



6-1958

# A Study of the Factors Involved in the Development and Establishment of Merit Rating Programs in Public School Systems

James L. Keeney

*University of Tennessee - Knoxville*

---

## Recommended Citation

Keeney, James L., "A Study of the Factors Involved in the Development and Establishment of Merit Rating Programs in Public School Systems." PhD diss., University of Tennessee, 1958.  
[https://trace.tennessee.edu/utk\\_graddiss/2930](https://trace.tennessee.edu/utk_graddiss/2930)

This Dissertation is brought to you for free and open access by the Graduate School at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact [trace@utk.edu](mailto:trace@utk.edu).

To the Graduate Council:

I am submitting herewith a dissertation written by James L. Keeney entitled "A Study of the Factors Involved in the Development and Establishment of Merit Rating Programs in Public School Systems." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Educational Administration.

John W. Gilliland, Major Professor

We have read this dissertation and recommend its acceptance:

Galen N. Drewry, Ralph B. Kimbrough, Earl M. Ramer, Lawrence M. DeRidder

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

---

May 12, 1958

To the Graduate Council:

I am submitting herewith a thesis written by James L. Keeney entitled "A Study of the Factors Involved in the Development and Establishment of Merit Rating Programs in Public School Systems." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Educational Administration and Supervision.

John W. Gilliland  
Major Professor

We have read this thesis and  
recommend its acceptance:

Galen N. Draney  
Ralph B. Kimbrough  
Carl M. Lamer  
L. M. DeRudder

Accepted for the Council:

Alvin M. Mantling  
Dean of the Graduate School

A STUDY OF THE FACTORS INVOLVED IN THE DEVELOPMENT  
AND ESTABLISHMENT OF MERIT RATING PROGRAMS  
IN PUBLIC SCHOOL SYSTEMS

---

A THESIS

Submitted to  
The Graduate Council  
of  
The University of Tennessee  
in  
Partial Fulfillment of the Requirements  
for the degree of  
Doctor of Education

---

by  
James L. Keeney

June 1958

## ACKNOWLEDGMENTS

It has been possible to proceed with this study only through the excellent cooperation of many individuals who have given the writer much encouragement and help. Sincere appreciation to all who have assisted in any way is hereby expressed.

A special acknowledgment is made to the members of the writer's sponsoring committee: Professor John W. Gilliland, Chairman, under whose guidance the study was initiated and written; Professor Galen N. Drewry, Professor Ralph B. Kimbrough, Professor Earl M. Ramer, and Professor Lawrence M. DeRidder, who gave much encouragement and assistance.

The writer is indebted to Michael Y. Nunnery and Helen Powers, graduate students, who gave their valuable time for proof reading and other assistance.

Special acknowledgment must be made to the writer's wife, who encouraged, proof read, and gave up many pleasures to enable the writer to pursue a graduate degree.

## TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION . . . . .	1
Statement of the problem . . . . .	4
Delimitations . . . . .	5
Basic assumptions . . . . .	5
Definitions . . . . .	6
Significance of the study . . . . .	7
Related studies . . . . .	14
Procedures . . . . .	22
Organization of the study . . . . .	26
II. HISTORICAL DEVELOPMENT OF MERIT RATING IN EDUCATION . . . . .	28
Early programs . . . . .	28
Merit rating programs since 1940 . . . . .	37
Summary . . . . .	44
III. PROBLEM AREAS IN MERIT RATING . . . . .	47
Implications from industrial research . . . . .	51
Purposes for merit rating . . . . .	58
Merit rating arguments, pro and con . . . . .	61
Common elements in merit programs . . . . .	66
Summary . . . . .	72

CHAPTER	PAGE
IV. CURRENT POLICIES AND PRACTICES IN	
MERIT RATING . . . . .	75
The programs examined . . . . .	81
Problem areas . . . . .	93
Purposes for merit program . . . . .	99
Rating instrument components . . . . .	99
Summary . . . . .	112
V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS .	115
Summary . . . . .	115
Conclusions . . . . .	124
Recommendations . . . . .	126
BIBLIOGRAPHY . . . . .	129
APPENDIX . . . . .	135

## LIST OF TABLES

TABLE	PAGE
I. Types of Salary Programs Available . . . . .	78
II. Use of Rating Results by Responding Systems .	86
III. Other Factors Used with Rating Instruments in Determining Merit . . . . .	88
IV. Estimated Percentages of Personnel and Administrators Approving, Disapproving and Indifferent to the Programs . . . . .	91
V. Frequency and Degree of Difficulty for Problem Areas in Development of Merit Rating Programs . . . . .	95
VI. Frequency and Degree of Difficulty for Problem Areas in Implementation of Merit Rating Programs . . . . .	98
VII. Instructional Skills Category, Components and Frequency of Use by the Responding School Systems . . . . .	102
VIII. Social Relations Category, Components and Frequency of Use by the Responding School Systems . . . . .	103
IX. Personal Characteristics Category, Components and Frequency of Use by the Responding School Systems . . . . .	105



## TABLE

## PAGE

X.	Professional Qualifications Category, Components and Frequency of Use by the Responding School Systems . . . . .	106
XI.	Habits of Work Category, Components and Frequency of Use by the Responding School Systems . . . . .	108
XII.	Non-Instructional School Service Category, Components and Frequency of Use by the Responding School Systems . . . . .	109
XIII.	Pupil Results Category, Components and Frequency of Use by the Responding School Systems . . . . .	111

## CHAPTER I

### INTRODUCTION

With the increasing shortage of teachers and pressures for more competent teachers, the question of more pay to attract trained personnel appears greatest upon the horizon. The mere mention of better pay seems to attract the attention of persons who offer panaceas for the teacher shortage problem. The best known plans are loosely catalogued under the general heading of merit rating.

Merit rating, so used, is simply a method of paying more money to more competent teachers and less money to less competent teachers and is usually based upon some form of evaluation of the teacher by one or more persons. For purposes of this thesis the term will be defined more clearly later; however, as merit rating is generally defined it is rating for pay purposes. In a broad sense any evaluation for any definite purpose is an evaluation of merit, becoming a rating only when a scoring device is used.

It seems to be commonly agreed that merit rating programs are formally developed and adopted plans for the purpose of determining the degree of efficiency in job performance through a device designed for that purpose. Authorities in the area of merit rating do not agree with such an assumption. Increasingly research presents statements to

the contrary, similar to that of Lawrence A. Appley:

Merit rating, like employee communication, is a continuing, day-to-day activity--whether we realize it or not, whether we do it formally or informally--it occurs in every situation where one person is responsible for, or has reason to be interested in, the work of another. It need not be a "program" in order to exist. In fact, as one author observes here, merit ratings probably are made as frequently in washrooms as in personnel offices. We are all constantly evaluating the performance of those about us.<sup>1</sup>

Those who fight for the cause of merit rating in industry present the cause as necessary because of the high degree of specialization. An atmosphere of uncertainty may develop when personnel performing special tasks are evaluated by persons only generally acquainted with the problems involved in the job performance. Schools have increased the scope of the curriculum by adding more and more courses with more and more specialization and have taken on this atmosphere which demands specialized evaluation. As this has occurred evaluation has logically been placed in the hands of professional rather than lay people, since the complexity of the task demanded that trained people perform the duty.

It is professionally sound and necessary to evaluate the teachers' work. Estimates of teachers' work serve the two-fold purpose of being a basis for administrative decisions

---

<sup>1</sup>Lawrence A. Appley, "Foreword," Rating Employee and Supervisory Performance (New York: American Management Association, 1951), p. 5.

and improving instruction. For purposes of administrative decisions administrators need some evaluative guides and skills to facilitate wise decisions concerning initial employment and retention of teachers. Administrators must apply tenure and salary laws, promote, demote, transfer, re-assign and retire teachers for best teacher utilization in order to facilitate pupil learning. Any change in position or status of personnel is usually caused by justifiable reasons; therefore, some system of evaluation appears necessary and sound. Supervisors must know strengths and weaknesses in order to help teachers improve their efforts for the benefit of the pupil's academic accomplishments as one of the purposes for evaluating teacher competency.

Because of these intra-professional reasons and the fact that industry is generally assumed to feel that merit rating is a successful means of evaluating for pay purposes, lay people are becoming interested in the possibilities of a merit pay program for public schools. Public officials who are elected to office, seeking means of maintaining low tax rates and at the same time having good school plants with equally good teachers, are interested in merit pay plans.

The above reasons might account for the great amount of materials published on the subject. There seems to be articles on nearly every aspect of merit rating plans, especially the pros and cons. There are those who find any

plan good and those who find any plan bad. School boards and other persons interested in the tax dollar seem to feel that merit rating can be successfully applied to educators. Teachers, through their professional organizations and privately published articles, seem to insist that before merit rating for pay purposes can be considered, a salary level must be developed which is more realistic and at least a living wage. Teachers believe that a plan of merit pay must be over and beyond this living wage.

With all the studies conducted on the reasons for merit rating, its successes and failures, there is still a great deal of misunderstanding concerning the term, its uses and possibilities. Authorities seem to feel that the problem of how to relate, fairly and successfully, the factor of meritorious competency to acceptable salaries is still one of the largest unsolved problems in the area of teacher personnel.

#### Statement of the Problem

The purpose of this study was to examine the factors involved in the establishment and development of merit rating programs in public school systems, to determine the present status of merit rating policies and practices in the field of education.

### Sub-problems

1. To trace the development of merit rating plans in public education in the United States.
2. To identify common elements of merit rating plans in the field of education.
3. To identify the problems involved in the development and implementation of policies and practices for merit rating plans in educational systems.

### Delimitations

1. This study included only literature covering the past two decades, except for a brief history of the materials to that time.
2. Only data within the field of education or having implications for education were used.
3. Problems in the development and implementation of merit rating were compiled from a check list completed by representatives of the school systems which responded.

### Basic Assumptions

1. There are certain common elements within most plans for merit rating.
2. The common elements will reveal the major ideas for consideration within merit rating.

## Definitions

Merit rating refers to an evaluation of the effectiveness of teaching, based on a definite scale or collection of items, usually accepted as legitimate measures for such purpose and used to determine or as a base for the determination of salary.

Merit evaluation refers to an evaluation of the effectiveness of teaching, based on a definite scale or collection of items, formally adopted for any purpose except as a base for salary.

Fixed salary schedule and permanent salary schedule refer to an established plan for paying salaries according to a definite scale of increases depending on length of service and/or professional preparation.<sup>2</sup>

Position salary schedule refers to a plan adopted for the adjustment of salaries according to the position held.<sup>3</sup>

Acceleration refers to a plan of salary increment, closely akin to the fixed salary schedule, by which individuals are pushed ahead of the regular salary increment for reward of meritorious service or extra duties.

Superior-service maximum and super maximum refer to

---

<sup>2</sup>Carter V. Good, Dictionary of Education (1st ed.; New York: McGraw-Hill Book Company, Inc., 1945), p. 258.

<sup>3</sup>Loc. cit.

plans of salary increments which allow individuals who have reached the maximum salary level to receive additional increments for meritorious service.

### Significance of the Study

A number of school systems throughout the United States have tried, or are trying, the merit rating with varying results. Other systems have other types of ratings which they do not term merit but in which merit seems to be present. Some school systems study merit rating as a possible means of determining pay, but conclude that no present plan is suitable or that valid means of recognizing merit are not available. North Carolina, after a lengthy study, concluded that merit rating was not feasible. Utah is presently making a study on a state-wide basis to determine the possibilities of adopting the merit rating plan for the state.

Business appears to be interested in merit rating for schools in the hope that better education for prospective employees will benefit all concerned. Political interests of the country are vitally concerned in merit rating, as it might apply to educators, possibly because of the tax dollar and a desire to gain the most for each dollar. Professional organizations and individual members within the field of education seem to be opposed to the idea. Before educators will accept the idea of merit rating much thought and work will



be necessary to find a way to determine what constitutes merit as well as a method to reward it. The great shortage of teachers demands that something be done to increase the holding power of the profession and that something be done to induce better qualified people to enter teaching. Those who favor merit rating put forth the argument that it will accomplish this two-fold purpose. Numerous educators question whether it will.

A perusal of the literature on merit rating points up a high degree of inconsistency with respect to the concept of merit rating and its utilization. Studies are presented which concluded that merit rating works, that it does not work, that one type is better than another, that no type is better than another, that students can do ratings, that students can not do ratings. To sum it up, articles can be found to reinforce nearly any opinion of merit rating.

A study is needed which would survey the literature and present ideas in use, the pros and cons and the similarities among the various systems, one which would present problem areas in developing and implementing a plan of merit rating. This study is intended to identify the common elements in merit rating, to point out the difficulties involved in implementing a merit rating program and to define a workable set of policies and practices for educational administration which would be consistent with democratic

concepts of leadership and supervision.

The importance of more studies concerning the nature, status and value of merit rating in education was stressed by the National Education Association, through its National Committee on Teacher Education and Professional Standards report.<sup>4</sup>

Over the years the number of teachers has risen until 1,127,845 were reported in the 1950 census.<sup>5</sup> The total estimated number of teachers for the 1957-58 school term was placed at over 1,300,000 by the National Education Association.<sup>6</sup> The total is so large that teachers can no longer bargain on an individual annual basis. There presently is a trend toward bargaining only at the policy level for the group. School systems are dealing with individuals at the group level, which has helped to revive interest in rating the effectiveness of teachers as a basis for paying salaries according to merit. One reason for this interest is that some school boards seem unwilling to go further in providing

---

<sup>4</sup>National Commission on Teacher Education and Professional Standards, The Professional Standards Movement in Teaching, 1956-1965 (Washington, D. C.: National Education Association of the United States, 1956), pp. 11-13.

<sup>5</sup>U. S. Bureau of the Census, "U. S. Summary," Bulletin P-C1 (Washington: Government Printing Office, 1950), p. 261.

<sup>6</sup>Research Division, Advanced Estimates of Public Elementary and Secondary Schools for the Year 1957-58 (Washington: National Education Association, November 1957), p. 4.

blanket salary increases to teachers because of a desire to hold down expenditures and the tax rate.

In the decades of the past, schools were under a loose system of districts with trustees responsible for hiring, evaluating and paying teachers within the districts. A merit-type evaluation based upon individual bargaining resulted. As teachers rebelled against inequities resulting from individual bargaining, school committees attempted to eliminate inequities but maintain the former basis of bargaining, that of payment for service rendered. The position or preparation salary schedule resulted from this compromise. To some, this also has not proved satisfactory, for no rewards for good service were included. Those who favor merit rating claim that under set plans teachers who have great potential have been reduced to mediocrity because no incentive to do better is provided.

The major concern of critics of the preparational schedule and kindred plans is due to the basic assumption that all persons with equal training are equal in ability. The psychological concept of individual differences would tend to belie this assumption, which is not supported by research.

In a study of the relationship between experience and teaching efficiency, Barthlemess and Boyer<sup>7</sup> studied

---

<sup>7</sup>Harriet M. Barthlemess and Philip A. Boyer, "A Study

2722 Philadelphia teachers. For purposes of the study entire school faculties were used. The principals ranked teachers on their staffs according to efficiency, from highest to lowest. All schools were then put together on the assumption that the distribution of teaching efficiency was the same for each school. For elementary teachers a correlation of .272 was found between experience and efficiency, while a correlation of .355 was found between the same items for junior high teachers. A similar study by Boyce<sup>8</sup> in 1911 resulted in a .43 correlation between teaching efficiency and experience.

In 1928 Davis and French<sup>9</sup> reported a correlation of .23 between teaching experience and ratings of teachers by superintendents and other qualified personnel. A rating scale recommended by the Pennsylvania State Council of Education was the device used.

---

of the Relation Between Teaching Efficiency and Amount of College Credit Earned While in Service," Educational Administration and Supervision, 14:521-535, 1928.

<sup>8</sup>A. C. Boyce, "Qualities of Merit in Secondary School Teachers," Journal of Educational Psychology, 3:144-157, 1912.

<sup>9</sup>S. B. Davis and L. C. French, "Teacher Rating," University of Pittsburgh, School of Education Journal, -3:57, 60-64, 1928.

Knight<sup>10</sup> in the same year reported a correlation of .04 between teaching efficiency and experience. The study was made with the use of ratings by pupils, teachers involved, and supervisors.

An earlier study by Ritter<sup>11</sup> in Indiana found a higher correlation between teaching experience and a success grade which for all purposes was teacher efficiency. Success grades were determined from a list of general characteristics and an estimation of teaching power from which the superintendents estimated teacher effectiveness.

Ackerman<sup>12</sup> summarized literature concerning teaching experience and pupil change in 1954. Of those studies found, the highest correlation of .638 was for that group of teachers having from one to twelve years of experience. After twelve years of experience, efficiency measurements, determined by pupil change, declined.

---

<sup>10</sup>F. B. Knight, "Qualities Related to Success in Teaching," Teachers College, Contributions to Education, No. 120 (New York: Columbia University, 1928).

<sup>11</sup>Elmer L. Ritter, "Rating of Teachers in Indiana," Elementary School Journal, 18:750-756, 1918.

<sup>12</sup>Walter I. Ackerman, "A Critical Evaluation of Pupil Changes as a Criterion of Teacher Competency" (Special paper, Harvard Graduate School of Education, 1954)..

Broom<sup>13</sup> reported that the correlation between the number of units taken in education courses, excluding practice teaching, and estimated classroom teaching worth, was .01 in a group of 243 teachers.

According to the Torgerson and Adams<sup>14</sup> scale of predictive efficiency of coefficients of correlation of varying magnitude, only two of the listed studies have significance. The predictive efficiency for Ritter's study with a correlation of .75 is slightly more than 30 per cent better than chance, while Ackerman's correlation of .638 is slightly more than 20 per cent better than chance.

The studies presented indicate that measurement of teacher efficiency has been difficult and shows little validity. Other studies concerning merit rating in education seem to deal with its overall value, its successes or failures and its sundry items' validity or worthlessness. Little seems to have been done concerning the administrative problems involved in establishing policies, purposes and developing and implementing a program of merit rating.

The present study was made to discover what the problem areas were in merit rating programs used in school systems,

---

<sup>13</sup>M. E. Broom, "A Note on Predicting Teaching Success," Educational Administration and Supervision, 18:64-87, 1940.

<sup>14</sup>Theodore L. Torgerson and Georgia S. Adams, Measurement and Evaluation (New York: The Dryden Press, 1954), p. 445. (Edited by Albert J. Harris)

that some light might be shed on the problem to guide systems considering the use of some type of plan by which teachers would be paid on the basis of meritorious service.

### Related Studies

Studies concerning merit rating are numerous. Of the many available, only a few were found related to this study. Many writers from 1915 to 1940 included in books, chapters devoted to merit rating and formal evaluations. Of those studies and chapters found to relate to this study, only the following have direct relationship.

Young<sup>15</sup> reviewed the area of merit rating by making a study of fifty-nine out of a total of seventy-seven school systems reporting to the National Education Association in 1928, that increments were based upon merit. He covered the following questions:

1. What principles governing the administration of merit rating in teachers salary schedules can be authoritatively justified?
2. How well are these principles being followed in the use of merit in salary schedules?
3. How are the merit-rating plans being accomplished?
4. What are some of the conditions accompanying the use of merit in salary schedules?<sup>16</sup>

---

<sup>15</sup>Lloyd P. Young, The Administration of Merit-Type Teachers Salary Schedules (New York: Teachers College, Columbia University, 1933).

<sup>16</sup>Ibid., p. 6.

For each of the sub-problems Young set up a series of questions with which it was assumed school administrators were necessarily concerned in evaluating their forms for merit rating. Each chapter was to answer the questions considered important. All questions were answered from the reviewed literature.

Twelve justifiable practices and sixteen unjustifiable practices were found.<sup>17</sup> Young<sup>18</sup> concluded with thirteen basic concepts or practices to guide in the establishment of a merit pay plan.

1. The rating plan should be specific and a definite part of the salary schedule.
2. Superior merit should be rewarded.
3. Continued incompetence should be penalized.
4. Merit should be only one of the factors considered in granting salary increments.
5. Teacher accomplishments should be the basis of the rating plan.
6. The scale should be documentary with a standardized system of scoring.
7. The scale should define, set up standards, and make mutually exclusive all the factors entering the rating.

---

<sup>17</sup>Ibid., pp. 96-98.

<sup>18</sup>Ibid., pp. 94-95.



8. A final score should result from at least four different ratings.
9. Not more than five groups or levels should be used.
10. The number in each group should approach the curve of normal probability.
11. Only administrative officers should rate teachers for pay purposes.
12. Rating plans and procedures should be a cooperative enterprise between teachers and administrators.
13. One official should be held responsible for the final rating.

Using the thirteen standards for a rule, Young concluded that (1) the methods of rating teacher efficiency were largely subjective, with a low degree of reliability, and (2) in order to justify the use of merit rating for salary purposes, every precaution should be taken to increase the reliability of the rating scale.<sup>19</sup> The majority of the systems studied made little or no attempt to improve their instruments and the administration of the program.

Young<sup>20</sup> recommended fourteen minimum procedures to be followed by an administrator in attempting to use a merit

---

<sup>19</sup>Ibid., p. 99.

<sup>20</sup>Loc. cit.

type salary schedule:

1. The salary schedule should contain all provisions showing the effect of ratings upon the amount of the teacher's salary.

2. Each recognized level of efficiency should receive a proportional amount for annual increment and maximum salary.

3. The amounts paid should be in line with comparable communities.

4. There should be the same number of increments at all levels with maximum amounts varying according to ratings.

5. Additional training, travel, and experimentation and research should justify salary increments.

6. The rating scale should be based upon teacher accomplishments rather than personality traits.

7. The rating should be a standardized form.

8. The items included should be capable of being measured with standards for each merit group.

9. At least two administrative officers should rate each teacher twice annually.

10. Persons doing the rating should confer to insure common understandings and purposes of goal.

11. Groups should be on three or five levels, with no more than five.

12. Groups should be designated by letters.

13. The number in each group should approach the curve

of probable distribution.

14. The cooperation of teaching personnel should be secured by democratic involvement.

Butsch<sup>21</sup> reviewed approximately fifty references on teacher ratings up to November of 1930. He compiled, under general headings, the various conclusions and findings of selected writers up to that date. The studies revealed the following:

1. Opinions of pupils listed fairness, kindness, instructional skill, good discipline and a sense of humor as the desirable traits for a good teacher.

2. Opinions of educators and others listed discipline, teaching skill, personality, cooperation, and other items as those things which were most important in teaching efficiency.

3. Traits appearing on rating blanks most generally had the following items listed first: (a) teaching technique, (b) discipline, (c) teaching results, (d) personality and many other items of lesser importance.

4. Causes of teacher failure were generally considered to be due to many items but heading the list were: (a) poor instruction, (b) lack of discipline, (c) inability to cooperate, (d) lack of scholarship, (e) lack of preparation, (f) personality and laziness, and (g) lack of sympathy.

---

<sup>21</sup>Russell L. Butsch, "Teacher Rating," Review of Educational Research, 1:99-107, April 1931.

A listing of statistical correlations of scores on rating scales from the various studies of rating reliability had a range of from .32 to .96. The lowest correlation was that of two supervisors and the highest was that of supervisors and other teachers. One study reported a range of correlations from .04 between general merit and health to a correlation of .56 between general merit and ability to maintain order. Another study reported a range of correlations from .18 between general merit and health to .90 between general merit and instructional skill. Half of the coefficients Butsch found were below .60. The same disagreement was found for studies correlating the following:

1. General and professional training.
2. Academic ability and teaching ability.
3. Intelligence and teaching ability.
4. Experience and teaching ability.
5. Age and teaching ability.
6. Salary and teaching ability.
7. Credits earned since the beginning of teaching and teaching ability.
8. Professional tests and teaching ability.

Butsch also found that the literature seemed to indicate that most raters rated too high, especially so where acquaintances were used as raters. More than 40 per cent of the cities in the nation used rating scales of some kind, with

large cities using a higher number.

Barr<sup>22</sup> reviewed published references on the measurement of teaching ability in 1940 for the three previous years and repeated the work for the next three years. Both studies were relatively short and the conclusions were essentially the same, indicating little progress in six years. Barr was interested in three areas: (1) approaches to measurement of teaching, (2) teacher attitudes and adjustment, and (3) analysis of teaching ability. Only twelve usable articles could be found and each was reviewed.

All twelve of the articles used showed an interest in the objective measurement of teaching but little about how to do it. The results in the area were far from satisfactory, because of such factors as inadequacy of the criteria employed in the validation of instruments in the field, the absence of any validated theory of organization of human ability and the fragmentary character of the instruments available.

Barr recommended, on the basis of the material found and the lack of material to indicate otherwise, that there was a need for more comprehensive measures of teaching ability, covering those qualities which compose efficiency, teacher performance and pupil change.

---

<sup>22</sup>A. S. Barr, "Measurement of Teaching Ability," Review of Educational Research, 10:182-184, June 1940.

The follow-up study of 1943 presented by Barr<sup>23</sup> found little additional data. There seemed to be more articles covering more areas or topics but essentially the same information was given and the same conclusions drawn.

Two similar studies were made by Torgerson in 1934<sup>24</sup> and 1937<sup>25</sup>. In the 1934 study Torgerson selected studies of the measurement of desirable teacher traits and gave a statistical evaluation of the ratings. In the 1937 study data were selected which dealt with the measurement of teaching ability for the three-year period following the first study. Both concluded with findings similar to those of Barr and Butsch.

Industrial and educational research on merit rating seem to agree that the rater is of necessity a person with whom rateses must reckon. They further agree that the rater must know what he is doing as well as why and how. One study of teacher rating found a general lowering of scores for anonymous ratings over signed ratings.<sup>26</sup> A correlation of

---

<sup>23</sup>A. S. Barr, "The Measurement and Prediction of Teaching Efficiency," Review of Educational Research, 13: 218-223, June 1943.

<sup>24</sup>T. L. Torgerson, "The Measurement and Prediction of Teaching Ability," Review of Educational Research, 4: 261-266, June 1934.

<sup>25</sup>T. L. Torgerson, "The Measurement and Prediction of Teaching Ability," Review of Educational Research, 7: 242-246, June 1937.

<sup>26</sup>C. H. Smeltzer and R. S. Harter, "Comparison of Anonymous and Signed Ratings of Teachers," Educational Outlook, 8:76-84, January 1934.

.63 to .79 was found between the anonymous and signed ratings.

Investigators of merit rating seem to agree that ratings have much dependence on rater skill and are of little value as instruments of supervision. A merit program seems more effective when administrators and personnel make it a cooperative program and revise the instruments with new ideas and data.

### Procedures

Since the purpose of this study was to determine the present status of merit rating problems and practices, a survey of literature in the area of merit rating and a review of programs now in existence was conducted to achieve the purpose.

Literature concerning merit rating in educational systems and industrial systems which might have implications for education programs were organized and analyzed. Those studies and writings which involved personal experiences in merit rating programs were the primary source. Business and industrial research had much that was considered important and of potential help to educators setting up a program of merit rating because the two have made the most extensive use and study of merit rating.

Published materials dealing with educational programs indicated that some school systems adopted merit rating and

discarded it every year. A group of twenty-five systems were found which were using merit rating plans within the past five years. Since literature seemed to indicate that most merit plans were found in systems of between ten thousand and one hundred thousand population and that larger systems were tending to drop merit programs, it was deemed necessary to contact all systems which might have a program of merit pay. Desiring to check all possible sources, inquiry was made of the twenty-five systems found in literature, and 846 other systems of more than ten thousand population.

The 846 systems were ascertained from Part 2 of the Educational Directory.<sup>27</sup> Postal cards (see appendix) were sent to the total 871 systems, stating the kind of study being conducted and asked the system superintendent: (1) whether or not his system had a merit plan for paying teachers, and (2) whether or not he would respond to a check list on the plan.

Of the 871 school systems contacted, 789 complied. Of the responding group there were fifty-four possible leads, of which only thirty indicated a definite merit rating plan for paying teachers. The remaining group indicated that the plans under which they operated had limitations which might

---

<sup>27</sup>U. S. Department of Health, Education and Welfare, Office of Education, Educational Directory, Part 2, 1956-1957 (Washington: U. S. Government Printing Office, 1957).



not fit the definition of merit accepted for the present study.

From the literature analyzed a check list (see appendix) of problem areas and policy statements was developed. The check list was designed for simple checking in either yes and no columns or checking of statements applicable to the responding system. Two copies of the check list were then mailed to each of the possible respondents along with a letter (for copy, see appendix) requesting that such other information as might be available concerning the individual system's plan be enclosed with the check list returned for the study. The requested information was to supplement and facilitate intelligibility of the check list as it concerned the individual programs.

The major problem of finding the present status of merit rating, specifically its problem areas in administration, development and policies, was divided into three sub-problems to facilitate analysis of the literature and the individual programs responding for study.

The first sub-problem was to trace the development of merit rating plans in school systems of the United States. Merit rating has had a somewhat dubious, fluctuating amount of success in both industry and education. The years since its first popularity and today are fraught with critical writings, many with more emotional appeal than fact.

In sub-problem two it was necessary to trace references to rating and evaluation of individuals for pay purposes in order to see the broad picture of purposes and results. This procedure assisted in the identification of common elements in merit rating plans. Although there are numerous plans of merit rating, it was assumed that there would be much in common among them. Data were gathered from various studies concerning elements of similarity. The similarity revealed among the sundry plans were in general areas of evaluation and in the components of the general areas. These components or items, as they were termed by some research persons, were so numerous that a listing of individual components could become unwieldy.

The third sub-problem was to identify the problems involved in the development and implementation of policies and practices for merit rating. The numerous articles written concerning merit rating seem to indicate little concerning problems of development and implementation for educational programs of merit rating. Problem areas were best analyzed by industry and rather definite implications for education could be drawn.

Educational research simply classed problem areas into the two general areas of administrative policy making and practice. The check list sent to those responding school systems was designed to get at these problem areas.

## Organization of the Study

Chapter I of the study dealt with the significance of the study, the research which has been done and the lack of information concerning areas which create problems for educational administrators needing information to implement a program of merit rating. A section defining terms to facilitate comprehension, a statement of the problem and sub-problems, delimitations and basic assumption were also included in Chapter I.

Chapter II dealt with a review of literature concerning research and general articles which dealt with merit rating in education. The chapter mainly was an historical development of merit evaluation and merit rating in education, broken into chronological order. The chapter was a general rather than specific one.

Chapter III was a specific chapter which dealt with the purposes of merit rating, the problems in development and implementation of a program of merit rating, the criticisms of merit rating and a listing of common elements in merit rating, all as found in literature covering the general area. The literature from which Chapter III is taken covered merit rating in both educational and industrial systems.

Chapter IV dealt with the data collected from the check list and other information available concerning the responding school systems. The material was organized and presented for

study. Common elements found in the systems which responded and the problem areas each of the systems indicated were presented.

Chapter V dealt with the materials presented in Chapters III and IV as a basis for conclusions and recommendations for the development and implementation of merit rating programs.

## CHAPTER II

### HISTORICAL DEVELOPMENT OF MERIT RATING IN EDUCATION

#### Early Programs

Articles appearing in educational journals at times leave the impression that teacher evaluation is a relatively new innovation. Such is not the case.

. . . since time immemorial teachers have been evaluated by the general impression method. They have been "hired and fired" on that basis; likewise they have been paid, promoted and demoted. During recent years, however, the need has been increasingly seen for a more accurate and objective method of evaluation. Consequently, many attempts have been made to supplant the general impression method of evaluation with more qualitative and objective methods. All these attempts have the general purpose of collecting and organizing information on the ability and accomplishments of the teacher and all, therefore, have some value.<sup>1</sup>

Each year teachers make an agreement with the public's school system representative, the school board. This agreement, formally drawn and signed by both teacher and the system, is called a contract and commits the system to pay the teacher a certain sum, usually according to a specified rate or salary schedule, in return for the services of the teacher in tasks assigned or enumerated by the school board or superintendent.

---

<sup>1</sup>Ward G. Reeder, The Fundamentals of Public School Administration (3d ed.; New York: The Macmillan Company, 1951), p. 220.

Even as far back as 1789 accusations of discrimination were thrown at school trustees concerning teacher pay. Benjamin Franklin<sup>2</sup> made the accusation by writing a complaint in which he discussed at length how the masters and trustees had cut the English teacher's pay and increased his duties while increasing the pay and decreasing by one-half the duties of the Latin teacher.

Butts and Cremin state "In general the status and salaries of teachers in the colonial period set the pattern that persists to the present day . . . ." <sup>3</sup> Teachers, then as now, were thought by the general public to need no more than a pittance for their duties. A certain amount of the reward for teaching was thought to be in the knowledge that the teacher had helped another person. This was not enough for the colonial teacher, he further had to control his life to the point of social self imprisonment. Philip Vickers Fithian wrote that it was advisable for teachers to attend church regularly, stay home, pursue the scholarly life, read books and stay totally away from women.<sup>4</sup> Such was the colonial merit pay system.

---

<sup>2</sup>Thomas Woody, Educational Views of Benjamin Franklin, quoted in R. Freeman Butts and Lawrence A. Cremin, A History of Education in American Culture (New York: Henry Holt and Company, 1954), p. 134.

<sup>3</sup>Butts and Cremin, loc. cit.

<sup>4</sup>Ibid., p. 135.

What Mr. Fithian might have said was that the town selectmen's visitations and observations at school were of secondary importance. These visits were to make sure that a teacher was carrying out the wishes of the community in teaching reading, ciphering and writing. Criteria used for evaluation of the teacher's performance in carrying out these duties, varied from one selectman to another. Even if approval of the teacher's efficiency in the classroom were given, the private life had to pass the rigid test set forth by Fithian. Teachers had to please everyone. The class visits usually centered upon the degree of pupil discipline and smoothness of the subject matter recitations during the time of the visit. Teacher's pay was dependent, as was the job, on the extent to which the selectmen were pleased with the teacher's performance.

It is only natural, if one considers the requirements and/or expectations of early teachers, that salary schedules were of the type now called merit-type pay scale. The school committee or trustees explicitly stated the quality of services expected and paid accordingly. Some teachers were forced to go to the courts to get the amount of pay promised.<sup>5</sup>

Teachers individually bargained with school trustees as late as the early 1900's and men always merited more than

---

<sup>5</sup>Reeder, op. cit., p. 220.

women. Even today some systems frankly admit that men have priority over women for summer work or special work of one type or another; however, this type of masculine priority is rapidly falling by the wayside because of pressures from teacher organizations. Rather than payment of a higher salary because of sex and family responsibilities, teachers of special subjects or in hard-to-get-a-teacher areas receive salaries above schedules presently.

Research into published data on teacher rating indicate that the desire for more objective evidence of teacher competency began before the time of the first World War. Alberty and Thayer<sup>6</sup> report the first attempts at around 1910. One of the more publicized plans was that of Philadelphia, introduced in 1920.<sup>7</sup> This plan consisted of three parts or main divisions: (1) Instruction, (2) Management and Cooperation, and (3) Professional Attitude. The weights assigned were fifty, thirty, and twenty, respectively.

The plan was revised after two years and used only with teachers having new or at least different experiences. Two years later, in 1924, the plan was further revised, removing numerical weights and substituting S for satisfactory

---

<sup>6</sup>H. B. Alberty and V. T. Thayer, Supervision in the Secondary School (New York: D. C. Heath and Company, 1931), p. 143.

<sup>7</sup>Loc. cit.



and U for unsatisfactory.

Another of the full scale merit pay programs was that of Gary, Indiana, in 1921. The plan called for all teachers to be classed in one of four categories labeled A, B, C, and D. Each of the four categories represented a level of efficiency. A teacher rated as a B teacher, who had the same training and experience as a teacher rated A, would receive \$500 less in salary. A teacher rated as C would receive \$1000 less than the teacher rated A. Teachers rated as D would be carried at a minimum salary level for two years then dismissed.

The Gary plan limited the number of teachers in each group. Only 25 per cent of the teachers could be in level A. Only 40 per cent could be classed at the B level. Thirty per cent had to be classed in the C level. The remaining 5 per cent had to be classed at D level. Every teacher was rated in absolute order ranging from one to eighty. The principal rated each teacher, deciding whether or not the music teacher should get one point over the mathematics teacher, or two points over the science teacher or perhaps less than either or both. Regardless of how the principal felt, teachers had to be limited to the preordained percentages. The plan was abandoned in 1941.

In 1920 Rugg<sup>8</sup> reported an adaptation of a scale developed by W. D. Scott and used by the United States Army. Rugg adapted this scale for use by teachers for self rating or for supervisors to use in teacher evaluation. It was a man-to-man or face-to-face situation and used a five degree level of grading which was common at the time; best, better than average, average, poorer than average, and poorest. This type of rating was revised by Brueckner Courtes and others in 1927.<sup>9</sup> They divided the original main headings into descriptions of teaching performance, the ideal teacher type description, for the sundry subject matter areas.

A third type of evaluation used in the early days of rating attempted to base its criterion upon objective study of teaching results. The results of teaching were those degrees of successful learning which the pupils had within a specific teacher's subject matter areas. The only known way to study objectively such phenomenon is through the use of standardized tests. Kent<sup>10</sup> presented the best example of this type of teacher evaluation. The plan gave teachers from

---

<sup>8</sup>H. O. Rugg, "Self-Improvement of Teachers Through Self-Rating: A New Scale of Rating Teachers' Efficiency," Elementary School Journal, 20:670-684, May 1920.

<sup>9</sup>Albert and Thayer, op. cit., p. 146.

<sup>10</sup>R. A. Kent, "What Should Teacher Rating Schemes Seek to Measure?" Journal of Educational Research, 2:802-807, December 1920.

fifty to seventy-five points based upon the standardized test results of pupil achievement in knowledge and skills for each area of subject matter and other achievements secured from measurement of the pupil's study habits, attitudes toward work, school government (discipline), school organization, moral questions and life preparation. To the pupil achievement qualities Kent added an evaluation of the teacher as a social worker and awarded from twenty to forty points with a like range in points for efficiency as a teacher. This plan attempted to cover every aspect of the teacher's work and the results of the work. Kent's program was, perhaps, too ambitious for success.

Knight's<sup>11</sup> research into the problem found that the types of merit rating plans used around 1920 were not able to produce an objective rating of a teacher's efficiency, judged either from the standpoint of traits or activities or pupil results. The study concerned the methods of rating teaching, of determining the significant factors in teaching ability and the measurement of such factors. A total of 129 teachers were rated by supervisors and administrators in New York, using a score card. Correlations of coefficients were run of teaching ability scores with intellectual ability,

---

<sup>11</sup>Frederic B. Knight, "Qualities Related to Success in Teaching," Teachers College, Contributions to Education, No. 120 (New York: Columbia University, 1922).

skill in class discipline, voice and many other traits for the effective teachers. After presenting the data the following conclusions were reached.

Common sense would tell us that the correlation between voice, defined on the score card as "voice-pitch, quality, clearness of schoolroom voice," and interest in community is probably zero, but here it was found to be  $+.500$ , while voice and discipline was  $+.438$ , and general intellectual capacity and voice was  $+.625$ . The sizes of the correlations do not correspond to the importance of the relationship.

In other words a judge has a certain opinion of a teacher in toto, and his opinion is given according to his general impression in answer to any significant question about the teacher. Thus, the general estimate may be taken to permeate all particular judgments, and conversely, particular judgments are simply defenses for, or justifications of, general opinion which has thus been held.<sup>12</sup>

Knight concluded that it would seem fair to say ". . . that in judging particular traits, general estimate influences the particular estimate to such a degree that judgments of particular traits are in themselves of little use."<sup>13</sup> Teaching is so complex a task that if one adds to it the complexities of personality and interpersonal relations, one arrives at a multiplicity of complexities almost impossible to measure with any one device.

Alberty and Thayer add to this another complexity, that of scientific methodology in research.

---

<sup>12</sup>Ibid., p. 60.

<sup>13</sup>Ibid., p. 10.

It is a well established principle of scientific method, when conducting an experiment and drawing inferences, that the values of the elements which participate be known. Thus if A, B, and C contribute to the result R, no valid inference can be drawn regarding the value of A until we know the influence exerted by B and C apart from and in conjunction with A.<sup>14</sup>

The supervisor who was required to rate for pay purposes found himself in a peculiar position, ". . . as supervisor he presumably should exercise some influence upon teaching situations. As an objective appraiser of teaching merit, he must scrupulously refrain from participation!"<sup>15</sup> The supervisor who was forced to rate was rating a situation in which he was a partner.

A study of published research in teacher growth and evaluation usually became involved with a discussion of the important relationship between pupil growth and the teacher's efficiency. Argument on this point was perhaps best summarized by Prescott as follows:

The teacher is the ultimate agent of education. No matter what appears in the official courses of study, it is he who sets the daily tasks for the pupils, or who helps them to develop a plan of work. It is he who sanctions or condemns their habits, their attitudes, their personality qualities. . . . It is his philosophy of education put into practice which really matters.<sup>16</sup>

---

<sup>14</sup>Alberty, op. cit., p. 163.

<sup>15</sup>Loc. cit.

<sup>16</sup>Daniel A. Prescott, "The Training of Teachers," Rutgers University Bulletin, Series IX, No. 8, p. 5, quoted in A. S. Barr, William H. Burton, and Leo J. Brueckner, Supervision (New York: D. Appleton-Century Company, Inc., 1947), p. 323.

## Merit Rating Programs Since 1940

Accepting the concepts of teacher importance, the complex task of rating and the difficulty in attempting to measure teacher efficiency, Barr covered the problems involved, using available research and practices up to 1947. The conclusion reached was that

. . . few have the judicial temperament and sense of evidence that one would like for a complex task of this sort. Evaluation, like improvement, is cooperative enterprise involving group action and individual initiative.<sup>17</sup>

In spite of the difficulties involved Barr inferred that merit evaluation was possible and should be done, which is the reverse of the position taken by Alberty and Thayer<sup>18</sup> in an earlier work.

The period following World War II to 1958 seemed to find merit rating in education in approximately the same dilemma as that preceding the war. The discussions presently center around the issues involved and the arguments pro and con, which are numerous. The criterion for evaluation has been refined somewhat and the types of plans can be placed into categories or general headings. Although each plan

---

<sup>17</sup>A. S. Barr, William H. Burton, and Leo J. Brueckner, Supervision (New York: D. Appleton-Century Company, Inc., 1947), p. 323.

<sup>18</sup>Alberty and Thayer, op. cit., p. 164.

will differ slightly from general types, all are similar in purpose and design.

Perhaps the most frequently used method of evaluation of teacher effectiveness or efficiency was through the use of score cards and scales or lists, similar to those of the earlier years of merit rating. There were as many different lists as systems using them. They did appear to have some common characteristics. Rating forms were designed to give the administration documentary evidence for rejecting probationary teachers, placing teachers on tenure, promoting and providing for additional or special tasks. Nearly all systems have some such device for measuring competency among teachers. Some go further than others and attach salary increments to these ratings as industry does and merit pay is then in existence.

Recent studies on merit rating did not show any significant changes in the approaches to merit rating. The differences seem to lie in the refinement of the rating tools. Studies did reveal a decline in the number of systems using merit rating plans. Reavis and Cooper<sup>19</sup> report that they found studies indicating that 57 per cent of the nation's school systems reported rating schemes in 1923 with only

---

<sup>19</sup>William C. Reavis and Dan H. Cooper, "Evaluation of Teacher Merit in City School Systems," Supplementary Educational Monographs, No. 59 (Chicago: University of Chicago, January 1945), p. 15.

46 per cent reporting such in 1930-31 and a decline to 40 per cent by 1940-41. By 1957 the number declined so that Davis<sup>20</sup> found only 7.3 per cent of the reporting systems with merit provisions in operation.

In 1945 Reavis and Cooper<sup>21</sup> completed a study on teacher evaluation dealing with merit in city school systems. Replies were received from 123 out of 488 cities polled. Seventeen replied that no merit rating plan was used within those systems. Two others rated systematically but did not send any forms. Six of the remaining 104 systems used the merit evaluation in determining pay while the other used such ratings for other purposes. From the 104 collected copies of forms and instructions, one of the better studies of teacher evaluation, still usable, was published.

Reavis and Cooper concluded that teacher evaluation must attain the objectiveness of an achievement test in order for any two judges (raters) to arrive at the same result.<sup>22</sup> The researchers held this to be true because it was apparent that teachers who have many outstanding qualities occasionally

---

<sup>20</sup>Hazel Davis, "Facts and Issues in Merit Salary Schedules," The Journal of Teacher Education, 8:131, June 1957.

<sup>21</sup>Reavis and Cooper, op. cit., p. 17.

<sup>22</sup>Ibid., p. 80.



lose their positions because of a single outstanding defect.<sup>23</sup> Based upon this conclusion the recommendation was made that "the cumulative personnel record system meets the criteria of a good method of teacher evaluation better than any known program now in use in city schools."<sup>24</sup> Other research in the field has given some support to this position.

Recent publications by teacher organizations, while agreeing that teachers need evaluation, object to connecting evaluation with pay. The National Education Association has held that there are both scientific (objective) and subjective criteria for judging any plan of rating.<sup>25</sup> Ratings should meet the criteria established for test standardization. There should be a minimum of embarrassment between rater and ratee. Supporting evidence should be given or available. Democratic principles should be adhered to in the establishment and application of any system. Any evaluation should be systematic because such evaluation of teaching makes it more probable that teachers will get more real help in improving themselves.

Teachers feel that any program of evaluation needs good administration, maintaining that "A good system badly

---

<sup>23</sup>Ibid., p. 83.

<sup>24</sup>Ibid., p. 103.

<sup>25</sup>Department of Classroom Teachers and Research Division, Teacher Rating (Washington: National Educational Association, 1954), p. 24.

administered could be more destructive than helpful."<sup>26</sup> Teachers felt that supervisors should not do ratings because the specter of a rating might dampen relations between supervisors and teachers by placing a premium on the absence of problems. Teachers recommend that more than one rating be made by more than one person, that teachers receive a copy of each rating and have a conference concerning each rating. Teachers also wanted to know the visits are for the purpose of an evaluation. Teachers want evidence for any judgments made. All things considered the Classroom Teachers Organization recommended as superior to all other forms of evaluation, the cumulative record file, the best, according to their research.

The Association for Supervision and Curriculum Development (A.S.C.D.) recommended a cooperative enterprise involving pupils, school people and lay citizens for curriculum development as the best solution to teacher evaluation plans, because such involvement will begin where teachers are, help them to evaluate, see errors and improve instruction effectiveness.<sup>27</sup>

---

<sup>26</sup>Ibid., p. 11.

<sup>27</sup>Commission on Teacher Evaluation, Association for Supervision and Curriculum Development, Better Than Rating (Washington, D. C.: National Education Association, 1950), p. 67.

The American Federation of Teachers (A.F.T.) concluded that merit rating creates a false salary maximum for a few, undermines morale, and pits teacher against teacher. The A.F.T. recommends a salary schedule set for all and based upon the idea that the best teaching is attainable by keeping teachers' loads at a reasonable minimum and by sympathetic and understanding supervision.<sup>28</sup>

The American Association of School Administrators held that teachers and other school personnel should be paid what they are worth.<sup>29</sup> Recognizing that up until the statement was issued, 1958, that no sufficiently valid instrument or procedure had been developed which would justify a general adoption of salary schedules based on individual merit ratings, the organization felt that it was within the realm of possibility and should not be discarded. The association strongly urged accelerated, systematic experimentation in teacher evaluation to the end that professional pay could be attached to professional ratings of merit. Along with the backing given to experimental programs of merit rating, the association cautioned the profession against adamant opposition to such experimentation lest it place the supposed interests of

---

<sup>28</sup> Mary Herrick, Merit Rating--A Dangerous Mirage (Chicago: American Federation of Teachers, 1956).

<sup>29</sup> "Looking Forward," The Nations Schools, 61:41, April 1958. (Editorial page)

the profession above those of the public. The association also cautioned lay groups against using a concept of merit pay as a subterfuge by which they could oppose paying teachers what they are worth.

Chandler and Mathis<sup>30</sup> after a study of 614 teachers, half of whom were in merit pay systems and half of whom were used for a control group on a matching basis, concluded that: (1) merit rating is not detrimental to teacher morale; (2) a significant difference in morale exists between schools as measured by the Chandler-Mathis Attitude Inventory; (3) the difference in morale was not significant when an arithmetic mean was found for merit rating schools as opposed to set salary schedule schools; and, (4) morale is a function of many variables and is not predictable.

Elsbree and Reutter<sup>31</sup> opposed merit rating and recommended the use of group evaluation, such an evaluation to be completed by a special team which would assess progress along a minimum of three lines: (1) staff growth; (2) pupil achievements; and (3) community school relationships. It would be an individual school team with only those system personnel who directly affect the particular school program

---

<sup>30</sup>B. J. Chandler, "Study Shows That Merit Rating is not Detrimental to Teacher Morale," The Nation's Schools, 61:4, April 1958.

<sup>31</sup>Willard S. Elsbree and E. Edmund Reutter, Jr., Staff Personnel in The Public Schools (New York: Prentice-Hall Inc., 1954), p. 243.

such as teachers, principal, and possibly supervisors who spend considerable time with the school, students and parents. This plan was similar to the plan proposed by the Association for Supervision and Curriculum Development mentioned previously.

### Summary

The use of ratings in conjunction with salary increments for teachers has decreased over the years from 57 per cent in 1923 to 7.3 per cent in 1956. As the larger urban areas of one hundred thousand or more population abandoned the use of merit rating in school systems, the smaller school systems, those with thirty thousand to 99,999 population, seemed to adopt programs.

More evidence was found to indicate that merit rating programs were unsuccessful than were successful. There seemed to be a lack of common ground among those groups who favor and those who do not favor the use of merit ratings in school systems. Administrators and school boards believe in the principles of merit rating. Teacher organizations disapprove of any evaluation which will serve as a basis for determination of pay or pay increments. Each has a different opinion concerning definition and feasibility of merit rating. Teachers claim that merit rating is bad for morale; at least one study finds that it has no detrimental

effects upon morale. Teachers imply that it can never be done, administrators feel that it is possible but needs careful experimental development.

Recommendations for teacher evaluation by those groups opposing merit rating seemed to fall into two different types: (1) cumulative files; and (2) group evaluation for an individual school. There seemed to be common agreement that whatever type of evaluation of teachers was used it should be based upon democratic principles; teachers should be notified of individual ratings and conferences held between the teacher who was rated and the person or persons who did the rating.

There was no agreement on what constituted merit, teacher effectiveness, and how to measure teacher ability for merit. Studies presented evidence that there were many concepts. Programs developed in the early twenties have since been abandoned.

There was agreement that in order for merit rating to be effective and acceptable, it must attain the objectiveness of a ~~standardized test~~. Statistical analysis of rating have not presented sufficient evidence to indicate reliable or valid instruments for teacher evaluation, but it has not produced sufficient evidence to indicate that merit rating is not within the realm of possibility.

There was agreement that one of the functions of administration is the evaluation of the services rendered by

the personnel within the organization. The disagreement seemed to center around how it should be done, for what purposes and by whom it should be done.

## CHAPTER III

### PROBLEM AREAS IN MERIT RATING

The preceding chapter presented a chronological development of early programs of merit rating in education. Evidence from studies was pointed out which indicated the early programs were troubled by many problems. The present chapter presents a review of educational and industrial literature concerning merit rating which points up a high degree of inconsistency with respect to concepts of evaluation and the utilization of evaluations; however, some indications of unity may be found. Many of the school districts across the country which reportedly claim to operate a merit rating program do so in a negative manner. Many of the reputed merit rating programs are in reality only programs of teacher evaluation, usually to determine placement on tenure. Others involve no system of rating, using instead an accumulation of credit points attainable from various sources to indicate professional growth upon which salary increments depend.

Authoritative research indicates that there are but two basic plans of teacher evaluation for pay purposes which might be termed merit rating.<sup>1</sup> The first of the two is the

---

<sup>1</sup>Hazel Davis, "Facts and Issues in Merit Salary Schedules," The Journal of Teacher Education, 8:131, June 1957.



acceleration method, which refers to the practice of giving either double or larger than normal increments to a teacher's pay based upon a formal rating device. By this method the teacher reaches the maximum salary faster than the average teacher; making the total life earnings slightly higher than that of the average teacher. This type of plan may be authorized with the board reserving the right to act when it desires, or it may be authorized for use at stated intervals within the pay scale.

Second of the two types is the superior-service maximum or super-maximum plan which provides certain promotional steps above the normal maximum salary schedule, in recognition of outstanding service. This plan is not a system of paying extra or above schedule rates for extra duty service such as coaching, sponsoring classes, school papers or clubs. Superior-service maximums are awarded only after teachers have reached the top salary offered through normal progression and have been adjudged superior through examination of performance or formal ratings.

Many references to merit rating programs are sometimes misleading. Davis<sup>2</sup> pointed out that plans sometimes considered or reputed to be merit pay schedules are not. The professional growth pay plan, such as found in Grosse

---

<sup>2</sup>Ibid., p. 129.

Point, Michigan, has been termed a merit plan by some authorities but is not according to Davis. The professional growth plan requires certain evidences of professional improvement as a basis for pay raises. Such growth or improvement is evidenced by extra college credits, travel, research, committee work or other specified areas of endeavor. Some merit plans do incorporate principles of professional growth but such evidence is used to supplement the formal ratings which are the primary basis for salary increases.

Another plan of pay sometimes confused with merit pay plans is that of pay penalties, commonly termed demerit. This type of plan provides that when service is considered unsatisfactory, increments are withheld or reduced to either penalize such service or warn of dismissal at the end of a set period unless improvement is made. Unless such a plan is coupled with a plan to reward superior service, it can not be considered a true merit plan. Davis<sup>3</sup> reports that if a negative type of plan were considered a merit plan then at least one in four teacher pay schedules would be merit plans since most school systems have provisions to allow withholding of increments for unsatisfactory service.

At the present time no state minimum salary law for teachers provides for a merit rating pay provision. Delaware repealed a merit provision in 1947. New York repealed

---

<sup>3</sup>Ibid., p. 128.

the 1948 merit provision in 1956. North Carolina appointed a commission to study the possibilities of merit rating for salary purposes but no plan was found workable. A school merit rating study commission was appointed in Utah in 1953 and plans to present its findings in the early part of 1959. Authorization by local boards of education constitute the legal authority for merit provisions.

Local authorization of merit rating was found in only thirty-seven systems by Davis.<sup>4</sup> Of those authorized eleven were making no use of the authorization; however, the authorization was maintained in the event that the system desired to make use of it. Six systems used the authorization for less than 1 per cent of the total personnel. One system with a super-maximum provision paid all teachers who reached the maximum that amount authorized for meritorious service, discriminating against no one; therefore, Davis concluded that merit pay did not exist, although it was authorized.

The number of systems abandoning or adopting merit rating programs has been unstable, changing from year to year. During the 1956-57 school term thirteen systems dropped the superior-service maximum program while seven added the provision.

---

<sup>4</sup>Loc. cit.

### Implications from Industrial Research

Industrial experiences in merit rating appear to have much to offer the field of education. Bittner<sup>5</sup> presented a series of steps by which a merit rating program should be developed. The steps followed a logical type of development, going from the establishment of purposes to instrument items. Mahler<sup>6</sup> presented a similar set of procedures and problems. Other authorities generally supported these concepts of merit development.

Authorities on industrial merit ratings did not pretend that merit rating has been successful in industry, at least to the degree that many lay people assume. The oft-repeated idea that industry makes merit rating work, so why can't the schools, appeared frequently in articles favoring merit rating programs for education. Evidence usually was found that the writer of such articles used more emotion than logic in the presentations.

Industrial merit rating was found to be nearly as controversial in personnel research as was educational merit

---

<sup>5</sup>Reigh Bittner, "Developing Employee Merit Rating Procedures," Rating Employee and Supervisory Performance, edited by M. Joseph Dooher and Vivienne Marquis (New York: American Management Association, 1950), pp. 20-36.

<sup>6</sup>Walter R. Mahler, "Lets Get More Scientific in Rating Employees." Ibid., p. 49..

rating. Mahler reported that

Employee rating plans have come and employee rating plans have gone. This is an oft-repeated sequence in many companies. Mention employee rating to 10 industrial relations executives, and nine of them will shake their heads and say, "My experience with employee rating has been discouraging. I realize it is important, but . . . ." The tenth one will say, "It certainly has worked for us," or "We wouldn't know how to get along without it."<sup>7</sup>

Mahler's<sup>8</sup> research presented sufficient data to conclude that the lack of a scientific approach to the entire problem is the root of discouraging failures. Authorities in education had presented similar evidence. The elimination of trial and error approaches in favor of an up to date scientific approach might help industry (and education) to find a type of rating which it can use effectively. Bittner<sup>9</sup> presented the following steps which should be accomplished before any forms of rating could be developed except by trial and error methods.

1. The aims and purposes of the anticipated program should be established.
2. A method to inform the persons rated of how they were rated should be established.

---

<sup>7</sup>Ibid., p. 50.

<sup>8</sup>Loc. cit.

<sup>9</sup>Bittner, op. cit., p. 20.

3. Persons should be designated to do ratings who have ability and are willing.

4. The persons doing the rating should have sufficient time to rate and perform regular duties.

5. The traits to be measured should be determined.

6. Criterion should be established by which traits may be determined.

7. The type of rating form to be used should be established.

8. Whether or not the traits would be weighted should be determined.

An area of great importance in industry seemed to be the degree of skill in rating within raters. Educational research revealed similar concern. The only way industry felt that rater skill could be insured was through a special training program for raters. Bittner<sup>10</sup> stated that "A merit rating program must include specific plans and procedures for training the raters." After completion of the preliminary steps outlined by Bittner, consideration should be given to implementation of the program.

Richardson<sup>11</sup> recommended six such considerations for implementing the program which education might use:

---

<sup>10</sup>Ibid., p. 28.

<sup>11</sup>Marion W. Richardson, "Forced Choice Performance Reports," Ibid., pp. 44-46.

1. Secure a qualified technician to be given specific responsibility for administering the program.
2. Secure an understanding and acceptance of the program by administration and general personnel.
3. Devise methods for assisting administration in formulating written standards of performance.
4. Develop methods of utilizing rating information in various phases of the personnel program.
5. Develop methods to handle rating grievances and appeals.
6. Evaluate the program at regular intervals for revision and clarity.

Mahler<sup>12</sup> presented four basic reasons for using a scientific approach in developing and implementing merit rating programs, based upon research in industrial programs, which could assist education. The reasons were:

1. To develop an understanding of the fundamentals underlying the rating process.
2. To develop a systematic procedure for the development of a rating plan.
3. To develop trained personnel to install and administer the program.
4. To attain the support of management (school administrators) during the developmental stages of the program.

---

<sup>12</sup>Mahler, op. cit., p. 50.

The desired understanding of the fundamentals included two basic premises: (1) the ability to rate; and (2) the willingness to rate. Ability to rate was based upon the extent of rater opportunity to observe ratee performance and the rater's consciousness of the prospective rating during the observation periods. Ability to rate should be judged by the personal characteristics of the rater, his training and experience in translating observations into judgments and the facility with which the rating form permitted the rater to record judgments.

Authorities seemed to believe that ability to rate was insufficient unless willingness to rate was also a qualification. Persons who have ability to rate might have no desire to perform the task and therefore be prone to make errors. Authorities seemed to believe that willingness would tend to create a more objective attitude within the rater and increase reliability and validity.

Willingness to rate should be based upon an understanding of the entire program. The rater must accept the stated purposes of the program and reflect the attitude of management (administration). The rater's experiences in applying the rating to personnel under his jurisdiction (supervision) should be considered in determining willingness to rate.

Systematic procedures usually begin with a purpose, so should the development of a merit program. After the purposes



for a merit rating program are established the components (items) of the rating form should be prepared tentatively. Authorities believe that instrument components should be basic and have a definite connection with purposes. In consideration of the components, a method for determining rank or degree of efficiency should be established in order that statements and questions might facilitate scoring. One of the types of ranking such as the check scale plan should be adopted.

Second in a systematic procedure for merit rating establishment should be a program title. Mahler<sup>13</sup> felt that a title descriptive of the program and acceptable to the personnel involved was necessary. This would logically lead to the third procedure which would be interpreting the program to all concerned and initiation or implementation of the program. Mahler<sup>14</sup> did not state but the implication was given that a democratic method of group involvement was desirable.

Fourth, in a systematic development should be the number and time of the ratings. Decisions should cover the time of the year best for evaluation, the number of ratings which should be made and whether or not rating done by more

---

<sup>13</sup>Ibid., p. 51.

<sup>14</sup>Loc. cit.

than one person should be averaged or that all should be responsible for only one rating.

Who would be rated, by whom, and the qualifications necessary for the rater should be determined next according to Mahler's<sup>15</sup> proposal for program development. Some plans should be initiated for training of raters and informing persons rated of their ratings. Since even the best of raters are honestly going to evaluate differently on the most objective forms, some adjustment for variations between raters should be made.

One essential of the systematic procedure was that those who develop a merit plan must realize that it is impossible to determine the variables or combinations of variables necessary for each individual program. All plans should be based upon study of existing successes and failures, research evidence, visits to systems using merit at the time, and the advice of trained personnel. Any plan resulting from such efforts should be tentative until it is tried and proven effective for the system which plans to use it.

Authorities believe that after the tentative program plan is tried, it should be evaluated. Evaluations of the program should center around the reliability and validity of the total and individual item scores. The distribution of the

---

<sup>15</sup>Ibid., p. 52.

individual ratings was considered important because some discrimination at the lower and upper levels would be vital. In short a complete analysis of the merit rating program should be made before final adoption of any specific plan.

### Purposes for Merit Rating

Before any other plans are developed for merit rating programs, research indicates that aims and purposes for a specific program must be developed. Industrial research specifically indicated that statement of purposes was desirable. Research into educational programs indicated purposes neglected or poorly developed; however, purposes did exist and have been developed. The following list is representative of the sundry purposes for educational programs of merit rating.

1. Merit rating will facilitate and serve as a basis for administrative decisions.
2. Merit rating will serve as a basis for improvement of instruction and supervision.
3. Merit rating will provide a reliable and valid measurement of teaching efficiency and as such will establish a higher morale level.
4. Merit rating will help the profession to hold good teachers by making salaries competitive with those in industry and other professions.

Merit rating programs appear to have several reasons for the individual purposes. Usually such reasons as were found turned out to be purposes for teacher evaluation or rationalization for an existing program rather than a foundation for such a program. The following list is representative of the many reasons given by the groups supporting merit rating:

1. The public demands that money be spent for the most efficient teaching service possible.
2. Merit salaries are in harmony with the principles underlying efficiency in public service.
3. Merit salaries stimulate quality work.
4. Merit salary schedules are flexible.
5. Merit salaries provide stimuli and recognition for teachers to go beyond the usually accepted plateau of maximum efficiency.
6. Merit salary programs stimulate teachers to be critical of their own work.
7. Merit salaries, to insure continued maximum efficiency, provide for those teachers who do not make hither promotional hurdles because of limited number of openings.

Even the best plans fail if a favorable environment is not provided. School boards and administrators need to provide sufficient funds and time to develop, implement, and evaluate a merit rating program. School boards and

administrators must give encouragement and moral support to the program. Administrators must assist personnel in comprehension of the program. Teaching personnel must be willing to test the program, and assist in the program evaluation. All persons involved must have an open mind concerning the program. Studies into merit programs in public schools indicated that there are about seven environmental conditions which help make a program successful. These seven conditions were summarized thusly:

1. Teachers must accept the premise that the program is designed to help teachers succeed and improve on the job.
2. Administrators must be well trained.
3. Administrators must be given sufficient time to work closely with teachers.
4. The principal or department head-teacher ratio must be about one to fifteen and the teacher-pupil ratio not over one to thirty.
5. All teachers must be provided with a good basic salary, sufficient for professional levels, with merit increments above the basic amount.
6. The school board should provide environmental conditions favorable to development of the individual system's program.
7. Teachers should be invited to develop and regularly evaluate the entire program.

## Merit Rating Arguments, Pro and Con

Engleman<sup>16</sup> presented three broad factors creating difficulties or problems for merit rating which separate educational programs from those of industry. The first of the three problem areas dealt with the exceedingly complex character of teaching, which was, perhaps, the most frequently used point by those who oppose merit rating programs for educational systems. The second problem area was the great range of specialization inherent in a modern school, the practice or need for specially prepared persons in each of the subject matter areas. The third problem area concerned the difficulty of finding raters of ability to do the ratings. The three problem areas were used in arguments pro and con concerning merit rating in many of the articles reviewed.

The issues involved in merit rating have been broken into many different sub-issues by writers, many seemingly with an emotional bias to pick on a specific pro or con involved in merit pay schedules. The difficulties, advantages and disadvantages are weighed and positions are taken by educational writers, similar to that of personnel men in industry. For this reason it was found that there were implications which could be drawn from industry for education concerning

---

<sup>16</sup>Finis E. Engleman, "Difficulties and Obstacles Inherent in Merit Rating for Teachers," The Journal of Teacher Education, 8:136, June 1957.

merit rating programs. Industrial research which deals with job analysis, purposes for rating, development of devices for rating and group involvement seem definitely to provide guidance for educators.

Authorities seem to agree with Davis<sup>17</sup> that there were but three main issues concerning merit rating for teachers. These issues appeared to be:

1. Can the quality of teachers' services be rated so as to give a valid basis for classification of teachers?

2. If the first issue is true, should salaries be based on the quality of teaching?

3. If salaries should not be based on quality of service, should salaries for all teachers be set at levels similar to other professions?

Problems involved in merit rating seem to be centered around these issues. How groups and individuals answer the above questions largely determines the basis for the pro and con discussions concerning merit rating.

Those who approve of merit rating and present argument and evidence in favor of rewards for meritorious service generally accept all of the reasons and purposes previously listed. The pro arguments were usually based upon one or more of the following points:

---

<sup>17</sup>Davis, op. cit., p. 113.

1. Merit ratings are just, because every teacher is paid according to the worth of individual contributions.
2. Payment of salary based upon merit evaluation keeps teachers alert and on their toes.
3. Merit ratings make higher maximum salaries possible since not every teacher is assured of reaching the maximum.
4. Merit rating conforms to practices established and considered inherent in a capitalistic economy.
5. Merit ratings work for industry and government services.
6. Merit ratings improve teacher morale because it recognizes a job well done, through objective evaluation of individual work and creates within individual teachers knowledge that each can increase his living standards through hard work.
7. Merit rating will give objective evidence of needs within the staff for in-service programs and stimulate experimentation.

Those who oppose merit rating reject the previously given reasons and purposes and develop a slightly different definition substituting the word subjective for objective when defining evaluation. Arguments against merit rating usually begin with

Merit rating is a subjective, qualitative judgment of a teacher, made administratively by one or more persons, with or without the participation or



knowledge of the person rated, for the purposes of determining salary.<sup>18</sup>

Those articles which do not argue from the standpoint taken by the Department of Classroom Teachers usually consider the complexities of teaching. The American Federation of Teachers considered merit rating as impossible because the American Federation of Teachers considered educational purposes and teaching techniques as being intangible, unmeasurable qualities and concluded thusly:

The child himself is the end product of teaching. The aims of the public school are: To impart to him knowledge--that part of the heritage of man which is important to his living; to teach him to think--to use the knowledge which he acquires; to instill proper attitudes and moral standards; to secure right social behaviour. All of these, with the exception of the first, to some degree, are intangibles. . . . The results of teaching are often not immediately apparent. . . . Unfortunately, there is neither complete agreement on what good teaching is nor on what it should achieve.

There is even greater disagreement regarding the methods and procedures by which the objectives of education can be achieved. . . . Teaching is a complex art in which no single method has been demonstrated to be right or best. Since there is no agreement on what is to be measured, there can be no accord on the criteria of the measurement.<sup>19</sup>

---

<sup>18</sup>Department of Classroom Teachers, Classroom Teachers Speak on Merit Rating (Washington: National Education Association, 1957), p. 5.

<sup>19</sup>Mary Herrick, Merit Rating--A Dangerous Mirage (Chicago: American Federation of Teachers, 1956), pp. 3-4.

Gale Rose,<sup>20</sup> Research Director for the Utah School Merit Study, reported that 75 per cent of the systems claiming to have merit programs, provide only for the penalization of unsatisfactory service which is usually done by withholding an increment on an automatic salary schedule. The conclusion was that such a program could be the reason teachers do not think much of merit rating programs.

To summarize the criticisms of merit rating in education, the following points are presented as representative of published opinions:

1. Teaching can not be evaluated out of the context of its environment and the environmental influences upon the pupil. These environmental influences consist of peer groups, home and religious concepts and academic ability.

2. Rating devices must provide a constant measure of what they purport to measure and present day devices are incapable of measuring the intangible aspects of teaching.

3. An effective program of teacher efficiency for merit increments would take more time than supervisors can give without harm to the program.

4. An effective program would be costly and time consuming for the entire system.

---

<sup>20</sup>Gale Rose, Research Director, Utah School Merit Study Commission, Quoted in Education U.S.A. (Washington: National Education Association, March 1958), p. 3.

5. Merit rating tends to lower morale, create professional jealousy, destroy the cooperative spirit between teachers and administrators, and place supervisors into the position of evaluating persons with whom they work, creating yes men not strong teachers.

6. Merit rating tends to produce mediocre teachers by forcing teachers of higher potential to conform to preconceived ideas of some one person or group of persons and discourages experimentation thereby.

7. Merit rating would set teacher evaluation and salary schedules back many years.

#### Common Elements in Merit Programs

Research into data presented concerning programs of merit rating used in public school systems reveal that each plan differs from others, but each is also similar to others in many ways. These similarities may be simply termed common elements found in educational programs of merit rating. Similarities were found in reasons, purposes, types of scales and items within scales, and weighting of items.

Reavis and Cooper<sup>21</sup> found rating instrument items for check scales to include a list of 1538 different items, which,

---

<sup>21</sup>William C. Reavis and Dan H. Cooper, "Evaluation of Teacher Merit in City School Systems," Supplementary Educational Monographs, No. 59 (Chicago: The University of Chicago, January 1945), pp. 25-28.

for the sake of clarity, were grouped under general headings according to relationships. The general grouping included:

1. Sixty different items to ascertain teacher co-operativeness.
2. Fifty-nine different items concerning social attractiveness.
3. Fifty-three different items under the general heading of discipline and guidance.
4. Fifty-two different items concerning professional growth.
5. Forty-four different items concerning participation in "extra-curriculum and other school activities."
6. Forty-two different items concerning the fitness state of the teacher, mental and physical.

The above items were further reduced to 168 different items due to the fact that many were different wordings intended to measure the same thing such as health and vitality, and health and physical vigor. Analysis of the 168 general items revealed seven general areas in which teachers were expected to ascertain desirable competencies. The seven areas which Reavis and Cooper<sup>22</sup> established were:

1. Personal Characteristics. This category headed a group of items designed to reveal the teacher's physical and

---

<sup>22</sup>Ibid., p. 29.

mental condition, personal appearance, ethical behaviour, command of language, judgment and many other different characteristics.

2. Social Relations. This area included a group of items designed to reveal the teacher's cooperativeness, social attractiveness, leadership, and other social qualities.

3. Professional Qualifications. This category included a group of items designed to reveal the teacher's command of subject matter, professional growth and preparation, philosophy of education and other professional ties.

4. Habits of Work. This category included a group of items designed to reveal the teacher's initiative, punctuality, industry, efficiency and other habits which affect the ability to begin a task and carry through to its completion.

5. Instructional Skill. This area included a group of items designed to reveal the teacher's ability to prepare, present and evaluate a lesson and an assignment. Other items included the use of teaching aids, pupil involvement, maintenance of a learning atmosphere and use of instructional materials.

6. Non-instructional School Service. The instruments include a group of items designed to reveal the teacher's extra-curriculum activities, discipline and guidance of pupils and frugal use of supplies and equipment.

7. Pupil Results. Of least consideration in merit

rating was the effect of the teacher upon the behaviour of pupils. The items included the pupil's scholastic and social achievement, attitudes toward school and teachers, work habits, ability to think, assume self-responsibility and other behavioral results.

Social relations contained the largest number of items with a total of 329. The smallest total number of items were the 146 which concerned pupil results.

All but one of the 104 school systems Reavis<sup>23</sup> studied relied upon some form of rating plan for part or all of the teachers' evaluations. Various combinations of plans were used for final evaluations for sundry purposes with the various teachers. The rating plans were grouped by the investigators into five general types of rating instruments:

1. The check scale, which was a listing of several attributes of teachers and their work, each of which was to be evaluated. The check scale was the most frequently used plan.

2. The guided comment form required the rater to write out his comments on a number of leading questions or suggested topics. The guided comment plan was used most frequently in combination with check scales. The combination was used by the second largest group.

---

<sup>23</sup>Ibid., pp. 18-19.

3. The characterization report ranked third. This type of rating form requested the rater to characterize his total impression of the teacher's merit with a single descriptive adjective.

4. The descriptive report plan was fourth in frequency of use. This type of rating required the rater to write a paragraph or two describing the teacher's merit.

5. Ranking, as a rating device, was last on the list. The ranking plan of rating required the rater to list the teachers in a school in order of excellence, from the best to the poorest.

The persons to be rated by the various scales proved difficult to decipher because sixty-five made no comment designating which persons would be rated. Sixteen were for probationary teachers and one each for tenure, non-degree personnel, unsatisfactory teachers and other levels.

The purposes for the evaluative scales were equally varied. The greatest number, forty-seven, had no stated purpose. In twenty-three systems the scales were used for re-employment purposes; fourteen were for change in assignment; eleven were for transfers; six were for salaries; five were for tenure; and, four were for promotion. Others were for maintenance of tenure, retirement, supervision and various other purposes.

Examination of the material submitted by the 104 systems revealed that raters were most frequently principals, supervisors second, and undesignated persons third. Examination of other materials revealed not much change for the ten years following the 1945 study.

Little was found as to required qualifications for raters. No evidence of a qualification requirement could be found in more than half of the scales studied. The one qualification which was most frequently required was the length of time the person rated had been under the rater's supervision.

Teachers were notified of results through various means. Some were required to sign the form, but the largest percentage (82 per cent) had no stated policy for informing teachers of the rating results.

In the 104 rating devices Reavis and Cooper<sup>24</sup> evaluated, much of the material was considered questionable. The material revealed that many of the criticisms of merit rating, at the time, had some basis. The material which was found of value and did not appear questionable was similar to cumulative records. The investigators concluded that the best type of record for teacher evaluation was the cumulative record file.<sup>25</sup>

---

<sup>24</sup>Ibid., p. 46.

<sup>25</sup>Loc. cit.



### Summary

Chapter III has presented an analysis of educational and industrial literature concerning merit rating. Evidence has shown that a high degree of inconsistency existed with respect to the recommendations for program development and implementation, and actual practices. Studies indicated that there were but two basic plans of teacher evaluation for merit pay: (1) the acceleration plan which allowed merit increments at specified periods throughout the teaching career; and (2) the super-maximum plan which allowed merit increments only when the teacher had reached the maximum salary level. A number of systems which use merit pay combine the two basic plans.

Studies seemed to indicate that merit rating programs should be based upon definite purposes; however, investigation into actual practices did not give evidence that this was done. Evidence indicated that a systematic procedure must be used to develop and implement a program of merit rating. Evidence did not indicate that systematic procedures had been used. Studies indicated that programs of merit rating should continually be revised. Evidence indicated that merit programs were not continuously revised.

Authorities indicated that raters should be required to meet certain qualifications and/or undergo special training and have sufficient time to rate and perform other

duties. Evidence did not indicate such was the case. Authorities believed that plans should be developed for specific groups but evidence indicated that usually one plan was used for all groups of teachers.

There were five different types of rating forms used in teacher evaluation; however, two were used by the majority either alone or in combination. The two types most frequently used were the check scale plan and the guided comment plan.

There were seven general categories of concern used in the various rating instruments designed for teachers. The general areas were: (1) Personal Characteristics; (2) Social Relations; (3) Professional Qualifications; (4) Habits of Work; (5) Instructional Skill; (6) Non-instructional School Services; and (7) Pupil Results.

Investigation of industrial studies on merit rating revealed that a systematic or scientific approach of pre-planning and experimentation was the best way to develop a program of merit rating; that there should be a minimum of subjectivity; and that approval of and belief in any program by the raters and the ratees seemed to be necessary for success.

Authorities seemed to agree that programs of merit rating, in order to insure success, would have to consider the following questions and arrive at answers acceptable to

all involved persons:

1. What constitutes merit?
2. What type of instrument was best for formal evaluation?
3. What weights should be awarded to individual items in the instrument?
4. What relationship did pupil achievement have to teacher efficiency?
5. What person or persons should rate teachers?
6. What educational purposes can be achieved by linking teacher efficiency and pay?
7. What reliability and validity can a set pattern of teacher evaluation have?

There appeared to be little evidence of common understanding and research among the groups opposing and approving merit rating programs. Studies indicated that evidence thus far neither prove nor disprove either side. Evidence seemed to indicate that a successful way to increase teachers' salaries by formal evaluation may be developed, but that careful study and experimentation will necessarily precede success.

Summaries of studies presented in this chapter have shown that recommended procedures for the establishment of a merit rating program do not agree with evidence showing actual practice. It has been shown that merit rating programs seem established on a trial and error basis rather than a scientific, systematic set of procedures as recommended by authorities.

## CHAPTER IV

### CURRENT POLICIES AND PRACTICES IN MERIT RATING

The preceding chapter presented the problem areas in development and implementation of merit rating programs. Material was presented which showed that criticisms of merit rating programs have had reasons to exist. The actual practices were more in line with trial and error procedures than with scientific procedures. This chapter deals with the same problems and is based upon investigation of materials concerning specific programs. The data were collected through correspondence.

A group of twenty-five school systems was found to have merit rating programs from a study of literature. Further study of literature concerning merit rating revealed that the larger city school systems were dropping merit programs and smaller city systems were using merit programs. Desiring to check as many programs as possible the writer felt that all school systems utilizing merit rating programs should be contacted. Such a task was greater than the possible results might warrant; therefore, only systems of ten thousand to two hundred thousand in school population and those known to have merit programs were contacted.

The method of initial contact was a double postal card which stated who was conducting the study and for what purpose it was being conducted. The postal card requested that

the school superintendent check yes or no to two questions. The responses would indicate whether or not the system had a merit plan as a basis for paying teachers and whether or not the superintendent would be willing to respond to a check list on the policies and practices involved in the program. A total of 871 postal cards was sent out.

Of the 871 school systems contacted by postal cards 789 responded. Of the total responding school systems, fifty responded no to both questions, 380 responded no to the first question and left the second question blank. Of the total 789 school systems 298 responded no to the first question but yes to the second. There were fifty-four systems which checked yes to both questions. Seven systems returned the postal cards unchecked and enclosed copies of salary policies.

Of the fifty-four positive responses seven qualified the first question, indicating that programs were in developmental stages or that a difference of opinion might exist between the writer and the system regarding the definition of merit. One of the fifty-four was eliminated on the basis of a definition difference. Since fifty-three had indicated a willingness to cooperate with the study a check list was sent to each one.

The check list contained items concerning the developmental, implemental and administrative procedures which might concern merit rating programs. The check list further

requested specific information concerning uses of rating instruments and persons involved with the programs from developmental stages to the date of the check list. Forty-two of the fifty-three systems returned completed check lists. Nine systems returned the check list unanswered explaining that the merit definition that seemed to apply for this study would not include a program such as theirs. Two did not respond.

Along with the check list a letter was sent to each of the fifty-three school systems which had responded positively to the postal card. The letter requested that the system send copies of salary policies, rating instruments, rating procedural policies and such other information which might be pertinent. The purpose of the request for such information was to give the researcher evidence to validate responses on the check list.

Of the forty-two systems which responded to the check list, nineteen were found to have a program which could not be classified as a merit program. The remaining twenty-three programs were considered to fulfill the accepted definition of merit rating. Table I, page 78, presents the programs submitted for study by types of salary program. Definitions for the types came from literature concerning teacher salary plans. As described in the footnote to the table, one each of the acceleration and super-maximum merit

TABLE I

## TYPES OF SALARY PROGRAMS AVAILABLE

<u>Type of Program</u>	<u>Number</u>
<u>Non-Merit Rating Programs</u>	
Demerit	4
Extra duty pay	2
Professional growth plan	6
Other non-merit plans	8
Insufficient data to classify	2
Plan authorized but not used	1
<u>Merit Rating Programs</u>	
Acceleration only	10
Super-maximum only	3
Both	8
Program in developmental stage	3
Insufficient data to classify	2

rating programs was in an experimental stage. Two systems definitely had merit rating programs but did not send sufficient data to classify by type.

As pointed out on Table I, one system had had a merit program authorized but was not using the authorization and was, therefore, considered to be operating without merit provisions. Twenty-two salary programs were considered meritorious by the school systems but did not conform to the definition established for this study. These twenty-two programs were various combinations of plans. Most of the non-merit plans were negative in approach, with four of the twenty-two being for penalization purpose only and were referred to as demerit plans. Only nine of the programs approached salary in a positive manner. The nine positive plans were defined as professional growth and extra duty plans. Professional growth plans based salary upon a system of credits allowed for basic training, travel, extra study, committee work and other specific activities. Extra duty plans allowed salary increments above the salary schedule on a definite scale, for the coaching of teams, plays, clubs, for sponsoring classes and other activities considered beyond a defined normal duty load.

Data from the check list were tabulated by item and sub-item. Those school systems which responded did not all respond to all items; therefore, totals did not always add



up to twenty-three respondents. Some systems had written comments on the check list, others had written letters which were returned with the check list. Two did not send sufficient materials with the check list to classify but that material which was sent was included wherever possible.

The materials from each respondent were kept together in separate files. Each file was numbered in consecutive order from one to forty-two, as received. The materials were studied and analyzed according to the definitions previously presented. Those which were not classified as programs of merit rating were so noted and eliminated from the analysis. The remaining group of twenty-three were tabulated item by item to find the total number of respondents which checked the various possible areas of concern.

The information returned with the check list was examined for clarification of check list items and to present evidence for the items checked. In some cases there was insufficient evidence to indicate that a check list item was correctly indicated, in some cases evidence indicated that the check list item had been misinterpreted; therefore such items were either corrected or omitted, as the case determined.

The information presented in this chapter was then compiled from the check list, as completed by the respondents and general information available concerning the responding

systems of merit rating.

### The Programs Examined

Programs of merit rating appear to vary according to individual system demands. Concepts of merit rating, as indicated by the data for this study, differed in statements of purpose, policies, and practices. These differences were more peripheral than central. There were many basic common groupings. The similarities were: (1) types of programs, (2) types of instruments, (3) items designed to indicate teacher effectiveness and general purposes, (4) raters, (5) problems, and (6) other characteristics.

#### Types of Program Instruments

Methods of evaluation have been classified by general type with various names and in various ways. In order to facilitate the procedure of categorization the instruments reviewed were classified according to the types presented by Reavis and Cooper.<sup>1</sup> The rating instruments used by the twenty-three responding school systems were then classified into six groups: (1) check scales, which were used by ten of the systems; (2) guided comment rating forms, used by five

---

<sup>1</sup>William C. Reavis and Dan H. Cooper, "Evaluation of Teacher Merit in City School Systems," Supplementary Educational Monographs, No. 59 (Chicago: The University of Chicago, January 1945), p. 18.

of the systems; (3) a combination of check scale and guided comment forms, used by one system; (4) characterization forms, used by two systems; and (5) ranking form plan, used by one system. The one using a conference method for teacher rating had no standardized form, using instead a guide for the supervisors of the individual being rated. The conference resulted in a composite evaluation determining the salary increment. One system required a certain score on a standardized test as well as the rating form for establishment of an increment based upon merit.

#### Persons Responsible for Ratings

Ratings were made most generally by the school principal; however, other personnel were held to be equally responsible for the initial rating in some systems. Six systems required the superintendent to rate teachers initially. Six systems held supervisory personnel responsible for evaluating. Seven systems required the department head to serve as an evaluator. Two systems required the individual teacher to assist in his own evaluation, while one system utilized a fellow teacher as an initial evaluator.

Persons held to be responsible for the final rating likewise differed. The superintendent of schools was held solely responsible for the final rating in six systems. In seventeen systems the superintendent and one or more persons were held equally responsible for final ratings. One system

used a systematic plan of responsibility with the superintendent being mainly responsible for final ratings with the supervisor and principal following. The remaining eight systems had various other combinations of position to be responsible for final ratings. One held that an agreement of three out of four positions, including department heads, supervisors, principals and the superintendent was necessary for a final rating. Another held a committee of administrators, teachers and school board members responsible for final evaluation. Three systems included board members as evaluators. One system held the administrative staff to be responsible for the final evaluation. One system held an administrative coordinator to be equally responsible with the principal, supervisor and superintendent for final evaluation.

#### Qualifications for Raters

Qualifications for raters were not defined in the materials covering nineteen of the responding systems. Evidence was found for only three systems which desired specific qualifications for raters. As stated above specific persons were designated to do the rating; however, the only general qualification was evidently a personal acquaintance with the ratee. One of the respondents required that raters be school administrators but did not define the term. The three systems which had requirements specified that the raters should have observed the teacher rated, or know the teacher

personally or both; however these would not indicate ability to rate.

### Manner of Performing Ratings

The initial ratings were based upon several different ways for raters to complete the rating instrument. The check list contained five possibilities for gaining information upon which to base ratings. When the data were studied it was found that several combinations of ways were used, but that the general overall manner was simply an acquaintance with the ratee.

The responding systems combined many of the means presented by the check list. Five of the systems used three different ways of completing ratings; four required two different methods; two required five different methods and one required four different methods. The methods used by the responding school systems were based upon personal knowledge of the individual being rated in nineteen of the total twenty-three. Seventeen systems rated after classrooms visits. Sixteen systems rated after a series of classroom visits for supervisory purposes. Five systems rated teachers after classroom visits and consultation with the teacher's peers. Four systems rated after consultation with colleagues and/or students. One system used a conference plan in which rater and ratee completed the rating

instrument together during a conference held for that purpose.

### Evidences to Validate Judgments

Evidences required to validate ratings were specified by seven systems, according to data made available. One system did require evidence but did not specify the kind of evidence. A final rating was acceptable with no validating evidence or general statements of behaviour by one system. Thirteen systems required objective evidence according to the check list. Fifteen systems required a general statement of behaviour. Three systems required other types of evidences for validation of evaluation. Objective evidence was considered to be examples of work or affidavits of work by authoritative sources such as the principal or the teacher.

### Utilization of Instruments

Final rating results were used with or without other information or evidence, for many purposes. Table II, page 86, presents a comparison of the different ways the instruments were used. To determine promotion of personnel, three systems used the rating instrument without other data, while fifteen used it with other data. To determine demotions, thirteen systems used the rating instrument alone while ten did not. To determine retention or rejection of probationary or non-tenure teachers, six systems used ratings alone while

TABLE II

## USE OF RATING RESULTS BY RESPONDING SYSTEMS

Utilization	Without Other Data	With Other Data
To determine promotion	3	15
To determine demotion	13	10
To determine retention or rejection of probationary or non-tenure personnel	6	11
To determine amount of salary increment	9	14
To determine full salary	1	2

eleven did not. Nine systems used ratings without other data to determine the amount of salary increment, while fourteen utilized other data. One system reported using rating results without other data to determine the full amount of salary. Two systems reported using the rating results with other data to determine the full amount of salary.

The data used with rating results to determine merit increments were utilized in various ways with different combinations. Table III, page 88, presents the analysis of responses concerning other factors used with rating instrument results for determination of merit. Those systems which used other data with rating results usually used more than one and usually allowed specific credits for each other factor which would determine the final decision. Individual school or special committee work was credited most frequently as data to be used with rating results for determination of merit. Equal use of credits based upon system-wide committee work and extra duty responsibilities was made by the responding school systems. Equal use was also made of credits based upon college credit hours, travel and work experiences within the field of education but other than teaching.

Non-credit study, experiences outside the field of education and experimentation, were less frequently used with ratings for determination of merit. Less than half of the total twenty-three responding school systems gave credit



TABLE III

OTHER FACTORS USED WITH RATING INSTRUMENTS  
IN DETERMINING MERIT

Factors	Number Using	Number Not Using
Individual school or special committee work	19	1
System-wide committee work	18	1
Extra duty (above normal expectations)	18	1
Special awards (outside school system)	16	1
In-service training	17	5
College credit hours	16	6
Travel	16	3
Experience with field of education but outside of classroom	15	4
Non-credit, private study	13	6
Experience outside field of education	12	5
Experimentation and research	10	0
Other data (such as tests and community life)	7	0

to teachers for experimentation and research. Other data such as standardized tests and judgments of observed behaviour in community life were used with rating results by some systems. As evidence that teachers' activities in communities were considered important in estimating teacher merit, the number of systems which gave consideration to special awards received outside the school system and to experiences outside the area of education, should be noted.

#### Program Development

The programs studied were developed by a committee of administrators, school board members and teachers in thirteen systems. One of the thirteen systems also had a representative of the local Parent-Teacher Association on the committee. In two systems the plan originated and was developed by system level administrators (administrative personnel other than school principals). Five systems had programs originated and developed by classroom teachers. System level administrators and school boards originated and developed the program in two systems. One system did not know who had originated the plan, which was an old one but did indicate that all personnel had helped revise the plan through the years.

Of the twenty-three responding systems, nineteen felt that provisions were spelled out sufficiently in statements of salary policy, while four did not. Sixteen included

levels of merit increments in reproduced copies of salary schedules which were available to personnel, while seven did not. Twenty systems made copies of salary schedules and merit increments available to all personnel while three did not.

Personnel within the system were not informed who received merit increments and the reasons for the awards, in eleven systems. Eleven systems did so inform the personnel. One system did not respond to the check list items which asked whether or not personnel were informed concerning which persons received merit increments for whatever reasons the system awarded merit. Several respondents stated that such information was considered confidential while others felt such action had no place in a merit program. Many noted that usually such information was learned through gossip. Those systems which informed personnel of merit awards granted, did so by bulletin, other publications and through school principals.

### Program Approval

The responses to two check list items concerning estimated percentages of staff and personnel approval of merit ratings brought two types of responses: (1) estimates; and (2) results of polls conducted to determine exact percentages. Table IV, page 91, presents the tabulated results of the two items. Sixteen systems estimated that 100 per cent of the

TABLE IV

ESTIMATED PERCENTAGES OF PERSONNEL AND ADMINISTRATORS APPROVING,  
DISAPPROVING AND INDIFFERENT TO THE PROGRAMS

System Re- spond.	Percentage of Administrators			Percentage of Teachers		
	Approv- ing	Disapprov- ing	Indiffer- ent	Approv- ing	Disapprov- ing	Indif- ferent
A	100	0	0	-	10	-
B	100	0	0	70	20	10
C	100	0	0	-	-	-
D	100	0	0	70	10	20
E	0	0	40	30	30	40
F	100	0	0	99	0	0
G	100	0	0	90	10	0
H	60	20	20	40	20	40
I	65	5	30	75	15	10
J	100	0	0	60	10	30
K	100	0	0	5	0	95
L	100	0	0	100	0	0
M	100	0	0	90	10	0
N	90	10	0	60	10	30
O	100	0	0	-	-	20
P	60	10	30	60	10	30
Q	100	0	0	25	45	30
R	100	0	0	67	33	0
S	90	0	10	50	25	25
T	100	0	0	85	45	0
U	100	0	0	85	5	10
V	90	0	10	75	15	10

administrators approved of the programs of merit rating used by the individual systems. Three systems estimated that 90 per cent of the administrators approved; two estimated that 50 per cent approved; and one estimated that 60 per cent approved. More administrators were indifferent to merit programs than disapproved of them. Estimated percentages of administrators disapproving of merit were: one estimation of 5 per cent; two estimations of 10 per cent; and one estimation of 20 per cent. Administrators who were indifferent to merit rating were estimated to number 40 per cent of one system; 20 per cent of one system; 12 per cent of one system; and 10 per cent in another system. The range of teaching personnel approving, disapproving and indifferent to merit rating programs was great. Fifteen respondents estimated from 50 to 100 per cent of the classroom teachers approved of the program. One system estimated only 5 per cent of the teaching personnel approved of merit rating; one estimated that 25 per cent of the teachers approved and one estimated that 40 per cent approved. The respondent which indicated only 5 per cent approval by teachers, estimated that 95 per cent of the teachers were indifferent to the program. All other estimated indifference was 40 per cent and below with five estimations at 10 per cent.

Those systems which did not respond were either in the experimental stages, and as a result felt that even an estimate

was impossible at the time or refused to give such an estimate. Reasons given by the respondents concerning estimates indicated that two systems had taken votes, one of which required approval by the personnel as a prerequisite to the establishment of the program. Another respondent stated, "In my opinion, most gripes came from the teachers who got the largest raises." Another felt that the disapproving percentage were those who did not show evidence of intent to improve professionally. One respondent felt indifference to be centered in that group of teachers with less than five years of experience. Three systems had no evidence to base estimates upon, while one responded that such a question was "impossible to answer."

#### Problem Areas

Evidence indicated that problem areas in developing a plan of merit rating differed with each system. Using problem areas found in the literature examined, as a basis, item fourteen on the check list requested the responding systems to check the degree of difficulty for such problem areas as were related to the individual system. The levels of difficulty were placed at: (1) major, or that which created great difficulty; (2) moderate, or that which created less than great difficulty; and (3) little, or that which created

enough difficulty to warrant special attention. There were seventeen problem areas listed concerning program development. There were nine problem areas listed concerning program implementation.

Table V, page 95, presents the problem areas concerning program development with the number of systems which had problems, as the difficulty level was checked by the respondents. Some respondents did not check some of the areas, indicating that there had been no problem or that the program had not developed that far. The tabulated results showed that none of the respondents had major difficulties with acceptance of the plan by school boards and administrators. Sixteen systems had little difficulty with acceptance by school administrators, twenty-one had little difficulty from school boards. The greatest difficulty seemed to be that of getting teaching personnel to accept a program of merit rating with five systems indicating major difficulty, ten indicating moderate difficulty and seven indicating little difficulty.

Determination of when ratings should be made and finding persons willing to rate seemed to have given little difficulty to most of the respondents. Only two school systems indicated that securing persons with ability to rate was a problem area of major importance. Two systems also indicated major difficulty with establishing a program to train

TABLE V

FREQUENCY AND DEGREES OF DIFFICULTY FOR PROBLEM AREAS IN  
DEVELOPMENT OF MERIT RATING PROGRAMS

Problem Areas	Degree of Difficulty		
	Major	Moder- ate	Little
Acceptance of the plan by the school administrators	0	6	16
Acceptance of the plan by the school board	0	1	21
Acceptance of the plan by the school personnel	5	10	7
Establishing purposes for merit rating program	1	11	10
Determining what should constitute merit	12	7	4
Establishing the items for the rating form	6	11	5
Weighing of the items on the rating form	6	8	7
Method of initiating the plan	3	7	13
Determining the time of ratings	4	5	13
Determining method of informing ratees of results	4	9	10
Establishing groups or levels of personnel rated	3	10	8
Determining the time of ratings	0	7	16
Securing persons with willingness to rate	0	6	17
Securing persons with ability or skill to rate	2	9	11
Developing a method of rater education for skill	2	15	4
Standardizing the form for minimum variation	5	11	3
Evaluation of ratings for revision	0	12	6
Other problem areas not mentioned above	4	0	0



raters while fifteen indicated a moderate problem concerning such rater education. One system considered the establishment of program purposes a major problem. None of the responding systems considered evaluation of the program to be a major problem area.

The problem of major importance was that of determining what should constitute merit. Establishing and weighing items for the rating instruments were the second major problem areas to most of the respondents. Establishing a program for education of raters and standardizing the rating forms, for the control of rater variation, to insure a minimum amount of reliability were problems of moderate concern. The respondents indicated that establishment of merit levels was also a problem of moderate difficulty.

The problem areas which created least difficulty for the responding school systems in order of frequency were: (1) acceptance of the program by the school board; (2) securing persons with willingness to rate; (3) determining the time to rate and acceptance of the program; (4) determining a method of initiation for the program; and (5) the frequency of ratings.

Problems not included in the check list, which troubled the responding school systems were: (1) the approval of local, state and national teacher organizations, and (2) finding sufficient time for raters to observe teaching.

Table VI, page 98, presents the problem areas and the degree of difficulty for each area involved in the implementation of the school merit rating programs. Tabulation of the frequency of the checked items indicated that the greatest problem area was that of adjustment for variation in rater's evaluations. The least amount of difficulty was in gaining cooperation of staff and administrators.

Examination of the other problem areas indicated rather close division among the levels of difficulty. Nearly half of the systems indicated little difficulty and the remainder indicated major or moderate difficulty in all areas except the two mentioned above. The implication seems to be that about half of the programs found difficulty of a greater degree in implementation of the program concerning utilization of results of the rating instrument and with the personnel involved as raters and ratees.

The two problem areas added to the group presented in item fifteen of the check list were directly concerned with the administration of the program. More time was needed for raters to do an effective job and for the staff as a whole to understand the program. Evidence was not found in the informational materials of the various programs to indicate that the programs were continuously evaluated and revised. The materials implied that programs once established continued as they were originally established.

TABLE VI

FREQUENCY AND DEGREE OF DIFFICULTY FOR PROBLEM AREAS IN  
IMPLEMENTATION OF MERIT RATING PROGRAMS

Problem Areas	Degree of Difficulty		
	Major	Moder- ate	Little
Rater knowledge of the program in general	2	6	17
Evaluation of effectiveness of rating form	3	10	4
Adjustments for variations in rater's evaluations	0	13	4
Cooperation of staff and administrators	1	2	17
Utilization of rating results for personnel considered ineffective	0	8	9
Consultations with ratee and rater after ratings	2	7	9
Revision and improvement of the program	2	8	10
Coordinating ratings with personnel data on file	2	5	13
Other problems not listed above	2	0	0

### Purposes for Merit Program

Data covering purposes for the merit rating programs of the responding systems were taken from the materials returned with the check list. Most of the systems stated purposes for the existing program were for pay, one program had no other stated purpose. Eleven of the responding systems had certain purposes other than pay, all of a very general nature, usually concerning improvement of instruction. Seven systems were for the specific purpose of improving instruction. There were insufficient data to determine whether one system had any purpose outside of the supposed purpose of determining the amount of salary increment teachers were to receive. One system maintained a salary difference between the sexes. Men received nearly \$200 more at each level than women, for which there was no stated purpose.

### Rating Instrument Components

All of the systems rated teachers while only three of the group which responded rated principals. The rating instruments indicated a wide range of concepts concerning what constituted efficiency for the teachers and the principals. The total different items found on the various rating instruments numbered 610. Analysis of the 610 different items

indicated that many were different statements of the same thing. The analysis presented evidence indicating that the items could be further divided into sixty-eight different sub-items.

By grouping the sub-items which seemed to be related into larger groups, termed categories of general concern, it was found that there were seven areas which the responding systems seemed to consider when rating personnel for merit awards. These areas of concern could be easier understood by comparing the intent of the area with the general categories presented by Reavis and Cooper.<sup>2</sup> Using the areas of concern in this manner, the responding systems were found to be concerned with: (1) instructional skills of teachers, first; (2) social skills possessed by teachers, second; (3) the teacher's personal characteristics, thirdly; (4) the teacher's professional qualifications, fourth; (5) the teacher's work habits, fifth; (6) the non-instructional school services which teachers must do were sixth; and (7) the results of the teacher's work as measured through the pupils were seventh.

The instruments varied in length and in category importance. Many rating instruments were only one or two pages in length, others were five or more. The definitions given for items and categories were usually found in policy

---

<sup>2</sup>Ibid., p. 29.

statements rather than on a specific instruction sheet or the evaluation instrument.

The general categories of concern found in the rating instruments of the responding twenty-three school systems are presented in Tables VII through XIII. The tables are presented according to the number of items found to be used.

Table VII, page 102, presents the instructional skill category. There were twenty systems which used 129 items concerning the skill teachers possessed in classroom instruction. The items of major concern as indicated by the rating instruments appeared to be the skills teachers possessed in the areas of classroom organization, achievement of growth in pupils, and the maintenance of a learning atmosphere within the classroom. About one-half of the responding system's rating instruments had items concerning pupil evaluation, formulation of objectives and utilization of psychological principles of learning. Of less concern were the teachers' ability to plan lessons, stimulate pupils, use instructional materials and present lessons.

Table VIII, page 103, presents the category of concern labeled social relations. Eighteen of the responding school systems used 129 different items which were placed under the general heading of social relations. The areas of major concern in social relations were the teachers' ability to deal effectively with parents and cooperate with coworkers. About

TABLE VII

INSTRUCTIONAL SKILLS CATEGORY, COMPONENTS AND FREQUENCY  
OF USE BY THE RESPONDING SCHOOL SYSTEMS

<u>Category and Components</u>	<u>Number of Systems Using</u>	<u>Frequency of Use in the Responding Instruments</u>
<u>Instructional Skill</u>	20	129
Classroom organization		19
Achieving pupil growth		18
Maintaining a learning atmosphere		15
Pupil evaluation		14
Use of psychological principles of learning		13
Formulating objectives		12
Lesson planning		11
Stimulating pupils		11
Use of instructional materials		9
Lesson presentation		7

TABLE VIII

SOCIAL RELATIONS CATEGORY, COMPONENTS AND FREQUENCY  
OF USE BY THE RESPONDING SCHOOL SYSTEMS

Category and Components	Number of Systems Using	Frequency of Use in the Responding Instruments
<u>Social Relations</u>	18	129
Relations with parents		16
Cooperativeness		15
Relations with pupils		13
Participation in community activities		12
Relations with colleagues		12
Command of social graces		12
Social attractiveness		11
Loyalty to system and school		10
Tactfulness		9
Attitudes toward pupils		6
Sense of justice and fair play		5
Leadership		4
Ability to accept criticism		2
Ability to give criticism		2



half of the responding school systems were concerned with the teachers' relations with pupils, participation in community activities, relations with colleagues, command of the social graces, and general social attractiveness. Less than half were concerned with the teachers' loyalty to the individual school and system and the tactfulness with which the teacher handled awkward situations. Of least concern to the responding school systems were teachers' attitudes toward pupils, sense of justice, social leadership and the ability to accept or give criticism.

Table IX, page 105, presents the general category of personal characteristics which was used by thirteen of the twenty-three responding school systems in eighty-nine different items. The major concern was the teacher's personal appearance. Slightly less than half of the systems were concerned with the teacher's command of English, voice, stability of character, ability to judge and sense of humor. Less than one-third of the instruments contained items concerning the teacher's physical health, enthusiasm, and general attitude toward life.

Table X, page 106, presents the category of concern labeled professional qualifications. Sixteen of the responding school systems used eighty-three different items to determine the professional qualifications of teachers. The areas of major concern were the teacher's professional

TABLE IX

PERSONAL CHARACTERISTICS CATEGORY, COMPONENTS AND FREQUENCY  
OF USE BY THE RESPONDING SCHOOL SYSTEMS

<u>Category and Components</u>	<u>Number of Systems Using</u>	<u>Frequency of Use in the Responding Instruments</u>
<u>Personal Characteristics</u>	13	89
Personal appearance		13
Mental health		11
Command of English		11
Voice		10
Stability of character		10
Ability to judge		10
Sense of humor		9
Physical health		7
Enthusiasm		5
Attitude toward life		3

TABLE X

PROFESSIONAL QUALIFICATIONS CATEGORY, COMPONENTS AND  
FREQUENCY OF USE BY THE RESPONDING SCHOOL SYSTEMS

Category and Components	Number of Systems Using	Frequency of Use in the Responding Instruments
<u>Professional Qualifications</u>	16	83
Professional attitude		12
Participation in professional groups		12
Utilization of up-to-date techniques		11
Knowledge of subject matter		11
Professional growth		8
General cultural knowledge		5
Experiences other than teaching		4
Professional preparation		4
Professionally ethical		3
Contributions to profession		3
Accepted philosophy of education		1

attitude, participation in professional groups, utilization of up-to-date teaching techniques and knowledge of subject matter. About one-third were interested in the amount of outside activities which could be termed professional growth. The teacher's knowledge of the cultural aspects of civilization, non-educational experiences, initial preparation and professional ethics were of lesser concern. Three systems were interested in specific contributions to the profession by individual teachers. One system desired that teachers have a certain philosophy of education.

Table XI, page 108, presents the category of concern labeled habits of work. Sixteen responding school systems used seventy-six different items concerning the teacher's ability to work systematically. The areas of major concern were the teacher's punctuality, ability to assume responsibility, resourcefulness and dependability. The areas of secondary concern were the teacher's ability to adapt in new situations of work and the initiative or aggressiveness necessary to forge ahead on projects. Of third concern were the teacher's ability to create, organize and be efficient in working situations.

Table XII, page 109, presents those determinants of teacher efficiency which were categorized under the general heading of non-instructional school services. Seventeen responding school systems had rating instruments which were

TABLE XI

HABITS OF WORK CATEGORY, COMPONENTS AND FREQUENCY  
OF USE BY THE RESPONDING SCHOOL SYSTEMS

<u>Category and Components</u>	<u>Number of Systems Using</u>	<u>Frequency of Use in the Responding Instruments</u>
<u>Habits of Work</u>	16	76
Punctuality		11
Responsibility		11
Resourcefulness		10
Dependability		10
Adaptability		8
Initiative		8
Creativeness		6
Organizational ability		6
Clerical efficiency		6

TABLE XII

NON-INSTRUCTIONAL SCHOOL SERVICE CATEGORY, COMPONENTS AND  
FREQUENCY OF USE BY THE RESPONDING SCHOOL SYSTEMS

<u>Category and Components</u>	<u>Number of Systems Using</u>	<u>Frequency of Use in the Responding Instruments</u>
<u>Non-Instructional School Service</u>	17	66
Discipline and guidance		17
Care and use of classroom equipment		11
Clerical efficiency		11
Direction of extra-curricular groups		10
Care and use of school property		9
Attention to routine details		8

concerned with sixty-six different items of non-instructional service to the schools. The major concern was that of the teacher's ability to maintain discipline and counsel students. Slightly less than half of the responding school systems were concerned with the teacher's care and use of classroom equipment, clerical efficiency and direction of extra-curricular activities. About one-third of the responding school systems were concerned with the teacher's care and use of school property and attention to routine details.

Table XIII, page 111, presents those components considered by the responding school systems to constitute measurement of pupils which determine teacher efficiency. Pupil results were the items which were least considered by the responding system's rating instruments. Fourteen school systems of the total twenty-three responding used thirty-eight different items to record pupil results. There were six different items to determine the results of teaching as measured by pupils which were concerned with the pupil's social and academic achievement, and attitudes toward school and teachers. There were five different items concerning the pupil's ability to study, work, think appraise his own work and to discipline himself.

The seven tables show that the rating instruments studied indicate that most evaluators of teacher's efficiency were first interested in instructional skills and social

TABLE XIII

PUPIL RESULTS CATEGORY, COMPONENTS AND FREQUENCY  
OF USE BY THE RESPONDING SCHOOL SYSTEMS

<u>Category and Components</u>	<u>Number of Systems Using</u>	<u>Frequency of Use in the Responding Instruments</u>
<u>Pupil Results</u>	14	38
Social achievement		6
Academic achievement		6
Attitudes toward school and teacher		6
Ability to study and work		5
Ability to discipline self		5
Ability to think		5
Ability to appraise own work		5



skills. Secondly, the responding school systems were interested in the teacher's personal characteristics. Of least concern to the evaluators of teacher merit were the items based upon pupil results.

### Summary

This chapter has presented the practices and policies in school merit rating programs. The data were collected through check lists and a review of materials concerning the responding school systems. Data were presented which indicated the responding school systems had problems and procedural practices in common.

Although the school systems which responded to the check list had many concepts of what constituted merit, there were enough similarities to conclude with seven general categories of concern. The seven areas used to determine teacher efficiency for merit purposes were:

1. Instructional skills of teachers
2. Social skills possessed by teachers
3. Personal characteristics of teachers
4. Professional qualifications of teachers
5. Teachers work habits
6. Non-instructional school services performed by teachers
7. Teacher efficiency as determined by pupil results.

Most of the responding school systems used either a check scale rating instrument or a guided comment instrument. A combination of the two types of rating instruments were used by one school system. Other school systems used characteristic and descriptive statements, ranking of personnel, and supervisor's conference to rate teachers.

Few of the responding school systems had definite purposes for rating teachers other than as a method of determining salary increment. Few also had requirements for raters other than that ratings should be made by administrators.

Teachers were rated after the rater had become acquainted with them personally through supervisory visits and after visits to the classrooms for the specific purpose of completing a rating. Ratings could be used without validating evidences. Most systems permitted submission of evidence with the rating instruments to validate judgments made by raters but few required it. Half of the responding school systems had no means of informing teachers of how the teachers were rated. Those school systems which did inform teachers of rating results did so through publications and conferences.

Most of the responding school systems had programs approved by the larger percentage of personnel. The programs were most frequently developed by a committee which involved all persons concerned with the school system, which could

cause greater approval by the group.

The following chapter will present a summary of the findings as reported in this chapter and compare the findings with the recommended procedures as found in literature and reported in Chapters II and III.

## CHAPTER V

### SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

#### Summary

The purpose of this study was to examine the factors involved in the establishment and development of merit rating programs in public school systems; to determine the present status of merit rating policies and practices in the field of education. To achieve the purpose it was necessary to:

(1) trace the development of merit rating plans in public education in the United States; (2) identify common elements of merit rating plans in the field of education; and (3) identify the problems involved in the development and implementation of policies and practices for merit rating plans in educational systems. Since little research had been conducted relating to the development and implementation of merit rating programs in education, a study was made of literature concerning industrial research which might have implications for education.

The study was limited to an investigation and analysis of literature pertaining to merit rating programs and to an analysis of information from check lists and policy statements furnished by cooperating school systems. The cooperating school systems were ascertained through correspondence. The

original contacts with public school systems numbered 871 postal cards from which there were 789 replies. From the 789 postal card returns a total of fifty-three possible merit rating programs were sent check lists and a letter requesting copies of policies, rating instruments, and other materials which might assist the researcher to interpret the check list properly. From the fifty-three sent check lists forty-two replies were received. Twenty-three of the replies had a merit program as defined for this study.

Chapters II and III reported the findings from an analysis of literature studied concerning educational merit rating programs and industrial research into merit rating. Chapter IV reported the findings concerning merit rating programs in the responding school systems.

#### Merit Rating in Education

Evidence was presented that teachers' salaries have evolved through three stages: (1) individual bargaining, (2) position-type schedules, and (3) single salary schedules. Historically it was not possible to draw lines of demarcation which would denote periods of the stages, but general eras have been identified. By the early 1900's merit rating became important in industry and began to spread to education.

Merit pay plans for teachers came to the forefront in the early 1920's. The early plans generally ranked personnel on the normal distribution curve. Some rigidly held to

limited percentages for each level. As a result some teachers were always rated as ineffective and were subsequently penalized. Only one system reported a merit program in which teachers were ranked, according to the data collected for the present study; however, many salary programs approached teacher evaluation negatively.

The complexities of teaching and specialization of subject matter areas have made agreement on what should constitute teacher efficiency somewhat difficult. Concepts of teacher efficiency seemed to be as varied and as numerous as programs. Studies have shown very little correlation between teacher competency and the various items considered important in good teaching. The only agreements seemed to be that merit rating for teachers, in order to be effective, should attain the objectiveness of a standardized test; and that administrators should evaluate the services of personnel within the school system. Authorities in the area of educational evaluation concluded that supervisors of instruction should evaluate only for purposes of improvement of instruction and that supervisors' evaluations should not be used for determination of salary.

The various groups within or concerned with the area of education fail to agree on the value and use of merit rating. School boards and administrators seem more willing to consider merit rating. The organizations which represent

school boards and administrators have published general supporting statements and expressed interest in merit rating. Teachers' organizations generally do not approve of connecting ratings of teacher efficiency with salary. Teachers have generally held that the complexities of teaching make any evaluation subjective and is therefore not a stable base for salary determination.

The arguments pro and con generally center around three main issues: (1) Can the quality of teachers' services be evaluated in such a way that a valid basis for classification according to efficiency level is attained? (2) Should teachers' salaries be based upon the quality of teaching efficiency ascertained by means of a rating instrument? (3) Should teachers' salaries be set at levels similar to those of the professions?

The arguments usually have been based upon personal experiences and generalizations drawn from those experiences, with little statistical evidence to prove or disprove either side. A case in point is that of teacher morale. Those in favor of merit rating have argued that merit rating improved teacher morale, those opposed have taken the opposite view. Research evidence has shown that merit rating has little if any effect upon teacher morale. Those who favor merit rating have argued that it would increase experimentation and research. No studies have been made to determine which

viewpoint is true; however, the data collected for this study showed that experimentation and research was given little consideration in determination of merit.

### Programs of Merit Rating

Evidence presented indicates that many school systems have been reputed to have merit rating programs. Further evidence was presented to indicate that the definition of merit was somewhat loosely used. Evidence was presented which indicated that if negative or general programs of teacher evaluation were considered merit programs then one in four programs in the United States would be so considered.

There were several programs of teacher evaluation which were reputed to be, or claimed to be, merit rating programs which cooperated with this study. Many of those systems which were reputed to have merit rating programs were in reality systems which penalized or awarded increments based on professional growth and extra duty.

There were but two basic programs of merit: (1) the acceleration program which allowed personnel to move ahead on a set schedule either at set periods or yearly so that the maximum salary was reached earlier; and (2) the super-maximum program which allowed personnel who had reached the maximum salary to be awarded additional increments based upon evaluation of merit. The plans can and have existed together and nearly as many use such a combination as used



either plan among the school systems responding to the check list for this study.

The basic plans were implemented by five different methods of evaluation which utilized seven different categories for instrument components. The five methods of evaluation were: (1) the check scale method, which was a listing of several attributes of teachers which were checked by the rater according to a set pattern; (2) the guided comment method which required the rater to write out his comments on a number of leading questions or topics; (3) the characterization method which required the rater to characterize his total impression of the teacher's merit with a single adjective; (4) the descriptive method which required the rater to write a paragraph or two describing the teacher's merit; and (5) the ranking method which required the rater to list the teachers in a school in order of excellence. One of the systems which cooperated with the present study used a conference of all the supervisors of any specific teacher rather than a set form to determine merit.

The seven different categories of concern which the rating instruments commonly used to determine meritorious efficiency were:

1. Personal characteristics of teachers
2. Social relations
3. Professional qualifications

4. Efficient habits of work
5. Classroom instructional skills
6. Non-instructional school services
7. Pupil results

Programs of merit rating were concerned with ability and willingness of raters; however, few of the responding school systems had specific requirements for raters. Most of the responding systems did require a specific person to do the rating. Usually raters were administrators, however, fellow teachers were raters in some of the responding systems.

The policies which were common to school systems having merit programs were found to be:

1. Designation of administrators as raters without required qualifications.
2. Allowance of evidence presentations to lend support to rater judgment but with few evidences required.
3. Allowance of credits toward a merit increment for evidences of professional growth, which were not considered on the rating instrument.
4. Utilization of the evaluation in irregular methods (either with or without other evidence or information) to make personnel changes and to supplement pay.
5. Development of programs by committees composed of administrators, teachers, and board members.

From an examination of the literature the following policies for merit rating programs were recommended:

1. Programs of merit rating should be designed to fit one individual system.
2. Programs should be developed only after much time has been spent studying other programs and research evidence.
3. Programs should be developed by group action.
4. Programs should be developed through a scientific approach which would reduce trial and error.
5. To implement the program the system should have cooperation and understanding among the school board, administrators and teachers.
6. To implement the program and attain cooperation, methods should be designed for handling grievances and appeals.
7. To develop and implement a program the system should realize that it is not possible to determine variables or combinations of variables. The results of a program must therefore be consistently and carefully evaluated and revised.

Recommendations from research to assist effective implementation of a program of merit rating were:

1. Personnel should accept the premise that the program is designed to help teachers succeed and improve on the job.
2. Administrators must be capable and have sufficient time to work closely with teachers.

3. The personnel supervised by a principal and the number of pupils assigned to teachers must be moderate to allow for maximum efficiency.

4. All personnel who are affected by merit ratings must be provided with a good basic salary.

5. The school board should provide environmental and economic conditions under which the program can develop to satisfaction.

6. Personnel affected by the program should be invited to assist in the program development and evaluation.

#### Problems in Merit Rating

Evidence presented indicated that there were many common problems in educational merit rating programs. Those problems which concerned school systems were:

1. Getting teachers to accept the program.
2. Determination of what should constitute merit.
3. Determination of purposes which were best attainable by merit rating.
4. Determination of an instrument which would fit the purposes.
5. Determination of persons who will and can rate.
6. Training raters.
7. Adjustments for variations among raters.
8. Standardization of the rating instrument.
9. Establishing levels of efficiency, once ratings

have been made.

Little difficulty seemed to occur in getting administrators and school boards to accept merit rating. Most of the programs cooperating with this study indicated very little difficulty in evaluating the program's success. More problems of greater degree were encountered in program implementation than in program development.

### Conclusions

Merit rating programs are not numerous. Those programs which do exist in the public school systems which are termed merit, or reputed to be merit, usually approach merit from the negative standpoint. These programs do not usually award salary increments based upon professional growth or on the basis of extra duty. A merit salary program approaches merit from a positive standpoint, considering evidences of satisfactory or better work as well as unsatisfactory work. From the evidence presented in this study the following conclusions were reached:

1. Merit rating has not been completely successful in either industry or education. Evidence presented showed that the use of trial and error methods were probably behind failures.

2. Programs which succeed in school systems are those which involve all personnel within the system.

3. The ratings of programs which appeared to be most satisfactory required evidence to support the judgment of the rater.

4. The successful programs seemed to be those which considered the many variables in teaching and did not set up a program to last indefinitely without revision.

5. Arguments that merit rating does or does not improve teacher morale do not stand up under examination. Evidence indicated that merit rating has no determinable effect upon teacher morale in either a positive or negative manner.

6. The argument that merit rating determines the worth of contributions made by teachers did not hold up under the light of research evidence. There were too many variables involved to credit one device with completeness.

7. Merit rating is costly and time consuming; however, so is any program of evaluation which is not superficial. Evidence did not indicate that sufficient financial assistance was given for development of the program.

8. Evidence did not indicate that merit rating tended to set back teacher evaluation, neither did it indicate that evaluations and salaries moved forward.

9. School systems have not given sufficient concern in merit rating programs to the development of purposes, qualifications, interest of raters, and conferences with

ratees concerning the evaluations made.

10. Findings of this study indicated that school systems utilizing merit rating did not use scientific procedures in developing or implementing the programs.

11. There was little evidence of revisions or plans of evaluation for program revision. Programs once completed and on paper seldom changed or grew, remaining rigid as established.

12. The practice of ranking has faded from the forefront. Teachers are not compared in merit ratings but evaluated in the light of individual accomplishments. Levels of merit no longer limit personnel to percentages based upon the normal distribution curve.

13. Teacher effectiveness is usually measured by evaluation of personal characteristics, social relations, work habits, instructional skills, non-instructional school services, professional qualifications and pupil results.

14. The practice of requiring evidence to support rater's judgments appears to be increasingly important in merit programs. Most programs permit submission of justifying evidence and some require it.

### Recommendations

Evidence has been presented to indicate that certain steps should be taken to insure a measure of success in

developing and implementing merit rating programs. The following recommended procedures might be used as guide lines for program development:

1. Teachers, principals, supervisors, superintendents, and school board members should be involved in preliminary research and development of a merit rating program.

2. Specific program purposes should be determined.

3. Agreement should be made concerning what constitutes teaching merit, based upon observability, universality, and distinguishability.

4. A type of instrument should be determined and if weights are to be used, decision should be made as to the amount and type of weight.

5. The frequency and time of ratings must be determined.

6. Sufficient time and money must be allocated to cover cost of the program.

7. Persons with ability and willingness to rate should be determined. Supervisors of instruction should not be raters.

8. A program to insure rater ability and reduce rater variation should be established.

9. A procedure to inform ratees of ratings should be established. An individual conference of rater and ratee is recommended.



10. The program should be an experimental one with all plans being tentative and subject to change after a trial run. A complete and statistical analysis of the program should follow the trial run.

11. The rating instrument's results should be used with other material in personnel files for award of meritorious salary increment.

12. All decisions should be liable to appeal.

13. Provisions should be established for continuous evaluation and revision of the established program.

14. The results of final ratings should be suitable for and used for other personnel practices in addition to salary purposes.

BIBLIOGRAPHY

## BIBLIOGRAPHY

- Ackerman, Walter I. "A Critical Evaluation of Pupil Changes as a Criterion of Teacher Competency." Special paper, Harvard Graduate School of Education, 1954, quoted in Teacher Competence and Its Relation to Salary, The NESDEC Merit-Salary Committee Report, Cambridge, Massachusetts: New England School Development Council, 1956.
- Alberty, H. B. and Thayer, V. T. Supervision in the Secondary School. New York: D. C. Heath and Company, 1931.
- Appley, Lawrence A. "Foreword," Rating Employee and Supervisory Performance. New York: American Management Association, 1950.
- Barr, A. S. "Measurement of Teaching Ability," Review of Educational Research, 10 (June 1940), 182-184.
- \_\_\_\_\_. "The Measurement and Prediction of Teaching Efficiency," Review of Educational Research 13 (June 1943), 218-223.
- Barr, A. S., Burton, William H., and Brueckner, Leo J. Supervision. New York: D. Appleton-Century Company, Inc., 1947.
- Barthlemess, Harriet M. and Boyer, Philip A. "A Study of the Relation Between Teaching Efficiency and Amount of College Credit Earned While in Service," Educational Administration and Supervision, 14 (1928), 521-535.
- Benge, Eugene J. Job Evaluation and Merit Rating. Part II. New York: National Forman's Institute, Inc., 1944.
- Bittner, Reigh. "Developing Employee Merit Rating Procedures," Rating Employee and Supervisory Performance. Edited by M. Joseph Docher and Vivienne Marquis. New York: American Management Association, 1950.
- Boyce, A. C. "Qualities of Merit in Secondary School Teachers," Journal of Education Psychology, 3 (1928), 144-157.
- Broom, M. E. "A Note on Predicting Teaching Success," Educational Administration and Supervision, 18 (1940), 64-87.

- Butsch, Russell L. "Teacher Rating," Review of Educational Research, 1 (April 1931), 99-107.
- Butts, R. Freeman and Cremin, Lawrence A. A History of Education in American Culture. New York: Henry Holt and Company, 1954.
- Chandler, B. J. "Study Shows that Merit Rating is not Detrimental to Teacher Morale," The Nation's Schools, 61 (April 1958), 4.
- Commission of Teacher Evaluation. Better Than Rating. Washington, D. C.: National Education Association, 1950.
- Correvot, H. Earle. "The Purposes and Methods Involved in Teacher Rating," Phi Delta Kappan, 21 (January 1939).
- Davis, Hazel. "Facts and Issues in Merit Salary Schedules," The Journal of Teacher Education, 8 (June 1957), 131.
- Davis, S. B. and French, L. C. "Teacher Rating," School of Education Journal, 3 (1928), 60-64.
- Department of Classroom Teachers and Research Division. Teacher Rating. Washington, D. C.: National Education Association, 1954.
- Driver, Randolph S. "Employee Performance Evaluation," Personnel Series No. 93. New York: American Management Association, 1945.
- \_\_\_\_\_. "A Case History of Merit Rating," Personnel, 16 (May 1940).
- \_\_\_\_\_. "The Validity and Reliability of Ratings," Rating Employee and Supervisory Performance. Edited by M. Joseph Dooher and Vivienne Marquis. New York: American Management Association, 1950.
- Elsbree, Willard S. and Reutter, E. Edmund. Staff Personnel in the Public Schools. New York: Prentice-Hall, Inc., 1954.
- Engleman, Finis E. "Difficulties and Obstacles Inherent in Merit Rating for Teachers," The Journal of Teacher Education, 8 (June 1957), 136.
- ood, Carter V. Dictionary of Education. First Edition. New York: McGraw-Hill Book Company, Inc., 1945.

- Herrick, Mary. Merit Rating--A Dangerous Mirage. Chicago: American Federation of Teachers, 1956.
- Hopkins, Jess T. "Some Fallacies and Virtues of Merit Rating," Production Series No. 124. New York: American Management Association, 1940.
- Kent, R. A. "What Should Teacher Rating Schemes Seek to Measure?" Journal of Educational Research, 2 (December 1920), 802-807.
- Kingsbury, Forrest A. "Analyzing Ratings and Training Raters," Journal of Personnel Research, 1 (December 1922, January 1923).
- Knight, Frederic B. "Qualities Related to Success in Teaching," Teachers College Contribution to Education, No. 120. New York: Columbia University, 1922.
- Knudsen, Charles W. Evaluation and Improvement of Teaching. New York: Doubleday, Doran and Company, Inc., 1932.
- "Looking Forward," The Nation's Schools, 61 (April 1958), Editorial.
- Lynch, James M. "Teacher Rating Trends Psychologically Examined," American School Board Journal, 104 (June 1942).
- Mahler, Walter R. "Lets Get More Scientific in Rating Employees," Rating Employee and Supervisory Performance. Edited by M. Joseph Docher and Vivienne Marquis. New York: American Management Association, 1950.
- National Commission on Teacher Education and Professional Standards. The Professional Standards Movement in Teaching, 1956-1965. Washington, D. C.: The National Education Association, 1956.
- Prescott, Daniel A. "The Training of Teachers," Rutgers University Bulletin, Series IX, No. 8, p. 5, quoted in A. S. Barr, William H. Burton and Leo J. Brueckner, Supervision. New York: D. Appleton-Century Company, Inc., 1947.
- Reavis, William C. and Cooper, Dan H. "Evaluation of Teacher Merit in City School Systems," Supplementary Educational Monographs, No. 59. Chicago: The University of Chicago, January 1945.

- Reeder, Ward G. The Fundamentals of Public School Administration. Third Edition. New York: The Macmillan Company, 1951.
- Research Division. Advanced Estimates of Public Elementary and Secondary Schools for the Year 1957-58. Washington, D. C.: National Education Association, 1957.
- Richardson, Marion W. "Forced Choice Performance Reports." Rating Employee and Supervisory Performance. Edited by M. Joseph Docher and Vivienne Marquis. New York: American Management Association, 1950.
- Rose, Gale. Research Director, Utah School Merit Study Commission, quoted in Education U.S.A., March 1958.
- Rugg, Harold O. "Self-Improvement of Teachers Through Self-Rating: A New Scale of Rating Teachers' Efficiency," Elementary School Journal, 20 (May 1920), 670-684.
- \_\_\_\_\_. "Is the Rating of Human Character Practicable?" Journal of Educational Psychology, 12 (November, December 1921) 13 (January, February 1922).
- Ryan, Thomas Arthur. "Merit Rating Criticized," Personnel Journal, 24 (May 1945).
- Smeltzer, C. H. and Harter, R. S. "Comparison of Anonymous and Signed Ratings of Teachers," Educational Outlook, 8 (January 1934).
- Tiffin, Joseph. "Merit Rating: Its Validity and Techniques," Personnel Series No. 100. New York: American Management Association, 1946.
- Torgerson, Theodore L. and Adams, Georgia S. Measurement and Evaluation. Albert J. Harris, Editor; New York: The Dryden Press, 1954.
- Torgerson, T. L. "The Measurement and Prediction of Teaching Ability," Review of Educational Research, 4 (June 1934), 242-246.
- \_\_\_\_\_. "The Measurement and Prediction of Teaching Ability," Review of Educational Research, 7 (June 1937), 261-266.
- U. S. Bureau of the Census, "U. S. Summary," Bulletin P-C1. Washington: Government Printing Office, 1950.

U. S. Department of Health, Education and Welfare, Office of Education. Educational Directory, Part 2, 1956-57.  
Washington: Government Printing Office, 1957.

Young Lloyd P. "The Administration of Merit-Type Teachers Salary Schedules," Teachers College Contribution to Education No. 552. New York: Columbia University, 1933.

**APPENDIX**





(Copy of double postal card sent  
to public school systems)

January 11, 1958

Dear Superintendent:

The Department of Educational Administration and Supervision, University of Tennessee, is conducting a study concerning policies and practices involved in merit rating for pay purposes in public schools. We hope to contact all systems having a merit plan. Please check the statement which applies to you, on the attached postal card and return it to us.

Sincerely yours,

John W. Gilliland  
Professor of Education

James L. Keeney  
Graduate Student

- 
1. Our system has a merit plan for paying teachers.

Yes \_\_\_\_\_

No \_\_\_\_\_

2. We will respond to a check sheet on our policies and practices and return same to you.

Yes \_\_\_\_\_

No \_\_\_\_\_

System \_\_\_\_\_

Superintendent \_\_\_\_\_

THE UNIVERSITY OF TENNESSEE  
KNOXVILLE  
College of Education

Department of  
Educational Administration  
and Supervision

February 3, 1958

Dear \_\_\_\_\_:

We appreciate your indication of interest in the study of merit rating which we are doing. The study should be completed by June of this year. A summary of our results will be sent to you upon completion of the study.

Attached to this letter are two copies of the check list which you indicated you would complete for us. One copy is for your files, the other should be returned to us in the addressed and stamped envelope.

We also request that you include any materials which you may have such as copies of evaluation forms, pay schedules and policies or other information which has any relationship to your merit rating plan.

Sincerely yours,

John W. Gilliland  
Professor of Education

James L. Keeney  
Graduate Student

JLK:ed  
Enclosures

## Check List on Merit Rating in Public Schools

School System reporting \_\_\_\_\_  
 Person reporting \_\_\_\_\_  
 Position \_\_\_\_\_  
 Address \_\_\_\_\_

Directions: Place a Check (✓) mark in the appropriate blank.

Yes No

- \_\_\_ \_\_\_ 1. Are your merit provisions spelled out in salary policy statements? (please enclose a copy)
- \_\_\_ \_\_\_ 2. Are levels of merit increments included in reproduced copies of your salary schedule? (please enclose a copy)
- \_\_\_ \_\_\_ 3. Are copies of the salary schedule and merit increments available to all personnel?
- \_\_\_ \_\_\_ 4. Do all personnel know the name and number of persons who receive merit increments in your system?  
 If so, how are they informed? \_\_\_\_\_  
 \_\_\_\_\_
- \_\_\_ \_\_\_ 5. Do all personnel know the reasons for merit increments for those who receive them?  
 If so, how are they informed? \_\_\_\_\_  
 \_\_\_\_\_
- \_\_\_ \_\_\_ 6. Do those persons who do not receive merit increments know the reasons for their not being awarded the increment?
- \_\_\_ \_\_\_ 7. Check the person responsible for the initial evaluation of teachers in which your rating instruments are used.
- \_\_\_ a. Superintendent  
 \_\_\_ b. Supervisor

- \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_
- c. Principal
  - d. Department head
  - e. Fellow teacher
  - f. If other please state position (s) \_\_\_\_\_
8. The initial rating is made
- a. through classroom visits for that purpose
  - b. through personal knowledge of the individual by raters
  - c. after a series of classroom visits for supervisory purposes
  - d. after classroom visits and consultation with teacher's peers
  - e. after consultation with colleagues and/or students.
9. A final rating is acceptable with
- a. objective evidence validating each item
  - b. general statements of teacher behaviour
  - c. no validating evidence or general statement of behaviour
  - d. if other please state \_\_\_\_\_
10. Check the person or persons held responsible for the final rating of personnel.
- a. Superintendent
  - b. Supervisor through the Superintendent
  - c. Principal through the Superintendent
  - d. Principal through the Supervisor and Superintendent
  - e. If other than above state position \_\_\_\_\_
11. What person or group originated the plan and developed it to the final stage of acceptance and use?
- a. Administrators at system level
  - b. School Board
  - c. Classroom teachers
  - d. Principals
  - e. Committee of all of above
  - f. If other state position (s) \_\_\_\_\_
- \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

12. In the blank spaces to the left please estimate the percentage of administrators who are

- \_\_\_\_\_ a. indifferent to your program of merit rating
- \_\_\_\_\_ b. approve of your program of merit rating
- \_\_\_\_\_ c. disapprove of your program of merit rating

13. In the blank spaces to the left please estimate the percentage of teachers who

- \_\_\_\_\_ a. are indifferent to your program of merit rating
- \_\_\_\_\_ b. approve of your program of merit rating
- \_\_\_\_\_ c. disapprove of your program of merit rating

14. Check, in the space provided, the difficulty level of those areas which confronted you with problems in developing the merit plan your system uses.

Difficulty Level  
Major Moderate Little

- |       |       |       |  |
|-------|-------|-------|--|
| _____ | _____ | _____ | a. Acceptance of the plan by the school administrators |
| _____ | _____ | _____ | b. Acceptance of the plan by the school board          |
| _____ | _____ | _____ | c. Acceptance of the plan by the school personnel      |
| _____ | _____ | _____ | d. Establishing purposes for your merit rating program |
| _____ | _____ | _____ | e. Determining what should constitute merit            |
| _____ | _____ | _____ | f. Establishing the items for the rating form          |
| _____ | _____ | _____ | g. Weighting of the items on the rating form           |
| _____ | _____ | _____ | h. Method of initiating the plan                       |
| _____ | _____ | _____ | i. Determining the frequency of ratings                |
| _____ | _____ | _____ | j. Determining method of informing ratees of results   |
| _____ | _____ | _____ | k. Establishing groups or levels of personnel rated    |
| _____ | _____ | _____ | l. Determining the time of ratings                     |
| _____ | _____ | _____ | m. Securing persons with willingness to rate           |
| _____ | _____ | _____ | n. Securing persons with ability or skill to rate      |
| _____ | _____ | _____ | o. Developing a method of rater education for skill    |

14. (continued)

Difficulty Level  
Major Moderate Little

- |       |       |       |  |
|-------|-------|-------|--|
| _____ | _____ | _____ | p. Standardizing the form for minimum variation among raters |
| _____ | _____ | _____ | q. Evaluation of ratings for revision                        |
| _____ | _____ | _____ | r. Other problem areas not mentioned above are:              |

---



---

15. Check level of difficulty for those problems which confronted you in the implementation of your program of merit rating.

- |       |       |       |   |
|-------|-------|-------|---|
| _____ | _____ | _____ | a. Rater knowledge of the program in general                          |
| _____ | _____ | _____ | b. Evaluation of effectiveness of rating form                         |
| _____ | _____ | _____ | c. Adjustments for variations in rater's evaluations                  |
| _____ | _____ | _____ | d. Cooperation of staff and administrators                            |
| _____ | _____ | _____ | e. Utilization of rating results for personnel considered ineffective |
| _____ | _____ | _____ | f. Consultations with ratee and rater after ratings                   |
| _____ | _____ | _____ | g. Revision and improvement of the program                            |
| _____ | _____ | _____ | h. Coordinating ratings with personnel data on file                   |
| _____ | _____ | _____ | i. Other problems not listed above are:                               |

---



---



---

16. Final results from the evaluation form are used without other data

Yes No

- \_\_\_ \_\_\_ a. to determine promotion
- \_\_\_ \_\_\_ b. to determine demotion
- \_\_\_ \_\_\_ c. to determine retention or rejection of proba-  
tionary or non tenure persons
- \_\_\_ \_\_\_ d. to determine amount of salary increment
- \_\_\_ \_\_\_ e. If other please state \_\_\_\_\_

17. Final results from the evaluation form are used with other personnel data on file. Together these data

- \_\_\_ \_\_\_ a. determine promotion
- \_\_\_ \_\_\_ b. determine demotion
- \_\_\_ \_\_\_ c. determine retention or rejection of probationary  
or non tenure persons
- \_\_\_ \_\_\_ d. determine the amount of salary increment
- \_\_\_ \_\_\_ e. If other please state \_\_\_\_\_

18. The data coupled with rating results to determine justification for merit increments are:

- \_\_\_ \_\_\_ a. College credit hours
- \_\_\_ \_\_\_ b. In-service training
- \_\_\_ \_\_\_ c. Non-credit, private study, either in groups or  
individually
- \_\_\_ \_\_\_ d. Experimentation and research
- \_\_\_ \_\_\_ e. Travel
- \_\_\_ \_\_\_ f. System-wide committee work

## 18. (continued)

Yes No

- \_\_\_ \_\_\_ g. Individual school or special committee work
- \_\_\_ \_\_\_ h. Extra duty, any type above the normal duty expectations
- \_\_\_ \_\_\_ i. Special awards (honors received from other than school system for outstanding work in any field)
- \_\_\_ \_\_\_ j. Experience within field of education but outside teaching
- \_\_\_ \_\_\_ k. Experience within areas other than field of education
- \_\_\_ \_\_\_ l. Other areas not covered by the above are:

---



---

19. If the materials you enclose with this check list (as requested in the letter attached to this form) do not cover the following points state briefly those things which will clarify your system's handling of:

(1) Procedures for processing data

(2) Method of completing evaluation

(3) Other: