instruction, Des Moines, Iowa, 32: 1960-1961.