

THE PURPOSE AND TECHNIQUES OF SUPERVISORY EVALUATION
OF INSTRUCTION WITH PARTICULAR REFERENCE
TO SECONDARY MATHEMATICS

APPROVED:

John F. Curry
Major Professor

Harlan C. Miller
Minor Professor

Watt Blain
Dean of the School of Education

Robert B. Toulouse
Dean of the Graduate School

THE PURPOSE AND TECHNIQUES OF SUPERVISORY EVALUATION
OF INSTRUCTION WITH PARTICULAR REFERENCE
TO SECONDARY MATHEMATICS

THESIS

Presented to the Graduate Council of the
North Texas State College in Partial
Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

James Hendrix Rollins, B. A.

Denton, Texas

August, 1958

TABLE OF CONTENTS

	Page
LIST OF ILLUSTRATIONS	iv
Chapter	
I. INTRODUCTION	1
Statement of the Problem	
Background of the Study	
Need for the Study	
Method and Scope	
Procedure for the Study	
II. SUPERVISORY EVALUATION OF INSTRUCTION.	6
Definition	
Purpose	
Criteria	
III. EVALUATION OF TEACHER ACTIVITIES	15
Techniques Employing Rating Instruments	
Techniques Employing Other Types of Records and Recording Devices	
Techniques Employing Classroom Visitation and Observation Without Rating	
Summary	
IV. EVALUATION OF PUPIL CHANGE	67
Techniques for Appraising Pupil Change	
Pupil Status and Change as a Criterion	
V. SUMMARY, CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS	90
Summary	
Conclusions	
Implications	
Recommendations	
BIBLIOGRAPHY	95

LIST OF ILLUSTRATIONS

Figure	Page
1. Excerpt from <u>The Teaching Evaluation Record</u> . . .	19
2. Excerpt from <u>Almy-Sorenson Rating Scale for Teachers</u>	21
3. Excerpt from "The Latham Description and Analysis of Student Teacher Performance"	23
4. Excerpt from <u>Scale for the Rating of Teaching Skill Type I</u>	25
5. Reproduction of <u>Teacher Rating Card</u>	28
6. Reproduction of <u>An Analysis Chart for Evaluating Some Observable Factors in the Teaching-Learning Act</u>	30
7. Excerpt from "Improvement Sheet for First Year Algebra"	32
8. Excerpt from <u>A Survey of Instructional Practices and Materials Used in Observed Lessons in the Social Studies in Grades 6, 7, and 8</u> .	33

CHAPTER I

INTRODUCTION

Statement of the Problem

This study was undertaken to investigate the purposes and techniques of supervisory evaluation of instruction with particular reference to secondary mathematics. To this end the following areas have been explored:

1. The purposes of supervisory evaluation of instruction.
2. The techniques of evaluation used by persons engaged in evaluation programs.
3. The limitations of these techniques.
4. The application of these techniques to the evaluation of instruction in secondary mathematics.

Background of the Study

During the last few decades the public schools of the United States have undergone a tremendous growth. Quite obviously the increasing population of the country has been a factor in this growth of the schools. But, as the American public schools have developed, the idea of an education for all the children of all the people has developed also. The

development of this idea has resulted in an increased percentage of the population attending the public schools.

As the percentage of the population attending the schools increases, the administrators and teachers are faced with the growing problem of providing the proper educational services for pupils with greatly differing abilities, needs, and purposes. This problem has special meaning for the mathematics supervisor, because, in addition to the problem of providing appropriate training in mathematics for these several groups of students, which is enough in itself, he is faced with an ever increasing shortage of properly trained mathematics teachers with whom to work. Consequently many mathematics classrooms are staffed with substandard teachers.

The problems of the mathematics supervisor do not stop here though. Because of today's increasingly technological society, mathematics is coming to play an increasingly important role. The degree to which mathematics is applied to the sciences and to non-scientific social activities has increased tremendously during the last decades and is increasing continually. More recently mathematical methods have been applied to industrial planning, medicine, biochemistry, biophysics, and sociology--even philosophy and linguistics.

The number and variety of mathematical disciplines have greatly increased in the last sixty years. Also mathematicians are creating new branches of pure mathematics. These new fields

of applied and pure mathematics have burst the existing compartments that house arithmetic, algebra, and geometry. They by their very nature have made the classical treatment of secondary mathematics obsolete. As a result the concepts and design of mathematics education are undergoing great changes, possibly the greatest changes they have undergone since the beginning of the public school movement. Quite recently the Russian advances in the scientific field have injected a note of urgency.

The task for the mathematics supervisor is cut out. He must provide the leadership through which mathematics teachers can meet the challenge of the increased numbers of students and the challenge of the emerging new curriculum. One facet of this supervisory leadership is leadership in the evaluation of instruction as a basis for improving teaching effectiveness.

Need for the Study

Good teachers and administrators are vitally concerned with the improvement of instruction. When these teachers set out to improve their teaching effectiveness, they soon become aware of the need for some dependable evaluation of their present effectiveness as a basis for planning for improvement. Numerous writers in the fields of supervision and evaluation have offered these teachers their opinions on improving teaching effectiveness. Some have conducted experimental studies concerned with evaluation of teaching effectiveness. However, as far as can be determined, few, if any, studies

have been made considering evaluation of instruction with particular reference to secondary mathematics. Since the evaluation of teaching effectiveness in secondary mathematics is as desirable as the evaluation of teaching effectiveness in any field, it appears that a study of supervisory evaluation of instruction with particular reference to secondary mathematics is warranted. This area is listed in the 1950 edition of the Encyclopedia of Educational Research as one in which additional research and study is needed.

Method and Scope

This study was conducted on the campus of North Texas State College, Denton, Texas, during the summer sessions of 1957 and 1958. The research was confined to the publications available in the library of North Texas State College. The bibliographical sources used include the card catalogue, the Encyclopedia of Educational Research, the Education Index, and numerous bibliographies in books and professional journals.

Procedure for the Study

The tasks encountered in conducting this study were:

1. Defining specifically evaluation of instruction.
2. Delimiting the investigation.
3. Obtaining a working bibliography from which to conduct the research.

4. Determining the purpose of supervisory evaluation of instruction as distinguished from the purpose of administrative evaluation of instruction.

5. Determining the criteria for supervisory evaluation of instruction.

6. Determining the techniques of supervisory evaluation of instruction.

7. Determining the limitations of these techniques.

8. Drawing conclusions and making recommendations for the application of these techniques to supervisory evaluation of instruction in secondary mathematics.

CHAPTER II

SUPERVISORY EVALUATION OF INSTRUCTION

Definition

The modern concept of evaluation involves much more than traditional rating, testing, or measuring alone. According to Remmers and Gage, "Evaluation includes measurement along with values or purposes" (12, p. 1). Beecher (5), Burton (6), McNerney (7), and Ostrander (8) describe it as determination of status plus concern with planning for growth and improvement. More specifically, Wiles (16, p. 292) defines the process of evaluation as:

1. Defining goals and establishing standards by which to judge the amount of change.
2. Collecting evidence of change.
3. Applying the criteria and making judgments about the worth of change.
4. Revising plans in terms of the judgments made.

According to Tyler (15, pp. 266-268) evaluation is:

1. Formulating a statement of educational objectives.
2. Defining each of these objectives in terms of behavior.

3. Selecting and trying promising methods for obtaining evidence regarding each type.

4. Selecting on the basis of this preliminary trial, the more promising appraisal methods for further development and improvement.

5. Devising means of interpreting and using the results of the various instruments of evaluation.

Remmers and Gage (12, pp. 50-51) describe evaluation as:

1. Stating the purpose and content of the evaluation.
2. Constructing or selecting an evaluating device.
3. Administering the evaluating device.
4. Interpreting the data yielded by the evaluating device.
5. Evaluating the evaluating device.

Barr, Burton, and Brueckner (4, p. 356) view the evaluation of instruction as a threefold operation involving:

1. The securing of adequate records of the purposes and conditions that prevail in the learning-teaching situation under consideration.
2. The collection of reliable data relative to teacher and pupil activities.
3. The evaluation of the data collected.

Purpose

According to Tiegs, "Most devices for the evaluation of teaching and teachers have been used as a basis for the selection, promotion, demotion, or dismissal of teachers, and for the modification of salaries"(14, p. 315). However, to determine what should be the purpose of supervisory evaluation of teaching and teachers, the modern concept of supervision must be considered.

In 1922 Burton (6, pp. 9-10) advanced the idea that the supervisor should be concerned with the following things:

1. The improvement of the teaching act.
2. The improvement of teachers in service.
3. The selection and organization of subject matter.
4. Testing and measurement.
5. The rating of teachers.

The next year Dunn focused more attention on the developing concept of supervision by stating that "instructional supervision . . . has the larger purpose of improving the quality of instruction" (3, p. 5). The present day concept of the role of supervision is expressed by Wiles as "assistance in the development of a better teaching-learning situation" (16, p. 8). To this end Wiles (16, p. 25) characterizes supervision as, among other things, skill in evaluation. McNerney says, "Supervision is the procedure of giving direction to and providing for critical

evaluations of the instructional process"(7, p. 1). In agreement, Alberty and Thayer (3) point out that supervisory evaluation of instruction justifies itself or fails to the extent that improvement is realized. Thus the purpose of supervisory evaluation of instruction develops as "the diagnosis of teaching difficulties and the improvement of teaching and learning" (14, p. 315). More briefly stated the purpose of supervisory evaluation of instruction is the improvement of instruction.

The above is not to be interpreted as implying that the uses of instructional evaluation mentioned in the opening paragraph of this section cannot be justified, or that they do not bear on instructional improvement. But, according to Reavis and Cooper (11) and others (1, 6, 16), since the role of the modern supervisor should be that of a co-worker with and helper of teachers, such administrative use of their evaluations could interfere with and even nullify the supervisor's usefulness and effectiveness. For this reason Reavis and Cooper (11, p. 93) maintain that the supervisor should be excused from evaluations that will be used for administrative purposes.

McNerney (7, p. 123) says that in his opinion professional teachers will welcome the modern concept of evaluation in which the actual use of the results appears in plans for both teacher and pupil growth. Alexander and

Halverson (1), Beecher (5), Burton (6), McNerney (7), and Wiles (16) feel that if instructional evaluation is to be most effective it must be a co-operative effort of all who are concerned in it. Also Beecher (5, p. 273) says that adequate evidence must be available; and therefore, the appraisal should be continuous rather than periodic. Patterson (9) and Ostrander (8) concur.

Tyler (12, pp. 268-271) says that this kind of evaluation is a powerful tool for bringing about improvement of instruction and points out that it can contribute to improvement by:

1. Providing a means for gaining objectivity and cooperation in working with teachers on the improvement of instructional acts.

2. Requiring that objectives be formulated and clearly defined.

3. Providing a means for the identification and the analysis of learning difficulties of individuals and of the group.

4. Providing a means of testing basic hypotheses on which curriculum and procedures of instruction are based.

5. Placing emphasis on the study of children.

6. Fixing the attention of the learner on immediate goals rather than on his own specific activities.

Beecher points out that:

Evaluation of the effectiveness of teaching is a basic, if not the most important function of the supervisor. Effective carrying out of other functions of supervision will depend to a large extent upon evaluative diagnosis of current practice (5, p. 270).

Criteria

According to Barr, Burton, and Brueckner (4, p. 356), few persons who have not attempted to make studies of teaching have any appreciation of its complexity and elusiveness. Nevertheless, following a study of criteria for the evaluation of instruction, Anderson (2, p. 41) says that the multitude of attempts that have been made to establish satisfactory criteria for the evaluation of teaching effectiveness may be placed in one of two categories. These categories are:

1. The appraisal of teachers in the light of methods and abilities thought essential to teaching success.
2. The measurement of pupil status and change.

The findings of Anderson's study back up the opinions of Barr, Burton, and Brueckner (4), Ryans (13), and McNerney (7). About this McNerney says:

Method objectives and subject matter objectives must have coexisting equality. It seems obvious that method, regardless of how democratically it might be conceived, would do little to promote the democratic way of life unless it implemented the student's understanding of the experiences and facts upon which the democratic way of life is

founded. Consequently any consideration of the evaluation of method must include in its scope the evaluation of the student's competencies and attitudes as they have evolved from his experiences with subject matter as well as method (7, p. 78).

In this study both approaches to the problem of instructional evaluation are investigated; but, in line with the earlier statement of the problem, the first approach has been accorded more extensive and intensive investigation.

CHAPTER BIBLIOGRAPHY

1. Alexander, William M, and Halverson, Paul M., Effective Teaching in Secondary Schools, New York, Rinehardt and Company, Inc., 1956.
2. Anderson, Harold M., "A Study of Certain Criteria of Teaching Effectiveness," Journal of Experimental Education, XXIII (September, 1954), 41-71.
3. Alberty, H. B. and Thayer, V. T., Supervision in the Secondary School, New York, D. C. Heath and Company, 1931.
4. Barr, A. S., Burton, William H., and Brueckner, Leo J., Supervision, 2nd ed., New York, Appleton-Century-Crofts, Inc., 1947.
5. Beecher, Dwight E., "Judging the Effectiveness of Teaching," Bulletin of the National Association of Secondary School Principals, XXXIV (December, 1950), 270-281.
6. Burton, William H., Supervision and the Improvement of Teaching, New York, D. Appleton and Company, 1923.
7. McNerney, Chester T., Educational Supervision, New York, McGraw-Hill Book Company, Inc., 1951.
8. Ostrander, R. H., "Evaluation in the Oak Ridge Schools," Educational Leadership, VIII (November, 1950), 86-90.
9. Patterson, Allen E., "The Essentials of a Program of Evaluation and Measurement of Student Teaching," National Business Education Quarterly, XXIII (December, 1954), 5-9.
10. Rautman, Arthur L., "We All Hate to be Evaluated," Educational Administration and Supervision, XXXV (December, 1949), 449-460.

11. Reavis, William C. and Cooper, Dan H., Evaluation of Teacher Merit in City School Systems, Supplementary Educational Monograph, No. 59, Chicago, University of Chicago Press, 1945.
12. Remmers, H. H. and Gage, N. L., Educational Measurement and Evaluation, New York, Harper and Brothers, 1955.
13. Ryans, David G., "The Criteria of Teaching Effectiveness," Journal of Educational Research, XLII (May, 1949), 690-699.
14. Tiegs, Ernest W., Tests and Measurements in the Improvement of Learning, Boston, Houghton Mifflin Company, 1939.
15. Tyler, Ralph W., "Evaluation as a Function of Supervision," Elementary School Journal, XLIV (January, 1944), 264-273.
16. Wiles, Kimball, Supervision for Better Schools, Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1955.

CHAPTER III

EVALUATION OF TEACHER ACTIVITIES

Methods and techniques for the evaluation of teaching have been given various names and have been classified in various ways by different writers. When these names and classifications are carefully examined, however, they seem to fall logically into three general groups. These groups are:

1. Techniques employing rating instruments.
2. Techniques employing other records and recording devices.
3. Techniques employing classroom visitation and observation without rating.

In this chapter each of these techniques is discussed.

Techniques Employing Rating Instruments

According to Knudsen (39, p. 206) and Reavis and Cooper (50, p. 2) rating is defined as a means of evaluating teaching through the use of a printed form designed to guide the user in the formation of a judgment. The results of a study by King (38) indicate the practice of rating was widespread in 1925. A survey conducted in 1944 by Reavis and Cooper (50, pp. 17-18) sought to determine the evaluation practices in

104 city school systems. The results of that survey indicate that 99 per cent of the schools surveyed practice rating in some form, either exclusively or in combination with some other technique. Since then Alexander and Halverson (3, p. 496) report that the use of rating scales and check lists has declined. Nevertheless, Ryans (53, p. 693) says that some form of rating has been and still is the most common form of teacher evaluation for both administrative and supervisory purposes.

Rating Instruments Classified According to Form

There are many variations in the nomenclature of rating instruments to confuse and bewilder the person attempting to study teacher evaluation. Boyce (17), an early investigator of evaluation of teaching, classifies rating methods as follows:

1. Descriptive reports,
2. A type "in which a series of questions is asked" (17, p. 17).
3. A type in which "teachers for a building are listed usually in order of general excellence" (17, p. 17).
4. A type "in which definite numerical values are given to the various qualities" (17, p. 18).

Boyce's second method is called a guided comment report by Reavis and Cooper (50, p. 78), who designate the third method as a ranking report and the fourth as a check scale.

Knudsen (30, p. 211) in describing rating scales notes that some use a number to indicate performance while others use phrases, but make no distinction between the two. This situation, however, is made the basis of further classification by McNerney (42, p. 121) and Barr, Burton, and Brueckner (11), who designate instruments that attempt to fix a score or give a numerical value as rating scales and those which do not as check lists. Barr, Burton, and Brueckner (11) make still a further breakdown of both rating scales and check lists according to the actual form of the instrument. This classification is meant for supervisory use and has served as a guide for the classification used in this paper.

Rating Scales

According to Barr, Burton, and Brueckner (11, p. 357), there are six types of rating scales in general use for evaluating the work of the teacher. These scales are:

1. Point scales.
2. Graphic scales.
3. Diagnostic scales.
4. Quality scales.
5. Man-to-man comparison scales.
6. Conduct or performance scales.

On the following pages descriptions and examples of each of these scales are given, since each is thought to be important in studying the teacher at work (11, p. 357).

Point scales.--Point scales or score cards as they are sometimes called are according to Alberty and Thayer (2, p. 143) the earliest device for rating and are, according to Reavis and Cooper (50, p. 2) the most widely used rating instrument. They ordinarily consist of a list of qualities commonly associated with good teaching to which point scores have been assigned according to the supposed contributions of each quality to teaching success (2, 11, 23, 39, 42). Beecher's Teaching Evaluation Record (15) is one of the most recent scales of this type. It contains thirty-two statements which are believed by its author to include all criteria of effective teaching commonly indicated in lists of cardinal objectives and pupil needs. An excerpt from this score card has been reproduced in Figure 1.

According to Barr, Burton, and Brueckner three problems confront those interested in developing and using point scales:

First, there is the problem of the selection of traits, characteristics, and qualities representative of teaching success. The traits chosen for use in the scale must be known to characterize good teaching. Second, there is the problem of the description of each trait in such terms that the judgments about it are made objective. The description of such traits is usually highly subjective. And third, there is the problem of the weighing of each trait and the degree of control over it in such a way that the teacher's total score correlates with his observed success as a teacher. This latter condition is not frequently attained (11, p. 361).

THE TEACHING EVALUATION RECORD

1. The teacher is fair and impartial.

The teacher's behavior is consistently unbiased.

0 1 2 3 4

Sample Evidences

Shows no favoritism or partiality; praise and criticism are based on fact; all criticism constructive; no pets; appraisal of pupils fair and reliable; no excessive criticism of individual pupils; maintains the confidence of children.

2. Pupils are happy and cheerful at work and play.

The teacher creates a happy situation so that pupils express a liking for class.

0 1 2 3 4

Sample Evidences

A spirit of shared enthusiasm; spontaneous pupil or parents comment; friendliness and cooperation of the pupils; pupils approach and visit with teacher during their free time; cheerful exchanges of greetings and conversations between pupils and teacher.

3. Pupils are met in a friendly and sympathetic manner.

Teacher is friendly in manner and tone of voice to all pupils; consistently gives attention to individual questions and apparent needs for individual help; sympathetic with failure due to difficulty; is a sympathetic and understanding listener as indicated by:

0 1 2 3 4

Sample Evidences

Teacher's cordiality, kindness, courtesy, and display of good manners is indicated by consideration of pupils' feelings in the presence of the class; minimizing accidents, unfortunate incidents or embarrassing situations; frequent requests for the teacher's help on personal and educational problems; teacher acceptance of and attention to pupils questions even if unrelated to the subject at hand; teacher gives time to help individual pupils.

4. Contributions and efforts of individual pupils are given recognition.

Teacher shows respect for pupil opinion and suggestions.

0 1 2 3 4

Sample Evidences

Expresses interest and gives appropriate commendation to pupil effort even if small. Attention is given to individual comments and problems.

Fig. 1--Excerpt from The Teaching Evaluation Record

Graphic scales.--The graphic scale is quite similar to a point scale or score card. The difference is that the degree of control exercised over each item is portrayed graphically (11, p. 363). The Almy-Sorenson Rating Scale for Teachers (4) is an example of the graphic scale. It is composed of twenty items as follows: resourcefulness, enthusiasm, leadership, co-operation, trustworthiness, honesty, fairness, sympathy, tact, patience, courteousness, love for children, progressiveness, poise, kindness, originality, good humor, helpfulness, promptness, and foresight. Figure 2 is an excerpt from that scale. According to Almy and Sorenson, the graphic aspects of such scales are interesting and worth while, but they in no manner lessen the necessity for carefully choosing, defining, and weighing the aspects of teachers and teaching considered in such instruments of measurement (5).

Diagnostic scales.--A diagnostic scale, as defined by Barr, Burton, and Brueckner (11, p. 363) is a point scale organized around the different aspects of teaching in such a manner as to reveal levels of attainment by reference to different characteristics ordinarily associated in varying degrees to teaching success. "The Latham Description and Analysis of Student Teacher Performance" (40) is a very recent example of the diagnostic scale. It has been developed

1. Resourcefulness—Means for meeting situations and overcoming them.										Basis for Judgment			Score	
10	9	8	7	6	5	4	3	2	1	0				()
Skillfully meets every difficulty.		Usually equal to every difficulty.			Successful in most situations.		Rather mechanical. Often overcome.		Unable to cope with difficulties. Easily "floored."		Definite	General	Inadequate	()
2. Enthusiasm—Lively manifestation of zeal and earnestness.														
10	9	8	7	6	5	4	3	2	1	0				()
Shows lively interest.		"Self starter." Usually interested.			Is moderately zealous.		Quite dead and indifferent.		Dead. Inanimate.		Definite	General	Inadequate	()
3. Leadership—Capacity or ability to instil into action.														
10	9	8	7	6	5	4	3	2	1	0				()
Children manifest whole-hearted response.		Very seldom does teacher fail to activate children.			Ordinarily children are responsive.		Ineffective in conducting children.		Children not responsive. Ignore teacher.		Definite	General	Inadequate	()
4. Cooperation—Collective and concurrent effort or labor.														
10	9	8	7	6	5	4	3	2	1	0				()
Works splendidly with others for common objective.		Works well with others.			Usually cooperative.		"Solo worker." Reluctant in common endeavor.		Completely individualistic. Neither gives nor takes.		Definite	General	Inadequate	()
5. Trustworthiness—Worthy of confidence—Can be depended upon.														
10	9	8	7	6	5	4	3	2	1	0				()
Always dependable. Does work 100%.		Very seldom fails to do work properly.			Generally to be trusted		Uncertain. Spasmodic.		Unreliable. Can't accept trust or duty.		Definite	General	Inadequate	()

Fig. 2--Excerpt from Almy-Sorenson Rating Scale for Teachers

with respect to performance of the student-teacher in the following areas:

1. The student-teacher as a classroom teacher.
2. The student-teacher as a citizen.
3. The student teacher as a member of the profession.

Under each major heading activities are listed that are considered to constitute the tasks of the student teacher in that area. An excerpt from the Latham scale is presented in Figure 3.

Quality scales.--Barr, Burton, and Brueckner (11, p. 364) describe a quality scale as one in which the different degrees of teaching merit are arranged at equal intervals according to a system of scale values from zero merit to perfection. On such a scale each of the different degrees of teaching merit are described in terms of its characteristic aims, methods, and procedures. Brueckner's Scales for the Rating of Teaching Skills (19) are examples of scales of this type. These scales were developed because it was felt that more reliable rating scales might result if they were constructed so as to differentiate between the teacher's method of teaching and his skill in utilizing his method. Barr, Burton, and Brueckner feel this is an important point because "one of the many sources of confusion has to do with the failure of many evaluators to differentiate between the

D. Directs Learning Activities.

Manner of directing learning activities.

(average score:)

(a)

- 1. Consistently utilizes multisensory learning aids; displays adequate preparation and follow up.
- 2. Utilizes multisensory learning aids frequently; usually displays adequate preparation and follow up.
- 3. Occasionally utilizes multisensory aids with inadequate preparation and follow up.
- 4. Seldom utilizes multisensory learning aids, but when used little or no preparation for the experience is evident.
- 5. Utilizes only text book materials in preparation of material.

Comments:

(b)

- 1. Always presents materials in lecture or "reporting" manner.
- 2. Rarely deviates from textbook presentation of materials.
- 3. Sometimes utilizes unique and interesting presentation techniques but more likely to be lecture or recitation.
- 4. Often utilizes unique and interesting presentation of methods. *Tries new ideas frequently.*
- 5. Consistently uses methods appropriate for the purposes of the class.

Comments:

(c)

- 1. Permits discussion by selected pupils or by representatives of specific levels of ability.
- 2. Usually provides for full discussion of subject and related topics.
- 3. Makes provision for limited discussion for an inadequate length of time.
- 4. Consistently provides opportunity for full discussion of subject and related material.
- 5. Makes no provision for discussion of subject and related areas; simply relates or has pupils recite material.

Comments:

(d)

- 1. Consistently utilizes community resources in presentation of materials and in problem solving.
- 2. Frequently utilizes available community resources in presentation of materials and in problem solving.
- 3. Sometimes utilizes community resources in the presentation of materials and in problem solving.
- 4. Utilizes library resources in conjunction with text books in presentation and problem solving.
- 5. Confines presentation of materials and problem solving to text books.

Comments:

Fig 3--Excerpt from "The Latham Description and Analysis of Student Teacher Performance"

method per se as distinguished from control over method" (11, p. 367). Figure 4 is an excerpt from one of Brueckner's scales (the compulsion type) for evaluating the teaching of geography in grades five and six.

This scale is designed to evaluate the teacher's skill in utilizing a method described by Brueckner as follows:

The subject matter is organized wholly in terms of logical arrangement, usually of text-book arrangement. It is presented either orally or by text, with or without some explanation by the teacher. Pupils are expected to study same and learn it by heart. The recitation consists in having the children give back what they have learned. Usually the form in which it is given must be exactly that of the text. Much dependence is placed on repetition, review, and drill. There is complete teacher domination and control, and almost perfect attention because of rigid discipline maintained by teacher by force. Results in terms of knowledge are emphasized. Respect and unquestioning obedience are demanded of children (19, p. 12).

Man-to-man comparison scales.--Man-to-man comparison scales, or human scales as they are called by Barr, Burton, and Brueckner (11, p. 367), were first utilized by industry and the military service in 1917; and according to Alberty and Thayer (2), the first man-to-man comparison scale for teacher evaluation was worked out by Rugg in 1920. The qualities selected for consideration may be identical in many cases to the qualities found on score cards or point scales. Barr, Burton, and Brueckner (11, p. 367) and Alberty and Thayer (2, p. 145-146) state that it differs

*Teacher A**Scale Value 10.0*

The teacher was a rigid disciplinarian. Every child was compelled to keep in perfect order, to sit rigidly in the standard position, to pay absolute attention to everything that was said, and to strive to acquire perfection in all his work.

Every child worked during his study period at his top speed, because the lessons assigned were generally sufficiently long to require it, and the compelling force back of the command made by the teacher to know these important facts served to make everyone sit up and concentrate on what he was doing. On the other hand, if the material was difficult, the lessons assigned were short, so that it was possible to learn them.

Papers were marked with care, every *i* not dotted and every *t* not crossed being noted and later corrected by the pupil. Answers to questions which were not in the exact language of the book were counted wrong, and there were no supplementary readings or discussions. Any child could ask any formal question he wished about anything he did not understand, but the question had to be asked during the study period, not during the recitation.

The teacher was absolutely fair and impartial, knew every pupil's weakness and success, held herself up to the standards set for the class. Deliberate misbehavior was sure to receive swift and vigorous corporal punishment; failure to learn meant additional drill.

There was much well-organized drill and review. Class questioning was vigorous and snappy and enjoyed by the entire class. When the study of France was concluded, the children could answer any question on the continuous list, which the teacher had given, without hesitation, and with no deviation from the words of the text.

*Teacher C**Scale Value 11.5*

The teacher has assigned the subject-matter on France, logically, according to the textbook, stating emphatically that the facts were to be memorized as they were found in their geographies. Cities, rivers, and mountains were to be located on their maps, and the list of questions in the book was to be used for drill work.

The next day the questions were asked rapidly and methodically with no explanations by the teacher. Children who timidly raised their hands for help were ignored. The drill and review work were enjoyed by most of the pupils, and although quite well organized, this part of the lesson was hurried through so rapidly that the slower pupils failed to profit by it. They became a source of annoyance until the most persistent of them were dismissed from the room.

During the class period most of the children were interested and alert, and were able to give back the main facts of the lesson with a good measure of accuracy. The posture of the children was excellent, and the lesson proceeded with snap and precision.

*Teacher E**Scale Value 9.8*

"For the next assignment take pages 118-119, and be ready to answer questions 10 to 20, particularly emphasizing 11, 14, 16, and 18. Look up difficult words in the dictionary and refer to the large map of France in the textbook in locating places wanted in your reading."

Three or four pupils whose inattention the teacher failed to check were required to get their assignment from their neighbors. No connection was made between the previous lesson and the new assignment.

The teacher deviated occasionally from the logical order due to lack of preparation on her part, thus confusing several of the pupils, and as a result time was wasted in getting back on the track. All questions were stressed alike in spite

Fig. 4--Excerpt from Scale for the Rating of Teaching Skill Type I

of the fact that she had asked the pupils to pay particular attention to certain definite ones. No reference was made to the map and dictionary assignment. She stated that answers must be in the exact words of the book, but in four or five instances let inaccuracies slip by. A fair amount of drill was given over part of the work.

She asked questions of most of the pupils, but never worried if she failed to reach three or four of the same pupils each day. Seven or eight of the pupils failed to answer the questions they were asked, and only in two instances did she find out their difficulties. Instead they were marked zero, and someone else was called upon to give the answer. Two pupils were corrected, one for not standing on both feet, the other for leaning on the desk, but no attention was given to incorrect sitting posture of the other children.

At least three-fourths of the class were attentive during the whole period and these learned some answers to most of the questions in the lesson. There was a strong bond of sympathy between the bright pupils and the teacher, but little attention was paid to the lower group, and as a result these pupils came to class reluctantly.

Teacher G

Scale Value 7.1

The teacher, after reminding her 6A geography class that this was their last lesson in the study of France said, "Complete yesterday's lesson, and begin with paragraph No. 1 on page 63, and finish the chapter."

During the recitation the pupils of the class who had recited the previous day and knew that they would not be called upon today slouched in their seats and made no attempt to follow the work. The teacher was constantly nagging at the pupils who failed to respond but gave them no help. Because of this a few pupils disliked her and created as many difficulties as they dared. She meant to be fair in her decisions but in her carelessness she blamed the children for things which they did not do. The drill was very ineffective because it met the needs of so few of the pupils.

The results of the work were general ideas about France and a large mass of vague and often inaccurate information.

Teacher I

Scale Value 5.0

The class had one more day to complete the study of France.

"Get out your books and begin where we left off." Several pupils who did not seem to know where the point was wasted most of the study period thumbing through their texts, because they were afraid to disclose this fact to the teacher and dared not ask a neighbor.

During the recitation that followed, the textbook map-question list furnished the line of least resistance for the teacher. She attempted to ask the questions in their logical order. Frequently she lost her place or asked the same question twice, because it was often necessary to stop the lesson to check disorder in the class, which occurred when she was off her guard. Then, to save time, she skipped two pivotal questions around which the subject was organized with the remark, "We haven't time to take that up now."

Not once was the map on the wall referred to by either teacher or pupils. No attempt was made to check the pupils' answers, as she scarcely waited for them to reply until another point was taken up. Hence many inaccuracies crept in.

Several pupils who failed to answer any questions were given no help, and her only comment was, "It's your own fault; you should never have been promoted to this grade anyway."

After many interruptions and outbursts of disorder the work was only partially covered.

Fig. 4--Continued

from the score card in that the judgments about the degree of control exercised by the teacher over the different qualities selected for consideration are derived by comparisons between the teachers being rated and named individuals previously judged by the raters to be average, inferior, superior, etc.

In general, merit ratings of the teachers' ability are arrived at by comparing the teacher to be rated with the rater's personal standards of teaching ability. Because of its direct comparison features, the human scale furnishes a fairly objective mode of rating teachers. Its chief limitations lie in the difficulty of administering such a scale because of the personal element and in the lack of commonly accepted standards (11, p. 367).

Figure 5 is an example of a man-to-man comparison scale. As is characteristic of such scales, the number of qualities considered is small. In this example the following ten qualities are considered: vitality, general personality, dynamic personality, growth and progressiveness, team work, attitude toward children, preparation, skill in control and management, skill in teaching (techniques), and skill in teaching (results)(9, p. 352).

Conduct or performance scales.--The conduct or performance scales represent another attempt to rate teaching and not teachers. These scales are of two distinct types. One has to do with observation of teacher performance and the other with pupil performance (11). At this point only

MICHIGAN EDUCATION ASSOCIATION
TEACHER RATING CARD—Long Form
HUMAN SCALE METHOD

		1	2	3	4	5	6	7	8	9	10
Quality Groups											
Letters or Scale Words Indicating Degree of Quality	Points Assigned to Scale Step	Vitality	General Personality	Dynamic Personality	Growth and Progressiveness	Team Work	Attitude Toward Children	Prepara- tion	Skill in Control and Man- agement	Skill in Teaching (Techni- que)	Skill in Teaching (Ability)
A or Very Superior.....	50										
B or Superior.....	40										
C or Average.....	30										
D or Inferior.....	20										
E or Very Inferior.....	10										

Teacher..... Experience Years..... Building or Department.....

DIRECTIONS: Use of the "human scale" is strongly urged. (See III and IV on the reverse side of this card). After determining the degree of merit in each quality group, place a dot in each vertical column opposite the proper degree of capability as indicated by the scale words in the left-hand column. (The horizontal lines, not the space between them, indicate the steps of the scale. No dot should ever be placed in the space above the line of the scale step "Very Superior," as that line represents a perfect score.) Connect the dots by lines. If it is desired to weigh qualities double the number of points assigned to each quadruple the number in groups 9 and 10. Interpret the total score as follows: 180-200, E, or very inferior; 200-250, D, or inferior; 250-300, C, or average; 300-350, B, or superior; 350-400, A, or very superior; 400, perfect score.

Total Numerical Rating { (1)..... Date { (1).....
 { (2)..... { (2).....
 { (3)..... { (3).....
 { (4)..... { (4).....
 { (5)..... { (5).....
 { (6)..... { (6).....
 { (7)..... { (7).....
 { (8)..... { (8).....
 { (9)..... { (9).....
 { (10)..... { (10).....
 { (11)..... { (11).....
 { (12)..... { (12).....
 { (13)..... { (13).....
 { (14)..... { (14).....
 { (15)..... { (15).....
 { (16)..... { (16).....
 { (17)..... { (17).....
 { (18)..... { (18).....
 { (19)..... { (19).....
 { (20)..... { (20).....
 { (21)..... { (21).....
 { (22)..... { (22).....
 { (23)..... { (23).....
 { (24)..... { (24).....
 { (25)..... { (25).....
 { (26)..... { (26).....
 { (27)..... { (27).....
 { (28)..... { (28).....
 { (29)..... { (29).....
 { (30)..... { (30).....
 { (31)..... { (31).....
 { (32)..... { (32).....
 { (33)..... { (33).....
 { (34)..... { (34).....
 { (35)..... { (35).....
 { (36)..... { (36).....
 { (37)..... { (37).....
 { (38)..... { (38).....
 { (39)..... { (39).....
 { (40)..... { (40).....
 { (41)..... { (41).....
 { (42)..... { (42).....
 { (43)..... { (43).....
 { (44)..... { (44).....
 { (45)..... { (45).....
 { (46)..... { (46).....
 { (47)..... { (47).....
 { (48)..... { (48).....
 { (49)..... { (49).....
 { (50)..... { (50).....
 { (51)..... { (51).....
 { (52)..... { (52).....
 { (53)..... { (53).....
 { (54)..... { (54).....
 { (55)..... { (55).....
 { (56)..... { (56).....
 { (57)..... { (57).....
 { (58)..... { (58).....
 { (59)..... { (59).....
 { (60)..... { (60).....
 { (61)..... { (61).....
 { (62)..... { (62).....
 { (63)..... { (63).....
 { (64)..... { (64).....
 { (65)..... { (65).....
 { (66)..... { (66).....
 { (67)..... { (67).....
 { (68)..... { (68).....
 { (69)..... { (69).....
 { (70)..... { (70).....
 { (71)..... { (71).....
 { (72)..... { (72).....
 { (73)..... { (73).....
 { (74)..... { (74).....
 { (75)..... { (75).....
 { (76)..... { (76).....
 { (77)..... { (77).....
 { (78)..... { (78).....
 { (79)..... { (79).....
 { (80)..... { (80).....
 { (81)..... { (81).....
 { (82)..... { (82).....
 { (83)..... { (83).....
 { (84)..... { (84).....
 { (85)..... { (85).....
 { (86)..... { (86).....
 { (87)..... { (87).....
 { (88)..... { (88).....
 { (89)..... { (89).....
 { (90)..... { (90).....
 { (91)..... { (91).....
 { (92)..... { (92).....
 { (93)..... { (93).....
 { (94)..... { (94).....
 { (95)..... { (95).....
 { (96)..... { (96).....
 { (97)..... { (97).....
 { (98)..... { (98).....
 { (99)..... { (99).....
 { (100)..... { (100).....

Fig. 5--Reproduction of Teacher Rating Card

the former type is discussed. Sister Higgins' Analysis Chart for Evaluating Some Observable Factors in the Teaching-Learning Act (35) is an example of a scale for rating teacher performance.

According to Barr, Burton, and Brueckner this scale is an attempt:

. . . to introduce more objectivity into the evaluation of teaching through the application of principles to concrete learning and teaching situations. Although the principles of good teaching considered in this application are described only briefly and in only moderately objective terms, the results secured from the investigation seem to indicate that a device such as that employed in this study will increase the objectivity in rating (11, p. 371).

Figure 6 is a reproduction of the Analysis Chart for Evaluating Some Observable Factors in the Teaching-Learning Act (35).

Check Lists

According to Barr, Burton, and Brueckner (11, p. 343) check lists are usually organized either as groups or lists of questions to be answered "yes" or "no" or as lists of activities to be checked as either present or absent. McNerney (42) says that as in the case of rating scales, check lists can be either general or specific in nature. That is, the traits, qualities, or activities that are to be appraised may have reference to teaching in general or to the teaching of some particular subject.

AN ANALYSIS CHART FOR EVALUATING SOME OBSERVABLE FACTORS IN THE TEACHING-LEARNING ACT
The Chart Shows the Five Levels of Each of the Seven Factors

Factors	EVALUATION				
	Inferior	Below Average	Average	Above average	Superior
A. Is there evidence of definite and carefully planned procedure for directing learning?	1 An indefinite procedure for providing necessary information and material	2 A definite usable plan for providing only necessary information and material.	3 A definite usable plan for providing necessary and supplementary information and materials.	4 A definite plan to achieve desirable, attainable goals.	5 A definite highly flexible plan to achieve desirable, attainable goals, and to investigate pupil needs and abilities.
B. Does the teacher take care quickly and correctly the necessary point of contact with the learner?	1 Insensitive to individual or group difficulties	2 Senses the group difficulties only.	3 Senses the group and also individual difficulties.	4 Senses the specific pupil difficulty of group and also of the individual.	5 Anticipates specific pupil difficulties through evaluation of previous pupil responses.
C. Does the teacher guide the learner in the solution of the difficulty?	1 Neglects both the individual and group difficulties.	2 Revises procedure on basis of group difficulties.	3 Revises procedure on basis of both individual and group difficulties.	4 Provides remedial work to aid in the solution of specific pupil difficulties.	5 Guides pupils in thinking ways and means to solve their own specific difficulties.
D. Is there evidence of desirable teacher-pupil relationships?	1 Absence of desirable teacher-pupil relationships.	2 Desirable teacher-pupil relationship present in a slight degree.	3 Desirable teacher-pupil relationship present in a moderate degree.	4 Desirable teacher-pupil relationship present in a considerable degree.	5 Desirable teacher-pupil relationship present in a marked degree.
E. Is there skill in the use of the vernacular?	1 Inaccurate in speech, or indistinct in enunciation.	2 Speaks correctly and distinctly, but without animation.	3 Speaks correctly, distinctly, and forcefully.	4 Speaks correctly, distinctly, with color and force.	5 Speaks fluently with precision, color, force, and simplicity.
F. Is there an established control technique?	1 Mechanical details over emphasized or totally neglected. No pupil initiative. Pupil initiative not balanced by social responsibility.	2 Mechanical details handled efficiently by teacher. Limited pupil initiative, but balanced by social discipline.	3 Mechanical details under teacher guidance but pupil control. Marked pupil initiative balanced by social discipline.	4 Mechanical details under pupil guidance and pupil control. Pupil initiative and responsiveness in efficient group work.	5 Mechanical details under efficient pupil guidance and pupil control. Pupil initiative and control in group work.
G. Is there evidence of pupil growth toward desirable and attainable goals?	1 Evidence of pupil growth in the acquisition of necessary information.	2 Evidence of pupil growth in the acquisition of necessary information and supplementary information.	3 Evidence of pupil growth in the ability to secure and to organize materials while acquiring information.	4 Evidence of pupil growth in analyzing problems and in locating and utilizing materials in solving problems.	5 Evidence of pupil growth in originating and interpreting problems and in utilizing learned techniques in new situations.

Fig. 6--Reproduction of An Analysis Chart for Evaluating Some Observable Factors in the Teaching-Learning Act

Question check lists.--Franzen's Improvement Sheets for the Teaching of High School Subjects (27) are examples of question check lists. An excerpt from his "Improvement Sheet for First Year Algebra" is reproduced as Figure 7.

The "Improvement Sheet for First Year Algebra" is divided into four sections:

1. General aims in the teaching of mathematics.
2. Specific aims in the teaching of algebra.
3. Pupil activity.
4. Teacher activity.

A bibliography for each section is provided at the end of the check list.

Activity check lists.--A slightly different approach has been made by Brueckner and others through what may be called an activity check list. Appraisal with the conventional check list is likely to be highly subjective, because observations and judgments are made in one act. Noting this fact, Barr, Burton, and Brueckner point out:

The idea behind the activity check list is that the evaluation and improvement of teaching might be enhanced if the descriptive facts in the case were first recorded as objectively as possible and then the interpretation of these facts made a separate operation (11, p. 346).

One of Brueckner's activity check lists for recording facts about classes in junior high school social studies is reproduced in Figure 8.

	Yes	No
C. In solving the problems, do the pupils:		
*1. Let the unknown term equal the number of items to be found?	_____	_____
2. Keep the equal signs in vertical order?	_____	_____
*3. Secure all possible results to the problem?	_____	_____
4. Indicate by the abbreviation "Ans." the answer or set of answers?	_____	_____
*5. Check all answers with original equations or equation?	_____	_____
*6. Discuss all results obtained?	_____	_____
*7. Use paper of uniform size for problem solving?	_____	_____
*8. Make neat arrangement of all written work?	_____	_____
D. In the treatment of exercises do pupils:		
1. Arrange literal terms in alphabetical order?	_____	_____
2. Arrange exponents in ascending-descending order?	_____	_____
IV. Teacher Activity		
A. In presenting the new assignment does the teacher:		
*1. Show the connection between the present lesson and those that have preceded?	_____	_____
*2. Encourage the discussion of the problem and its solution?	_____	_____
*3. Lead the class discussion of new terms, words, and symbols?	_____	_____
*4. Explain new terms and symbols in terms of arithmetic knowledge?	_____	_____
*5. Place the solution upon the blackboard as the pupils direct?	_____	_____
*6. Make demonstrations a working-model for the pupils?	_____	_____
*7. Encourage class selection and dictation of demonstration problems?	_____	_____
8. Permit further pupil demonstration with class assistance?	_____	_____
*9. Consider class suggestions in the selection of home-work?	_____	_____
*10. Secure class judgment relative to the number of problems for solution?	_____	_____
*11. Provide suitable problems for retarded pupils?	_____	_____
*12. Provide extra work for the accelerated pupils?	_____	_____
B. In taking up the day's assignment does the teacher:		
1. Encourage pupil cooperation in problem solving?	_____	_____
*2. Have pupils work problems with him rather than for him?	_____	_____
3. Encourage free discussion of each problem explained?	_____	_____
4. Review forgotten fundamentals of arithmetic?	_____	_____
*5. Give "chalk-talks" to members of the class?	_____	_____
*6. Introduce short-cuts wherever possible?	_____	_____
*7. Encourage pupils to challenge the usefulness of problems?	_____	_____
*8. Divide the class into competitive groups for drill purposes?	_____	_____
9. Encourage pupils to help each other in finding mistakes?	_____	_____
10. Formulate simple problems to help overcome difficulties?	_____	_____
C. In securing algebraic motivation does the teacher:		
1. Put homework on a voluntary basis?	_____	_____
*2. Have pupils graph their own progress in problem solving?	_____	_____

Fig. 7--Excerpt from "Improvement Sheet for First Year Algebra"

A SURVEY OF INSTRUCTIONAL PRACTICES AND MATERIALS USED IN OBSERVED
LESSONS IN THE SOCIAL STUDIES IN GRADES 6, 7 AND 8 ²⁶

School City Grade Observed
Years of Experience of Teacher Number of Pupils in Class
Teacher's Training, Normal or T. C., 1, 2, 3, 4. College or University 1, 2, 3, 4, 5.
Directions: Observer please check items below as seen in one social-studies lesson in grades 6, 7, 8. Space is provided for addition of items that seem vital.

I. Objectives (Check one most apparent)

- 1. To determine how completely pupils have mastered facts in text
- 2. To develop effective habits and methods of study
- 3. To develop understanding of current social order
- 4. Finding and using facts for development of topic in unit
- 5. To develop interest of pupils in social study through activities planned and executed by pupils under teacher guidance
- 6. Others such as

II. Instructional procedures used (Check all those occurring)

- 1. Discussion by teacher such as overview, preview, etc
- 2. Extensive questioning by teacher
- 3. Pupils volunteer personal experiences related to topic
- 4. Teacher illustrates topic by reference to personal experience
- 5. Application of general principles to local study situation (Constant Change, etc.)
- 6. Dramatization of materials studied
- 7. Class debate under formally adopted rules
- 8. Discussion of current events
- 9. Reports given on assigned topics studied independently
- 10. Visual stimuli presented
- 11. Listing activities involving mechanical devices (Radio, Instruments etc.)
- 12. Construction of models, mountings, other items

Fig. 8--Excerpt from A Survey of Instructional Practices and Materials Used in Observed Lessons in the Social Studies in Grades 6, 7, and 8

Check lists such as these are meant simply to supply a convenient means of collecting data judged to be significant about the happenings in a learning-teaching situation. When these lists are employed to indicate merely the presence or absence of some important activity, they are referred to as qualitative; when the records are extended to include both the frequency with which the various activities occur and the time consumed by each, they are referred to as quantitative. In activity check lists such as those described above, evaluation is treated, as it should be, as a separate operation (11, p. 348).

Other Rating Instruments

There are four other rating instruments that are, according to Boyce (17) and Reavis and Cooper (50), used in evaluating teachers.

Guided-comment reports.--A guided-comment report asks the rater to write out his comments on a number of leading questions or suggestive topics. According to Reavis and Cooper (50, p. 18) the guided-comment rating is frequently used in connection with point scales, graphic scales, and diagnostic scales.

Characterization reports.--On a characterization report the rater is simply requested to characterize his total impression of the teacher's efficiency with a single descriptive adjective or letter. Reavis and Cooper (50, p. 19) state that sometimes these reports require the rater to justify his mark by explanatory statements, but generally no additional statements are desired. They say further that

the characterization report may also be used in connection with point scales, graphic scales, and diagnostic scales.

Descriptive reports.--On a descriptive report the rater is required to write a paragraph or two describing the efficiency of the teacher. This report is also sometimes used in connection with other rating instruments (50, p. 19).

Ranking reports.--A ranking report is one on which the rater lists the teachers of a school or department in order of excellence, placing the best teacher at the top of the list and continuing until all teachers are included (50, p. 19). Boyce (50, p. 17) reports that the ranking report is sometimes used in conjunction with the characterization report and the descriptive report.

Criticisms of Rating Instruments

According to Reavis and Cooper (50, p. 80), teacher evaluation is a form of testing; and as such, it should be expected to meet standards of validity, reliability, and objectivity which testing specialists have established. Alberty and Thayer (2, p. 142) note that the motive for introducing teacher rating schemes is to substitute for the old method of general impressions of objective and accurate appraisal; and thus, they must be judged in part according

to their success in achieving objectivity and accuracy in results. Much criticism of rating has been in these areas.

In discussing rating, Burton (23, p. 350) points out that ratings may differ with individuals and that ratings by the same individual made at different times may also differ. He attributes this to:

1. Lack of definite meaning attached to some of the elements of the scale.
2. Difficulty of objectifying grades.
3. Human differences in judgment.
4. Errors in the formulation and application of the scales.

According to Burton (23) these criticisms are not logical reasons for the abolition of rating but rather reasons for greater attention to the formulation and application of the scales. Scales worked out co-operatively and discussed freely and frankly will be free from many of these defects. Johnson (37, p. 44) reports that observation with the aid of a rating device or check list appears to be a valuable tool for the estimation of teaching effectiveness.

Reavis and Cooper (50, p. 83) hold that the validity and reliability of rating instruments depends on the nature of the abilities which it is designed to measure. They summarize their findings by saying:

Rating forms, since they fail to control the wide variations which exist in evaluators' personal judgments, cannot be used as valid and reliable measures of either general or specific teaching ability. On the other hand, any form of rating reveals the rater's approval or disapproval of the teacher and is therefore a valid measure of the evaluator's personal reactions to the teacher (50, pp. 88-89).

Barr (10, p. 215), on concluding his latest study of the measurement of teaching efficiency, says that more attention is being given to reliability and validity than was given in earlier studies and that in the devices employed for collecting data the reliabilities seem to be relatively high but the validities relatively unknown. The validity of rating instruments as measures of general teaching ability is challenged by the results of study of eighty-five scales by Reavis and Cooper.

While 168 distinct elements are rated by the different scales, no single scale contains more than 52 items, and 40 per cent of the scales contain not more than ten items. If 168 elements of a teacher's personality and work must be evaluated in order to obtain a complete picture of the teacher, it is evident that no scale contains a sufficient number of items. On the other hand, if fewer than 168 elements are involved in teaching success, many of the scales contain irrelevant material which tends to invalidate them (50, p. 83).

Often times rating instruments are justified on the ground that it is sufficient to select a sample of the attributes needed for successful teaching to obtain a valid measure of teaching success, but again Reavis and

Cooper (50) note exceptions which they consider render this concept useless.

Sometimes a single defect is so serious that it outweighs any accumulation of good qualities in other respects. An adequate evaluation technique must provide for the identification of such isolated but crucial defects. A scale which depends on a sample of abilities cannot do this (50, p. 83).

Along this same line, Rautman (49) cautions that any evaluation based on samplings will be valid only to the extent that the individual samplings upon which it is based are representative of the total situation.

Reavis and Cooper (50, pp. 83-84) note the ambiguity of many items appearing on rating devices, the frequent combinations of unrelated terms, and the failure in many cases to provide definitions for items. They point out that these serve still further to throw doubt on the validity of the devices.

Since the elements of teaching success on many rating devices represent entirely the views of one individual or those of a small group of persons, Reavis and Cooper (50, p. 84) hold that they can be expected to be valid only to the extent of the competence of the individuals who make the instruments. This could be remedied by determining the characteristics of general teaching ability from a wider selection of judges; however, attempts by Charters

and Waples (24), and Hart (34) have yielded a confusion of elements.

In the area of specific teaching ability, Reavis and Cooper (50, p. 85) state that while a particular teacher could not expect similar ratings in different communities using different scales, any one scale should control results regardless of who applied it. However, research conducted by Barr (8) has shown that equally competent persons obtain entirely different results in applying the same scale to the same teachers.

In that study sixty supervisors were asked to rate a teacher on the quality of two demonstration lessons. The instrument used contained twelve items and a summary item to be scored on a ten-point scale. Since ratings were made for each lesson, twenty-six scores were made for the teacher by each supervisor.

In fourteen of the twenty-six ratings these supervisors spread their ratings over the entire ten point scale; in eleven instances their ratings covered nine points, and in only one instance did they show any agreement whatsoever. In rating motivation, for example, twenty supervisors (second observation) said that the motivation was superior, and twenty-one supervisors said that it was very poor. In general merit thirteen of these supervisors rated this teacher as superior (second observation), but thirteen other supervisors rated this same teacher as very poor After the demonstration was over, one group of supervisors commented upon the very poor quality of teaching exhibited; in another group a superintendent of schools made the remark that he wished he might

employ this teacher for the coming school year
If these supervisors had closed their eyes, stopped up their ears, and then had rated these recitations at random upon the twelve items which composed the recitation score card used in this demonstration, their ratings would have been only five per cent poorer than they were when rated according to conventional standards of classroom supervision (8, pp. 6-10).

Broom and Ault (18), Hardesty (33), Nanninga (44), and Sheils (56) have found their rating scales equally inadequate. The findings of these surveys tend to support Anderson (6, p. 14) who says that not only do rating scales tend to emphasize the qualities and behaviors the scale producers have selected as the attributes of good teaching, but they are further affected by the raters own personal ideas as to the behaviors and qualities that are essential to teaching success.

Some investigators have been hopeful that training of the raters might produce better results. According to Ryans, (53, p. 693) observers can, with training and experience, recognize and analyze their prejudices and biases and can eliminate to a large extent the influence of such factors on their interpretation of teaching performance. Van Denburg (58, p. 698) says that a scale is only as good in its final results as the skill of the persons who use it. Shannon, after completing an experiment comparing attention scores, score cards, and informal estimates as means of

measuring teaching efficiency, reports that "on the whole, agreement between team members using the score card was gratifying" (55, p. 507). In this experiment the raters had received training in the use of the score card before the experiment was undertaken. Also, Sister Higgins (35) succeeded, by means of a training program, in reducing the variability of rating by seventy supervisors. Nevertheless, according to Reavis and Cooper:

Training cannot be considered a method of increasing validity of rating devices. It is only a demonstration that a group of evaluators can acquire the same biases and prejudices. There is no guaranty that these biases and prejudices constitute a true or valid standard of teaching ability (50, p. 86).

Baird and Bates (7) compared ratings of 517 reading teachers with the teachers' success in promoting pupil achievement in reading and obtained a correlation of .135. Barr, Torgerson, Johnson, Lyon, and Walvoord (13) obtained correlations ranging from \nearrow .36 to $-$.15, none statistically significant, between ratings for sixty-six teachers on seven scales and gain in pupil achievement. Reavis and Cooper (50, p. 87) point to these as still further evidence that rating instruments are lacking in validity. However, Barr in summarizing a later study says that "where several criteria are employed to measure success in a given respect, . . . the intercorrelations are low.

Apparently the several criteria . . . measure different things"(10, p. 224).

The criticisms which have been brought out in the foregoing paragraphs have been directed mainly toward point scales, graphic scales, and diagnostic scales. However, according to Reavis and Cooper (50, p. 88) the criticisms apply with equal justice to all types of ratings, even though to each type belong certain minor advantages and disadvantages.

For example Monroe and Clark say:

Neither score cards nor man-to-man comparison scales may be expected to yield highly accurate measures of teaching efficiency. Even under the most favorable conditions the probable error of measurement will be so large that serious limitations must be placed on the measure secured. It is, however, worth while to note that the measures yielded by the man-to-man comparison scale will ordinarily be more accurate than those secured by the usual score (45, p. 11).

According to Reavis and Cooper:

A summary review of the various forms of rating discloses no form which can be used as the sole measure of teaching success. Neither does it disclose any form that measures more than the rater's personal opinion of the teacher. As measures of the prestige of the teacher with other persons, rating devices may be used advantageously (50, p. 93).

Rating by Supervisors

Traditionally, according to Reavis and Cooper (50, p. 93), administrators and supervisors are the persons who

have rated teachers. Ryans (53, p. 693) too says the most frequent source of ratings has been supervisors. He believes this is so because it has been common to assume that the teacher's immediate supervisor has the background and acquaintance with the teacher to enable relatively valid and reliable ratings.

The use of teacher rating by supervisors has received much criticism. Not only has rating been criticized from the standpoint of the reliability, validity, and objectivity of the rating instruments themselves, but it has been criticized in theory and practice as a result of the modern concepts of evaluation and supervision which characterize the supervisor as the helper and co-worker of the teacher.

Wiles (61, p. 293) says rating is not satisfactory evaluation because it is the passing of judgment by someone who assumes superior knowledge about the teaching process and the activities being conducted by the teacher. Burton (23) notes that in the hands of the wrong person rating offers unlimited opportunity for abuse. Both Burton (23) and Wiles (61) feel that rating tends to establish a pattern to which the teacher must conform, but Burton feels this might be overcome by introducing into the rating process co-operation between the supervisor and the teacher. According to Grotke (32) there is a direct relationship between the divergency in the points of view held by a

supervisor and a teacher as to what constitutes good teaching and the rating that teacher will receive from the supervisor.

According to Alexander and Halverson (3), Burton (23), Reavis and Cooper (50), and Wiles (61), rating may keep the supervisor from helping the teacher with his problems because to admit weakness or a problem might decrease the possibility of a good rating. To overcome this Knudsen states, "A supervisor's purpose in rating should have no connection with matters of demoting or dismissing teachers until it has been demonstrated that a teacher cannot or will not improve in some essential activity"(39, p. 235).

Rating is also criticized because it is said that teachers do not like it. A study involving sixty New York city schools and schools in twelve other cities reported by Reavis and Cooper (50, p. 12) reveals that 75 per cent of those surveyed were opposed to rating while only 20 per cent were not and 5 per cent did not care.

Although writers on the subject of rating are almost universal in their criticism of it, they do not all go as far as Gans (28) who says that teacher rating scales disquiet or intimidate teachers and no longer befit an informed profession. Beecher (14), Burton (23), and others (9, 11, 42, 57) hold that rating can be constructive, if its

purpose of instructional improvement is understood and accepted by those involved. This understanding and acceptance will be obtained according to Spears (57) if rating systems are developed co-operatively. Beecher (14, p. 271) says that the teachers and supervisors should together decide on the actual behavior and practices of teachers which tend to result in effective teaching. Also he feels they should decide co-operatively which of these evidences can be observed, how they shall be recorded, and what relative importance or weight should be attributed to each. Beecher (14, p. 273) then suggests that any scales which result should be considered as guides for determining presence or absence of the recognized objectives and ideals. McNerney (42) and Burton (23) concur. Spears (57, p. 414) summarizes by saying that the fact that the problem of judging teaching is extremely difficult is not reason that it should be rationalized out of existence with the comment that such rating is undemocratic.

Rating by Pupils

"Of all persons who may apply rating devices to the teacher, pupils appear in the best position to use them to good advantage"(50, p. 94). Spears (57, p. 421) says that pupils will constantly evaluate instruction and that this is quite natural, for they are the ones who actually

experience the instruction. They live with it day in and day out and feel its effectiveness or ineffectiveness. Accordingly, pupil rating was one of the first topics in teacher evaluation to be investigated by research workers, and it has been the topic receiving the greatest attention in recent years, say Reavis and Cooper (50, p. 94). Ryans (53, P. 694) points out that since the teachers' function is to influence pupil behavior, the reactions of those pupils to the teacher should provide data of considerable interest. Following a study of pupil evaluation of classroom procedures, Hirschi (36, p. 356) reports that while the responses do not provide "yes" or "no" answers to the question of whether the pupils are learning a certain subject, they can give a definite impression that the subject is or is not being learned. Albert conducted a study involving 1588 pupil ratings of seventy-eight high school teachers and reports the following conclusions:

1. Pupil ratings of teachers are reliable, valid, practical, and inexpensive.
2. Teachers can be benefited by pupil rating.
3. Pupils are sufficiently consistent in the rating of teachers for the results to be meaningful (1, p. 274).

Bryan (20) and Max (41) also report of profit gained from pupil ratings.

Even though there appears to be wide agreement as to the worth of pupil evaluation of teachers, it is not without

criticism. According to Reavis and Cooper (50, p. 94), the most common objections made against pupil ratings are that they will promote the currying of favor from pupils and that a teacher's prestige with pupils is not a measure of the teacher's true worth or effectiveness. Where the teacher's ability to motivate students is concerned there can be no objection, but it is the fear that teachers will seek pupil favor in unprofessional ways that raises the objection. However, Reavis and Cooper maintain that this is not a real danger because:

Favoritism to certain pupils would no doubt, be reflected in critical ratings from the less favored pupils. Hence any tendency toward favoritism is self-correcting. The winning of pupil favor by lax discipline or by "easy" assignments may well be equally self-correcting, for many pupils are quick to perceive and criticize these subterfuges (50, p. 95).

To those who point out that a teacher's prestige with pupils is not a valid measure of the teacher's worth or effectiveness, Bryan and Yntema have written:

It should be clearly understood that the evaluation of student reactions through ratings does not imply that students are judges of teacher merit or that students know enough about teaching methods to be able to tell the teacher how to teach . . . the rating results are no more of a rating on teacher merit than the results of other paper-and-pencil tests are a test of teacher merit. Of course, if an instructor has striven for a term to teach certain ideas, principles, facts, and abilities and finds that few of his students have mastered these, he may be justified in concluding that his teaching during that period

should be labeled "poor and ineffective." In the same manner, when ratings reveal that most students reacted unfavorably to the teacher's personality and procedures during a given term he may be justified in concluding that changes are needed (50, p. 8).

It may appear that the halo effect (an emotional or affective constant or attitude which causes a judge to rate a given individual in terms of his like or dislike for the given person) would render pupil ratings of little use. The results of a study by Remmers (51, p. 630) discount this fear, however, and point out that high school pupils will invest their teachers with less halo than college students will their instructors.

Also, it may appear that teachers will inevitably be rated high by those students who do well in school and low by those who do not succeed in school. Ward, Remmers, and Schmalzried (59) found, however, that pupils' scholastic standings to not affect their ratings. They computed the correlations between pupils' ratings of teachers and the pupils' scholastic standings, and the results are as follows:

The resultant coefficients . . . ranged from - .214 to .402. The median of 80 coefficients (one for each of two ratings of 40 practice teachers) was .05, a relationship of practically no significance There is practically no difference between ratings of teachers by students in the upper half of the class scholastically and students in the lower half of the class scholastically (59, p. 191).

In agreement with these findings, Eckelberry (25) reports that a study of student ratings at the University of Washington has shown that students with higher grade point ratios show no tendency to assign higher ratings than other students.

According to Reavis and Cooper (50, p. 95), there is the possibility that pupil ratings may become routinized to the point where pupils approach them with frivolous attitudes and to the point where teachers' merits become topics of excessive conversational comment among pupils. These risks, may be minimized, they feel, by avoiding pupil ratings at regular and too frequent intervals. Research by Remmers (51) and Ferguson and Hovde (26) indicates that ratings from twenty-five to forty-nine pupils are sufficient to provide ratings with a high degree of reliability.

Green (31) reports that pupil ratings of teachers can be of inestimable help in improving teaching procedures and practices and course content. Bryan (20) and Ward, Remmers, and Schmalzried (59) agree and report research findings to support their contention. Wiles (61, p. 301) says that pupil evaluation will be of greatest value if it is conducted during the term when changes can be made to improve class operation for the pupils making the suggestions.

One serious drawback to the use of pupil ratings in evaluation programs, according to Reavis and Cooper (50, p. 96), is that their accuracy depends on the willing cooperation of the teacher. While pupil ratings can be obtained without the consent of the teacher, to secure them this way would invite extremely unpleasant relations between the supervisor and the teacher. Green (21, p. 67) feels that results of the pupil ratings should be considered as personal property of the teacher involved, because only when the teacher's dignity and personal worth are fairly secure is he likely to give his full cooperation and examine criticism at its face value.

In the opinion of Reavis and Cooper:

When teachers understand the place of pupil rating in an evaluation program and appreciate its value, they cannot sincerely oppose it. Teachers have so much to gain from evidence which only pupils can supply that pupil rating might well be demanded by teachers as an essential part of . . . evaluation (50, p. 97).

Rating by Associates

Ryans (53, p. 694) reports that the rating of teachers by their colleagues is occasionally used in public schools. According to Reavis and Cooper (50, p. 97), it has received less attention than other types of rating. They suggest that:

Certainly there is little reason for expecting teachers to rate one another on teaching ability, classroom management, or similar items. Teachers cannot observe one another at work frequently enough to have accurate knowledge of these matters. They can, however, indicate whether or not they work successfully together on committees and in the daily routine of the school. Consequently there may be situations in which associate teacher ratings would be useful (50, pp. 97-98).

The result of a study conducted by Albert (1, p. 274) indicates that teachers can report more accurately than administrators pupil reaction to individual teachers.

Self-Rating

According to Alexander and Halverson, "Improvement of teaching occurs as teachers improve themselves, and it cannot be forced by other persons . . ." (3, p. 495). For this reason they, together with Gray (30) and Wiles (61), consider self-evaluation as the basis for all improvement. Wiles explains it as follows:

Participation in the evaluation develops more mature and responsible teachers. When a supervisor makes a judgment about teaching, the responsibility for improving instruction rests with him. He knows what is wrong, and it is his duty to improve it. When teachers make judgments and find themselves unsatisfactory, they are responsible to themselves for improvement.

Self-evaluation centers the full attention of the teacher on the learning situation. Time need no longer be devoted to fooling the supervisor (61, p. 299).

Gray (30, p. 57) points out that self-rating is good evaluation because:

1. It leads to a careful analysis of strong and weak points.
2. It requires resourcefulness usually leading to rapid growth.
3. It establishes a means of effective co-operation between the supervisor and teacher.
4. It leads to a continuous critical study of the problems of teaching frequently omitted.

Ruediger says:

There is only one place where a general rating scale or something that looks like one may be good and that is in the hands of the teacher for self-criticism. Any teacher needs to check up on himself occasionally to see that he is not getting into a rut or letting down at some vital point. A supervisor may also find this function of a scale helpful in observing teachers and in conferring with them (52, pp. 267-268).

Berger (16, p. 101) reports that he has found self-rating especially valuable with beginning teachers and older teachers who are sensitive to criticism and supervision. Burton (23), Knudsen (39), and McNerney (42) also believe that teachers should participate through self-evaluation.

However, if self-evaluation is to be effective, certain conditions must prevail. According to Wiles (61, pp. 299-300) the teacher must see self-evaluation as a procedure for getting more satisfaction out of his job. He maintains,

that no teacher will want to evaluate his teaching if he feels that the results can harm him; and therefore, self-evaluations should never be filed away by someone who might use them as a basis for salary increases or promotions.

Reavis and Cooper (50, p. 73) point out that self-evaluation is sometimes employed in conjunction with supervisory rating. When this is practiced, both forms are usually considered at the same time at a conference between the supervisor and the teacher. According to Beecher (14, p. 272) the co-operative atmosphere of modern supervision and evaluation may be maintained through the encouraging of teachers to use the supervisor's instruments for their own self-appraisal.

Alexander and Halverson (3, p. 502) and Wiles (61, p. 301) are of the opinion that rating scales and check lists used for self-evaluation are of greatest value if they have been developed with the participation of those persons who use them. McNerney (42) concurs in this belief. Alexander and Halverson (3, p. 498) see the greatest weakness in the use of published forms as the lack of guidance for the teacher in understanding the philosophy and practices of good teaching on which the form is based. Despite this weakness they feel that prepared forms are still to be desired over no forms at all.

A study of supervisor ratings, pupil ratings, and self-ratings involving fifty-one military instructors conducted by Webb and Nolan (60) reveals that the self-ratings and the pupil ratings were highly correlated. Webb and Nolan also noted that the more intelligent and more educated instructors were more self-critical.

Techniques Employing Other Types of Records and Recording Devices

For some time before the advent of rating scales and check lists, supervisors were accustomed to making written notes for the purpose of recording happenings in the classroom. Then as demands for more accurate information grew, these written reports included more and more of the happenings of the recitation until a number of fairly elaborate recording devices came into use. The stenographic report is the most elaborate of these. Somewhat less elaborate and more practical is the written diary. More recently the anecdotal record has come into use (11, p. 354).

Stenographic Reports

According to Barr, Burton, and Brueckner (11, p. 354), a stenographic report is a recording of the happenings of a classroom period made by stenographic means. A collection

of stenographic reports of class work has been published by Peterson and Turner (47).

Diaries

The diary record is another device that has been used somewhat extensively for collecting data about the happenings of a class period (11, p. 354). In a diary happenings of a class period are recorded seriatim as a kind of running account of events. The diary record thus retains the element of continuity in the data recorded, a factor which is, according to Barr, Burton, and Brueckner (11, p. 355), frequently important.

Anecdotal Records

According to Barr, Burton, and Brueckner (11, p. 355), much has been written and said about anecdotal records in recent years. However, Przychodzin (48, p. 14) reports that their development and use in the field of teacher evaluation is lagging far behind, and Patterson (46, p. 8) suggests that they should be given consideration over so called objective instruments. As in a diary or any partial record of the happenings observed, there is always a large amount of personal judgment involved in the choice of the items to be recorded on an anecdotal record (11, p. 355). The staff of the College of Education, Ohio State University, Columbus,

Ohio, has developed an anecdotal record form (45). This form begins by supplying a plan for gathering facts on which to base a description of the learning-teaching situation. Then, following a brief statement of the nature and use of the form, the authors supply a list of questions directing the observers' attention to important aspects of the events observed and also supply ruled space for recording anecdotal evidence for three observations under each of eight major headings:

1. The Materials of Instruction.
2. Purposes.
3. Methods.
4. Effectiveness.
5. Pupil Problems.
6. Use of Community Resources.
7. Fostering of Democratic Attitudes and Relationships.
8. Unique Competencies Suggested by the Field of Specialization.

Barr, Burton, and Brueckner say, "The form is an admirable attempt to think more systematically than is usually done about the complexities of teaching" (11, p. 355). Budd (22, p. 87) feels the importance of the anecdotal record as a statement of behavior cannot be over emphasized.

Techniques Employing Classroom Visitation and
Observation Without Rating

According to Wiles (61, p. 303), classroom observation without rating can be used as an evaluational technique for improving instruction, if it is a co-operative undertaking by the teacher and the supervisor. When undertaken in this manner, it becomes a basis for analysis of specifics with which the teacher needs help. Nevertheless, Wiles (61, p. 305) cautions that classroom observations should not be used until rapport has been established between the supervisor and the teacher--until the teacher knows the supervisor and feels secure with him.

When a supervisor does undertake to make classroom visits, Goodwin (29) says that he should make sure that his visit cannot be termed an interruption by the teacher or the pupils. To lessen the possibility of this, the visit should be scheduled in advance and the entry into the classroom made quietly. Wiles (61, p. 306) suggests that the visitor should plan to be in the classroom when the class starts and to stay in the background as much as possible.

While the supervisor is observing the teacher-pupil interaction, Wiles suggests that he ask himself such questions as:

Is the classroom one in which children feel secure in their relationships with each other and with the teacher?

Do the children see purpose in what they are doing?

Are children seeking ways of carrying out their purposes or are they seeking to discover what the teacher wants done?

Is there opportunity for creative thinking and activity in the classroom?

Is co-operation encouraged?

Are children stimulated to evaluate their ways of working and to plan revision of procedure that will make their work more effective?

Are the classroom equipment and materials organized to increase the efficiency with which the group achieves its purposes (61, p. 307)?

Satlow (54) says that a sensible supervisor will compare a teacher with what that teacher is potentially capable of becoming, and not with another teacher. Neither will he hold himself up as an example of how everything is done. According to Goodwin (29), note-taking during visits is permissible if the teacher is given a copy of the notes. He also suggests that the observation period be followed by an informal conference.

Wiles (61, p. 307) agrees that an informal conference should follow observation. It is here the actual evaluation takes place. In addition he says:

The supervisor must not make value judgments concerning what has gone on. His function is to assist the teacher in analyzing the situation and in formulating procedures for improving the work of the class The supervisor should give his opinion when it is called for and he should offer suggestions when the teacher asks for them, but it is not his function to tell the teacher

what should be done. He is a resource person and not a director. He is there to help the teacher grow in self-direction and professional maturity, not to increase the dependence of the teacher on someone else's judgment (61, p. 308).

Wiles (61,p. 308) does not believe that classroom observation is a technique that should be used alone.

Summary

There is a great variety of rating instruments available with which to study the teacher at work, and their use appears to be rather widespread. Research has on the whole shown them to be of doubtful validity and reliability. Modern writers in the field of supervision deplore the administrative use of supervisors' ratings and generally feel they are of little value in the improvement of instruction if imposed from above. The principal value appears to arise from their co-operative development and use. It is generally agreed that pupil-ratings offer a wealth of information if properly obtained. Some writers report the use of rating by associate teachers in such areas as co-operation. There is an increasing emphasis on self-evaluation.

The use of diaries, stenographic reports, anecdotal records, and mechanical recording devices is not so widespread. Relatively little research on their value and use is available.

Classroom observation, when co-operatively carried out, appears as a recognized technique for providing a basis for improvement of instruction.

CHAPTER BIBLIOGRAPHY

1. Albert, H. R., "An Analysis of Teacher Rating by Pupils in San Antonio," Educational Administration and Supervision, XXVII (April, 1941), 267-274.
2. Albery, H. B. and Thayer, V. T., Supervision in the Secondary School, New York, D. C. Heath and Company, 1931.
3. Alexander, William M. and Halverson, Paul M., Effective Teaching in Secondary Schools, New York, Rinehardt and Company, Inc., 1956.
4. Almy, H. C. and Sorenson, Herbert, Almy-Sorenson Rating Scale for Teachers, Bloomington, Illinois, Public School Publishing Company, 1930.
5. _____, "A Teacher Rating Scale of Determined Reliability and Validity," Educational Administration and Supervision, XVI (March, 1930), 179-186.
6. Anderson, Harold M., "A Study of Certain Criteria of Teaching Effectiveness," Journal of Experimental Education, XXIII (September, 1954), 41-71.
7. Baird, James and Bates, Guy, "The Basis of Teacher Rating," Educational Administration and Supervision, XV (March, 1929), 175-183.
8. Barr, A. S., Characteristic Differences in the Teaching Performance of Good and Poor Teachers of the Social Studies, Bloomington, Illinois, Public School Publishing Company, 1929.
9. _____, An Introduction to the Scientific Study of Classroom Supervision, New York, D. Appleton-Century Company, Inc., 1931.

10. _____, "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," Journal of Experimental Education, XVI (June, 1948), 203-283.
11. Barr, A. S., Burton, William H., and Brueckner, Leo J., Supervision, 2nd ed., New York, Appleton-Century-Crofts, Inc., 1947.
12. Barr, A. S. and Jayne, C. D., "The Use of Sound Recording Equipment in the Study and Improvement of Teaching," Journal of Experimental Education, IV (March, 1936), 279-286.
13. Barr, A. S., Torgerson, Theodore L., Johnson, Carl E., Lyon, Vergil E., and Walvoord, Anthony C., "The Validity of Certain Instruments Employed in Measurement of Teaching Ability," The Measurement of Teaching Efficiency, edited for Kappa Delta Pi Research Publications by Helen M. Walker, New York, Macmillan Company, 1935.
14. Beecher, Dwight E., "Judging the Effectiveness of Teaching," Bulletin of the National Association of Secondary School Principals, XXXIV (December, 1950), 270-281.
15. _____, The Teaching Evaluation Record, Buffalo, Educators Publishing Company, 1956.
16. Berger, Max, "Check List for Self-Evaluation in Teaching Skills," Bulletin of the National Association of Secondary School Principals, XXXVI (October, 1951), 101-104.
17. Boyce, A. C., Methods for Measuring Teachers' Efficiency, Fourteenth Yearbook, Part II, Chicago, National Society for the Study of Education, 1915.
18. Broom, M. E. and Ault, J. W., "How May We Measure Teaching Success," Educational Administration and Supervision, XVIII (April, 1932), 250-256.
19. Brueckner, Leo J., Scales for the Rating of Teaching Skill, Educational Research Bulletin, No. 12, Minneapolis, University of Minnesota, 1927.

20. Bryan, Roy C., "Benefits Reported by Teachers Who Obtained Written Student Reactions," Educational Administration and Supervision, XXVIII (January, 1942), 69-75.
21. Bryan, Roy C. and Yntema, Otto, A Manual on the Evaluation of Student Reactions in Secondary Schools, Kalamazoo, Western State Teachers College, 1939.
22. Budd, George F., "The Anecdotal Record as a Technique for Gathering Data on Professional Laboratory Experiences in Student Teaching," The Evaluation of Student Teaching, Twenty-Eighth Yearbook of the Association for Student Teaching, Ann Arbor, Michigan, Edwards Brothers, Inc., 1949.
23. Burton, William H., Supervision and the Improvement of Teaching, New York, D. Appleton and Company, 1923.
24. Charters, W. W. and Waples, Douglas, The Commonwealth Teacher Training Study, Chicago, University of Chicago Press, 1929.
25. Eckelberry, R. H., "The Evaluation of Teaching," Journal of Higher Education, XXVI (April, 1955), 220-221.
26. Ferguson, Harold and Hovde, Herman O., "Improving Teacher Personality by Pupil Rating," School Review, L (June, 1942), 439-443.
27. Franzen, Carl G. F., Improvement Sheets for the Teaching of High School Subjects, Dubuque, Iowa, William C. Brown Company, 1951.
28. Gans, Roma, "How Evaluate Teachers?" Educational Leadership, XIII (November, 1950), 77-81.
29. Goodwin, Jack C., "Techniques of Classroom Visitation," California Teachers Association Journal, XLVIII (April, 1952), 33.
30. Gray, William S., "Rating Scales, Self-Analysis, and the Improvement of Teaching," School Review, XXIX (January, 1921), 49-57.

31. Green, Helen H., "A Workable Teacher Evaluation Practice," Journal of Business Education, XXVII (October, 1957), 67-69.
32. Grotke, Earl M., "Professional Distance and Teacher Evaluation," Phi Delta Kappan, XXXIV (June, 1953), 127-130.
33. Hardesty, Cecil D., "Can Teaching Success Be Rated?" Nation's Schools, XV (January, 1935), 27-28.
34. Hart, Frank W., Teachers and Teaching, New York, Macmillan Company, 1936,
35. Higgins, Sister M. Xavier, Reducing the Variability of Supervisors' Judgment: An Experimental Study, Johns Hopkins University Studies in Education, No. 23, Baltimore, The John Hopkins Press, 1936.
36. Hirschi, L. Edwin, "A Student Evaluation of Classroom Procedures," School Review, LXII (September, 1954), 354-356.
37. Johnson, Granville B., Jr., "An Evaluation Instrument for the Analysis of Teacher Effectiveness," Journal of Experimental Education, XXIII (June, 1955), 331-344.
38. King, L. A., "The Present Status of Teacher Rating," American School Board Journal, LXX (February, 1925), 44-46.
39. Knudsen, Charles W., Evaluation of Teaching, New York, Doubleday, Doran, and Company, Inc., 1932.
40. Latham, D. L., "The Latham Description and Analysis of Student Teacher Performance," unpublished rating scale, School of Business, North Texas State College, Denton, Texas, 1957.
41. Max, Herbert J., "An Experience in Appraisal," Educational Administration and Supervision, XXIX (November, 1943), 505-507.
42. McNerney, Chester T., Educational Supervision, New York McGraw-Hill Book Company, Inc., 1951.

43. Monroe, W. S. and Clark, J. A., Measuring Teaching Efficiency, Educational Research Circular, No. 25, Urbana, University of Illinois, 1925.
44. Nanninga, S. P., "Estimates of Teachers in Service Made by Graduate Students as Compared with Estimates Made by Principal and Assistant Principal," School Review, XXXVI (October, 1928), 622-626.
45. The Ohio Teaching Record, Columbus, Ohio, Ohio State University, 1941.
46. Patterson, Allen D., "The Essentials of a Program of Evaluation and Measurement of Student Teaching," National Business Education Quarterly, XXIII (December, 1954), 5-9.
47. Peterson, H. A. and Turner, E. A., Stenographic Reports of Eighteen Lessons in the Elementary School, Bloomington, Illinois, Public School Publishing Company, 1928.
48. Przychodzin, Joe, "Methods of Evaluating Student Teaching," National Business Education Quarterly, XXIII (December, 1954), 10-15.
49. Rautman, Arthur L., "We All Hate to Be Evaluated," Educational Administration and Supervision, XXXV (December, 1949), 449-460.
50. Reavis, William C. and Cooper, Dan H., Evaluation of Teacher Merit in City School Systems, Supplementary Educational Monograph, No. 59, Chicago, University of Chicago Press, 1945.
51. Remmers, H. H., "Reliability and Halo Effect of High School and College Students' Judgments of Their Teachers," Journal of Applied Psychology, XVIII (October, 1934), 619-630.
52. Ruediger, W. C., "Rating Teachers," School and Society, XX (August 30, 1924), 267-268.
53. Ryans, David G., "The Criteria of Teaching Effectiveness," Journal of Educational Research, XLIII (May, 1949), 690-699.

54. Satlow, I. David, "What Does the Supervisor Look for in His Classroom Visits?" High Points, XXXVI (May, 1954), 19-22.
55. Shannon, J. R., "A Comparison of Three Means for Measuring Efficiency in Teaching," Journal of Educational Research, XXIX (March, 1936), 501-508.
56. Shields, Albert, "The Rating of Teachers in the New York City Public Schools," School and Society, II (November 20, 1915), 752-754.
57. Spears, Harold, Improving the Supervision of Instruction, New York, Prentice-Hall, Inc., 1953.
58. Van Denburg, Joseph K., "An Objective Rating Scale for Teachers of Biology in High School," School and Society, XLVI (November 27, 1937), 698-704.
59. Ward, William D., Remmers, H. H. and Schmalzried, N. T., "The Training of Teaching Personality by Means of Student-Ratings," School and Society, LIII (February 8, 1941), 189-192.
60. Webb, Wilse B. and Nolan, C. Y., "Student, Supervisor, and Self-Ratings of Instructional Proficiency," Journal of Educational Psychology, XLVI (January, 1955), 42-46.
61. Wiles, Kimball, Supervision for Better Schools, Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1955.

CHAPTER IV

EVALUATION OF PUPIL CHANGE

According to Spears (25, p. 428) the first law of evaluation teaches through the appraisal of pupil progress is to know exactly what one is trying to measure. This concept is in accord with the ideas about the process of evaluation discussed in Chapter II. Remmers and Gage (17, p. 124) state that at this point the evaluator should be equipped with a set of objectives stated in the form of grouped, unitary, understandable, and observable changes in pupils, toward which instructional effort has actually been directed, which involve both subject matter and mental processes, and which have been determined by community and individual needs.

The second law, according to Spears (25, p. 428) is that one must know what he is trying to measure to secure the most promising means of doing the job. To accomplish this Remmers and Gage say the following questions must be answered:

1. At what objective or group of objectives is the evaluation device aimed?
2. Which of the following three general types of devices is best suited to evaluate the achievement of these objectives?

a. Devices involving language products, either verbal or mathematical.

b. Devices involving the direct observation of behavior or performance.

c. Devices involving non-language products.

3. If a device involving language is chosen, should it be an essay test or a short-answer [objectively scored] test?

4. If a short-answer test is chosen should it be an externally made, standardized, purchasable test or a teacher-made test?

5. If a teacher-made test is chosen, what types of questions should be used and how should they be apportioned and composed?

6. If an externally-made standardized short-answer test is used, how shall it be chosen from among the many available?

7. If a non-language product or behavior device is used how shall it be constructed?

8. If an essay test is used how shall it be constructed (17, pp. 124-125)?

Since the purpose of this chapter is to consider pupil change as a criterion for judging the effectiveness of teaching and not to explore the complex area of test construction, to consider the answers to all of the above questions would be beyond the scope of this discussion. Nevertheless, an understanding of the technique of evaluating pupil change is necessary to a consideration of pupil change as a criterion for the evaluation of instruction. Therefore, consideration of the first, second, third, fourth, and seventh questions appear to be within the province of this chapter.

Techniques for Appraising Pupil Change

The first question in the quotation above points to the necessity of knowing specifically at what objectives of instruction the evaluation device is to be aimed. As pointed out by Remmers and Gage (17, p. 125) and Schwartz and Tiedeman (21, p. 141), different objectives require different evaluation devices. Consequently they say that the first step in selecting or constructing an evaluation device after the objectives have been formulated is to single out the objective or objectives at which the device is to be aimed. According to Remmers and Gage (17, p. 126) the devices for evaluating these instructional objectives fell into one of three general categories. These categories are:

1. Devices which involve language, either verbal or mathematical.
2. Devices which involve non-language products.
3. Devices which involve direct observation of performance.

Techniques Using Devices Involving Language

According to Remmers and Gage (18, p. 54), devices which involve language are by far the most frequently used type of device for the evaluation of pupil achievement. They further say that "obviously, in the majority of cases,

teachers will find devices involving language most suited to their instructional objectives" (18, p. 55). Barr, Burton, and Brueckner (3), Remmers and Gage (17, 18), Ross (19), and Schwartz and Tiedeman (21) classify language devices into two general groups:

1. The essay test.
2. The new-type or objective test.

Wrightstone, Justman, and Robbins (32, p. 101) add a third group to their classification. It is the oral test, which they feel is very frequently employed by teachers and, therefore, should not be overlooked.

Essay Tests

An essay test is one in which the pupil's response is in the form of a complete sentence or a series of sentences (17, p. 127). Wrightstone, Justman, and Robbins (32, p. 101) point out, that in spite of the widespread use of the newer objective tests, the essay test still constitutes an important aspect of the evaluation of pupil performance. According to Stalnaker (26) the essay examination calls for a relatively free written response to a problem situation in which the written answer, when properly analyzed by the scorer, reveals information regarding selected aspects of the pupil's mental life.

Advantages of essay tests.---Wrightstone, Justman, and Robbins (32, p. 102) state that there are valid educational outcomes which do not readily lend themselves to testing by the new-type objective tests. Schwartz and Tiedeman (21, p. 147) point out that essay tests have been found to be most useful with respect to the following:

1. Objectives which stress the pupil's ability to draw upon, organize, integrate, and/or evaluate their store of knowledge and experience.

2. Objectives which deal with creative writing or originality of expression.

3. Objectives which specify actual writing competence, such as one might find in English and journalism classes.

4. Objectives which have to do with the application and interpretation of facts and principles.

Wrightstone, Justman, and Robbins (32, p. 103) also note the creative character of the essay examination, but they add that until a more precise concept of creative ability is advanced it appears to be unwise to place too much stress on this value of the essay test. Sims (23) holds that essay test answers can furnish clues to the dynamics of the pupil's mental functioning. Wrightstone, Justman, and Robbins (32, p. 103) say that unfortunately the number of instances in which a pupil's responses to the usual essay question lend

themselves to interpretation in terms of psychological adjustment is very small.

Limitations of essay tests.--According to Wrightstone, Justman, and Robbins (32, p. 103) the principal limitation of essay tests is their low reliability. The experiments of Starch and Elliott (27), Hartog (8), and Brown (4) all report a marked variability in the marks assigned to the same paper by two or more readers. Schwartz and Tiedeman (21, p. 143) say that essay tests are subject to bluffing from students who say nothing but say it well. Remmers and Gage (18, p. 57) report that studies by James (11) and Sheppard (22) show that the penmanship of the person being tested has a positive relationship to the score that person received on an essay test. Another criticism, according to Schwartz and Tiedeman (21, p. 141), Remmers and Gage (17, p. 129), and Wrightstone, Justman, and Robbins (32, p. 104) is that the sampling of content, or range of information tested is narrower than it is in objective examinations. Wrightstone, Justman, and Robbins (32, p. 104) point out that this limited scope, in terms of sampling, places a disproportionate emphasis upon correct interpretation of each question by the pupil. Schwartz and Tiedeman (21, p. 141) say the limited number of questions that may be asked also make the essay test an inefficient method of measurement

when time is a factor. Other limitations pointed out by Schwartz and Tiedeman (21, p. 142) are:

1. The amount of the teacher's time that the essay test requires for adequate grading.
2. The likelihood of personal bias affecting the grade despite the most careful precautions by the teacher to prevent it.

Objective Tests

According to Remmers and Gage (17, p. 127) an objective test is one in which the response is a single word, phrase, number, or mark. Ross (19, p. 127) lists the most common types of objective test items as:

1. Simple recall.
2. Completion.
3. Alternate response (true-false).
4. Multiple choice.
5. Matching.

Advantages of objective tests.---Wrightstone, Justman, and Robbins (32, p. 79) state that objective tests, in comparison with essay tests, possess certain definite advantages. One of these advantages is in the area of sampling. Since the response in an objective test is quickly given, it is possible for the pupil to answer more questions in the same amount of time; and as a result, objective tests generally afford much better coverage of

total course content (32, p. 80; 21, p. 141). Remmers and Gage (18, p. 57) note that this greater sampling reduces the danger of chance variations in student preparation and achievement having a great effect on test scores. Another advantage of objective tests is their reliability of scoring, which according to Wrightstone, Justman, and Robbins (32, p. 80) and Remmers and Gage (18, p. 56) is very high. Also, according to Wrightstone, Justman, and Robbins (32, p. 80) and Remmers and Gage (17, p. 134) objective tests may be very easily scored by any person equipped with a key. Remmers and Gage (17, p. 134) note the possibility of scoring objective tests with the aid of an electronic scoring machine. The identification of pupil weaknesses is an easier task too, when objective tests are used. By counting the number of errors made on each question or item of the test, the teacher can readily ascertain the particular elements of course content which show inadequate mastery and arrange for reteaching (32, p. 80). Wrightstone, Justman, and Robbins (32, p. 80) state that the better sampling and identification of pupil errors make it possible to use an objective test as a pretest prior to starting on a new unit of work. Objective tests also have instructional value when used at the end of a unit. The studies of Curtis and Woods (5) and Plowman and Stroud (15) indicate that returning corrected papers to

pupils or having them correct their own papers prior to a discussion of errors results in better ultimate achievement.

Limitations of objective tests.--Wrightstone, Justman, and Robbins (32, p. 81) note that in objective tests where the pupil is called upon to select one of a number of possible alternative answers, a series of fortunate guesses will markedly increase the pupil's score. Remmers and Gage (17, p. 130) say this is particularly true of alternate response or true-false tests, while in the completion type test it is negligible. Various formulas have appeared from time to time designed to reduce the effect of guessing, but Wrightstone, Justman, and Robbins (32, p. 81) feel they are of little value in the course of classroom teaching. They point to a study by Holzinger (10) which shows:

When pupils respond to every item of a test (the usual classroom practice), it can be shown that the relative rank of pupils will be the same whether their scores are computed by simply counting their correct answers or by use of any of the common scoring formulae which correct for guessing (32, p. 81).

Another limitation of the objective test is difficulty of construction. According to Remmers and Gage (17, p. 134), construction of an objective test requires more effort and more time than the construction of an essay test. Wrightstone, Justman, and Robbins (32, p. 81) feel more resourcefulness is required to select the most appropriate objective

form and to apply it to a given question then is required to formulate a few essay questions. Remmers and Gage (17, p. 134) point out though that the amount of effort required by the objective test and the essay test becomes more equal when the scoring or grading is considered. The large number of items which are included on an objective test generally make it necessary to use some mechanical means for reproducing enough copies of the test for class use (32, p. 81). Remmers and Gage (17, p. 135) and Wrightstone, Justman, and Robbins (32, p. 81) note that although most schools now have such mechanical reproducers, some do not. In this case it is necessary to resort to oral presentation of the test. According to Wrightstone, Justman, and Robbins (32, p. 81) several studies (12, 24, 28) have indicated that oral presentation does not significantly decrease the validity or reliability of objective tests. Another limitation of objective tests, according to Wrightstone, Justman, and Robbins (32, p. 82) is in the area of testing complex processes. Remmers and Gage (17) do not agree. They state that "on all other counts except the ability to write essays the short-answer test has been found either equal or superior to the essay test" (17, p. 138).

Oral Tests

In the past, teachers relied very heavily on the oral work of pupils in order to arrive at an estimate

of the extent to which they had mastered the work of the course.(32, p. 110). Even though modern classroom procedure has markedly reduced the opportunities for testing and marking the oral work of pupils, Wrightstone, Justman, and Robbins (32, p. 111) feel the oral test is still significant.

Values of oral tests.--According to Wrightstone, Justman, and Robbins (32, p. 111), oral testing of an individual pupil constitutes an excellent means of following the thought processes. Used this way, the oral test becomes a valuable tool for diagnosis of pupil difficulties. They point out further that skillful questioning by a teacher may help the pupil to apply known scientific information to a new situation or to see implications. In this way the oral test has instructional value. Remmers and Gage (17, p. 498) and Wrightstone, Justman, and Robbins (32, p. 111) agree that another advantage to oral tests is their obvious economy of the time and expense usually involved in reproducing a test. Other advantages of oral testing are:

1. Orally administered tests insure that every pupil will try every item and hence will be measured by every item. Pupils will not waste time on or be stopped by the more difficult test items, thus failing to attempt those at the end of the test.

2. Pupils are constantly stimulated by the oral presentation so that each item constitutes a new challenge more distinctly set off from the preceding items in the test.

3. Oral comprehension ability is tested independently of the pupil's reading ability. Oral comprehension is only slightly if at all less important than reading comprehension to the pupil's success as a worker and citizen. Tests which put a premium on this ability may serve as valuable complements to the usual stress placed upon reading ability in school (17, p. 498).

Limitations of oral tests.--Remmers and Gage (17, p. 498) consider the major limitation of the oral test to be the variable introduced through faulty reading and speaking of the test administrator. Another limitation of the oral test lies in the time it takes to administer it (17, p. 499). According to Wrightstone, Justman, and Robbins:

Oral examinations take too much time to administer; like the essay test, they present the same difficulties of poor sampling and high subjectivity in markings; comparability of questions is difficult to obtain (32, p. 111).

Teacher-Made Tests Versus Standardized Tests

If it is decided that an objective test is best suited to the instructional objectives being evaluated, it must be decided whether that test should be a teacher-made form or a standardized form. According to Knudsen (13, p. 288), the term standardized when applied to a test means that certain careful steps were followed in its construction

that are not followed as well by teachers in the construction of their own tests. Schwartz and Tiedeman say that the term standardized test signifies a measuring instrument with the following six major characteristics:

1. The test is designed to measure important common outcomes of representative courses of study, avoiding items which are likely to be taught in relatively few schools.

2. Specific directions for administering the test have been worked out and are stated in detail, usually providing even the exact words to be used by the examiner and specifying exact time limits. By adhering to the instructions, teachers in many schools can administer the test in essentially the same way.

3. Specific directions are provided for scoring. Usually a scoring key is supplied which reduces scoring to merely comparing the answers with the key; little or nothing is left to the judgment of the scorer. Sometimes carefully selected samples are provided with which the child's product is compared.

4. Norms are supplied to aid in interpreting the scores. Norms, based on administration of the test to large numbers of children, provide a basis for comparing a child's score with representative scores for different ages and grades; or with representative scores of children of his own age or grade.

5. Information needed for judging the value of the test is provided. Before the test becomes available for purchase, research is conducted to produce satisfactory information about the test's reliability and validity.

6. A manual of directions is supplied which explains the purposes and uses of the test, describes briefly how it was constructed, provides specific directions for administering, scoring, and interpreting results, contains tables of norms, and summarizes available research data on the test (21, pp. 262-263).

Remmers and Gage suggest that the decision as to the kind of test best suited to local needs should be influenced by a comparison of the merits and limitations of standardized and teacher-made tests in the light of the following considerations:

1. Closeness of fit to instructional objectives.
2. Refinement of construction.
3. Interpretations possible with each type of test.
4. Expenditures and gains of the teacher (17, p. 139).

Following such a comparison of standardized tests and teacher-made tests, Remmers and Gage summarize as follows:

The chief advantages of standardized tests are their possession of norms and their greater technical refinement. On the other hand, teacher-made tests fit the instructional objectives better, yield greater benefits to the teacher, and are more adaptable to continuous evaluation of achievement through a semester. The more extensive interpretations possible with some standardized tests can be achieved only through careful selection of the tests according to the meaningfulness of their norms; they are most useful in the tool subjects, where the instructional objectives vary least from classroom to classroom and where achievement is to be viewed strictly as a aptitude for prediction of future achievements for guidance purposes. Only where intra-pupil differences in achievement from subject to subject and from year to year are to be studied do standard tests acquire an insurmountable advantage over teacher-made tests. For general comparisons of one pupil with another and of one classroom with another, teacher-made tests can easily be supplied with norms so as to enable worth while comparisons. Our general conclusion must be that, for the major part of evaluation of achievement, teacher-made tests should be used (17, p. 143).

Techniques Using Devices Involving Non-Language Products

According to Remmers and Gage (17, p. 213) and Schwartz and Tiedeman (21, p. 156) there remains a large sphere of educational activity which remains untouched by either objective tests or essay tests. This is the area in which the pupil's achievement is expressed by means of a product, something that is a direct indication of his application of information, skill and understanding. Remmers and Gage (17, pp. 213-213) point out that although need for this type of evaluation is more acute in such fields as industrial arts and home economics, products may constitute an important aspect of achievement in English, science, social studies, and to a lesser degree mathematics. The most common devices for the evaluation of this area of achievement are varied forms of check lists, rating scales, score cards, etc. These devices differ from those discussed in Chapter III mainly in content. Since these instruments are subject to much the same limitations as the check lists, rating scales, score cards, etc. previously discussed, similar care must be exercised in their construction and use (3, 17, 18, 21, 32). Placing instruments of this type in the hands of the pupils for their own self-evaluation would extend the process of evaluation to include the pupils as suggested by McNerney (14).

Techniques Using Devices Involving Direct

Observation of Performance

According to Remmers and Gage (17, p. 127) there are areas in which the achievement of instructional objectives are best evaluated through direct observation of pupil performance or behavior. Some of the objectives that are best evaluated by this technique are the pupil's achievement of desirable attitudes, work habits, and ways of moving and speaking (3, 17, 18, 21, 32). Check lists and rating scales are mentioned by Remmers and Gage (17, p. 127) as being useful as a guide in evaluation through observation. Schwartz and Tiedeman (31, pp. 190-212) and Wrightstone, Justman, and Robbins (32, pp. 123-155) suggest the use of anecdotal records and inventories which are similar to check lists and rating scales. In addition, Barr, Burton, and Brueckner (3, p. 227) suggest that diaries are useful in this type of observational evaluation. Again the instruments employed are essentially the same as the check lists, diaries, and anecdotal records discussed in Chapter III. The main difference is, of course, that they are structured for observing pupils instead of teachers.

Pupil Status and Change as a Criterion

Following a study of teacher evaluation, Anderson (1, p. 44) reports that teacher evaluation experts are almost

universally agreed that the measure of true effectiveness as a teacher is the change that is produced in the pupils taught by the teacher. According to Remmers and Gage (18, p. 479) the general scheme for evaluating the effectiveness of teaching through measurement of pupil change is to secure a measure of the status of the pupils before and after they come under the teacher's influence. Ryans (20, p. 695) says that pupil change would seem to be a criterion in which one could justifiably put his faith. Ross (19, p. 493) asserts it to be the only valid approach. Remmers and Gage (18, p. 479) too point to its validity and state that there can be no argument against the thesis that the teacher who produces desirable changes in his pupils in the most aspects and to the greatest degree is the best teacher. Nevertheless, there are, in the opinion of many writers, serious drawbacks to the use of pupil change in the evaluation of instruction.

Anderson (1, p. 44), while agreeing that pupil change is the only valid criterion of teaching success, points out that it is not reliable. According to Ryans (20, p. 695), the factors that affect pupil change are not at all clear. Remmers and Gage (18, pp. 480-481) give ten factors other than the teacher which they contend affect pupil change. They are:

1. General mental ability of the pupils.
2. Special mental abilities of the pupils.

3. Past educational experiences.
4. The pupil's socioeconomic background and environment.
5. Pupil motivation.
6. The instructional materials which the teacher can use.
7. The amount and quality of supervisory assistance and leadership provided the teacher.
8. The general attitude toward work that characterizes the school as a whole or particular segments of the community.
9. The quality of instruction in areas of the curriculum other than the one for which a given teacher is responsible, because all of a pupil's teachers and instruction in all areas have effect on a pupil's achievement of any specific objectives.
10. The pupil's achievement of objectives other than those evaluated by the test or other devices.

However, Remmers and Gage (18, p. 481) say that before-and-after examinations of pupils do have value. They feel they are reliable and effective for revealing the kinds and amounts of changes that have been produced in pupils over a period of time by all the change producing agencies that affect them.

The evaluation of teaching through the measurement of pupil change is also criticized according to Anderson (1),

Egner (7), McNerney (14), Rautman (16), and Wrightstone (31) because many of the desired outcomes of the modern school are intangible and frequently completely overlooked when such programs are undertaken. They point out though that some of these can be measured to an extent through the use of some of the newer check lists, anecdotal records and the like.

Wiles (30, p. 297) rejects the use of tests in evaluation on the grounds that intelligent pupils will learn in spite of a poor teacher, make a good showing on the test, and consequently make the teacher appear good. Conversely, he notes that, if a class has had poor teachers before or is below average in ability, superior teaching may not result in the attainment of a score high enough to reflect the teacher's ability. Several techniques for equating factors such as varying ability have appeared, but according to Tiegs (29) they have proved unreliable and have fallen into disrepute.

Barr, Burton, and Brueckner (3, p. 204), Anderson (1, p. 44), and Wiles (30, p. 298) have noted that when test results are used in the evaluation of teaching, teachers have a tendency to prepare their students for them. Hartung (9, p. 140) says that if the tests used are good and really measure the instructional objectives of the course, this will not really be harmful. However, he feels that if the teachers understand that the purpose of the testing is to provide a

basis for improvement, the incidence of special preparation will be very low.

In summarizing the value of pupil change as a criterion of teacher effectiveness, Remmers and Gage say:

Because of the difficulty of taking into account all the factors in pupil achievement, this approach to teacher evaluation cannot readily be used either by school administrators or by teachers themselves for self-evaluation (18, p. 481).

Ross (19, p. 493) feels that most competent observers today agree. He is supported by Domas and Tiedeman (6) who report the results of experiments in the attempt to evaluate teacher effectiveness through pupil changes have been largely disappointing. However, a recent summary by Barr notes encouraging progress:

The influence of any particular teacher is deeply enmeshed in a host of other school, pupil, and community factors. While very definite progress has been made in this area, it is not easy to isolate the effects of particular teachers in particular situations. There is reason to be optimistic about the use of more precise instruments of measurement in the management of the teaching personnel, but for the time being, discretion is the best part of valor.

CHAPTER BIBLIOGRAPHY

1. Anderson, Harold M., "A Study of Certain Criteria of Teaching Effectiveness," Journal of Experimental Education, XXIII (September, 1954), 41-71.
2. Barr, A. S., "The Use of Measurement in the Management of Teacher Personnel," Education, LXVI (March, 1946), 431-435.
3. Barr, A. S., Burton, William H., and Brueckner, Leo J., Supervision, 2nd ed., New York, Appleton-Century-Crofts, Inc., 1947.
4. Brown, Clara M., Evaluation and Investigation in Home Economics, New York, Appleton-Century-Crofts, Inc., 1941.
5. Curtis, F. D. and Woods, G. G., "A Study of the Relative Teaching Values of Four Common Practices in Correcting Examination Papers," School Review, XXXVII (October, 1929), 615-623.
6. Domas, S. J. and Tiedeman, D. V., "Teacher Competence: An Annotated Bibliography," Journal of Experimental Education, XIX (December, 1950), 101-218.
7. Egner, Robert E., "Changing Concepts in Evaluation," North Central Association Quarterly, XXIX (April, 1955), 348-350.
8. Hartog, P. and Rhodes, E. C., An Examination of Examinations, International Institute Examinations Inquiry, London, Macmillan and Company, Ltd., 1935.
9. Hartung, Maurice L., "Evaluating Instruction in Mathematics," Bulletin of the National Association of Secondary School Principals, XXXVIII (May, 1954), 138-143.
10. Holzinger, Karl J., "On Scoring Multiple Response Tests," Journal of Educational Psychology, XV (October, 1924), 445-447.

11. James, A. W., "The Effect of Handwriting on Grading," English Journal, XVI (March, 1927), 180-185.
12. Jensen, M. H., "An Evaluation of Three Methods of Presenting True-False Examinations: Verbal, Oral, and Visual-Oral," School and Society, XXXII (November 15, 1930), 675-677.
13. Knudsen, Charles W., Evaluation and Improvement of Teaching, New York, Doubleday, Doran, and Company, Inc., 1932.
14. Mc Nerney, Chester T., Educational Supervision, New York, McGraw-Hill Book Company, Inc., 1951.
15. Plowman, Letha and Stroud, J. B., "The Effect of Informing Pupils of the Correctness of Their Responses to Objective Test Questions," Journal of Educational Research, XXXVI (September, 1942), 16-20.
16. Rautman, Arthur L., "We All Hate to Be Evaluated," Educational Administration and Supervision, XXXV (December, 1949), 449-460.
17. Remmers, H. H. and Gage, N. L., Educational Measurement and Evaluation, New York, Harper and Brothers, 1943.
18. _____, Educational Measurement and Evaluation, rev. ed., New York, Harper and Brothers, 1955.
19. Ross, C. C., Measurement in Today's Schools, 2nd ed., New York, Prentice-Hall, Inc., 1947.
20. Ryans, David G., "The Criteria of Teaching Effectiveness," Journal of Educational Research, XLII (May, 1949), 690-699.
21. Schwartz, Alfred and Tiedeman, Stuart C., Evaluating Student Progress, New York, Longmans, Green, and Company, 1957.
22. Sheppard, E. M., "The Effect of the Quality of Penmanship on Grades," Journal of Educational Research, XIX (February, 1929), 102-105.

23. Sims, Verner M., "The Essay Examination is a Projective Technique," Educational and Psychological Measurement, VII (January, 1948), 15-31.
24. Sims, Verner M. and Knox, L. B., "The Reliability and validity of Multiple-Response Tests When Presented Orally," Journal of Educational Psychology, XXIII (December, 1932), 656-662.
25. Spears, Harold, Improving the Supervision of Instruction, New York, Prentice-Hall, Inc., 1953.
26. Stältnaker, John M., "The Essay Type of Examination," Educational Measurement, edited by E. F. Lindquist, Washington, American Council of Education, 1951.
27. Starch, D. and Elliott, E. C., "Reliability of the Grading of High School Work in English," School Review, XX (September, 1912), 442-457.
28. Stump, N. F., "Listening Versus Reading Method in the True-False Examination," Journal of Applied Psychology, XV (December, 1931), 552-562.
29. Tiege, Ernest W., Tests and Measurements in the Improvement of Learning, Boston, Houghton Mifflin Company, 1939.
30. Wiles, Kimball, Supervision for Better Schools, Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1955.
31. Wrightstone, J. Wayne, "Trends in Evaluation," Educational Leadership, VIII (November, 1950), 91-95.
32. Wrightstone, J. Wayne, Justman, Joseph, and Robbins, Irving, Evaluation in Modern Education, New York, American Book Company, 1956.

CHAPTER V

SUMMARY, CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS

Summary

This study was undertaken to investigate the purpose and techniques of supervisory evaluation of instruction with particular reference to secondary mathematics. The research was confined to the library of North Texas State College.

Modern concepts of evaluation and the purpose and criteria of evaluation of instruction were considered. The evaluation of instruction by the appraisal of teacher activities through the use of the activities listed below was investigated:

1. Rating instruments.
2. Other records and recording devices.
3. Classroom visitation without rating.

An investigation was made of the evaluation of instruction through appraisal of pupil status and change through the use of:

1. Devices involving language,
2. Devices involving non-language products.
3. Devices involving direct observation of pupil performance.

Conclusions

There are two conclusions which were drawn from the research that was reviewed. These conclusions are:

1. That, although recent studies report some progress, there are at present no rating instruments available with scientifically acceptable reliability or validity; and therefore, rating is not at present scientifically acceptable as a criterion of teacher effectiveness.

2. That, although more studies are being devoted to measurement of pupil change with some increasing success, it is not at present a scientifically acceptable criterion of teacher effectiveness.

Implications

The opinions expressed by the writers, whose works were reviewed, in the fields of supervision and evaluation lead to the formulation of the following implications:

1. The sole purpose of supervisory evaluation of instruction should be improvement of instruction.

2. Supervisors should have no part in administrative rating schemes.

3. The rating of teachers by supervisors alone is not constructive.

4. The rating of teachers by pupils can be constructive.

5. The rating of teachers in the area of co-operation by their associates can be constructive.

6. Self-rating by teachers can be constructive.

7. Although their validity and reliability are low, rating instruments should not be abandoned as means for appraising the activities of teachers. Instead continuous effort should be directed toward their improvement.

8. Maximum benefit will result from the use of rating instruments that are evolved and applied by both teachers and supervisors co-operatively.

9. Diaries and stenographic records are very time consuming and therefore of lesser value in appraising teacher activities.

10. The use of anecdotal records can be constructive in certain problem areas.

11. The use of mechanical recording devices can be constructive.

12. Classroom observation can be constructive if undertaken with the support and the co-operation of the teacher being observed.

13. The measurement of pupil status and change can, if undertaken cautiously, provide basis for evaluating teacher effectiveness.

Recommendations

To the end that mathematics education in the public schools of the United States may better meet the demands and the challenge of our modern technological age, the following recommendations are made:

1. The mathematics supervisor should exercise his leadership to enlist the entire department in an active, continuous program of instructional improvement.
2. Part of the in-service education program conducted by the mathematics supervisor should be devoted to the cooperative development of rating scales or check lists to be used for improving instruction in each of the course areas.
3. Part of the in-service education program conducted by the mathematics supervisor should be devoted to the cooperative development of local achievement tests to be used for improving instruction in each of the course areas.
4. Part of the in-service education program conducted by the mathematics supervisor should be devoted to the cooperative selection of standardized achievement tests to be used for improving instruction in each of the course areas.
5. The mathematics supervisor should encourage mathematics teachers to use pupil-rating in a constructive manner.
6. The mathematics supervisor should encourage mathematics teachers to engage in constructive self-rating.

7. The mathematics supervisor should use a mechanical recording device in improving instruction.

8. The mathematics supervisor should provide an opportunity for frequent informal departmental conferences at which teachers may freely discuss the progress of the instructional improvement program.

BIBLIOGRAPHY

Books

- Alberty, H. B. and Thayer, V. T., Supervision in the Secondary School, New York, D. C. Heath and Company, 1931.
- Alexander, William M. and Halverson, Paul M., Effective Teaching in Secondary Schools, New York, Rinehardt and Company, Inc., 1956.
- Almy, H. C. and Sorenson, Herbert, Almy-Sorenson Rating Scale for Teachers, Bloomington, Illinois, Public School Publishing Company, 1930.
- Barr, A. S., Characteristic Differences in the Teaching Performance of Good and Poor Teachers of the Social Studies, Bloomington, Illinois, Public School Publishing Company, 1929.
- _____, An Introduction to the Scientific Study of Classroom Supervision, New York, D. Appleton-Century Company, Inc., 1931.
- Barr, A. S., Burton, William H., and Brueckner, Leo J., Supervision, 2nd ed., New York, Appleton-Century-Crofts, Inc., 1947.
- Beecher, Dwight E., The Teaching Evaluation Record, Buffalo, Educators Publishing Company, 1956.
- Brown, Clara M., Evaluation and Investigation in Home Economics, New York, Appleton-Century-Crofts, 1941.
- Brueckner, Leo J., Scales for the Rating of Teaching Skill, Educational Research Bulletin, No. 12, Minneapolis, University of Minnesota, 1927.
- Bryan, Roy C. and Yntema, Otto, A Manual on the Evaluation of Student Reactions in Secondary Schools, Kalamazoo, Michigan, Western State Teachers College, 1939.

- Burton, William H., Supervision and the Improvement of Teaching, New York, D. Appleton and Company, 1923.
- Charters, W. W. and Waples, Douglas, The Commonwealth Teacher Training Study, Chicago, University of Chicago Press, 1929.
- Franzen, Carl G. F., Improvement Sheet for the Teaching of High School Subjects, Dubuque, Iowa, William C. Brown Company, 1951.
- Hart, Frank W., Teachers and Teaching, New York, Macmillan Company, 1936.
- Higgins, Sister M. Xavier, Reducing the Variability of Supervisors' Judgment: An Experimental Study, Johns Hopkins University Studies in Education, No. 23, Baltimore, Maryland, Johns Hopkins Press, 1936.
- Knudsen, Charles W., Evaluation and the Improvement of Teaching, New York, Doubleday, Doran, and Company, Inc., 1932.
- McNerney, Chester T., Educational Supervision, New York, McGraw-Hill Book Company, Inc., 1951.
- Monroe, W. S. and Clark, J. A., Measuring Teaching Efficiency, Educational Research Circular, No. 25, Urbana, Illinois, Bureau of Educational Research, University of Illinois, 1925.
- The Ohio Teaching Record, Columbus, Ohio, Ohio State University, 1941.
- Peterson, H. A. and Turner, E. A., Stenographic Reports of Eighteen Lessons in the Elementary School, Bloomington, Illinois, Public School Publishing Company, 1928.
- Reavis, William C. and Cooper, Dan H., Evaluation of Teacher Merit in City School Systems, Supplementary Educational Monograph, No. 59, Chicago, University of Chicago Press, 1945.
- Remmers, H. H. and Gage, N. L., Educational Measurement and Evaluation, New York, Harper and Brothers, 1943.
- _____, Educational Measurement and Evaluation, rev. ed., New York, Harper and Brothers, 1955.

- Ross, C. C., Measurement in Today's Schools, 2nd ed., New York, Prentice-Hall, Inc., 1947.
- Schwartz, Alfred and Tiedeman, Stuart C., Evaluating Student Progress, New York, Longmans, Green, and Company, 1957.
- Spears, Harold, Improving the Supervision of Instruction, New York, Prentice-Hall, Inc., 1953.
- Tiegs, Ernest W., Tests and Measurements in the Improvement of Learning, Boston, Houghton Mifflin Company, 1939.
- Wiles, Kimball, Supervision for Better Schools, Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1955.
- Wrightstone, J. Wayne, Justman, Joseph, and Robbins, Irving, Evaluation in Modern Education, New York, American Book Company, 1956.

Articles

- Albert, H. R., "An Analysis of Teacher Rating by Pupils in San Antonio," Educational Administration and Supervision, XXVII (April, 1941), 267-274.
- Almy, H. C. and Sorenson, Herbert, "A Teacher-Rating Scale of Determined Reliability and Validity," Educational Administration and Supervision, XVI (March, 1930), 179-186.
- Anderson, Harold M., "A Study of Certain Criteria of Teaching Effectiveness," Journal of Experimental Education, XXIII (September, 1945), 41-71.
- Baird, James and Bates, Guy, "The Basis of Teacher Rating," Educational Administration and Supervision, XV (March, 1929), 175-183.
- Barr, A. S., "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," Journal of Experimental Education, XVI (June, 1948), 203-283.
- _____, "The Use of Measurement in the Management of Teacher Personnel," Education, LXVI (March, 1946), 431-435.

- Barr, A. S. and Jayne, C. D., "The Use of Sound Recording Equipment in the Study and Improvement of Teaching," Journal of Experimental Education, IV (March, 1936), 279-286.
- Beecher, Dwight E., "Judging the Effectiveness of Teaching," Bulletin of the National Association of Secondary School Principals, XXXIV (December, 1950), 270-281.
- Berger, Max, "Check List for Self-Evaluation in Teaching Skills," Bulletin of the National Association of Secondary School Principals, XXXVI (October, 1951), 101-104.
- Broom, M. E. and Ault, J. W., "How May We Measure Teaching Success," Educational Administration and Supervision, XVIII (April, 1932), 250-256.
- Bryan, Roy C., "Benefits Reported by Teachers Who Obtained Written Student Reactions," Educational Administration and Supervision, XXVIII (January, 1942), 69-75.
- Curtis, F. D. and Woods, G. G., "A Study of the Relative Teaching Values of Four Common Practices in Correcting Examination Papers," School Review, XXXVII (October, 1929), 615-623.
- Domas, S. J. and Tiedeman, D. V., "Teacher Competence: An Annotated Bibliography," Journal of Experimental Education, XIX (December, 1950), 101-218.
- Eckleberry, R. H., "The Evaluation of Teaching," Journal of Higher Education, XXVI (April, 1955), 220-221.
- Egner, Robert E., "Changing Concepts in Evaluation," North Central Association Quarterly, XXIX (April, 1955), 348-350.
- Ferguson, Harold and Hovde, Herman O., "Improving Teacher Personality by Pupil Rating," School Review, L (June, 1942), 439-443.
- Gans, Roma, "How Evaluate Teachers?" Educational Leadership, VII (November, 1950), 77-81.
- Goodwin, Jack C., "Techniques of Classroom Visitation," California Teachers Association Journal, XLVIII (April, 1952), 33.

- Gray, William S., "Rating Scales, Self-Analysis, and the Improvement of Teaching," School Review, XXIX (January, 1921), 49-57.
- Green, Helen H., "A Workable Teacher Evaluation Practice," Journal of Business Education, XXVII (October, 1957), 67-69.
- Grotke, Earl M., "Professional Distance and Teacher Evaluation," Phi Delta Kappan, XXXIV (June, 1953), 127-130.
- Hardesty, Cecil D., "Can Teaching Success Be Rated?" Nation's Schools, XV (January, 1935), 27-28.
- Hartung, Maurice L., "Evaluating Instruction in Mathematics," Bulletin of the National Association of Secondary School Principals, XXXVIII (May, 1954), 138-143.
- Hirschi, L. Edwin, "A Student Evaluation of Classroom Procedures," School Review, LXII (September, 1954), 354-356.
- Holzinger, Karl J., "On Scoring Multiple Response Tests," Journal of Educational Psychology, XV (October, 1924), 445-447.
- James, A. W., "The Effect of Handwriting on Grading," English Journal, XVI (March, 1927), 180-185.
- Jensen, M. H., "An Evaluation of Three Methods of Presenting True-False Examinations: Verbal, Oral, and Visual-Oral," School and Society, XXXII (November 15, 1930), 675-677.
- Johnson, Granville B., Jr., "An Evaluation Instrument for the Analysis of Teacher Effectiveness," Journal of Experimental Education, XXIII (June, 1955), 331-344.
- King, L. A., "The Present Status of Teaching Rating," American School Board Journal, LXX (February, 1925), 44-46.
- Max, Hebert J., "An Experience in Appraisal," Educational Administration and Supervision, XXIX (November, 1943), 505-507.
- Nanninga, S. P., "Estimates of Teachers in Service Made by Graduate Students as Compared with Estimates Made by Principal and Assistant Principal," School Review, XXXVI (October, 1928), 622-626.

- Ostrander, R. H., "Evaluation in the Oak Ridge Schools," Educational Leadership, VII (November, 1950), 86-90.
- Patterson, Allen E., "The Essentials of a Program of Evaluation and Measurement of Student Teaching," National Business Education Quarterly, XXIII (December, 1954), 5-9.
- Plowman, Letha and Stroud, J. B., "The Effect of Informing Pupils of the Correctness of Their Responses to Objective Test Questions," Journal of Educational Research, XXXVI (September, 1942), 16-20.
- Przychodzin, Joe, "Methods of Evaluating Student Teaching," National Business Education Quarterly, XXIII (December, 1954), 10-15.
- Rautman, Arthur L., "We All Hate to Be Evaluated," Educational Administration and Supervision, XXXV (December, 1949),
- Remmers, H. H., "Reliability and Halo Effect of High School and College Students' Judgments of Their Teachers," Journal of Applied Psychology, XVIII (October, 1934), 619-630.
- Ruediger, W. C., "Rating Teachers," School and Society, XX (August 30, 1924), 267-268.
- Ryans, David G., "The Criteria of Teaching Effectiveness," Journal of Educational Research, XLII (May, 1949), 690-699.
- Satlow, I. David, "What Does the Supervisor Look for in His Classroom Visits?" High Points, XXXVI (May, 1954), 19-22.
- Shannon, J. R., "A Comparison of Three Means for Measuring Effectiveness in Teaching," Journal of Educational Research, XXIX (March, 1936), 501-508.
- Sheppard, E. M., "The Effect of the Quality of Penmanship on Grades," Journal of Educational Research, XIX (February, 1929), 102-105.
- Shields, Albert, "The Rating of Teachers in the New York City Public Schools," School and Society, II (November 30, 1915), 752-754.
- Sims, Verner M., "The Essay Examination is a Projective Technique," Educational and Psychological Measurement, VIII (January, 1948), 15-31.

- Sims, Verner M. and Knox, L. B., "The Reliability and Validity of Multiple Response Tests When Presented Orally," Journal of Educational Psychology, XXIII (December, 1932), 656-662.
- Starch, D. and Elliott, E. C., "Reliability of the Grading of High School Work in English," School Review, XX (September, 1912), 442-457.
- Stump, N. F., "Listening Versus Reading Method in the True-False Examination," Journal of Applied Psychology, XV (December, 1931), 552-562.
- Tyler, Ralph W., "Evaluation as a Function of Supervision," Elementary School Journal, XLIV (January, 1944), 264-273.
- Van Denburg, Joseph K., "An Objective Rating Scale for Teachers of Biology in High School," School and Society, XLVI (November 27, 1937), 698-704.
- Ward, William D., Remmers, H. H., and Schmalzried, N. T., "The Training of Teaching Personality by Means of Student-Ratings," School and Society, LIII (February 8, 1941), 189-192.
- Webb, Wilse B. and Nolan, C. Y., "Student, Supervisor, and Self-Ratings of Instructional Proficiency," Journal of Educational Psychology, XLVI (January, 1955), 42-46.
- Wrightstone, J. Wayne, "Trends in Evaluation," Educational Leadership, VIII (November, 1950), 91-95.

Reports

- Hartog, P. and Rhodes, E. C., An Examination of Examinations, International Institute Examinations Inquiry, London, Macmillan and Company, Ltd., 1935.

Publications of Learned Organizations

- Barr, A. S., Torgerson, Theodore L., Johnson, Carl E., Lyon, Vergil E., and Walvoord, Anthony C., "The Validity of Certain Instruments Employed in Measurement of Teaching Ability," The Measurement of Teaching Efficiency, edited for Kappa Delta Pi Research Publications by Helen M. Walker, New York, Macmillan Company, 1935.

Boyce, A. C., Methods for Measuring Teachers' Efficiency,
Fourteenth Yearbook, Part II, Chicago, National
Society for the Study of Education, 1915.

Budd, George F., "The Anecdotal Record as a Technique for
Gathering Data on Professional Laboratory Experiences
in Student Teaching," The Evaluation of Student
Teaching, Twenty-Eighth Yearbook of the Association
for Student Teaching, Ann Arbor, Michigan, Edwards
Brothers, Inc., 1949.

Stalnaker, John M., "The Essay Type of Examination,"
Educational Measurement, edited by E. F. Lindquist,
Washington, American Council on Education, 1951.

Unpublished Materials

Latham, J. L., "The Latham Description and Analysis of
Student Teacher Performance," unpublished rating scale,
School of Business, North Texas State College, Denton,
Texas, 1957.